

Chapter 1

Introductory Concepts

This book is devoted to the subject of Nonlinear Dynamics and the use of Lie Methods for the description and study of Nonlinear Dynamics. Where appropriate, special attention will be given to the application of these methods to the field of charged-particle electromagnetic optics in general and Accelerator Physics in particular.¹ The purpose of this chapter is to provide introductory background material that will be needed throughout the book.² The first four sections of this chapter provide an introduction to the history and use of *maps*, and their relation to differential equations, Hamiltonian dynamics, and Lie theory. Some terms in these sections may not be completely familiar. (If they are, perhaps you need not read further save as a cure for insomnia.) They will be explained and/or properly referenced subsequently. The remaining five sections treat some required aspects of Hamiltonian dynamics and the Theory of Special Relativity.

1.1 Transfer Maps

The use and analysis of maps now plays a major role in nonlinear dynamics and accelerator physics. Much of the material of this book will be dedicated to a map approach. The current use of maps arises from the confluence of two mathematical/physical streams of thought. The first of these streams originates in Geometry, and dates back to the ancient Greeks. The second is related to Dynamics, and originates largely in the discoveries of Isaac Newton (1642-1727).³

¹Lie Methods can also be applied to Light Optics. See Appendix X.

²We also confess to making, here and there, excursions into fascinating related material.

³Newton published his first edition of *Philosophiae Naturalis Principia Mathematica* in 1687, and subsequent editions in 1713 and 1726. Concerning Newton, Laplace said “There is but one law of the cosmos, and Newton has discovered it.” Vladimir Arnold was asked: “Mathematics is a very old and important part of human culture. What is your opinion about the place of mathematics in cultural heritage?” Arnold replied: “The word ‘mathematics’ means science about truth. It seems to me that modern science (i.e., theoretical physics along with mathematics) is a new religion, a cult of truth, founded by Newton three hundred years ago.”

1.1.1 Maps and Dynamics

Prediction is very difficult, especially about the future.

Niels Bohr (1885-1962), Yogi Berra (1925-2015)

Nature and Nature's laws lay hid in night:
God said, Let Newton be! and all was light.

Alexander Pope (1688-1744)

And from my pillow, looking forth by light
Of moon or favoring stars, I could behold
The antechapel where the statue stood
Of Newton with his prism and silent face,
The marble index of a mind forever
Voyaging through strange seas of thought, alone.

William Wordsworth (1770-1850)

Then ye who now on heavenly nectar fare,
Come celebrate with me in song the name
Of Newton, to the Muses dear; for he
Unlocked the hidden treasures of Truth:
So richly through his mind had Phoebus cast
The radiance of his own divinity.
Nearer the gods no mortal may approach.

Edmond Halley (1656-1742)

So few went to hear him, and fewer understood him, and oftentimes he did, for want of hearers, read to the walls. He usually stayed about half an hour; when he had no auditors he commonly returned in a quarter of that time.

Teaching Evaluation of Professor Newton (circa 1690)

Let us begin with the second stream, the stream of Dynamics. Newton's basic and most remarkable discovery was that motion is governed by *mathematical laws*, and the nature of these laws is such that the *future* can be determined/predicted from a knowledge of the *present*¹⁴. We illustrate this fact with the sketch in Figure 1.1. Suppose we think of the

¹⁴Roger Cotes (1682-1716), Newton's student, wrote the preface to the second edition of Newton's *Principia*. Much of this preface is devoted to defending the thesis that the ability to generate and respond to gravity in proportion to its mass is a *natural* property of every object (Cavendish did his experiment 70 years after Newton's death), and not an *occult* property as many critics complained, and to criticizing Descartes' rival theory of vortices. He also writes, with regard to what we call *natural laws*, "The business of true philosophy is to derive the nature of things from causes truly existent; and to enquire after those laws on which the Great Creator actually chose to found this most beautiful Frame of the World; not those by which he might have done the same, had he so pleased. . . . Without all doubt this World, so diversified

present as a set of *initial conditions*, and regard the future as a set of *final conditions*. Newton's laws, when appropriately formulated, can be regarded as a set of first-order ordinary differential equations. Indeed, Newton viewed differential equations and their applicability to describing nature as one of his fundamental discoveries, so important that he kept it secret initially by revealing it [in a 1676 letter (via Oldenburg) to his calculus rival Leibniz (1646-1716)] only in the form of a cypher/anagram/cryptogram:

6accdae13eff7i3l9n4o4qrr4s8t12ux

which Newton's friend Wallis years later (in his 1693 book *Algebra*, second edition) disclosed stood for

with that variety of forms and motions we find in it, could arise from nothing but the perfectly free will of God directing and presiding over all. From this fountain it is that those laws, which we call the laws of Nature, have flowed; in which there appear many traces indeed of the most wise contrivance, but not the least shadow of necessity. . . . He who thinks to find the true principles of physics and the laws of natural things by the force alone of his own mind, and the internal light of his reason must either suppose that the World exists by necessity, and by the same necessity follows the laws proposed; or if the order of Nature was established by the will of God, that himself, a miserable reptile, can tell what was fittest to be done. . . . He must be blind who from the most wise and excellent contrivances of things cannot see the infinite Wisdom and Goodness of their Almighty Creator, and he must be mad and senseless who refuses to acknowledge them." Newton himself wrote (in his book *Opticks*): "The main Business of natural Philosophy is to argue from Phenomena without feigning Hypotheses, and to deduce Causes from Effects, till we come to the very first Cause, which certainly is not mechanical."

With regard to the concept of *necessity*, it is interesting that centuries later Einstein (perhaps with his reptilian brain?) wrote: "What I am really interested in is whether God could have created the world in a different way; that is, whether the necessity of logical simplicity leaves any freedom at all? . . . I would like to state a theorem which at present cannot be based on anything more than faith in the simplicity, i.e., intelligibility, of nature: there are no *arbitrary* constants . . . that is to say, nature is so constituted that it is possible logically to lay down such strongly determined laws that within these laws only rationally determined constants occur."

Newton himself wrote the preface to the first edition of the *Principia*, and laid out his goals as follows: ". . . for the whole burden of philosophy seems to consist of this - from the phenomena of motions to investigate the forces of nature, and then from these forces to demonstrate the other phenomena; and to this end the general propositions in the first and second Books are directed. In the third Book I give an example of this in the explication of the System of the World; for by the propositions mathematically demonstrated in the former books, in the third I derive from the celestial phenomena the forces of gravity with which bodies tend to the sun and the several planets. Then from these forces, by other propositions which are also mathematical, I deduce the motion of the planets, the comets, the moon, and the sea. I wish we could derive the rest of the phenomena of Nature by the same kind of reasoning from mechanical principles, for I am induced by many reasons to suspect that they may all depend upon certain forces by which the particles of bodies, by some causes hitherto unknown, are either mutually impelled towards one another, and cohere in regular figures, or are repelled and recede from one another. These forces being unknown, philosophers have hitherto attempted the search of Nature in vain; but I hope the principles here laid down will afford some light either to this or some truer method of philosophy."

With but a few editorial changes Newton's words could equally well serve today as justification for the support of contemporary basic research! If, for the sake of argument, we identify the aims of "basic research" with those of High Energy Elementary Particle Physics, at the risk of offending a few colleagues, then we see that the goal remains the same, and even the subject matter has changed relatively little. Under the rubric of *bound states* and *scattering theory*, we still wonder about fundamental "forces" and "particles", and why they "cohere in regular figures, or are repelled and recede from one another." And as Newton hoped, his "principles have afforded some light on the truer methods" of Quantum Mechanics, Quantum Field Theory including the Standard Model of Particle Physics, General Relativity including the Standard Model of Cosmology, and the mysteries of Dark Matter and Dark Energy.

*Data aequatione quotcunque fluentes quantitates involvente, fluxiones invenire;
et vice versa*

and means

Given an equation involving any number of fluent quantities, find the fluxions; and vice versa.

In effect, as summarized by Arnold in the first paragraph of his book *Geometrical Methods in the Theory of Ordinary Differential Equations*, Newton said it is useful to formulate and solve differential equations.⁵

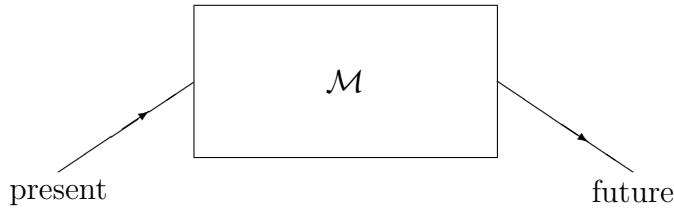


Figure 1.1.1: In Dynamics the future can be determined by performing a certain operation, called a mapping \mathcal{M} , on the present.

As will be described in Section 1.3, there are mathematical theorems about first-order ordinary differential equations to the effect that they generally have solutions. (That is, solutions *exist*.) Moreover, under quite general circumstances, these solutions are *unique* and are completely determined by the initial conditions. Thus there is a rule, or *mapping* \mathcal{M} , that sends the initial conditions (the present) into the final conditions (the future): one simply integrates Newton's equations in first-order form, perhaps numerically on a computer.⁶

In the same era, on the continent across the Channel from Newton, Leibniz wrote (in the context of a problem for which the future depends very sensitively on the present):

That everything is brought forth through an established destiny is just as certain as that three times three is nine. . . . If, for example, one sphere meets another sphere in free space and if their sizes and their paths and directions before

⁵For the Leibniz-Newton calculus controversy, see the Web link https://en.wikipedia.org/wiki/Leibniz–Newton_calculus_controversy. With regard to their rivalry, there is equity in the universe. Most modern calculus notation such as dy/dx and $\int y dx$ is due to Leibniz. He also coined the term *calculus*. Moreover, there might appear to be parity on the cookie front. There are the Fig Newton (1891) and the Leibniz Butterkeks (also 1891). But, alas for I. Newton, the Fig Newton is named for a Massachusetts town whose original name was Newtown.

⁶Given the final conditions (the future), one can equally well integrate backwards in time to find/retrodict the initial conditions (the present), or even farther back to find the past. Thus, we may equally well say that the future determines the present and even the past. The conditions at any instant determine the conditions at all other instants, both future and past. Mathematically, this means that the transfer map \mathcal{M} associated with any set of first-order ordinary differential equations is *invertible*.

collision are known, we can then foretell and calculate how they will rebound and what course they will take after the impact. Very simple laws are followed which also apply, no matter how many spheres are taken or whether objects are taken other than spheres. From this one sees then that everything proceeds mathematically -that is, infallibly- in the whole wide world, so that if someone could have a sufficient insight into the inner parts of things, and in addition had remembrance and intelligence enough to consider all the circumstances and to take them into account, he would be a prophet and would see the future in the present as in a mirror.

That this concept (in the context of motion) was generally understood in scholarly circles a generation after Newton and Leibniz is evident from the work of the Serbian Jesuit scholar Boscovich (1711-1787). In 1763 he wrote:

Any point of matter ... must describe some continuous curved line, the determination of which can be reduced to the following general problem. Given a number of points of matter, and given, for each of them, the point of space that it occupies at any given instant of time; also given ... the tangential velocity ...; and given the law of forces ...; it is required to find the path of each of the points, that is to say, the line along which each of them moves. How difficult this mechanical problem may become, how it may surpass all powers of the human mind, can be easily understood by anyone who is versed in Mechanics and is not quite unaware that the motion of even three bodies only, and those possessed of a perfectly simple law of force, have not yet been completely determined in general Now although a problem of such a kind surpasses all the powers of the human intellect, yet any geometer can easily see thus far that the problem is determinate Now, if the law of forces were known, and the position, velocity and direction of all the points at any given instant (were known), it would be possible for a mind of this type to foresee all the necessary subsequent motions and states, and to predict all the phenomena that necessarily followed from them.

Laplace (1749-1827) subsequently stated this concept equally explicitly in 1814 when he wrote:

We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it—an intelligence sufficiently vast to submit these data to analysis—it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes. The human mind offers, in the perfection which it has been able to give to astronomy, a feeble idea of this intelligence. Its discoveries in mechanics and geometry, added to that of universal gravity, have enabled it to comprehend in the same analytical expressions the past and future states of the system of the world. Applying the same method to some other objects of its

knowledge, it has succeeded in referring to general laws observed phenomena and in foreseeing those which given circumstances ought to produce. All these efforts in the search for truth tend to lead it back continually to the vast intelligence which we have just mentioned, but from which it will always remain infinitely removed. This tendency, peculiar to the human race, is that which renders it superior to animals; and their progress in this respect distinguishes nations and ages and constitutes their true glory.

Note the similarity in language!⁷ The same perception is echoed by Thomasina Coverly in Tom Stoppard's 1993 play *Arcadia*. In Act I she says:

If you could stop every atom in its position and direction, and if your mind could comprehend all the actions thus suspended, then if you were really, really good at algebra you could write the formula for all the future; and although nobody can be so clever as to do it, the formula must exist just as if one could.

In modern terminology, Leibniz, Boscovich, Laplace, and Thomasina (Stoppard) were describing what we call a *transfer map*.⁸

All this would have pleased the ancient Greek Stoic philosophers both in buttressing their belief in determinism and in addressing their desire to divine the future. As Cicero explained in his 44 B.C. work *On Divination*,

Besides, since everything happens by fate, as will be shown elsewhere, if there could be any mortal who could observe with his mind the interconnection of all causes, nothing indeed would escape him. For he who knows the causes of things that are to be necessarily knows all the things that are going to be. But since no one but God could do this, what is left for man is that he should be aware of future things in advance by certain signs which make clear what will follow. For the things which are going to be do not come into existence suddenly, but the passage of time is like the unwinding of a rope, producing nothing new but unfolding what was there at first.

Newton showed that what was needed to determine the future was a knowledge of the initial conditions and the universal force laws (the inverse square law for gravity in his case), followed by the integration of his equations of motion. And integration of the equations of motion, particularly when carried out time-step by time-step numerically (see Chapter 2),

⁷Later commentators and philosophers of science sometimes refer to Laplace's *vast intelligence* by the (what might be understood as pejorative) term Laplace's *demon*, perhaps in analogy to Maxwell's demon. Laplace never used that term, and based on his usage above it could be argued that he was envisioning an admirable/exalted transcendent/omniscient being. Actually, Maxwell didn't use the term demon for his being either. It was first introduced by Kelvin in 1874, and he implied that he intended the mediating, rather than malevolent, connotation of the word.

⁸The use of the terminology *transfer map* in this context is not to be confused with the use of the same terminology in other contexts including computer graphics, statistical mechanics, various aspects of group theory, and the articulation of courses between different universities and colleges. Our usage is motivated by terminology in (light) ray optics. In ray optics the linear (paraxial) approximation of what we call a transfer map is called a *transfer matrix*.

does resemble, in some ways, the unwinding of a rope. Whatever is produced is not “new”, but rather already inherent in the initial conditions.

Let us continue our historical narrative: Lagrange (1736–1813) and others discovered that for many systems of physical interest all the differential equations of motion could be generated by (derived from) a *single* master function now called the Lagrangian L .⁹ These equations of motion were second order. Building on this work, Hamilton (1805–1865) showed that it was possible to write a related set of first-order equations, and that all these equations could also be generated by a single master function now called the Hamiltonian H .¹⁰

Being well aware of the aforementioned properties of first-order differential equations, Hamilton made a detailed study of the nature of the relation between initial and final conditions (the transfer map \mathcal{M}) for Hamiltonian systems. In modern language, he showed that such maps must be *symplectic* (canonical). He also discovered mixed-variable generating functions, and showed that they can be used at will to produce symplectic maps. Finally, he and Jacobi (1804–1851) studied how symplectic maps could be used to transform Hamiltonians with the aim of simplifying them, and thus also the differential equations and flows they generate. In modern terminology, their work was the beginning of the Theory of Normal Forms for differential equations, Hamiltonians, and Symplectic Maps.

Poincaré (1854–1912) was the next person to champion the use of maps and explore their properties: He introduced what we now call stroboscopic maps and Poincaré surface-of-section maps. He showed that the existence of an infinite number of periodic orbits in the gravitational 3-body problem would follow from proving the existence of two fixed points for a certain symplectic map of an annulus (in the plane) into itself.¹¹ He discovered what we now call the Poincaré invariants and showed that they are preserved by symplectic maps. He studied normal forms for differential equations and showed that attempts to use symplectic maps to bring certain classes of Hamiltonians to a certain kind of normal form, which if successful would prove the existence of integrals of motion, seemed fraught with intractable difficulties due to the appearance of so-called *small denominators* that potentially spoil the convergence of the series designed to construct the desired normal form.¹² He also discovered what are now called *homoclinic tangles*, emphasized their generic existence, and demonstrated that their presence destroys integrability and leads to chaos.

⁹Together Lagrange and Euler (1707–1783), and later Hamilton, also developed variational calculus and showed that Lagrangian and variational formulations are equivalent. Presently it is commonly assumed that any fundamental theory of nature will be Lagrangian in form. See Section 5.

¹⁰The function H , its relation to L by way of a Legendre (1752–1833) transformation, and the resulting equations of motion were actually introduced earlier by Lagrange when Hamilton was still a child. Lagrange used the letter H to honor Huygens (1629–1695). Hamilton wrote definitive papers on light optics and dynamics in which he introduced characteristic (generating) functions and also employed the H of Lagrange. See Appendix X. To his great fortune, after that H became known as the Hamiltonian. With regard to Lagrange, Hamilton wrote “Lagrange has perhaps done more than any other to give extent and harmony to such deductive researches by showing that the most varied consequences . . . may be derived from one radical formula, the beauty of the method so suiting the dignity of the results as to make his great work a kind of scientific poem.”

¹¹The conjecture that the symplectic map of the annulus into itself must have two fixed points is called Poincaré’s last geometric theorem. He in fact knew that the existence of one fixed point already entailed the existence of two fixed points, and therefore it is only necessary to prove the existence of one fixed point.

¹²Poincaré was unable to prove either the convergence or divergence of the series in question, but inclined toward the opinion that such series were generally divergent.

Birkhoff (1884-1944), in addition to making other outstanding contributions to mathematics, extended the program of Poincaré. In a celebrated early paper he was able to prove what, despite considerable effort, had eluded Poincaré: the map for the 3-body problem developed by Poincaré did indeed have two fixed points. (He proved that the assumption that Poincaré's annulus map had no fixed point entailed a contradiction.) He also studied the possibility of using symplectic maps to bring certain classes of Hamiltonians to what is now called Birkhoff normal form. Again, he found that the appearance of small denominators potentially destroyed convergence, and left the convergence question unanswered. Finally, Birkhoff made other significant contributions to Dynamics including fundamental work on ergodic theory and the areas we now call bifurcation theory and symbolic dynamics.

Siegel (1896-1981) was the first to master the small denominator problem in the context of analytic maps of the complex plane into itself. That is, despite the appearance of small denominators in the series he developed to treat his problem, he was able to prove convergence provided the frequencies appearing in his problem were sufficiently incommensurate. Subsequently, Moser (1928-1999) overcame this problem for “twist” and area-preserving (symplectic) maps of the plane into itself under the assumption of only sufficiently high-order differentiability. Kolmogorov (1903-1987) and Arnold (1937-2010) handled the small denominator problem for the case of symplectic maps/Hamiltonian systems in any number of dimensions under the assumption of analyticity. Together their work proved, under suitable assumptions, the existence of KAM (Kolmogorov-Arnold-Moser) tori for symplectic/Hamiltonian systems. Major advances/extensions in KAM theory were made subsequently by Aubry, Mather, Nekhoroshev, Chirikov, and others.

Smale (1930-) greatly extended symbolic dynamics, invented his horseshoe construction which he described using symbolic dynamics, and showed that Poincaré's homoclinic tangle contained a horseshoe.¹³

Recent advances in nonlinear dynamics include bifurcation and chaos theory, symplectic differential geometry and symplectic topology, and special numerical integration methods often referred to as *geometric/structure-preserving/symplectic* integration.

1.1.2 Maps and Accelerator Physics

Let us momentarily turn our attention to accelerator physics. Ernest Courant (1920-2020) and Hartland Snyder (1913-1962) pioneered the use of matrices to characterize transverse beam behavior in the linear (first-order or paraxial) approximation. These matrices were enlarged by Samuel Penner (1930-) to include chromatic (energy dependent) effects.¹⁴ In subsequent work Karl Brown (1925-2002) made the important step of extending the linear matrix formalism to include nonlinear effects through second order. From the perspective of maps, we may view the use of a matrix as making a linear approximation to the underlying transfer map \mathcal{M} , and the inclusion of second-order effects as introducing the first nonlinear

¹³Smale also has made important contributions to many other areas of mathematics, and famously/scandalously claimed that some of his best early work, including his horseshoe construction, was done “on the beaches of Rio” while supported by a National Science Foundation Fellowship.

¹⁴Time-dependent effects were first included in the Lie algebraic code *MaryLie*.

terms that appear in a *Taylor* expansion of \mathcal{M} about some design orbit.¹⁵ It is now relatively easy to compute the terms in a Taylor expansion of \mathcal{M} to very high order. This computation is made possible by two tools. The first is the use of Lie methods. The second consists of *Truncated Power Series Algebra* (TPSA) and/or *Automatic Differentiation* (AD) computer programs that manipulate very high-order polynomials and various other familiar functions in several variables.¹⁶ Both these topics will be described extensively in subsequent chapters.

¹⁵Brook Taylor (1685-1731) was an English mathematician. The importance of Taylor's formula/series was not fully recognized until after his death when, in 1772, Lagrange realized its usefulness and termed it "the main foundation of differential calculus". Also, we apologize to the Scottish mathematicians James Stirling (1692-1770) and Colin Maclaurin (1698-1746). Series expansions about the origin were introduced by Stirling and are often called Maclaurin series since he studied such series extensively. For simplicity, we will refer to expansions about any point, including the origin, as Taylor expansions.

¹⁶Some authors refer to AD as *Algorithmic Differentiation*. Yet others refer to AD as *Differential Algebra* (DA).

1.1.3 Maps and Geometry

Euclid alone has looked on Beauty bare.
 Let all who prate of Beauty hold their peace,
 And lay them prone upon the earth and cease
 To ponder on themselves, the while they stare
 At nothing, intricately drawn nowhere
 In shapes of shifting lineage; let geese
 Gabble and hiss, but heroes seek release
 From dusty bondage into luminous air.

O blinding hour, O holy, terrible day,
 When first the shaft into his vision shone
 Of light anatomized! Euclid alone
 Has looked on Beauty bare. Fortunate they
 Who, though once only and then but far away,
 Have heard her massive sandal set on stone.

Edna St. Vincent Millay (1892-1950)

We now consider the first stream, the stream of Geometry. Much of geometry can be traced to the assembly of Euclid's *Elements* (c. 300 B.C.). In addition to providing extensive material of his own, Euclid documented the work of many of his predecessors and contemporaries. Moreover, he set the course for essentially all subsequent mathematics: Beginning with a small number of axioms (now commonly *naive* or *axiomatic Zermelo-Fraenkel* set theory often supplemented by the axiom of choice), seek to prove from these axioms by logical deduction/demonstration an astonishingly large number of both profound and enormously useful results.¹⁷

A fundamental notion in geometry as conceived by Euclid is that of *congruence*. Roughly speaking, we regard two triangles as congruent if one can be placed over the other with a resulting perfect fit. From the perspective of maps, we have in mind the operations of translations and rotations which map Euclidean space into itself. Together these operations form a group, the *Euclidean group*. Thus, following Felix Klein (1849-1925), we may say that two triangles are congruent if one can be *transformed* into the other under the action of the Euclidean group. And two triangles are *similar* if one can be transformed into the other under the action of the Euclidean group augmented by scale transformations.

¹⁷Logical demonstration was also important to one of America's greatest Presidents. While an aspiring lawyer, Abraham Lincoln kept a copy of Euclid in his saddlebag, and studied it late at night by lamplight. He related that he said to himself, "You never can make a lawyer if you do not understand what demonstrate means; and I left my situation in Springfield, went home to my father's house, and stayed there till I could give any proposition in the six books of Euclid at sight".

President Garfield too had some mastery of Euclidean geometry. While a U.S. Congressman for the state of Ohio, prior to serving as President, he developed and published in the 1 April 1876 issue of the *New-England Journal of Education* an independent proof of the Pythagorean theorem. In an introductory paragraph to Garfield's article the journal editor remarks that Garfield's proof "is something on which the members of both houses can unite without distinction of party". Alas, like Lincoln, Garfield too was assassinated while President.

The concepts underlying the Euclidean group were subsequently broadened by Klein (as part of his Erlangen program) and others to include the idea of general transformation groups that map various kinds of spaces or various classes of objects into themselves.¹⁸ Sophus Lie (1842–1899), and others both before and after him (including Poincaré), studied transformation groups for their applications to both geometry and function theory, and (in what amounts to a systematic procedure for transforming variables) the simplification and perhaps even solution of certain classes of differential equations.¹⁹ Lie studied in particular the properties of what we now call Lie groups: groups that can be *generated* by near-identity operations. The generators of these near-identity operations form algebras which we now call Lie algebras. For example, in the case of the rotation group (a subgroup of the Euclidean group) there exist small (infinitesimal) rotations, and any group element can be constructed (infinitesimally generated) from these near-identity operations. When a matrix representation is used (and assuming a three-dimensional space), the generators of the infinitesimal rotations are three matrices, call them L_x , L_y , and L_z , that obey the commutation (Lie algebraic multiplication) rules

$$\{L_x, L_y\} = L_x L_y - L_y L_x = L_z, \text{ etc.} \quad (1.1.1)$$

The elements of the set of all continuous and invertible maps of a space into itself are called *homeomorphisms*. Topology (another area pioneered largely by Poincaré) is the study of those properties of spaces, and objects in these spaces, that are invariant under homeomorphisms. Homeomorphisms that are differentiable are called *diffeomorphisms*. The set of all diffeomorphisms forms a group that is a Lie group. Differential geometry is the study of those properties of spaces, and objects in these spaces, that are invariant under diffeomorphisms.

The set of all symplectic maps (sometimes called *symplectomorphisms*) also forms a Lie group, and this Lie group is a subgroup of the Lie group of diffeomorphisms. In both the group of all diffeomorphisms and the group of all symplectic maps, *Lie transformations* are those group elements produced by a single generator. Hori (1932–) and Deprit (1926–2006) were the first (in the context of Dynamics) to use Lie transformations for the production of symplectic maps. They employed these maps to try to bring to normal form various Hamiltonians that arise in celestial mechanics, and showed that the use of Lie transformations is often much more convenient than the method of mixed-variable generating functions developed earlier by Hamilton and Jacobi. As will be described in subsequent chapters, Lie algebraic methods also have important applications to Accelerator Physics. In this case Lie transformations, and products of Lie transformations, can be used to represent symplectic transfer maps, and Lie algebraic formulas (the Baker-Campbell-Hausdorff and Zassenhaus formulas) can be used to multiply and factorize maps. Lie methods can also be used to

¹⁸The importance of groups was not always universally appreciated. In 1910 a board of experts including Oswald Veblen and Sir James Jeans, upon reviewing the mathematics curriculum at Princeton, concluded that group theory ought to be thrown out as useless. And, in the early days of Quantum Mechanics, the work of those physicists/mathematicians who sought to apply group theory to this new field was referred to as *Gruppenpest*.

¹⁹By developing a theory of continuous groups, Lie aspired to do for differential equations what Galois had done for algebraic (polynomial) equations using finite groups.

bring transfer maps to normal form. Among other things, normal form theory generalizes Courant-Snyder theory to the nonlinear regime.

1.2 Map Iteration and Other Background Material

There are important situations where it is desirable to know the effect of a map when it is applied a large number of times. Consider, for example, the case of a charged-particle storage ring. Such rings can be characterized by a one-turn map; call this map \mathcal{M} . Since storage rings are intended to hold particles for long periods of time and correspondingly a large number of turns, we find we are interested in properties of \mathcal{M}^n for values of n in the range 10^8 to 10^{10} .

We observe that the concept of map iteration, or equivalently the study of \mathcal{M}^n for large n , introduces an *infinity* into the game. Consequently, we might anticipate that phenomena arising from the iteration of maps could be very complicated. This is indeed the case.

1.2.1 Logistic Map

Consider, as a simple example, the biological subject of insect population growth. Let P_n be the population in year n (of some insect species), and let P_{n+1} be the population the following year. Then we might imagine that there is some kind of rule (or map) \mathcal{M} that relates the population in two successive years as shown schematically in Figure 2.1.

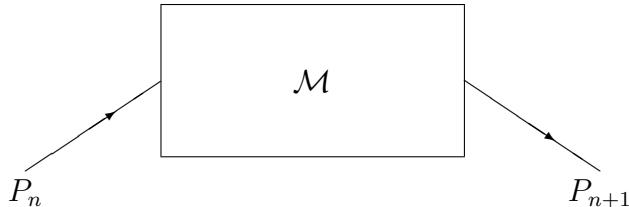


Figure 1.2.1: The insect populations in two successive years are related by a map \mathcal{M} .

The simplest form for the map \mathcal{M} is a relation of the kind

$$P_{n+1} = \alpha P_n, \quad (1.2.1)$$

where α is viewed as some *fixed* growth rate. However, depending on the size of α , the recursion relation (2.1) has only exponentially damped or exponentially growing solutions; and both these possibilities are unphysical – the actual insect population is neither dropping to zero nor growing indefinitely.

An improved model would be to assume that the growth rate itself depends on the current population. For example, we might imagine that if the population were small, then food would be plentiful, and the growth rate should be high. Conversely, if the population were

at some maximum value P_{\max} , then food might be in such short supply that there would be no reproduction at all. A simple form for α having this property is obtained by writing

$$\alpha(P) = \beta(P_{\max} - P). \quad (1.2.2)$$

With this improved model the map \mathcal{M} takes the form

$$P_{n+1} = \beta(P_{\max} - P_n)P_n. \quad (1.2.3)$$

Finally, for mathematical convenience, let us introduce the *fractional* population x defined by the rule

$$x = P/P_{\max}. \quad (1.2.4)$$

In terms of this variable the relation (2.3) takes the form known as the *logistic map* or *Verhulst process*,

$$x_{n+1} = f(\lambda, x_n) = \lambda x_n(1 - x_n). \quad (1.2.5)$$

(Here $\lambda = \beta P_{\max}$.) Note that (2.5) has the physically desirable property that $x_{n+1} \in [0, 1]$ if $x_n \in [0, 1]$ provided $\lambda \in [0, 4]$.

Let us solve (2.5) for an *equilibrium* value (*fixed point*) x_e . By definition, and using map notation, this value must satisfy the relation

$$\mathcal{M}x_e = x_e, \quad (1.2.6)$$

from which we find the result

$$x_e = \lambda x_e(1 - x_e) \quad (1.2.7)$$

with the solutions

$$x_e = 0, \quad (1.2.8)$$

$$x_e = (\lambda - 1)/\lambda. \quad (1.2.9)$$

Suppose we select some value x_0 for an initial (fractional) population and apply the map \mathcal{M} repeatedly for a total of m times to find the result

$$x_m = \mathcal{M}^m x_0. \quad (1.2.10)$$

That is, we carry out the operation (2.5) for a total of m times. Then we might wonder what happens in the limit of large m . (The set of all points x_m for all integer m is called the *orbit* of x_0 under the action of \mathcal{M}). For example, do the x_m approach x_e (in which case x_e is called an *attractor*), or does something else happen?

Figure 2.2 shows the values x_m as a function of m starting with $x_0 = 1/2$ for the case $\lambda = 2.8$. Other starting values of x_0 give similar results as m becomes large. Evidently the x_m converge to the value x_e given by (2.9) as $m \rightarrow \infty$, and x_e is an attractor. All points (starting values) x_0 such that the associated x_m converge to x_e are said to be in the *basin of attraction* of x_e . Let x_f be an attracting fixed point for some map \mathcal{M} ,

$$\mathcal{M}x_f = x_f. \quad (1.2.11)$$

In set theoretic language, $B(x_f)$, the basin of x_f under the action of \mathcal{M} , is defined by the rule

$$B(x_f) = \{x \mid \lim_{n \rightarrow \infty} \mathcal{M}^n x = x_f\}. \quad (1.2.12)$$

By contrast, Figure 2.3 shows the values x_m as a function of m starting with $x_0 = 1/2$ for the case $\lambda = 3.01$. Again other starting values of x_0 give similar results as m becomes large. Now we see that x_e , while still a fixed point, is no longer an attractor. Instead, as m becomes large, the successive values of x_m settle down to *two alternating* values; and it now takes *two* years for each of these values to repeat itself. We say that *period doubling* has occurred so that for $\lambda = 3.01$ the map \mathcal{M}^2 has two attracting fixed points, and \mathcal{M} itself sends each into the other. Insects living in this regime experience alternating fat and lean years! Since the map \mathcal{M}^2 has two attracting fixed points, there will be two basins of attraction for \mathcal{M}^2 , one for each fixed point.

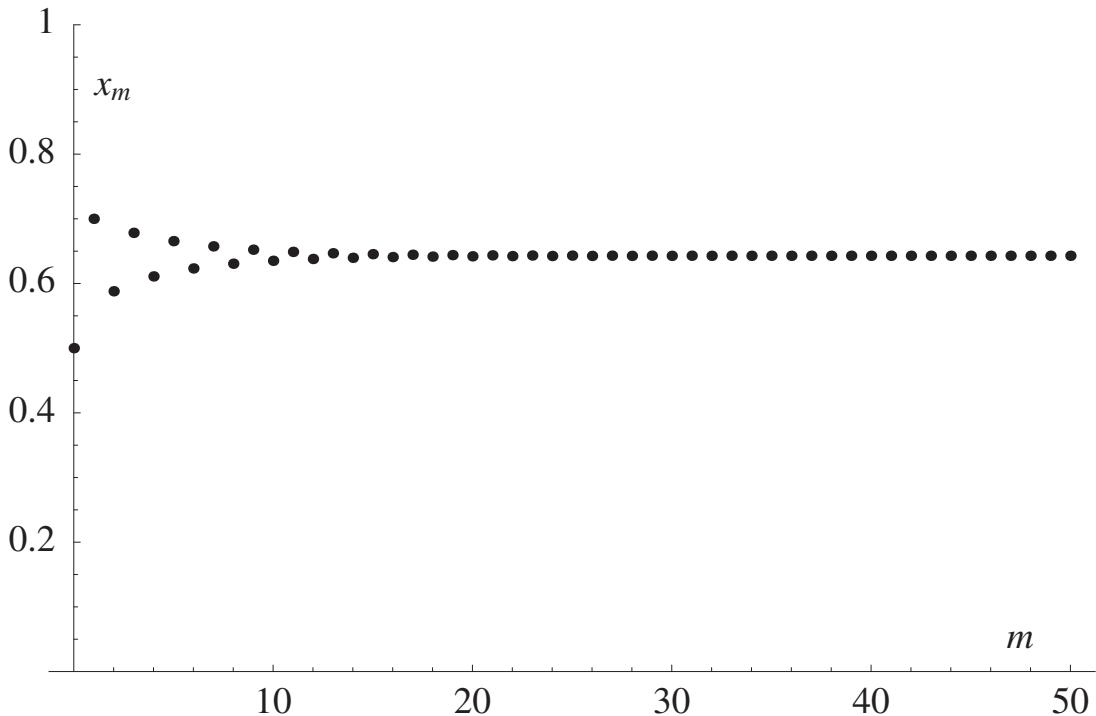


Figure 1.2.2: The values x_m as a function m for the case $\lambda = 2.8$.

Figure 2.4 shows, as a function of λ , the limiting values, called x_∞ , that occur as $m \rightarrow \infty$. Such a graphic is often called a *final-state* or *Feigenbaum diagram*.²⁰ The calculations for this graphic were again made using $x_0 = 1/2$, but other choices in the interval $(0,1)$ would have given the same result. We see that x_∞ is unique for $1 < \lambda < 3$, and can be verified to have the value x_e given by (2.9). That is, this fixed point x_e is *attracting* (stable) for $1 < \lambda < 3$. However, this x_e is *repelling* (unstable) for $\lambda > 3$ and, although it still is a

²⁰Mitchell Feigenbaum (1944-2019) was an American mathematical physicist whose pioneering studies in chaos theory led to the discovery of the universal Feigenbaum constants.

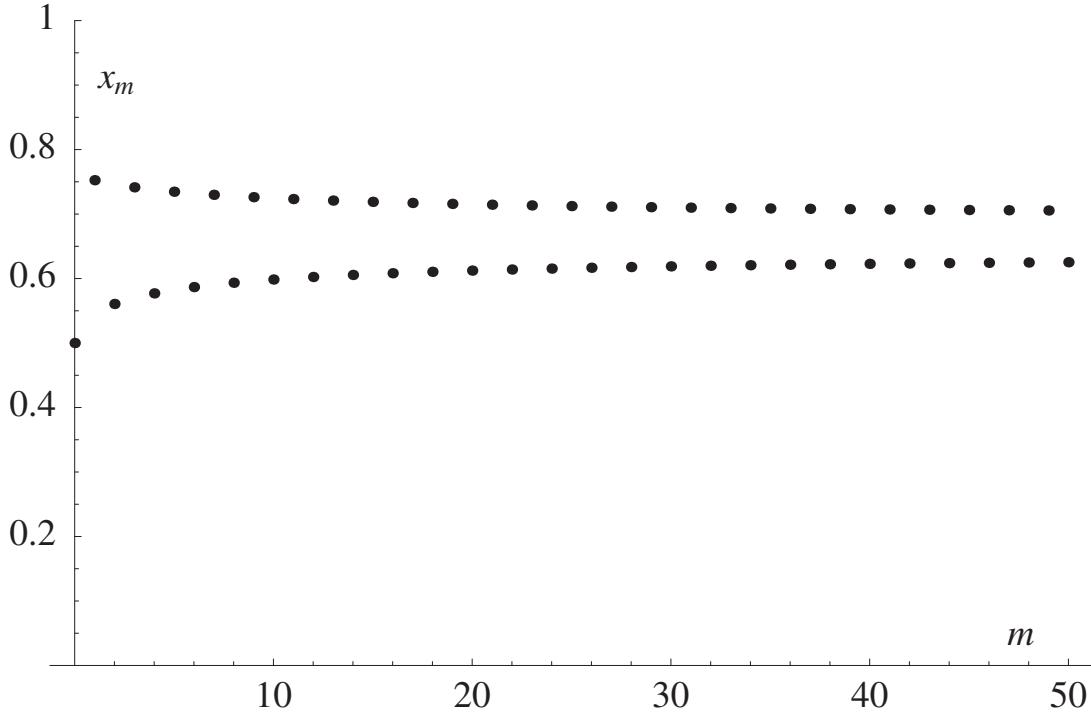


Figure 1.2.3: The values x_m as a function m for the case $\lambda = 3.01$.

fixed point, it no longer appears in the figure for these λ values.²¹ (A fixed point is called a *repeller* if points near it move away under repeated application of \mathcal{M} .) Instead *bifurcation* (period doubling) occurs at $\lambda = 3$ so that, as seen in Figure 2.3, \mathcal{M}^2 has two stable fixed points for λ slightly larger than 3.

Inspection of Figure 2.4 shows that there is a cascade of period doublings as λ increases beyond 3. For example, for λ slightly larger than $3.449\cdots$, there are four fixed points of \mathcal{M}^4 . Application of \mathcal{M} cyclically permutes these points among themselves, and it takes four years for each of these points to repeat itself. Moreover, further inspection shows that an *infinite* number of doublings have occurred by the time λ reaches the *critical* value $\lambda_{\text{cr}} \simeq 3.569$. Let $\lambda_1, \lambda_2, \dots$ denote the λ values at which successive period doublings occur. The first few values are given by the relations

$$\lambda_1 = 3, \quad \lambda_2 = 1 + \sqrt{6} = 3.449\cdots, \quad \lambda_3 = 3.544\cdots. \quad (1.2.13)$$

Let us also write $\lambda_\infty = \lambda_{\text{cr}} \simeq 3.569$. Then it can be shown that (for sufficiently large j) the λ_j converge to λ_∞ as $j \rightarrow \infty$ in the fashion

$$\lambda_j = \lambda_\infty + \gamma \delta^{-j} + \text{higher-order terms}, \quad (1.2.14)$$

²¹Sometimes Feigenbaum diagrams are called *bifurcation* diagrams. However, strictly speaking, bifurcation diagrams should also display the unstable fixed points, and Feigenbaum diagrams generally do not. The use of the term *bifurcation* in the context of Dynamics is due to Poincaré.

with

$$\begin{aligned}\lambda_\infty &= 3.569 \dots, \\ \gamma &= -2.66 \dots, \\ \delta &= 4.6692016 \dots.\end{aligned}\tag{1.2.15}$$

The values of λ_∞ and γ are specific to the logistic map. However, the quantity δ , called the *Feigenbaum constant*, is *universal*. Examination of a graph of the right side of (2.5) shows that the logistic map is produced by a function with one hump (an inverted parabola in this case), and the second derivative of the function does not vanish at the top of the hump. It can be shown that all maps with this property undergo an infinite cascade of period doublings as some appropriate parameter is varied, and there is a relation of the form (2.14) with the *same* (Feigenbaum's) value of δ . [Strictly speaking, what is required is that the *Schwarzian* derivative of the function be negative. If f is any function, its Schwarzian derivative, denoted by Sf , is defined by the rule

$$Sf = \frac{f'''}{f'} - \frac{3}{2} \left(\frac{f''}{f'} \right)^2.\tag{1.2.16}$$

The condition $Sf < 0$ is true for the logistic map, for example, since in this case $f''' = 0$.]

Many systems in nature exhibit a cascade of period doublings, and it is often found experimentally that these cascades behave according to (2.14), again with Feigenbaum's value. See, for illustration, the case of the Duffing equation treated in Chapter 23. Finally, we remark that there are maps for which the Feigenbaum period-doubling cascade begins as some parameter is varied, but does not complete. Rather, as the parameter is further increased after some finite number of period doublings have occurred, the cascade undoes itself. See Appendix J. There are also systems of physical interest that exhibit this kind of behavior. Again see Chapter 23.

Yet more can be said. Figure 2.5 shows an enlargement of the bifurcation cascade for the logistic map. Suppose d is the distance between two forks just as they themselves are about to bifurcate ($d = 0.409 \dots$ for the first fork in the logistic map, see Exercise 2.2). Then (to ever better approximation the farther one proceeds in the cascade), the distances between the next two forks when they are about to bifurcate are d/α and d/α^2 where

$$\alpha = 2.5029\,07875 \dots\tag{1.2.17}$$

Moreover, there is an explicit splitting rule for determining which distance will be d/α and which will be d/α^2 . For example, consider the *upper* fork after the first bifurcation, and let d^U be the distance between the two new forks produced when this fork bifurcates. Then, see Figure 2.5, one has the relation $d^U \approx d/\alpha^2$. Similarly, let d^L be the corresponding distance when the *lower* fork bifurcates. Then one has the relation $d^L \approx d/\alpha$. Next, let d^{UL} be the distance for the lower fork of the preceding upper fork. Then one has the relation $d^{UL} \approx d^U/\alpha$, etc. Again consult Figure 2.5.

The splitting rule and the scaling factor α are also universal for all one-hump maps (with negative Schwarzian derivative), and α is sometimes called the second Feigenbaum constant.

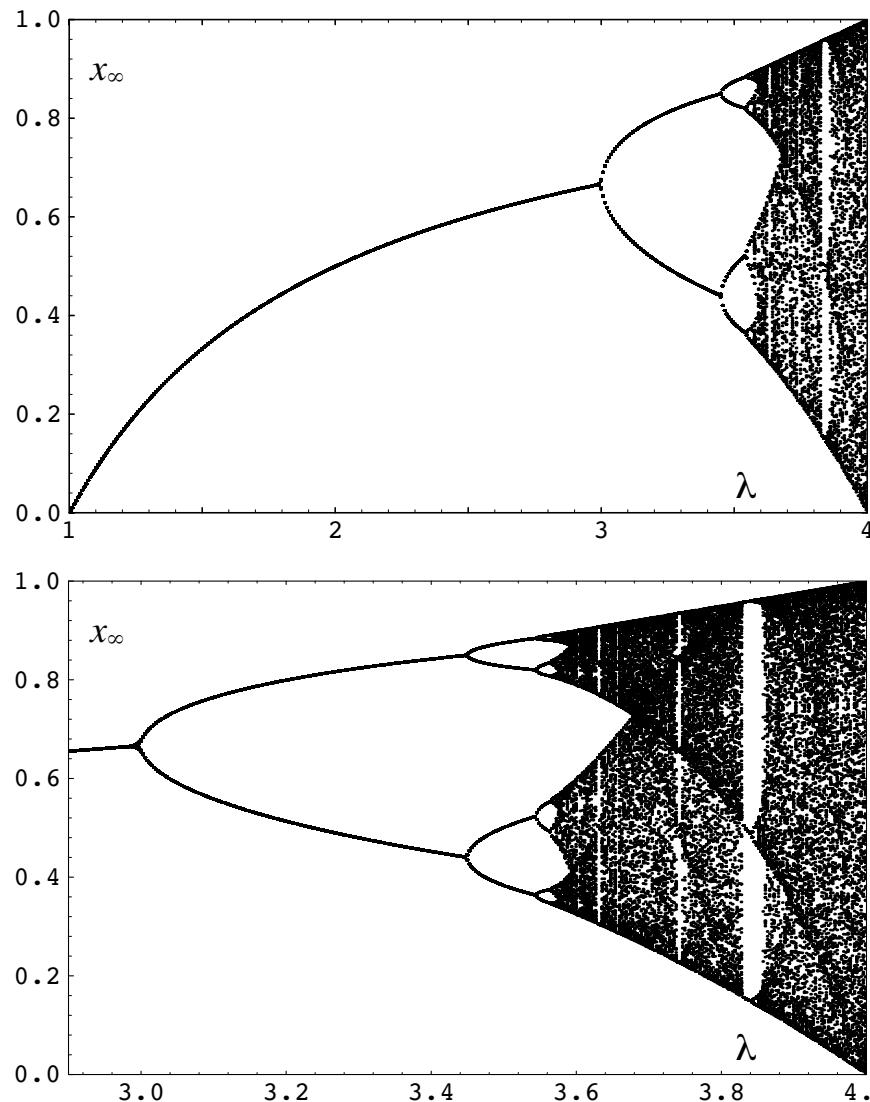


Figure 1.2.4: Feigenbaum diagram showing limiting values x_∞ as a function of λ for the logistic map.

How does this universality arise? Feigenbaum found an explanation that involves a study of certain maps acting on *function* space. The explanation is deep, and we will only be able to sketch part of it. Inspired by the observation of scaling, let \mathcal{R} be a map that acts on functions $\psi(x)$ according to the rule

$$\mathcal{R} : \psi \rightarrow \bar{\psi} \quad (1.2.18)$$

with

$$\bar{\psi}(x) = -a\psi(\psi(-x/a)). \quad (1.2.19)$$

In words, \mathcal{R} scales the argument x , lets ψ act twice on this scaled argument, and then rescales the result. Operations of this kind occur elsewhere in physics, and are called *renormalization*. It can be shown that the map \mathcal{R} has a “fixed point” in the space of *analytic* functions *if and only if* a has the Feigenbaum scaling value α ,

$$a = \alpha, \quad (1.2.20)$$

and this fixed point (function) is unique up to a normalization. Specifically, for $a = \alpha$, there is a unique analytic function $g(x)$ such that

$$g(x) = -ag(g(-x/a)) \quad (1.2.21)$$

provided g is normalized so that

$$g(0) = 1; \quad (1.2.22)$$

and there is no analytic function satisfying (2.21) for $a \neq \alpha$. Indeed, it can be shown that g has a convergent Taylor expansion of the form

$$g(x) = 1 - (1.5276329 \dots)x^2 + (0.1048151 \dots)x^4 + (0.0267056 \dots)x^6 - (0.0035274 \dots)x^8 + \dots \quad (1.2.23)$$

We have been informed that the second Feigenbaum constant α is a property of \mathcal{R} . We will next learn that the first Feigenbaum constant δ is also a property of \mathcal{R} . Let \mathcal{L} be the linear part of \mathcal{R} about the fixed point g . It is defined by the relation

$$\mathcal{R}[g(x) + \epsilon h(x)] = g(x) + \epsilon \mathcal{L}[h(x)] + O(\epsilon^2) \quad (1.2.24)$$

for ϵ small and h any function. It follows from (2.19) and (2.24) that \mathcal{L} is given explicitly by the relation

$$\mathcal{L}[h(x)] = -\alpha h(g(-x/\alpha)) - \alpha[g'(g(-x/\alpha))]h(-x/\alpha). \quad (1.2.25)$$

It can be shown that \mathcal{L} , which evidently and as expected is a linear operator, has eigenfunctions and eigenvalues. Moreover, there is an eigenfunction, call it h_δ , that has the Feigenbaum constant δ as its eigenvalue,

$$\mathcal{L}h_\delta = \delta h_\delta. \quad (1.2.26)$$

All other eigenvalues of \mathcal{L} (there are an infinite number of them) lie inside the unit circle of the complex plane. Thus, \mathcal{L} has a unique eigenvalue that lies outside the unit circle, and

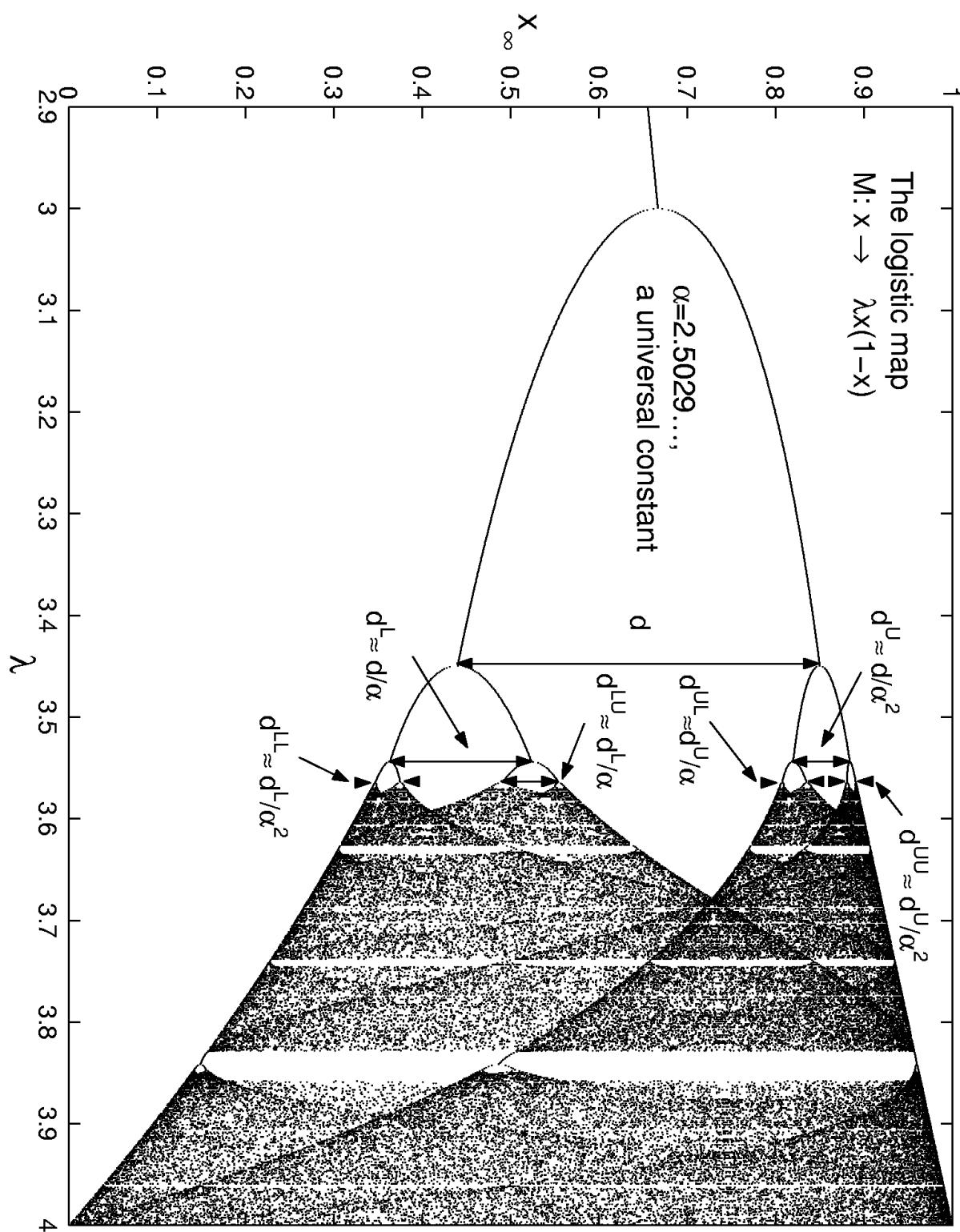


Figure 1.2.5: An enlargement of Figure 2.4 exhibiting how successive bifurcations scale.

this eigenvalue is δ . (Note that $\delta > 1$.) Put another way, in language that will become clearer later, \mathcal{L} has *one* repelling “direction” (eigenfunction) in function space associated with the eigenvalue δ and *all* other directions are attracting.

We have learned that both α and δ are properties of \mathcal{R} , and have told the part of the story that is easy to relate, if not to prove. What remains to be shown is that there is a connection between the set of maps that exhibit infinite period doubling cascades as some parameter is varied and the operator \mathcal{R} . For example, if $f(\lambda, x)$ is a function that produces any such map by the rule

$$\bar{x} = f(\lambda, x), \quad (1.2.27)$$

and the parameter λ has the critical value λ_∞ for which an infinite period doubling cascade has just occurred, then it can be shown that (with $a = \alpha$)

$$\lim_{n \rightarrow \infty} \mathcal{R}^n[f(\lambda_\infty, x)] = g(x). \quad (1.2.28)$$

For the whole story, the reader is referred to the references at the end of this chapter.

Let us, having made this pleasant detour through function space, return to a further discussion of the logistic map. We have sketched the behavior of \mathcal{M} as λ approaches λ_{cr} . For λ slightly beyond λ_{cr} the set of x_∞ points is infinite, and the action of \mathcal{M} on these points is chaotic. Then, remarkably, as λ is increased still further, there are occasional *windows of stability* again followed by period doublings and subsequent chaotic regimes. For example, there is a period-three window (a regime having three values for x_∞) beginning at $\lambda = 1 + \sqrt{8} = 3.828\dots$. Note that, by construction, only *stable* periodic orbits are displayed in Figures 2.4 and 2.5. Thus, as mentioned earlier, the x_e given by (2.9) no longer is shown for $\lambda > 3$. It can be demonstrated that, while there are only a finite number of stable periodic orbits in the windows of stability (as Figures 2.4 and 2.5 indicate), there are an infinite number of unstable periodic orbits. (By the way, all this behavior is also universal.)

1.2.2 Complex Logistic Map and the Mandelbrot Set

According to Paul Painlevé (1863-1933) and popularized by Jacques Hadamard (1865-1963),

The shortest path between two truths in the real domain passes through the complex domain.

In a similar vein, Gaston Julia (1893-1978) frequently instructed the students in his class, one of whom was Benoit Mandelbrot (1924-2010),

To simplify, you should ‘complexify’. That is, when you have a complicated problem and wish to simplify it, it is a good idea to replace all reals by complex numbers.

For example, the behavior of power series is understood more simply using complex variables rather than real variables.²²

²²Caspar Wessel (1745-1818), a Danish-Norwegian mathematician and cartographer, was the first person to describe (in 1797) complex numbers as points in a plane (which we call the *complex plane*) and their addition in terms of vector addition, the parallelogram rule “head to tail” composition of line segments in the complex plane. Indeed, some authors (e.g. Roger Penrose) refer to the complex plane, sometimes without explanation, as the Wessel plane. The same results were independently rediscovered by Jean-Robert Argand in 1806 and Carl Friedrich Gauss in 1831.

With this lesson in mind, and following Mandelbrot, suppose we extend both x and λ in (2.5) to complex values. Then the map \mathcal{M} takes the form

$$z_{n+1} = \mathcal{M}z_n = f(\gamma, z_n) = \gamma z_n(1 - z_n) \quad (1.2.29)$$

where z is the complex extension of x , and γ is the complex extension of λ . (See Exercise 2.5.) Associated with the map (2.29) are two complex planes. One of these, the z plane, will be called the *mapping* plane since the map sends this plane into itself. The other, the γ plane, will be called the *control* plane.

The nature of what happens in the mapping plane under repeated iteration depends sensitively on where γ is in the control plane. For example, Figure 2.6 shows the nature of the map for $\gamma = 2.55268 - 0.959456i$. Points in the black area of the mapping plane remain there indefinitely under repeated application of (2.29). By contrast, any point launched in the white area eventually iterates away to infinity. (We may view the point $z = \infty$ as an attractor for \mathcal{M} . See Exercise 2.6.) In Accelerator Physics language, we would call the black area the *dynamic aperture*. (Mathematicians call it the *filled Julia* set.²³) It can be shown that the boundary of the dynamic aperture (the Julia set) is fractal. Remarkably, it is nevertheless possible to name in a precise way every point on the boundary.

If γ is changed, the dynamic aperture also is changed. Figure 2.6 shows what is called *Douady's rabbit*; for some other values of γ the dynamic aperture disintegrates into a cloud of isolated points called *Fatou* dust.²⁴ Since the nature of what happens under repeated iteration in the mapping plane depends sensitively on the location of γ in the control plane, we may turn the matter around. That is, we may characterize points in the γ plane by the behavior (under repeated iteration) of points in the mapping plane. Suppose we consider those points M in the control plane for which the dynamic aperture in the mapping plane is a *connected* set. This set M in the control plane is called the *Mandelbrot* set.²⁵ It is shown in Figure 2.7.

There is another definition of the Mandelbrot set that is more computationally tractable, and which can be shown to be equivalent to that just given. The function $f(\gamma, z)$ has a *critical* point (a point where $\partial f / \partial z = 0$) at $z = 1/2$. Now consider the points $\mathcal{M}^n(1/2)$. They form the orbit of $(1/2)$ under the action of \mathcal{M} . If, for a particular value of γ , this orbit goes to infinity, then γ is *not* in the Mandelbrot set M . If the orbit of $(1/2)$ does *not* go to infinity for a particular value of γ , then this value of γ *is* in the Mandelbrot set. Technically, we say that $(1/2)$ is in the basin of attraction for the attractor $z = \infty$ if its orbit goes to infinity. Thus, γ is in the Mandelbrot set if $(1/2)$ is not in the basin of $z = \infty$; and γ is not in the Mandelbrot set if $(1/2)$ is in the basin of $z = \infty$. Finally, we remark that it is not necessary to follow an orbit to infinity by iterating infinitely often. See Exercise 2.6 to learn that a point z is in the basin of infinity, i.e. will go to infinity under infinite iteration, if z lies outside the disk specified by $|z| = 1 + 1/|\gamma|$. See (2.109). Therefore, if any point on the

²³Gaston Julia (1893-1978) and Pierre Fatou (1878-1929) began the study of complex dynamics during the early 20th century.

²⁴Adrien Douady (1935-2006) made significant contributions to the fields of analytic geometry and dynamical systems.

²⁵Elsewhere in this book the symbol M will commonly be used to denote the linear part of a map \mathcal{M} . But here it is used to honor Mandelbrot.

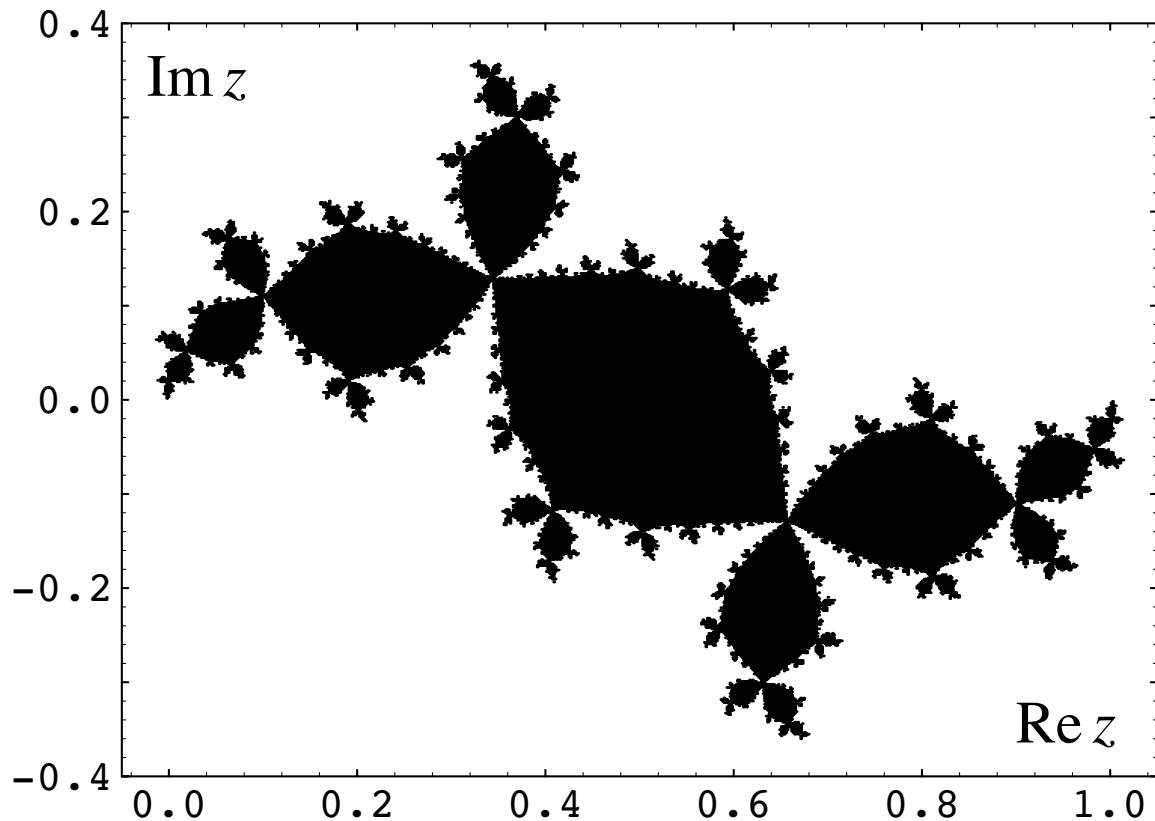


Figure 1.2.6: Douady's rabbit, the dynamic aperture in the mapping plane z for the case $\gamma = 2.55268 - 0.959456i$.

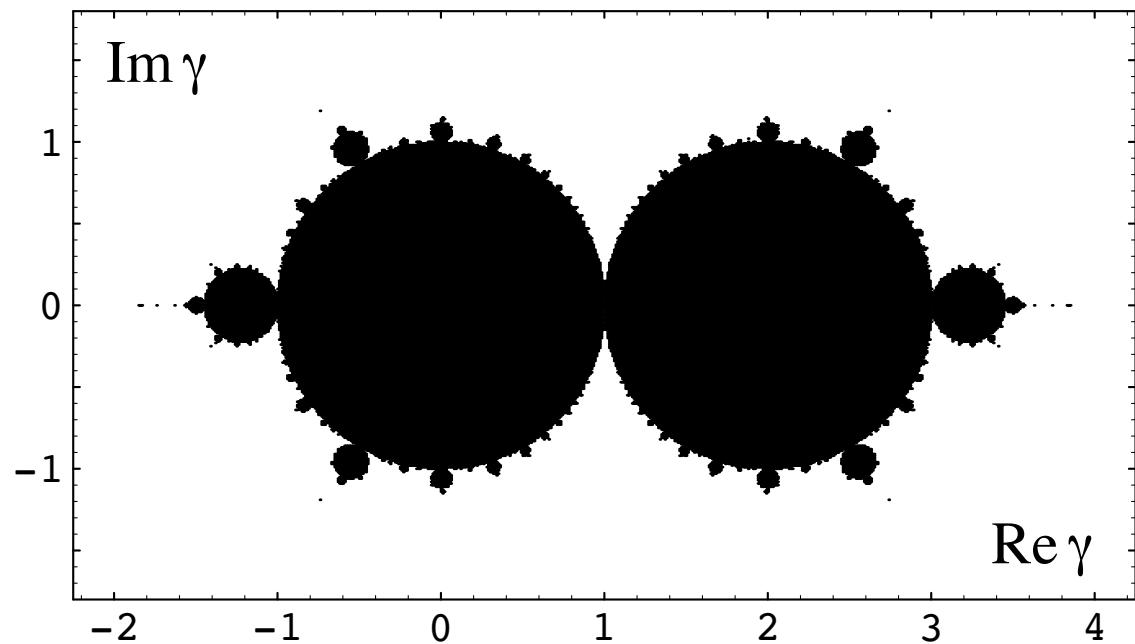


Figure 1.2.7: The Mandelbrot set M in the control plane γ .

orbit of $(1/2)$ falls outside this disk, it is not necessary to iterate further to determine the ultimate fate of points on the orbit.

When viewed from a distance, the Mandelbrot set M appears to be a mainland consisting of two back-to-back discs with sprouts. The discs are tangent at the point $\gamma = (1, 0)$, and M has reflection symmetry about both the lines $\text{Re } \gamma = 1$ and $\text{Im } \gamma = 0$. Closer examination reveals the presence of what appear to be very small islands around the mainland. (In fact these islands, when suitably magnified, resemble the mainland, and the whole structure of the Mandelbrot set is fractal.) Since γ is the complexification of λ , one can see that λ values in the range $(1, \lambda_{\text{cr}})$ correspond to *real* γ values lying in the right disc and its sprouts and its subsprouts. In addition, it can be shown that λ values for the windows of stability seen in Figure 2.4 correspond to real γ values lying in small islands on the real γ axis to the right of the mainland. Finally, contrary to superficial appearances, it can be shown that the Mandelbrot set is *connected* (and, indeed, *simply connected*). There are thin filaments/tendrils, too small to be seen in Figure 2.7, that connect the visible apparent islands to the mainland. Thus, there is really only a mainland (and this mainland has no holes)!

Consider the value of γ for Douady's rabbit. It lies in the sprout located at the five-o'clock position of the right disc in Figure 2.7. For this value of γ the complexified version of (2.8) and (2.9) yields for \mathcal{M} the fixed points $z_f = 0$ and $z_f = .656747 - .129015i$. These fixed points are both repellers. Also, there is a fixed point at ∞ , and it is attracting. See Exercises 2.6 and 2.11. See also Exercise 5.5 of Chapter 22 where the machinery is developed to deal with the nature of fixed points in 2-dimensional maps.

Moreover, it can be shown that for this γ value the map (2.29) has three *attracting* complex period-three fixed points. Indeed, Douady's rabbit turns out to be the basins of attraction for these fixed points. The three attracting fixed points of \mathcal{M}^3 have the locations

$$z^1 = 0.499997032420304 - (1.221880225696050E-006)i \quad (\text{red}), \quad (1.2.30)$$

$$z^2 = 0.638169999974373 - (0.239864000011495)i \quad (\text{green}), \quad (1.2.31)$$

$$z^3 = 0.799901291393262 - (0.107547238170383)i \quad (\text{yellow}). \quad (1.2.32)$$

The action of \mathcal{M} on these fixed points is given by the relations

$$\mathcal{M}z^1 = z^2, \quad (1.2.33)$$

$$\mathcal{M}z^2 = z^3, \quad (1.2.34)$$

$$\mathcal{M}z^3 = z^1. \quad (1.2.35)$$

Figure 2.8 shows Douady's rabbit again, this time in color. The red, green, and yellow points lie in the basins $B(z^1)$, $B(z^2)$, and $B(z^3)$ of \mathcal{M}^3 , respectively. The white points lie in the basin $B(\infty)$ of \mathcal{M} . Corresponding to the relations (2.33) through (2.35) there are the results

$$\mathcal{M}B(z^1) = B(z^2) \quad \text{or} \quad \mathcal{M} \text{ red} \subseteq \text{green}, \quad (1.2.36)$$

$$\mathcal{M}B(z^2) = B(z^3) \quad \text{or} \quad \mathcal{M} \text{ green} \subseteq \text{yellow}, \quad (1.2.37)$$

$$\mathcal{M}B(z^3) = B(z^1) \quad \text{or} \quad \mathcal{M} \text{ yellow} \subseteq \text{red}. \quad (1.2.38)$$

Note the marvelous fractal structure at the basin boundaries.

In addition to the attracting fixed points of \mathcal{M}^3 , there exist another three *repelling* complex period-three fixed points that lie on the boundary of the rabbit. Now continuously vary the value of γ until it enters the island for the period-three window, and eventually takes on a real value corresponding to a λ value lying in the period-three window of Figure 2.4. As γ varies, the period-three fixed points move. They may change their nature, (*e.g.* they all become repellers when γ leaves the sprout), but they *cannot* disappear. (See, for example, Exercise 2.2.) It can be shown that in this case, as γ changes from the Douady-rabbit value in the sprout to a real value in the period-three window, all the associated period-three fixed points of Douady's rabbit move from their original complex values to the real line. Furthermore, the three period-three fixed points that begin as repellers when γ lies in the sprout become the three attractors x_∞ when γ reaches the island. The other three period-three fixed points, which begin as attractors when γ lies in the sprout and become repellers when γ leaves the sprout, remain repellers when γ reaches the island. Thus, by extending the logistic map to the complex domain, we have learned that seemingly isolated phenomena are in fact related.

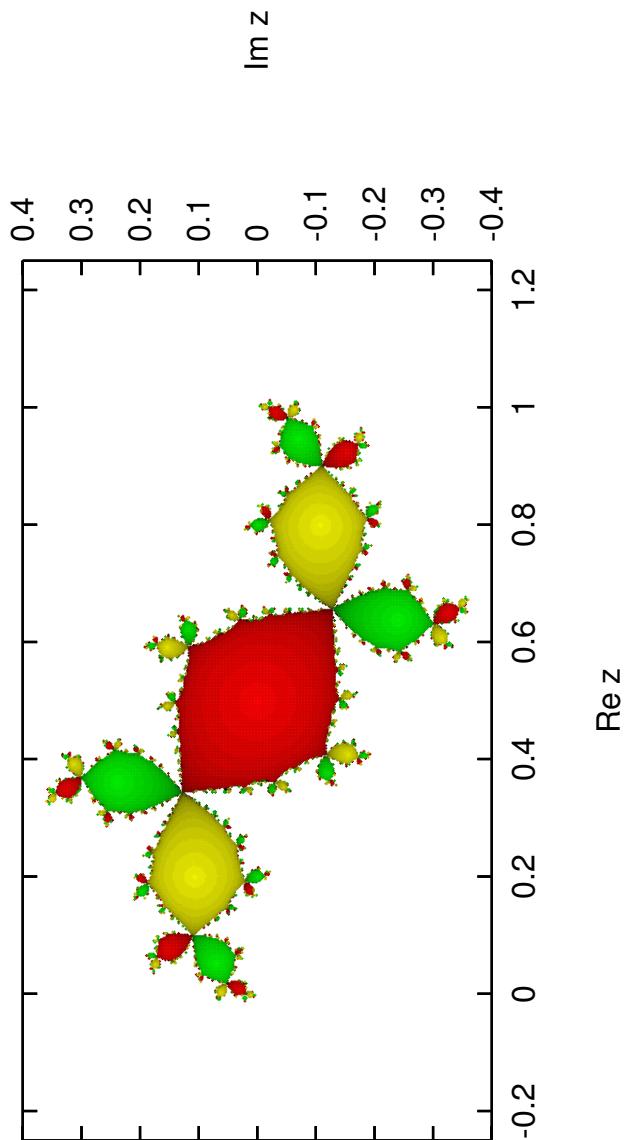


Figure 1.2.8: Douady's rabbit in color. The white points lie in the basin of ∞ under the action of \mathcal{M} . The origin is a repelling fixed point of \mathcal{M} . The other repelling fixed point has the location $z_f = .656747 - .129015i$. Under the action of \mathcal{M}^3 , red points lie in the basin of z^1 , green points lie in the basin of z^2 , and yellow points lie in the basin of z^3 .

1.2.3 Simplest Nonlinear Symplectic Map

The complex logistic map (2.29) may be viewed as the simplest nonlinear two-dimensional *analytic* map. In the same spirit, the simplest nonlinear two-dimensional *symplectic* map is the *Hénon* map. It, too, is a quadratic map. We take the opportunity here to describe it in Lie algebraic terms.²⁶ To do this, we will need to introduce some Lie algebraic tools. These tools will be described briefly below. Their full exposition is given in subsequent chapters.

We begin by redefining the symbol z ; it will now stand for a canonically conjugate pair of position and momentum variables q and p ,

$$z = (q, p). \quad (1.2.39)$$

Next, let $f(z)$ denote any function of q, p . We will associate with each such function a *differential* operator, called a *Lie operator* and denoted by the symbol $:f:$, by making the definition

$$:f: \stackrel{\text{def}}{=} (\partial f / \partial q)(\partial / \partial p) - (\partial f / \partial p)(\partial / \partial q). \quad (1.2.40)$$

Then if g is any other function of the phase-space variables z , we have the result

$$:f:g = (\partial f / \partial q)(\partial g / \partial p) - (\partial f / \partial p)(\partial g / \partial q) = [f, g], \quad (1.2.41)$$

where $[*, *]$ denotes the familiar Poisson bracket. (See Section 1.7.) Powers of $:f:$ are defined by repeated application of (2.40) or (2.41),

$$\begin{aligned} :f:^2 g &= [f, [f, g]], \\ :f:^3 g &= [f, [f, [f, g]]], \text{ etc.} \end{aligned} \quad (1.2.42)$$

Finally, we define $:f:^0$ to be the identity operator,

$$:f:^0 = \mathcal{I} \Leftrightarrow :f:^0 g = g. \quad (1.2.43)$$

Now that powers of Lie operators have been defined, we can also define power series. Of particular interest is the power series for the exponential function,

$$\exp(:f:) = \sum_{k=0}^{\infty} :f:^k / k!. \quad (1.2.44)$$

This object is referred to as a *Lie transformation*, and $:f:$ (or f) is called its *generator*. Specifically, if g is any function, we have the result/action

$$\exp(:f:)g = g + [f, g] + [f, [f, g]]/2! + \dots \quad (1.2.45)$$

With regard to its action on the phase-space coordinates q and p , it can be shown that any Lie transformation produces a symplectic map. See Section 7.1.

²⁶Michel Hénon (1931-2013), a French mathematician and astronomer, invented this map to model a Poincaré map. Although not originally intended for the purpose of Accelerator Physics, it can be described and applied in that context.

At this point the reader should verify the results

$$\exp(:q^3:)q = q, \quad (1.2.46)$$

$$\exp(:q^3:)p = p + 3q^2. \quad (1.2.47)$$

In Accelerator Physics terminology, the Lie transformation $\exp(:q^3:)$ produces the phase-space mapping associated with a *thin sextupole kick*. See Section 13.10.

Similarly, the reader should verify the results

$$\exp(-(\phi/2) :p^2 + q^2:)q = q \cos \phi + p \sin \phi, \quad (1.2.48)$$

$$\exp(-(\phi/2) :p^2 + q^2:)p = -q \sin \phi + p \cos \phi. \quad (1.2.49)$$

This verification requires the summation of an infinite series. In Accelerator Physics terminology, the Lie transformation $\exp[-(\phi/2) :p^2 + q^2:]$ produces the phase-space mapping for a *simple phase advance* (rotation in phase space) of amount ϕ .

With this background in mind, let us consider the map \mathcal{M} given by the product

$$\mathcal{M}(\theta) = \exp(-(\theta/4) :p^2 + q^2:) \exp(:q^3:) \exp(-(\theta/4) :p^2 + q^2:). \quad (1.2.50)$$

The map consists of a $\theta/2$ phase advance, followed by a sextupole kick, followed again by a $\theta/2$ phase advance. Figure 2.9 illustrates this map schematically. In Accelerator Physics terminology, it may be viewed as describing horizontal betatron motion in an idealized storage ring with a single thin *sextupole* insertion S , and an *observation* point O (Poincaré surface of section) located diametrically across the ring from the sextupole insertion. As seen from (2.46) through (2.49), the map (2.50) does indeed consist of linear and quadratic terms, as advertised. Since Lie transformations produce symplectic maps when acting on phase-space coordinates, and symplectic maps form a group, it follows that (2.50) is a symplectic map. Finally, it can be shown that this map is a variant of the usual Hénon map, and differs from it only by a linear change of variables. See Chapter 29 for a study of general quadratic maps in two dimensions. We also remark that, unlike the logistics map (real or complex), the Hénon map, like all symplectic maps, is invertible.

The Hénon map has been studied in detail. As simple as it appears, it is known to have very complicated properties: these include homoclinic points, chaotic behavior, and period bifurcations. Figure 2.10 shows the dynamic aperture for our variant of the Hénon map for the case $\theta/2\pi = 0.22$. Points in the black area of the q, p (mapping) plane remain there under repeated application of the map. [Actually, the points shown remain there for at least 10,000 iterations (\mathcal{M}^n with $n \leq 10,000$).]²⁷ By contrast, any point launched in the white area eventually iterates away to infinity. Inspection of the figure suggests, and it can in fact be proved, that the dynamic aperture for our variant of the Hénon map is symmetrical about the q axis.

Figure 2.11 illustrates how the size and shape of the dynamic aperture for our variant of the Hénon map depend on the total phase advance θ . As is evident from examination

²⁷We remark that the dynamic aperture is not known for the Hénon map, or any other nontrivial symplectic map for that matter, when $n = \infty$. See Section 20.10.

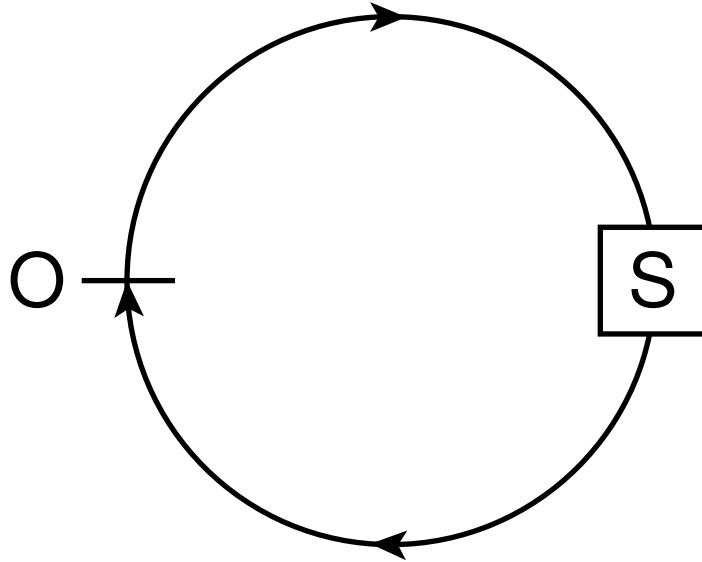


Figure 1.2.9: Schematic representation of the map (2.50).

of (2.46) through (2.50) and Figure 2.11, the dynamic aperture shrinks to the phase-space origin as θ goes to zero. By contrast, when $\theta = \pi$, one has the results

$$\mathcal{M}(\pi)q = -q + 3p^2, \quad (1.2.51)$$

$$\mathcal{M}(\pi)p = -p, \quad (1.2.52)$$

$$\mathcal{M}^2(\pi) = \mathcal{I}. \quad (1.2.53)$$

Correspondingly, the dynamic aperture in this case is *all* of phase space. For general θ it can be shown that the dynamic aperture for the map $\mathcal{M}(-\theta)$ is the same as that for the map $\mathcal{M}(\theta)$ save for a 180° rotation about the phase-space origin. Moreover, the dynamic aperture for the map $\mathcal{M}(\pi + \phi)$ is the same as that for $\mathcal{M}(\pi - \phi)$. Finally, the dynamic aperture for $\mathcal{M}(\theta)$ is periodic in θ with period 4π . It follows that the information presented in the figure is sufficient to deduce the dynamic aperture for all (real) values of θ .

The study of phenomena arising from the iteration of symplectic maps is still in its infancy, and much remains to be done in even the very simplest of cases. For example, in analogy with what has been learned in the case of the logistic map, one might wonder if further insight could be gained by complexifying the Hénon map, i.e. by making both q and p complex. Then (2.50) would become a mapping of \mathbb{C}^2 (the space of two complex variables) into itself. Also, the control parameter θ could be made complex. By such a study one might hope, for example, to better understand the boundary of the dynamic aperture. *Hubbard* and *Oberste-Vorth* have begun this exploration, and results to date indicate that the complex Hénon map is a remarkably complicated object. This should be a sobering thought to accelerator physicists, because they are interested in knowing the behavior of far more complicated symplectic maps in more (four and six) dimensions. When complexified, four- and six-dimensional phase spaces become \mathbb{C}^4 and \mathbb{C}^6 . Thus it is no wonder that questions of dynamic aperture for realistic accelerators are so complicated. Nor, in analogy to the properties of the Mandelbrot set, should we be surprised that the

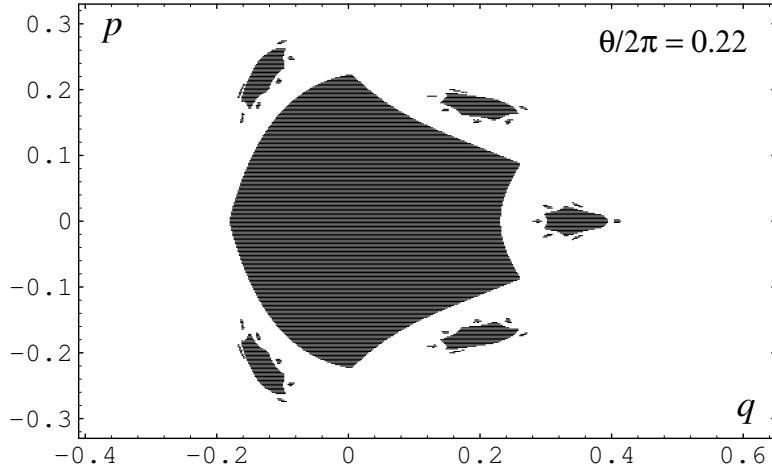


Figure 1.2.10: The dynamic aperture of the Hénon map for the case $\theta/2\pi = 0.22$.

dynamic aperture depends sensitively on the choice of accelerator parameters such as tunes, local phase advances, multipole strengths, etc. What we are observing in all these instances is that complicated properties can arise as a result of an infinite process, namely that of indefinite iteration.

1.2.4 Goal for Use of Maps in Accelerator Physics

In some areas of nonlinear dynamics, e.g. celestial/galactic dynamics, the Hamiltonian is dictated by Nature and the goal is to understand/predict the dynamics arising from this Hamiltonian. In Accelerator Physics, the Hamiltonian can, more or less, be engineered; and the goal is to engineer the Hamiltonian in such a way that particles will be accelerated, stored, and directed to achieve various desired ends. In particular, in the context of Accelerator Physics, the long-term goal of map methods is to be able to describe, predict, and control nonlinear properties with the same facility with which we now handle linear properties. Much has been accomplished in this direction, particularly with regard to single-pass systems and short-to-moderate-term behavior in circulating systems.

It is known that once-differentiable symplectic maps (and probably even analytic symplectic maps) *generically* have simultaneously hyperbolic fixed points, elliptic fixed points, and homoclinic points that are all *everywhere dense* in phase space. (The meaning of the terms *hyperbolic*, *elliptic*, and *homoclinic* will be defined subsequently.) Consequently, the detailed long-term behavior of most symplectic maps under repeated iteration must be complicated beyond comprehension.²⁸ However, there is still the hope that it may be possible to

²⁸Thus, the properties of the Hénon map are vastly more complicated than those of the already very complicated complex logistic map. For example, apart from the behavior of the Julia set which is sent into itself in a complicated way, the behavior at most points in the mapping plane for the complex logistic map is governed by a few attractors. By contrast, we will see in Chapter 3 that symplectic maps have no attractors (and also no repellers). Therefore the orbits produced by a symplectic map never settle down, and something new should always be expected. But this “newness” would not be surprising to the vast intelligence described by Laplace. It is surprising only to those who have not done enough computation to see how results depend on the initial conditions and the number of iterations.

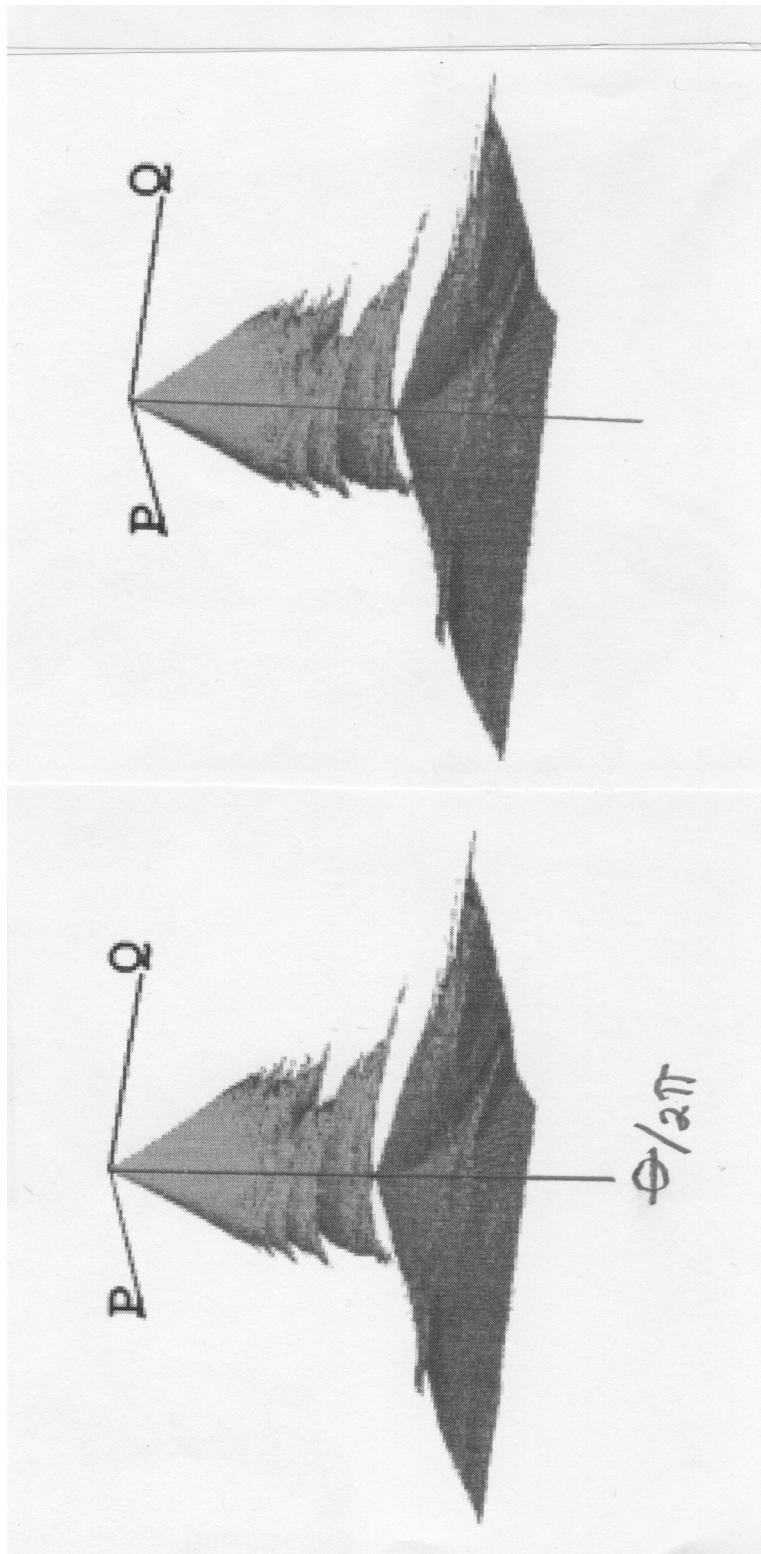


Figure 1.2.11: Stereographic view of the dynamic aperture of the Hénon map as a function of the parameter θ . The region shown is $q \in [-.8, .8]$, $p \in [-.7.7, .7.7]$, $\theta/2\pi \in [0, .5]$.

compute gross long-term properties: the rough size of the dynamic aperture, approximate (but useful) lower bounds on the life time for some sizable fraction of a circulating beam, etc.

We now know that generically Hamiltonian motion is *chaotic* in the sense that final conditions (in the long-term) generally depend very sensitively on initial conditions. (And, we know that final conditions can also depend very sensitively on parameter values.)²⁹ This possibility was already envisioned by Maxwell, and subsequently by Poincaré. In 1873 Maxwell wrote:

When the state of things is such that an infinitely small variation of the present state will alter only by an infinitely small quantity the state at some future time, the condition of the system ⋯ is said to be stable; but when an infinitely small variation in the present state may bring about a finite difference in the state of the system in a finite time, the condition of the system is said to be unstable. It is manifest that the existence of unstable conditions renders impossible the prediction of future events, if our knowledge of the present state is only approximate, and not accurate ⋯ It is a metaphysical doctrine that from the same antecedents follow the same consequences. No one can gainsay this.³⁰ But it is not of much use in a world like this, in which the same antecedents never occur, and nothing ever happens twice.

Strictly speaking, if continuity holds as we know it does for solutions of differential equations under quite general circumstances, Maxwell was not correct in the assertion that infinitesimal changes in initial conditions could produce (in finite time) a finite change in final conditions. But his ideas were correct in spirit. In 1903, in the same spirit and with more precision, Poincaré wrote:

If we knew exactly the laws of nature and the situation of the universe at some initial moment, we could predict exactly the situation of that same universe at a succeeding moment. But even if it were the case that the natural laws had no longer any secret for us, we could still only know the initial situation approximately. If that enabled us to predict the succeeding situation with the *same approximation*, that is all we require, and we should say that the phenomenon had been predicted, that it is governed by laws. But it is not always so: it may happen that small differences in the initial conditions produce very great ones in the final phenomena. A small error in the former will produce an enormous error in the latter. Prediction then becomes impossible, and we have the fortuitous phenomenon.

²⁹The word *chaotic* can have a variety of meanings. The least stringent is sensitive dependence on initial conditions. A more stringent definition is to require in addition that for a map \mathcal{M} to exhibit chaotic behavior in some domain \mathcal{D} it must be *transitive* in \mathcal{D} in the sense that if \mathcal{E} and \mathcal{F} are any two subdomains in \mathcal{D} , then there is some point in \mathcal{E} such that applying \mathcal{M} enough times to this point yields some point in \mathcal{F} . Finally, we require that the set of periodic points of \mathcal{M} and its powers be dense in some subdomain \mathcal{G} of \mathcal{D} . According to Exercise 2.9 the logistic map is chaotic, following this more stringent definition, when $\lambda = 4$ and for some $\lambda < 4$.

³⁰Note that quantum mechanics does gainsay this.

One of the goals of accelerator design is to minimize chaotic behavior and its effects, and to minimize sensitive dependence on parameter values.³¹

For the most part we will restrict our attention to single-particle dynamics. To the extent that multiparticle dynamics is considered, we will generally assume that interactions between individual particles can be neglected, or that we are interested only in single-particle dynamics occurring in the presence of an already specified multiparticle background. That is, we will not attempt a *self-consistent* treatment of many-particle effects such as wake fields, space-charge forces, and strong-strong beam-beam interactions. As Newton already realized, the self-consistent inclusion of even relatively *few*-particle effects raises a whole new set of complications:

The orbit of any one planet depends on the combined motion of all the planets, not to mention the actions of all these on each other. To consider simultaneously all these causes of motion and to define these motions by exact laws allowing of convenient calculation exceeds, unless I am mistaken, the forces of the entire human intellect.

In the case of the solar system, the “forces that exceed those of the entire human intellect” have recently been provided by special-purpose super computers running special-purpose integration algorithms (based, as it turns out, on map methods). And, by following orbits for sufficiently long times, it has been found that solar-system dynamics is chaotic.

Routine detailed treatment of long-term *many*-particle effects in the context of Accelerator Physics awaits the advent of readily accessible super computers routinely operating at or exceeding petaFLOPS speed.

1.2.5 Some Highlights of the *N*-Body Gravitational Problem

Under the assumption of an inverse square gravitational force law for point masses, Newton was able to show that the gravitational forces between rigid extended (macroscopic) spherical distributions of point masses (assuming they do not collide) are the same as if they were point masses with the mass of each distribution (body) concentrated at its center. Next, Newton was able to show for two point masses that, under their mutual gravitation, their center of mass would move with constant velocity, and the motion of each about their center of mass would be an ellipse (more generally a conic section). This conclusion of Newton [about what is now called the Kepler (1571-1630) problem] had to be extracted from him by Edmond Halley (of cometary fame) after Robert Hooke (of spring-force law fame) had failed to deliver on a promised proof that an inverse square force law led to Kepler’s laws of planetary motion. When subsequently asked by Halley, Newton claimed that he had proved it four years earlier, but then, because he had apparently lost his notes, was able to produce

³¹In the context of chaotic behavior, “sensitive dependence on initial conditions” is now generally taken to mean that, to achieve a given accuracy in the final conditions after a given time or, in the case of maps, a given number of map iterations, the required accuracy in the initial conditions ultimately grows *exponentially* in time or the number of map iterations. Since parameter values may also be viewed as dynamical variables and therefore as initial conditions, see Section 10.12, the same is also possibly true of parameter values. See Exercise 2.9 for an example of how sensitive dependence on initial conditions can occur in the case of the logistic map.

a new and improved proof only after three months delay. Although Newton had invented the basics of calculus, his actual armamentarium of mathematical concepts and tools was quite limited by modern standards. After this, and at the urging of Halley, Christopher Wren (of architectural fame), and others, he began to work in earnest on writing his *Principia*. When completed, it was edited by Halley and published at Halley's expense.

We remark in passing, as already understood by Kepler, that specification of an orbit by its *shape*, e.g. specification of $r(\theta)$ where r and θ are polar coordinates, is only part of the problem. For an elliptical orbit there is, according to his first law, the orbit relation

$$r(\theta) = a(1 - \epsilon^2)/(1 + \epsilon \cos \theta) \quad (1.2.54)$$

where a is the semi-major axis and ϵ is the eccentricity. What are ultimately desired as well are the functions $r(t)$ and $\theta(t)$ where t is the time. In the treatment of conic sections from ancient times it was common to define a quantity ψ , called the *eccentric anomaly*, by the relation

$$r = a(1 - \epsilon \cos \psi). \quad (1.2.55)$$

It can be verified that eliminating the quantity r between (2.54) and (2.55) yields the result

$$\tan(\theta/2) = [(1 + \epsilon)/(1 - \epsilon)]^{1/2} \tan(\psi/2), \quad (1.2.56)$$

a relation between ψ and θ also known to the ancestors. In terms of the quantities just discussed, and beyond specifying orbit shape, the further insight Kepler brought to the problem is that he deduced the relation

$$\omega t = \psi - \epsilon \sin \psi, \quad (1.2.57)$$

which is commonly known as *Kepler's equation*. Here

$$\omega = 2\pi/T \quad (1.2.58)$$

where T is the period of the elliptical orbit. It can be shown that Kepler's equation arises from his second law, his observation that "equal areas are swept out in equal times". Finally, according to Kepler's third law, the period T is given by

$$T = 2\pi a^{3/2}/[G(M + m)]^{1/2} \approx 2\pi a^{3/2}/[GM]^{1/2} \quad (1.2.59)$$

where G is the gravitational constant, M is the mass of the sun/star, and m is the mass of the planet.

How are we to use Kepler's equation? Once the quantities ωt and ϵ have been specified, the relation (2.57) is to be solved for ψ . Note that (2.57) is *nonlinear*, and hence must be solved graphically, or numerically, or by some other means. Once ψ is known, $r(t)$ and $\theta(t)$ are given by (2.55) and (2.56). For more detail, see for example the book of H. Goldstein listed in the Bibliography at the end of this chapter under the heading "Classical/Celestial/Galactic..."

This was the state of affairs with regard to Kepler's equation until the advent of Friedrich Wilhelm Bessel (1784-1846). Among many things, he developed the theory of Bessel functions and showed, as an application, that the solution to Kepler's equation is given by the formula

$$\psi = \omega t + \sum_{n=1}^{\infty} (2/n) J_n(n\epsilon) \sin(n\omega t). \quad (1.2.60)$$

Here it is assumed, as is the case, that Bessel functions can be computed to high accuracy with relative ease and that the series (2.60) is rapidly convergent. Note that (2.60), remarkably, states that ψ is given as ωt plus a *Fourier* series in $\sin(n\omega t)$ with essentially Bessel function coefficients. A further surprise is that one of the earliest applications of Bessel functions was in the area of Astronomy rather than the areas of vibrations of circular drum heads or electromagnetic fields in cylindrical wave guides, etc., with which they are more commonly associated.³²

Let us return to the main discussion. In the approximation that all the planets have very small masses compared to that of the sun, and with neglect of mutual interactions among the planets, the orbits of all the planets would be ellipses. Correspondingly, in this approximation, the solar system would be *stable* for *all* time. But what happens if the very small mass approximation is not made for the planets and if mutual planetary interactions are included? This so-called *gravitational N-body problem* is difficult for two reasons: First, the consideration of arbitrarily long times introduces an infinity into the problem. Second, the idealization of treating extended macroscopic bodies as point masses means that bodies can become arbitrarily close with their associated gravitational potential energies possibly supplying an unbounded amount of kinetic energy to other bodies that could lead to their ejection from the system. Note that gravitational forces are always *attractive*. (In fact, in some cases even relatively close encounters might provide enough kinetic energy for the ejection of others.) Thus, the singular nature of the idealized $1/r^2$ gravitational force introduces additional possible infinities.

As indicated by the quotation above, Newton apparently viewed the *N*-body gravitational problem as humanly intractable. Nevertheless he attempted to estimate the effects of mutual interactions and concluded that they would rapidly become noticeable and detrimental to stability. Since he believed that the solar system had and should continue to exhibit regular motion for a long period of time (based on his Biblical studies, to which he devoted more time than to physics, he believed that the world would last at least until 2060), he concluded that *divine intervention/reformation* was required from time to time to correct the effect of these mutual interactions:

. . . By the help of these principles, all material things seem to have been composed of the hard and solid particles above-mentioned, variously associated in the first creation by the counsel of an intelligent agent: for it became Him who created them to set them in order. And if He did so, it is unphilosophical to seek for any other origin of the world, or to pretend that it might arise out of chaos by the mere laws of Nature; though, being once formed, it may continue by these laws for many ages. For while comets move in very eccentric orbs in all manner of positions, blind fate could never make all the planets move one and the same way in orbs concentric, some inconsiderable irregularities excepted, which may have arisen from the mutual actions of comets and planets on one another, and which will be apt to increase, till this system wants a reformation. Such a wonderful uniformity in the planetary system must be allowed the effect of choice; . . .

³²Note that Fourier's dates are (1768-1830), and therefore Bessel and he were near contemporaries.

With regard to the solar system itself and God, Newton (in the General Scholium that appears as an appendix to the second edition of the Principia) wrote:

This most beautiful system of the sun, planets, and comets, could only proceed from the counsel and dominion of an intelligent and powerful Being. And if the fixed Stars are the centers of other like systems, these, being form'd by the like wise counsel, must be all subject to the dominion of One; especially since the light of the fixed Stars is of the same nature with the light of the Sun, and from every system light passes into all the other systems. And lest the systems of the fixed Stars should, by their gravity, fall into each other mutually, he hath placed these Systems at immense distances from one another. . . . This Being governs all things, not as the soul of the world, but as Lord over all; and on account of his dominion he is wont to be called Lord God *παντοκράτωρ*, or Universal Ruler. . . . He is eternal and infinite, omnipotent and omniscient; that is, his duration reaches from eternity to eternity; his presence from infinity to infinity; he governs all things, and knows all things that are or can be done.

In the terminology of the philosophy of religion or natural theology, Newton's invoking divine action to "reform" (adjust) from time to time the solar system is an early example of *God of the gaps*: When something is not understood or a theory appears to fail, direct action by God is invoked as an explanation. For further discussion of Newton's Biblical and historical studies see the book of Jed Z. Buchwald and Mordechai Feingold and the book of Rob Iliffe cited in the Bibliography at the end of this chapter.

Kepler's discoveries of elliptical planetary orbits also posed unanswered questions. Like his contemporaries he initially believed, based on philosophical grounds dating back to Greek/Platonic ideas, that circular motion was the most perfect of all motions, and therefore the planets might naturally be expected to move in circular orbits. What then is the explanation for the small transverse deviations from circular motion associated with elliptical motion? Rather than invoking supernatural agents or unphysical powers, he eventually came to the physical hypothesis that both the underlying circular motions and the deviations from it arose from magnetic effects associated with a rotating sun. Of course, with Newton's discovery that the orbits associated with an inverse-square force law must be conic sections, the need for further explanation vanished, and one need not think that circular motion is the most perfect of all motions.

We digress to note that Kepler made other scientific/mathematical contributions in addition to his laws of planetary motion. He discovered that the eye has a lens, and that the action of this lens forms an (inverted) image on the back of the eye. He also studied sphere packing, and calculated the packing fraction for a particular configuration that has since been conjectured to be optimal (the so called *Kepler conjecture*). In 2014 Thomas Hales, leader of the Flyspeck Team, announced that this conjecture was finally proved. The proof involved 300 pages of text, about 3 gigabytes of computer programs and data, and about 5000 processor-hours.

With much improved mathematical tools and a century later Laplace, in his books *Exposition du Système du Monde* and *Mécanique Céleste*, claimed to show that the effects of mutual interactions of the planets and the sun essentially average to zero over large times,

and therefore no “reformation” is required. He also studied solar system formation mechanisms for which the planets would be expected to orbit in essentially the same plane and in the same direction. Laplace’s claims might actually have pleased Newton because Newton also maintained that, “No more causes of natural things should be admitted than are both true, and sufficient to explain their phenomena.”

There is a story, perhaps apocryphal/embellished, to the effect that Napoleon met Laplace and said, “I understand you have written a large book on the system of the universe and have not mentioned its creator.” To this comment Laplace replied, “I had no need of that hypothesis.” Napoleon, greatly amused by this response, later related this interchange to Lagrange. Lagrange reportedly replied, “Ah, it is a beautiful hypothesis; it explains many things.” Subsequent versions of the Laplace-Napoleon event claim that Laplace was not denying the existence of God or his ability to intervene should he so desire, but only denying that it was necessary for God to intervene from time to time to set the planets back on a regular course. [In *Exposition du Système du Monde*, Laplace quotes Newton’s assertion that “This most beautiful system of the sun, planets, and comets, could only proceed from the counsel and dominion of an intelligent and powerful Being.” This, says Laplace, is a “thought in which he (Newton) would be even more confirmed, if he had known what we have shown, namely that the conditions of the arrangement of the planets and their satellites are precisely those which ensure its stability”. Laplace originally trained for the priesthood before taking up mathematics, and received last rites at his death. But there are also indications that Laplace was very skeptical about the occurrence of miracles in general and transubstantiation in particular.] In effect, there was no gap that needed special filling.

Leibniz had thought, from the beginning and on philosophical grounds, that Newton’s view was ill conceived because surely God could and therefore necessarily would create a universe that did not constantly require maintenance. In fact, Leibniz held, this world (universe) is the *best* of all possible worlds: “In whatever manner God created the world, it would always have been regular and in a certain general order. God, however, has chosen the most perfect, that is to say, the one which is at the same time the simplest in hypothesis and the richest in phenomena.”

It is now known that Laplace’s stability calculations are inconclusive for long-term stability (although presumably satisfactory to show stability through the year 2060) because in his perturbative method he neglected some important high-order terms. Moreover, he did not consider the possibility, now known to be generically likely, that the perturbative series he was generating would ultimately be divergent and therefore useless for determining stability.

Early in his career Poincaré also crossed swords with the N -body gravitational problem in the form of determining stability in the restricted 3-body approximation. His work won the King Oscar II of Sweden prize. But when it came time for publication a year later, Poincaré found he had made a major error, stopped the presses, paid for the printing costs himself, and wrote a corrected manuscript that was published yet a year later. See the book by June Barrow-Green cited in the Bibliography at the end of this chapter. The question of 3-body stability and solar-system stability remained unresolved.

It is now believed possible that one or more planetary ejections from the solar system may have indeed occurred in the distant past. (Numerical and analytical studies of the gravitational N -body problem indicate that there are indeed solutions for which one or

more bodies escape to infinity. Moreover, numerical simulations of stellar globular clusters indicate that they routinely “boil off” individual stars.) Thus, in its early history, the solar system may have been unstable. However, long-term numerical integrations indicate that the solar system we now observe should survive far into the future. (It takes approximately 50 million years, when integrating forward or backward in time, for the uncertainties in orbital positions to grow to substantial values due to chaotic sensitivity to initial condition and parameter value uncertainties.) Of course, such calculations do not rule out collisions with small unknown objects such as asteroids. But, as locally damaging as such collisions might be to various planets and moons, they would not seriously perturb the solar system as a whole. Google *solar system stability* or *stability of the solar system*. Look for, among others, the Web sites <https://www.ias.edu/about/publications/ias-letter/articles/2011-summer/solar-system-tremaine>, and http://www.scholarpedia.org/article/Stability_of_the_solar_system. See also the link <https://arxiv.org/abs/2302.06641> for a discussion of what happens if another earth-like planet is “inserted” in the solar system. Essentially, the solar system is full as it stands. Finally, see the book of Dumas on The KAM Story cited in the Classical/Celestial . . . section of the bibliography at the end of this chapter.

We close this section with a brief discussion of what is known about the effect of the $1/r^2$ singularity on orbits in the gravitational N -body problem. First we remark that even in the two-body problem and in the case of elliptic orbits so that all body coordinates and velocities/momenta are well defined for all *real* time, the “virtual possibility” of a two-body collision (thus bringing the $1/r^2$ singularity into evidence) appears in the form of singularities in the *complex* time plane. If the orbits are highly elliptic/eccentric so that very close encounters are possible, these complex singularities lie very close to the real time axis thereby making numerical integration very difficult near times of close encounters. This problem is generally treated by *regularization* of the equations of motion prior to numerical integration. See Subsection 2.7.4 and the regularization references at the end of Chapter 2 for a discussion of regularization. When regularized, which involves a change of independent variable, the Kepler problem becomes the harmonic oscillator problem, and solutions for the harmonic oscillator have no singularities in the complex plane of its independent variable save at infinity.

The equations of motion in the case of two-body collisions can be solved exactly. Collisions occur after a finite time, and at the collision time, call it t^* , the coordinates \mathbf{r}_1 and \mathbf{r}_2 of the colliding particles are equal and the associated momenta are *infinite*. Since the momenta are infinite when $t = t^*$, we may say that there is a singularity in the t plane at the point t^* . However, it can be shown that for the gravitational N -body problem the regularized equations can be uniquely integrated through a two-body collision. Thus if at worst only two-body collisions occur in an attempt to integrate trajectories in the gravitational N -body problem, trajectories can be extended arbitrarily far forward in time.

What about 3-body collisions? These are solutions of the gravitational N -body problem for which at time t^* the coordinates \mathbf{r}_1 , \mathbf{r}_2 , and \mathbf{r}_3 of three colliding particles are equal and the associated momenta are *infinite*. It can be shown that there is no satisfactory way of extending such solutions beyond $t = t^*$. Thus in this case the $1/r^2$ singularity has a definitive/disastrous effect. However, it can be shown from the equations of motion that such 3-body collisions *cannot* occur if the total angular momentum of the 3-body system is nonzero. Moreover, if there are only three bodies so that we are dealing with

the gravitational 3-body problem, then the 3-body angular momentum is the total angular momentum, and we know that the total angular momentum is conserved for the gravitational 3-body problem. Therefore, for the gravitational 3-body problem, all trajectories can be extended arbitrarily far forward in time except those whose initial conditions are such that the total angular momentum is zero.

Painlevé (already referred to earlier in the context of complexification) was a distinguished French Mathematician as well as a distinguished French Statesman.³³ One of his primary interests in Mathematics was to understand the singularity properties of the solutions of ordinary differential equations. Singularities are characterized as being of two types, those (called *not movable*) whose location in the complex time plane is independent of the initial conditions, and those (called *movable*) whose location depends on the initial conditions.³⁴ Painlevé and others were able to classify all second-order nonlinear ordinary differential equations whose singularities satisfy what is called the *Painlevé property*. This is the property that the singularities are either not movable or are movable and are *poles*. Painlevé and others were able to show that there are 50 types of second-order nonlinear differential equations whose singularities satisfy the Painlevé property. Moreover, the equations that belong to 44 of these types have the feature that they can be solved in terms of previously known functions. There are 6 remaining types called Painlevé I through Painlevé VI. Differential equations of these types can be solved in terms of previously unknown (but now more or less studied) functions called the Painlevé transients. (Curiously, all the differential equations of Painlevé types I through VI arise from Hamiltonians.)³⁵

Presumably because of his general interest in the singularity properties of the solutions of ordinary differential equations, Painlevé also studied the singularity properties of the solutions to the gravitational N -body problem. Evidently for the gravitational N -body problem all singularities are moveable. Let $\mathbf{r}_{jk} = \mathbf{r}_j - \mathbf{r}_k$ be the vector separation of particles j and k . The right sides of the relevant differential equations of motion are analytic when, for $j \neq k$, $|\mathbf{r}_{jk}| \neq 0$. As the equations of motion are integrated forward in time, no singularity can occur as long as all bodies remain a finite distance apart. (See Theorem 3.3 that you will eventually encounter in Section 3.) And the times at which single or simultaneous collisions might occur can clearly be adjusted by varying the initial coordinates and momenta prior to the expected single or simultaneous collision times.

We have described simultaneous collisions of two and three bodies, and it is easy to envision N -particle generalizations of such collisions in the N -body case. All these collisions lead to singularities since all the momenta \mathbf{p}_j must become infinite at the moment of collision. Painlevé called these singularities *collisional* singularities. In the course of t increasing to the value $t = t^*$ it is assumed that the functions $\mathbf{r}_j(t)$ are continuous and take on the limiting values $\mathbf{r}_j(t^*)$ in such a way that at least some $\mathbf{r}_{jk}(t^*) = 0$ and the $\mathbf{r}_{jk}(t) \neq 0$ for $j \neq k$ and $t < t^*$. Correspondingly, by energy conservation, there will be the limiting values $\lim_{t \rightarrow t^*} \mathbf{p}_j(t) = \infty$. Painlevé was able to show that for $N = 2$ and $N = 3$ all singularities are

³³He served twice as Prime Minister of France, and is buried in the French Panthéon. For a description of his career, see https://en.wikipedia.org/wiki/Paul_Painlevé. For a description of the French Panthéon, inspired in part by the ancient Roman Pantheon, and a list of those who have received the honor of burial there, see the Web site <https://en.wikipedia.org/wiki/Panth%C3%A9on>

³⁴See https://en.wikipedia.org/wiki/Movable_singularity.

³⁵See https://en.wikipedia.org/wiki/Painlevé_transcendents.

collisional.

However, Painlevé also conjectured that when $N \geq 4$ there are also what he called *noncollisional* singularities. His conjecture has now been proved. Noncollisional singularities correspond to exotic “solutions” to the gravitational N -body problem for which one or more bodies escapes to *infinity* in *finite* time with *infinite* velocity. Moreover, *prior* to that finite time t^* , there are *no* collisions. Thus the motion is called “noncollisional”. Such “solutions/scenarios”, constructed with great ingenuity, exploit the singular nature at $r = 0$ of the $1/r^2$ idealized model for the gravitational force in that they require arbitrarily close encounters and thereby entail arbitrarily large forces. But in so doing they violate the “finiteness” conditions to be presented in Theorem 3.1. For example they do not occur if $1/r^2$ in the gravitational force law is replaced by $1/(r^2 + a^2)$ for any nonzero but arbitrarily small value of a , for then there are no infinite forces and the conditions of Theorem 3.1 are met. These exotic “solutions” are sometimes cited as evidence for instances in which *determinism* in classical mechanics is violated. This is a misunderstanding. Their true nature is that they are instances where singularities arising from idealizations are allowed to play a hidden but nonetheless decisive role. They have no deep philosophical significance. They are, however, of great mathematical interest because they clarify/prove some long-open conjectures (e.g. the Painlevé conjecture) about the nature of singular “solutions” in the gravitational N -body problem. Moreover, they have heuristic value, for they suggest that there may be nearby true solutions for which no infinities arise (forces remain bounded) but for which large (but finite) excursions may occur. Thus, for example, there are instances in which it is possible to employ relatively close encounters to achieve deep-space satellite missions with a minimum expenditure of fuel.

See the Web site https://en.wikipedia.org/wiki/Painleve_conjecture for a qualitative description of the Painlevé conjecture in a case for which $N = 5$. The illustration shown there suggests how the complicated required motion might come about in a case of 5 bodies. It presents the 5-body positions and velocities at any given time, and from this illustration one is supposed to infer how the 5 bodies might be expected to move as $t \rightarrow t^*$. More detail is provided at <https://www.jstor.org/stable/2946572>. See the Web site <https://arxiv.org/abs/1409.0048> for a detailed description with proofs for a case where $N = 4$. See also the references at the end of this chapter listed under the heading “Solar System Stability; Singularities in the Newtonian N -Body Problem”.

1.2.6 Maps from Hamiltonian Differential Equations

There is one last set of motivational remarks to be made. Often, as already described and to be illustrated subsequently in Section 1.4 and later, we are interested in maps produced by integrating differential equations. In the case that these differential equations arise from a *time-independent* Hamiltonian H , the associated map $\mathcal{M}(t^f, t^i)$ that takes initial conditions q^i, p^i at time t^i to final conditions q^f, p^f at time t^f can formally be written as the Lie transformation

$$\mathcal{M}(t^i, t^f) = \exp(-(t^f - t^i) : H(q^i, p^i) :). \quad (1.2.61)$$

This result is proved in Section 7.4. How to capitalize on this result, and what to do in the time-dependent case, are discussed in subsequent sections. There are related results for

non-Hamiltonian differential equations. One can then work with exponentials of what are called non-Hamiltonian vector fields.

Exercises

1.2.1. The purpose of this exercise is to examine the stability of the fixed points x_e given by (2.8) and (2.9). Re-express the logistic map (2.5) by using the notation

$$\bar{x} = \mathcal{M}x = f(\lambda, x) = \lambda x(1 - x). \quad (1.2.62)$$

Introduce *deviation* variables δ and $\bar{\delta}$ about the fixed point x_e by the relations

$$x = x_e + \delta, \quad \bar{x} = x_e + \bar{\delta}. \quad (1.2.63)$$

Show that in terms of these deviation variables the logistic map (2.44) takes the form

$$\bar{\delta} = \mu\delta - \lambda\delta^2 \quad (1.2.64)$$

where

$$\mu = \lambda(1 - 2x_e). \quad (1.2.65)$$

The first term on the right side of (2.57) is called the *linear* part of \mathcal{M} about x_e , and μ is called the eigenvalue of the linear part. Evidently, unless $\mu = 0$, the behavior of (2.57) under repeated iteration, and for δ sufficiently small, is governed by the linear part, which in turn is described by μ . That is, we may neglect the δ^2 term in (2.57). Show that if $|\mu| < 1$, then x_e is stable; and if $|\mu| > 1$, then x_e is unstable. In particular, suppose (2.57) is rewritten in the form

$$\delta_{n+1} = \mu\delta_n - \lambda(\delta_n)^2 \quad (1.2.66)$$

and assume $|\mu| < 1$ but $\mu \neq 0$. Show that, for sufficiently small δ_0 , (2.59) yields the asymptotic behavior

$$\delta_n \simeq \mu^n \delta_0. \quad (1.2.67)$$

Show that if x_e is given by (2.8), then μ is given by the relation

$$\mu = \lambda. \quad (1.2.68)$$

Show that if x_e is given by (2.9), then μ is given by the relation

$$\mu = 2 - \lambda. \quad (1.2.69)$$

For $\lambda \in (0, 1)$, verify that the fixed point given by (2.8) is stable, and that given by (2.9) is unstable. Show that their stability roles are reversed for $\lambda \in (1, 3)$. Show that when $\lambda = 1$, $\mu = 1$ for both values of x_e , and show that the two fixed points then also coincide. Show that the x_e given by (2.9) is especially attractive when $\lambda = 2$. You will have to retain the δ^2 terms in (2.57) because now $\mu = 0$. In particular, show that (2.59) now yields the asymptotic behavior

$$\delta_n \simeq -(1/\lambda)(-\lambda\delta_0)^{2^n}. \quad (1.2.70)$$

When $\mu = 0$, the associated fixed point x_e is called *super attractive* or *super stable*. For $\lambda > 2$ show that μ as given by (2.62) is negative, $\mu < 0$. Use this fact to explain the behavior of the x_m in Figure 2.2. Show that μ as given by (2.62) has the value $\mu = -1$ when $\lambda = 3$, and that the fixed point given by (2.9) is unstable for $\lambda > 3$. Verify from Figures 2.4 and 2.5 that period doubling occurs when $\lambda = 3$. See also Exercise 2.2. That is, period doubling for a fixed point occurs when the associated value of μ passes through the value $\mu = -1$.

1.2.2. For $\lambda \geq 3$ the maps \mathcal{M} and hence \mathcal{M}^2 continue to have the x_e given by (2.8) and (2.9) as fixed points. Show that, for $\lambda > 3$, \mathcal{M}^2 also has the two additional fixed points ${}^2x_e^\pm$ given by

$${}^2x_e^\pm = [(\lambda + 1)/(2\lambda)] \pm [(\lambda - 3)(\lambda + 1)]^{1/2}/(2\lambda), \quad (1.2.71)$$

and that these points are mapped into each other under the action of \mathcal{M} . (In point of fact, \mathcal{M}^2 also has these fixed points for $\lambda \leq 3$, but then they are complex. For an analytic map fixed points cannot be created or destroyed.) Verify that x_e as given by (2.9) and ${}^2x_e^\pm$ agree when $\lambda = 3$. Verify also that

$$\partial({}^2x_e^\pm)/\partial\lambda = \pm\infty \text{ at } \lambda = 3. \quad (1.2.72)$$

Thus the curves ${}^2x_e^\pm(\lambda)$ have infinite slope at $\lambda = 3$. See Figure 2.5. Finally, verify that

$$d = ({}^2x_e^+ - {}^2x_e^-)|_{\lambda=3.449\dots} = 0.409\dots. \quad (1.2.73)$$

Again see Figure 2.5.

1.2.3. It has already been mentioned, and in Section 1.4 we will see in more detail, that differential equations produce maps. Moreover, in Chapter 10 and Section 24.12 we will learn how to compute these maps, how to find their fixed points, and how to expand them in deviation variables (see Exercise 2.1). Suppose a map has been expanded up to some order in deviation variables about a fixed point. Can this expansion be used to predict period doubling and other bifurcation phenomena? If so, to what order must the map be expanded to make such predictions? The purpose of this exercise is to explore these questions for the simplest case of one-dimensional maps.

Let \mathcal{M} be a one-dimensional map and suppose [in analogy to (2.57)] that it has an expansion, in deviation variables about a fixed point, of the form

$$\bar{\delta} = a\delta + b\delta^2 + c\delta^3 + d\delta^4 + e\delta^5 + \dots. \quad (1.2.74)$$

Suppose we employ the notation

$$\bar{\delta} = \mathcal{M}\delta \quad (1.2.75)$$

and

$$\bar{\delta} = \mathcal{M}\bar{\delta}. \quad (1.2.76)$$

Show that \mathcal{M}^2 , the square of \mathcal{M} , then has an expansion about the same fixed point of the form

$$\bar{\bar{\delta}} = \alpha\delta + \beta\delta^2 + \gamma\delta^3 + \sigma\delta^4 + \tau\delta^5 + \dots, \quad (1.2.77)$$

where

$$\alpha = a^2, \quad (1.2.78)$$

$$\beta = ab + a^2b = ab(1 + a), \quad (1.2.79)$$

$$\gamma = 2ab^2 + ac + a^3c, \quad (1.2.80)$$

$$\sigma = b^3 + 2abc + 3a^2bc + ad + a^4d, \quad (1.2.81)$$

$$\tau = 2b^2c + 3ab^2c + 3a^2c^2 + 2abd + 4a^3bd + ae + a^5e. \quad (1.2.82)$$

Evaluate α , β , γ , σ , and τ for the logistic map, see (2.57), and show that in this case the terms beyond order 4 in (2.70) vanish. Now let δ_e be a fixed point of \mathcal{M}^2 . According to (2.70) it must satisfy the equation

$$\delta_e = \alpha\delta_e + \beta\delta_e^2 + \gamma\delta_e^3 + \sigma\delta_e^4 + \tau\delta_e^5 + \dots. \quad (1.2.83)$$

One solution to (2.76), which we already know about because it is also a fixed point of \mathcal{M} , is $\delta_e = 0$. Upon dividing both sides of (2.76) by δ_e , we see that any *nonvanishing* solution must satisfy the relation

$$1 - \alpha = \beta\delta_e + \gamma\delta_e^2 + \sigma\delta_e^3 + \tau\delta_e^4 + \dots. \quad (1.2.84)$$

Show for the logistic map that the terms beyond order 3 in (2.77) vanish. Show in fact that, for the logistic map, (2.77) becomes the relation

$$Q(\delta_e) \stackrel{\text{def}}{=} \delta_e^3 - (2\mu/\lambda)\delta_e^2 + [\mu(\mu+1)/(\lambda^2)]\delta_e - [(\mu^2 - 1)/(\lambda^3)] = 0. \quad (1.2.85)$$

For the logistic map we also know from (2.57) that

$$\delta_e = (\mu - 1)/\lambda \quad (1.2.86)$$

is a second fixed point of \mathcal{M} . Verify this assertion. The quantity δ_e given by (2.79) is therefore also a fixed point of \mathcal{M}^2 , and consequently is also a solution of (2.78). Indeed, verify that

$$P(\delta_e) \stackrel{\text{def}}{=} Q(\delta_e)/[\delta_e - (\mu - 1)/\lambda] = \delta_e^2 - [(1 + \mu)/\lambda]\delta_e + (1 + \mu)/(\lambda^2). \quad (1.2.87)$$

Solve the equation $P(\delta_e) = 0$ and use (2.62) to find the results

$$\delta_e^\pm = (3 - \lambda)/(2\lambda) \pm (1/2\lambda)[(\lambda - 3)(\lambda + 1)]^{1/2}. \quad (1.2.88)$$

Check that these results agree with (2.64).

At this point it is convenient to introduce the quantity ϵ defined by the relation

$$\epsilon = -(\mu + 1). \quad (1.2.89)$$

Evidently ϵ will be small when $\mu \simeq -1$, namely when μ is near the bifurcation value $\mu = -1$. Show that in terms of the quantity ϵ , see (2.62), the relation (2.81) has the expansion

$$\delta_e^\pm = \pm (1/3)(\epsilon)^{1/2} - (1/6)(\epsilon) \mp (5/72)(\epsilon)^{3/2} + (1/18)(\epsilon)^2 + \dots. \quad (1.2.90)$$

For the general one-dimensional map (2.67), we do not have at our disposal a second fixed point besides the first fixed point $\delta_e = 0$. Therefore we cannot solve (2.77) directly by factorization. However, we may still proceed as follows: We see from (2.83) that for the logistic map the δ_e of interest are small when ϵ is small. We might therefore try to solve (2.77) perturbatively under the assumption that in the general case the desired δ_e are small near a bifurcation, and consequently sufficiently high powers of δ_e may be neglected. Suppose we neglect all powers of δ_e in (2.77) beyond the first. Then (2.77) has the tentative solution

$$\delta_e \stackrel{?}{=} (1 - \alpha)/\beta = (1 - a^2)/[ab(1 + a)] = (1 - a)/ab. \quad (1.2.91)$$

Here we have used (2.71) and (2.72). However, since the parameter a in (2.67) plays the role of μ in (2.57), near a bifurcation we expect that $a \simeq -1$. Therefore (2.84) does not produce a solution near 0, and our assumption about being able to neglect terms in (2.77) beyond the first is unjustified. The quantity $(1 - \alpha = 1 - a^2)$ is small, which is expected and desirable, but the quantity $[\beta = ab(1 + a)]$ that multiplies δ_e is also small. Therefore the product $\beta\delta_e$ is not large compared to higher powers of δ_e .

We need to make a careful expansion in small quantities. To do so, in analogy with the case of the logistic map, now define ϵ by the relation

$$\epsilon = -(a + 1). \quad (1.2.92)$$

Presumably the quantities a, b, \dots in (2.67), and correspondingly the quantities α, β, \dots in (2.70), depend analytically on some common parameter, and it is the change in this parameter that causes bifurcation. Without loss of generality, we may replace this parameter with the quantity ϵ using (2.85). The quantities α, β, \dots may then be expanded in terms of ϵ to yield relations of the form

$$\alpha - 1 = a^2 - 1 = 2\epsilon + \epsilon^2, \quad (1.2.93)$$

$$\beta = ab(1 + a) = \beta_1\epsilon + \beta_2\epsilon^2 + \dots, \quad (1.2.94)$$

$$\gamma = \gamma_0 + \gamma_1\epsilon + \dots, \quad (1.2.95)$$

$$\sigma = \sigma_0 + \sigma_1\epsilon + \dots, \quad (1.2.96)$$

$$\tau = \tau_0 + \tau_1\epsilon + \dots, \text{ etc.} \quad (1.2.97)$$

Here we have made explicit use of (2.71) and (2.72).

Now we are ready to proceed. Write (2.77) in the form

$$\delta_e^2 = (1 - \alpha)/\gamma - (\beta/\gamma)\delta_e - (\sigma/\gamma)\delta_e^3 - (\tau/\gamma)\delta_e^4 + \dots. \quad (1.2.98)$$

Suppose now we neglect all powers of δ_e in (2.91) beyond the second. Then (2.91) has the tentative solution

$$\delta_e^{\pm} \stackrel{?}{=} -\beta/(2\gamma) \pm (1/2)[(\beta/\gamma)^2 - 4(\alpha - 1)/\gamma]^{1/2}. \quad (1.2.99)$$

Verify (2.92) and show that inserting (2.86) through (2.88) into it yields the expansion

$$\delta_e^{\pm} \stackrel{?}{=} \pm [2/(-\gamma_0)]^{1/2}(\epsilon)^{1/2} + [\beta_1/(-2\gamma_0)](\epsilon) + \dots. \quad (1.2.100)$$

According to (2.93), δ_e is now of order $(\epsilon)^{1/2}$. Assuming this to be true, let us examine the orders of the various terms on the right side of (2.91): The term $(1 - \alpha)/\gamma$ is of order ϵ . See (2.86) and (2.88). The term $(\beta/\gamma)\delta_e$ is of order $(\epsilon)^{3/2}$. See (2.87) and (2.88). Moreover, the term $(\sigma/\gamma)\delta_e^3$ is also of order $(\epsilon)^{3/2}$. See (2.88) and (2.89). Finally, the terms $(\tau/\gamma)\delta_e^4$ etc. are of order ϵ^2 and higher.

With these estimates in mind, we will now seek to solve (2.91) by iteration. For the zeroth iteration we will first write

$$(\delta_e^{(0)})^2 = (1 - \alpha)/\gamma \quad (1.2.101)$$

with the solution

$$\delta_e^0 = [(1 - \alpha)/\gamma]^{1/2} = \pm [2/(-\gamma_0)]^{1/2}(\epsilon)^{1/2} \pm (1/4)[2/(-\gamma_0)]^{1/2}[1 - 2(\gamma_1/\gamma_0)](\epsilon)^{3/2} + \dots \quad (1.2.102)$$

More simply, for our purposes, it suffices to start with the approximation

$$\delta_e^0 = \pm [2/(-\gamma_0)]^{1/2}(\epsilon)^{1/2}. \quad (1.2.103)$$

For subsequent iterations we will rewrite (2.91) in the form

$$(\delta_e^{(n+1)})^2 = (1 - \alpha)/\gamma - (\beta/\gamma)\delta_e^{(n)} - (\sigma/\gamma)(\delta_e^{(n)})^3 - (\tau/\gamma)(\delta_e^{(n)})^4 + \dots \quad (1.2.104)$$

Verify that carrying out this iterative solution yields the expansion

$$\delta_e^\pm = \pm [2/(-\gamma_0)]^{1/2}(\epsilon)^{1/2} + [(\sigma_0)/(\gamma_0^2) - (\beta_1)/(2\gamma_0)]\epsilon \pm (**)(\epsilon)^{3/2} + \dots \quad (1.2.105)$$

As a sanity check on this procedure, verify that (2.98) yields (2.83) for the case of the logistic map.

We conclude that finding the leading behavior of δ_e^\pm , the coefficient of $(\epsilon)^{1/2}$ in (2.98), requires a knowledge of γ_0 . This knowledge in turn, according to (2.73), requires a knowledge of the quantities a through c in (2.67). We see that \mathcal{M} must be known through *third* order, that is through terms of order δ^3 , to find the leading bifurcation behavior. And finding subsequent terms in the expansion of δ_e^\pm requires knowing \mathcal{M} to successively higher orders. For example, finding the order ϵ term in (2.98) requires a knowledge of σ_0 , which in turn according to (2.74) requires a knowledge of the *fourth*-order coefficient d .

This is the result for the case of one-dimensional maps. Since one-dimensional maps can be parts of many-dimensional maps, we conclude that a *necessary* condition to find the leading bifurcation behavior of a many-dimensional map is also that we know its expansion in deviation variables (about a fixed point) through *third* order. We speculate that this information is also *sufficient*. See Section 24.12.

1.2.4. Assuming that (2.14) is asymptotically correct, show that δ can be determined by the limiting process

$$\lim_{j \rightarrow \infty} [(\lambda_j - \lambda_{j-1})/(\lambda_{j+1} - \lambda_j)] = \delta. \quad (1.2.106)$$

Suppose a map is re-parameterized by introducing the parameter $\mu = g(\lambda)$, where g is any invertible differentiable function. Show that the $\mu_j = g(\lambda_j)$ also satisfy (2.99).

1.2.5. For the complex logistic map in the form (2.29), write

$$z = x + iy, \quad (1.2.107)$$

$$\gamma = \alpha + i\beta. \quad (1.2.108)$$

Show that in terms of these quantities the complex logistic map in the form (2.29) is equivalent to the two-dimensional real quadratic map given by the relations

$$x_{n+1} = \alpha x_n - \beta y_n - \alpha(x_n^2 - y_n^2) + 2\beta x_n y_n, \quad (1.2.109)$$

$$y_{n+1} = \beta x_n + \alpha y_n - \beta(x_n^2 - y_n^2) - 2\alpha x_n y_n. \quad (1.2.110)$$

1.2.6. Consider the transformation

$$z = 1/v \Leftrightarrow v = 1/z, \quad (1.2.111)$$

which interchanges the origin and the point at infinity. That is, we make the correspondence $v = 0 \leftrightarrow z = \infty$. Show that under this change of variables the logistic map (2.29) takes the form

$$v_{n+1} = -(1/\gamma)(v_n)^2/(1 - v_n). \quad (1.2.112)$$

Evidently $v = 0$ is a fixed point. Is it an attractor and, if so, can something be said about its basin?

Consider the number $|\gamma|/(1 + |\gamma|)$. Suppose that v_n is sufficiently close to the origin so that

$$|v_n| = \tau|\gamma|/(1 + |\gamma|) \text{ with } \tau < 1. \quad (1.2.113)$$

Show that then there is the inequality

$$|v_{n+1}| \leq \tau|v_n|. \quad (1.2.114)$$

Thus, $v = 0$ is an attractor, and its basin, at the very least, contains the open disk

$$|v| < |\gamma|/(1 + |\gamma|). \quad (1.2.115)$$

Show, in fact, that $v = 0$ is super attractive. See Exercise 2.1.

Show that all points z that satisfy

$$|z| > 1 + 1/|\gamma| \quad (1.2.116)$$

iterate to ∞ under the action of \mathcal{M} as given by (2.29). Thus $z = \infty$ may be viewed as an attractor of \mathcal{M} as given by (2.29), and points that satisfy (2.115) lie in its basin.

We remark that this exercise shows that the complex logistic map can better be viewed as a mapping into itself of the Riemann sphere rather than the complex plane. We also remark that the Julia set may be viewed as the *boundary* of the basin of attraction for the attractor $z = \infty$. That the Julia set is fractal is an instance of the theorem that basin boundaries are generally fractal.

1.2.7. Show that, under the change of variables

$$z = -(w/\gamma) + (1/2) \Leftrightarrow w = \gamma[(1/2) - z] \quad (1.2.117)$$

and the parameter change

$$\mu = (\gamma^2/4) - (\gamma/2) = (\gamma - 1)^2/4 - (1/4) \Leftrightarrow \gamma = 1 \pm \sqrt{1 + 4\mu}, \quad (1.2.118)$$

the logistic map (2.29) takes the form

$$w_{n+1} = w_n^2 - \mu. \quad (1.2.119)$$

Note that the change of variables relation (2.117) connecting z and w is *globally invertible* and *linear*. Therefore, whatever features occur in a mapping plane, say a rabbit, will have a *similar* appearance whether they are described in terms of z or in terms of w . By contrast, the change of parameters relation (2.118) connecting γ and μ is nonlinear and not one-to-one. Therefore we may expect, and indeed we will find, that the Mandelbrot set will have *different* appearances whether described in terms of γ or in terms of μ .

Show that the logistic map is two-to-one, and therefore not globally invertible. Show that it is, however, locally invertible in the neighborhood of each fixed point. Verify the symmetry claimed for the Mandelbrot set shown in Figure 2.7. Figure 2.12 shows the Mandelbrot set in the complex μ plane. Verify that μ is unchanged under the substitution $\gamma \rightarrow 2 - \gamma$. Verify that (2.111) maps the two disks in Figure 2.7 into a cardioid. See Figure 2.12. Verify that the point $\gamma = 1$ in Figure 2.7 corresponds to the point $\mu = -(1/4)$ in Figure 2.12, and that this point is at the cusp of the cardioid. Verify that the points $\gamma = 2, \gamma = 3, \gamma = \lambda_{\text{cr}}$, and $\gamma = 4$ correspond to the points $\mu = 0, \mu = (3/4), \mu = \mu_{\text{cr}} \simeq 1.40$, and $\mu = 2$. Find μ for Douady's rabbit, and describe the location of this μ value in Figure 2.12.

Show that the map (2.119) has the equilibrium (fixed) points

$$w_e^\pm = (1/2) \pm [\mu + (1/4)]^{1/2}, \quad (1.2.120)$$

and relate these points to the x_e given by (2.8) and (2.9). Show that w_e^- is stable for μ real and in the interval $(-1/4, 3/4)$, and w_e^+ is unstable.

Figure 2.13 is the analog of Figure 2.4 for μ real and in terms of the variable w . Only the trail of w_e^- , as μ is varied, is shown because w_e^+ is unstable. However, if both were shown and according to (2.120), verify that the trails $w_e^\pm(\mu)$ would together comprise a parabola lying on its side and extending to the right with vertex $\mu = -(1/4), w = (1/2)$. Note that since $w_e^\pm(\mu)$ are complex for $\mu < -(1/4)$, these fixed points do not appear in Figure 2.13. Thus, from the perspective of one living in the real world, two fixed points have appeared “out of the blue” at $\mu = -(1/4)$. There are no (real) fixed points for $\mu < -(1/4)$ and there are two, one stable and one unstable, for $\mu > -(1/4)$. This appearance of two fixed points out of nowhere is an example of a *blue sky* or *saddle-node* bifurcation. Finally, verify that this behavior is not manifest when x is used as a variable and λ is used as a parameter, see (2.5), because according to (2.118) γ and hence λ is complex when $\mu < -(1/4)$.

1.2.8. The general one-variable *analytic* quadratic map is of the form

$$\zeta_{n+1} = a + b\zeta_n + c\zeta_n^2 \quad (1.2.121)$$

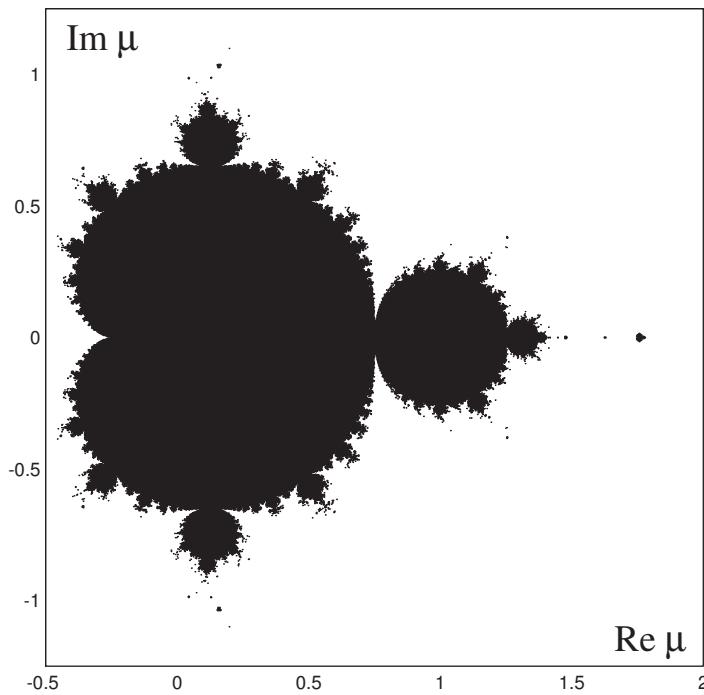


Figure 1.2.12: The Mandelbrot set in the μ plane. The “plate” has been somewhat “overexposed” compared to Figure 2.7 to bring out the island chains.

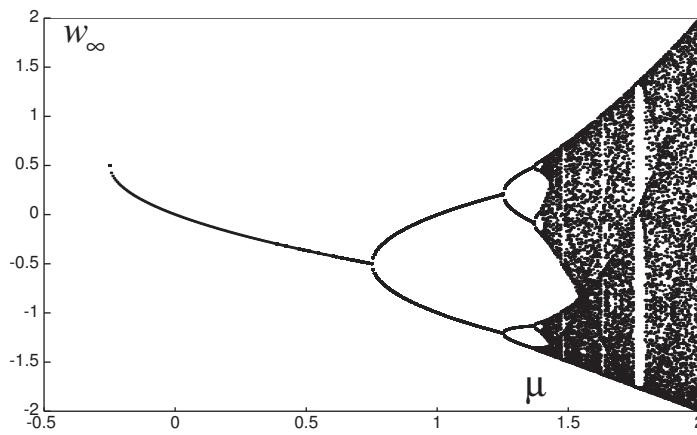


Figure 1.2.13: The analog of Figure 2.4 for μ real and the variable w .

with $c \neq 0$. Show that, under the change of variables

$$\zeta = w/c - b/(2c) \Leftrightarrow w = c\zeta + b/2, \quad (1.2.122)$$

this map takes the form (2.119) with

$$\mu = b^2/4 - b/2 - ac. \quad (1.2.123)$$

That is, verify that

$$\begin{aligned} w_{n+1} &= c\zeta_{n+1} + b/2 \\ &= c(a + b\zeta_n + c\zeta_n^2) + b/2 \\ &= (ca + b/2) + cb\zeta_n + c^2\zeta_n^2 \\ &= (ca + b/2) + cb[w_n/c - b/(2c)] + c^2[w_n/c - b/(2c)]^2 \\ &= (ca + b/2 - b^2/2) + bw_n + c^2[w_n/c - b/(2c)]^2 \\ &= (ca + b/2 - b^2/4) + w_n^2 \\ &= w_n^2 - \mu. \end{aligned}$$

1.2.9. The behavior of the *real* logistic map (2.5) can be analyzed fully in the case $\lambda = 4$. This analysis also provides a simple example of *symbolic dynamics*.

Suppose x_0 is some number in the interval $[0, 1]$,

$$x_0 \in [0, 1]. \quad (1.2.124)$$

Define a related angle ϕ_0 by the rule

$$x_0 = (1/2) - (1/2) \cos \phi_0. \quad (1.2.125)$$

Show that (2.118) has a unique solution satisfying

$$\phi_0 \in [0, \pi]. \quad (1.2.126)$$

Now define a sequence $\{x_0, x_1, x_2, \dots\}$ by the rule

$$x_n = (1/2) - (1/2) \cos(2^n \phi_0). \quad (1.2.127)$$

Show that these points satisfy the recursion relation (2.5). Define α_0 by the rule

$$\alpha_0 = \phi_0/\pi \quad (1.2.128)$$

and verify that

$$\alpha_0 \in [0, 1]. \quad (1.2.129)$$

Next define a map \mathcal{B} , called the *Bernoulli shift*, that acts on a sequence $\{\alpha_0, \alpha_1, \alpha_2, \dots\}$ by the rule

$$\alpha_{n+1} = \mathcal{B}\alpha_n \stackrel{\text{def}}{=} 2\alpha_n \mod 2. \quad (1.2.130)$$

Show that this recursion relation, with the initial condition α_0 , has the solution

$$\alpha_n = 2^n \alpha_0 \pmod{2}. \quad (1.2.131)$$

Verify that, because of the periodicity of the cosine function, we may rewrite (2.120) in the form

$$x_n = (1/2) - (1/2) \cos(\pi \alpha_n). \quad (1.2.132)$$

If we call the points α_n the *orbit* of α_0 under the action of the Bernoulli shift, and call the points x_n the orbit of x_0 under the action of the logistic map, then we see that the logistic orbit is the image of the Bernoulli orbit under the relation (2.125). Show, by drawing a suitable graph, that the relation (2.125) is two to one.

Suppose α_n for some n is written in *binary* form. Then we get an expression of the form

$$\alpha_n = a_1.a_2a_3a_4\cdots \quad (1.2.133)$$

where the entries a_i are 0 or 1. For example, there are the relations

$$\begin{aligned} 0 &= 0.000\cdots, \\ 3/2 &= 1.100\cdots, \\ 1 &= 1.000\cdots = 0.11111\cdots, \\ 1/2 &= 0.1000\cdots, \\ 1/4 &= 0.01000\cdots, \\ 2/5 &= 0.0110011001100\cdots. \end{aligned} \quad (1.2.134)$$

Show that α_{n+1} then has the binary expansion

$$\alpha_{n+1} = a_2.a_3a_4\cdots. \quad (1.2.135)$$

That is, the binary sequence for α_{n+1} is gotten by *shifting* the binary sequence for α_n one entry to the left and then discarding the first term. In the language of symbolic dynamics, the quantities 0 and 1 are called *symbols* (or letters from an alphabet if letters are used in place of digits) and the sequences (2.126) are called *words*. The Bernoulli map is an example of a dynamical operation on symbols.

By using (2.126) and (2.128) one can show that the Bernoulli map has many more or less evident properties that are reflected, in turn, in the behavior of the logistic map (when $\lambda = 4$). As a simple example, suppose α'_n is a number whose binary expansion is the same as that given for α_n in (2.127) save that the first entry, the one before the binary point, is different from a_1 . Then, according to (2.128), the result of \mathcal{B} acting on α'_n is the same as the result of \mathcal{B} acting on α_n . We immediately see that \mathcal{B} , and hence \mathcal{M} , is two to one.

Next consider some more complicated examples. To begin, suppose α_0 has a repeating binary expansion. Then \mathcal{B} acting repeatedly on α_0 produces a periodic orbit, and so will \mathcal{M} acting repeatedly on x_0 . Verify that, when α_0 has the value

$$\alpha_0 = .0100100100\cdots = 2/7, \quad (1.2.136)$$

the map \mathcal{B} has a 3-cycle (period three orbit) consisting of the values $2/7, 4/7, 8/7$. Correspondingly, the map \mathcal{M} has a 3-cycle when acting on the associated x_0 given by the relation

$$x_0 = (1/2) - (1/2) \cos(2\pi/7) = .188255099 \dots \quad (1.2.137)$$

Verify that when α_0 has the value $2/5$, the map \mathcal{B} has the 4-cycle $2/5, 4/5, 8/5, 16/5 = 6/5 \bmod 2$. See the expansion given for $2/5$ in (2.127). Correspondingly, one might expect that \mathcal{M} has a 4-cycle when acting on the associated $x_0 = .345491503 \dots$. However, it actually has a 2-cycle because the relation (2.125) is two to one.

Conversely, if α_0 does not have a binary expansion that eventually repeats, then the α_n will never repeat and the corresponding x_n given by (2.125) will never repeat. As a special case of this circumstance, suppose the successive a_j in the binary expansion for α_0 are determined by tossing a coin with $a_j = 1$ if the j th toss gives a head, and $a_j = 0$ if the j th toss is tails. Then we may say that α_0 is a random number, and the successive α_n and their corresponding x_n will also reflect this randomness. Thus, in this sense we can say that the long-term behavior of some orbits of \mathcal{M} is as random as a coin toss.

Next, suppose x^a and x^b are any two points in $[0, 1]$. Let the associated α^a and α^b have the binary expansions $a_1.a_2a_3 \dots$ and $b_1.b_2b_3 \dots$. Define a number α^ξ by the rule

$$\alpha^\xi = a_1.a_2a_3 \dots a_N b_1 b_2 b_3 \dots \quad (1.2.138)$$

Note that in (2.131) the sequence for α^a has been truncated after N terms and the full sequence for α^b has been appended at the end. Let x^ξ be the point associated with α^ξ using (2.125). Show that x^ξ can be made arbitrarily near x^a by making N large enough. That is, study how $|x^a - x^\xi|$ goes to 0 for large N . Next show that

$$\mathcal{M}^N x^\xi = x^b. \quad (1.2.139)$$

Thus, in any vicinity of an arbitrary point x^a there are points x^ξ , and these points can be sent to any other point x^b by a sufficiently high power of \mathcal{M} .

This construction also illustrates that the long-term behavior of an orbit generated by \mathcal{M} depends very sensitively on the initial condition x_0 . Indeed, we see that to determine the effect of \mathcal{M}^N on x_0 we must know at least the first few digits beyond the first N digits of the binary expansion of α_0 . Thus, to achieve a given accuracy in the final condition $\mathcal{M}^N x_0$, the required accuracy in α_0 , and hence also in x_0 , grows exponentially in N . Verify this claim. Moreover, this construction reveals that chaotic behavior in the orbit x_n , if any, arises from random behavior, if any, in the binary expansion of α_0 .

Extend the construction just given to an arbitrary sequence of points x^a, x^b, x^c, \dots and show that there are points arbitrarily near x^a which, when taken as initial conditions, have orbits that pass arbitrarily near (and in sequence) the remaining points x^b, x^c, \dots . You have demonstrated that there are orbits of \mathcal{M} that are *ergodic*.

As one last observation, suppose α^d is the number having the binary expansion

$$\alpha^d = \{[0][1]\} \{[00][01][10][11]\} \{[000][001][010][011][100][101][110][111]\} \{[\dots] \quad (1.2.140)$$

Here the curly and square brackets $\{\}$ and $[]$ are to be removed. They simply guide the eye to indicate that α^d consists first of all one-letter words, then all two-letter words, then all

three-letter words, etc., with the words for each fixed length listed, when viewed as binary numbers, in ascending order.³⁶ Evidently, under sufficiently many Bernoulli shifts acting on α^d , it will happen that any finite string will eventually occur as the leading string in the shifted α^d . Let x^d be the point associated with α^d using (2.125). Show that the orbit of x^d under the action of \mathcal{M} is *dense* on the interval $[0, 1]$. That is, it comes arbitrarily close to any point in the interval. In fact, show that it does so infinitely often. Show that if one wishes to minimize (to any finite degree) the effect, on a word, of nearby words, one can separate adjacent words by strings of 0's of any desired (but finite) lengths so that α^d is of the general form (2.133) except for strings of 0's inserted between the words.

Remark: When $\lambda = 4$ you have shown that the logistic orbit is the image of the Bernoulli orbit. Let (2.125) define a map \mathcal{T} so that we may write

$$x_n = \mathcal{T}\alpha_n. \quad (1.2.141)$$

Then the relation between the two orbits is equivalent to the equation

$$\mathcal{M}\mathcal{T}\alpha_n = \mathcal{T}\mathcal{B}\alpha_n \quad (1.2.142)$$

or, more abstractly,

$$\mathcal{M}\mathcal{T} = \mathcal{T}\mathcal{B}. \quad (1.2.143)$$

We say that \mathcal{M} is *conjugate* to \mathcal{B} under the action of \mathcal{T} . (See Section 19.2.) Thus, you have shown that the logistic map is conjugate to the Bernoulli map when $\lambda = 4$. The same can be proved (although with considerable more difficulty) for some λ values less than 4. Of course, when $\lambda \neq 4$, the conjugating map \mathcal{T} is no longer given by (2.125).

1.2.10. Show that it follows from the fixed-point property (2.21) and the normalization condition (2.22) that

$$g(1) = -1/\alpha. \quad (1.2.144)$$

Evaluate the series (2.23) at $x = 1$ and compare your result with (2.137).

1.2.11. Verify (2.25) using (2.19) through (2.21) and (2.24).

1.2.12. This exercise studies the complex logistic map (2.29). The complexified version of (2.8) gives

$$z_f = 0 \quad (1.2.145)$$

as a fixed point of \mathcal{M} . Locate this point in Figure 2.8. Let z' be the point

$$z' = 1. \quad (1.2.146)$$

Locate it in Figure 2.8. Show analytically that

$$\mathcal{M}z' = z_f. \quad (1.2.147)$$

Find points z'' such that

$$\mathcal{M}z'' = z' \quad (1.2.148)$$

³⁶Constructions of this kind were first made by D. G. Champernowne.

and hence

$$\mathcal{M}^2 z'' = z_f. \quad (1.2.149)$$

Can you find points z''' such that

$$\mathcal{M} z''' = z'', \quad (1.2.150)$$

and hence

$$\mathcal{M}^3 z''' = z_f, \text{ etc.?} \quad (1.2.151)$$

Verify that the complexified version of (2.9) gives (for Douady's γ value) the second fixed point

$$z_f = 1 - 1/\gamma = .656747 - .129015i. \quad (1.2.152)$$

To an uninformed botanist, Douady's rabbit, particularly in color, might look more like a *cactus*.³⁷ Again see Figure 2.8. Adopting this terminology, verify, by examining Figure 2.8, that this fixed point z_f is located at the point where the three lobes containing the period-three fixed points z^1 , z^2 , and z^3 meet. Define a point z' by the relation

$$z' = 1/\gamma = .343253 + .129015i. \quad (1.2.153)$$

Verify, again by examination, that three lobes also meet at this point. Show analytically that

$$\mathcal{M} z' = z_f. \quad (1.2.154)$$

Can you again find points z'' , z''' , etc., such that (2.141) through (2.144), etc. hold for z_f given by (2.145)?

Next consider the yellow lobe containing the point $z^{\text{in}} = .2 + .1i$. View z^{in} as an *initial* condition. Find the successive lobes that the orbit of z^{in} belongs to under successive applications of \mathcal{M} , and list their colors. Carry out the same exercise for the green point $z^{\text{in}} = .05 + .08i$ and the red point $z^{\text{in}} = .08 + .15i$. Suggestion: Study Exercise 2.5, and write and execute a suitable computer program.

1.2.13. As described in Subsection 2.2, the Mandelbrot set is connected. It is also *bounded*. Let M be the Mandelbrot set in the control plane γ . See Figure 2.7. Also, let σ be a complex variable. Define in the complex plane C a (filled) disk $D(r)$ of radius r , and centered on the origin, by writing

$$D(r) = \{\sigma \in C \mid |\sigma| \leq r\}. \quad (1.2.155)$$

We will find that, for

$$r > r^* = 2 + \sqrt{5} = 4.2360679 \dots, \quad (1.2.156)$$

there is the result

$$M \subset D(r). \quad (1.2.157)$$

That is, in the control plane γ , the *entire* Mandelbrot set (apparent "mainland", island chains, tendrils, and all) is contained within a *fixed* disk surrounding the origin.

How does one see this result? Review Exercise 2.6. Next, as described earlier, consider the *orbit* of $z = 1/2$, the set of points $\mathcal{M}^n(1/2)$ with $n = 0, 1, \dots$. We have been informed that, if

$$\mathcal{M}^n(1/2) \rightarrow \infty \text{ as } n \rightarrow \infty, \quad (1.2.158)$$

³⁷In fact, it is sometimes called a cactus fractal.

then the γ defining \mathcal{M} is *not* in M . That is, if $z = 1/2$ is in the basin of the fixed point $z = \infty$, then the γ defining \mathcal{M} is *not* in M . Now consider the point in the mapping plane given by $\mathcal{M}(1/2)$. From (2.29) we find

$$\mathcal{M}(1/2) = \gamma(1/2)[1 - (1/2)] = \gamma/4. \quad (1.2.159)$$

The point $1/2$ will be in the basin of ∞ if $\mathcal{M}(1/2)$ is in the basin of ∞ . According to (2.159) and (2.116), $\mathcal{M}(1/2)$ will be in the basin of ∞ if

$$|\gamma/4| > 1 + 1/|\gamma| \Leftrightarrow |\gamma| > 4 + 4/|\gamma|. \quad (1.2.160)$$

Verify that (2.160) is satisfied if

$$|\gamma| > 2 + \sqrt{5} = 4.2360679 \dots . \quad (1.2.161)$$

Therefore γ is *not* in the Mandelbrot set M if (2.161) holds. Correspondingly, in accord with (3.157), the Mandelbrot set M must be within the bounded set $D(r)$ provided (2.156) holds. As a sanity check, print or Xerox Figure 2.7. On this copy of Figure 2.7 draw, with a compass, the boundary of $D(r^*)$ and verify that (2.157) appears to hold.

What can be said about the Mandelbrot set in the μ plane? See Figure 2.12). It is also bounded. In accord with (2.118), write

$$\mu = (\gamma^2/4) - (\gamma/2) = (\gamma - 1)^2/4 - (1/4). \quad (1.2.162)$$

Find the image of the interior and boundary of $D(r^*)$ under this mapping. Verify that it is bounded because $D(r^*)$ is bounded and (2.162) is polynomial

1.2.14. Verify (2.46) and (2.47).

1.2.15. Verify (2.48) and (2.49).

1.2.16. Verify (2.51) through (2.53).

1.2.17. Show that the dynamic aperture for the map (2.50) is periodic in θ with period 4π .

1.3 Essential Theorems for Differential Equations

Among all the disciplines of mathematics, the *theory of differential equations* is the most important one. All areas of physics pose problems which lead to the integration of differential equations. In fact, it is the theory of differential equations which shows the way to understanding all time-dependent phenomena. If, on the one hand, the theory of differential equations has extreme *practical* significance, then, on the other hand, it attains a corresponding *theoretical* importance because it leads in a rational way to the study of new functions or classes of functions.

Sophus Lie (1894)

In this book we shall be concerned primarily with processes and maps that are described by or arise from differential equations. When all is said and done, the Laws of Motion for a Newtonian Dynamical System, however formulated, reduce to a set of second-order ordinary differential equations of the form

$$\begin{aligned}\ddot{q}_1 &= h_1(q_1, q_2, \dots; \dot{q}_1, \dot{q}_2, \dots; t), \\ \ddot{q}_2 &= h_2(q_1, q_2, \dots; \dot{q}_1, \dot{q}_2, \dots; t), \\ &\text{etc.}\end{aligned}\tag{1.3.1}$$

where the quantities $q_j(t)$ refer directly or indirectly to the instantaneous coordinates of various particles, and (following William Jones' and Newton's convention) a dot above a letter denotes differentiation with respect to time.³⁸ Do differential equations such as (3.1) actually contain information about trajectories? If so, how much? To these questions mathematicians have given answers in the form, as is their custom, of theorems. Actually, their theorems apply to sets of first-order differential equations. But that is no problem. We can easily convert a set of n second-order equations such as (3.1) into a set of $2n$ first-order equations. We define $2n$ variables $y_j(t)$ by the rule

$$\begin{aligned}y_1(t) &= q_1(t) \\ &\vdots \\ y_n(t) &= q_n(t), \\ y_{n+1}(t) &= \dot{q}_1(t) \\ &\vdots \\ y_{2n}(t) &= \dot{q}_n(t).\end{aligned}\tag{1.3.2}$$

The equations (3.1) are then equivalent to the first-order set

$$\dot{y}_j = y_{n+j}, \quad j \leq n$$

³⁸Surprisingly, nowhere in Newton's *Principia* does Newton's second law of motion appear in the familiar equation forms $F = ma$ or $a = F/m$, not to mention in the then unavailable concise vector notation form $\mathbf{a} = \mathbf{F}/m$. He writes no equation, but employs only the words "A change in motion is proportional to the motive force impressed and takes place along the straight line in which that force is impressed".

$$\dot{y}_j = h_{j-n}(y_1, \dots, y_{2n}, t) \quad n < j \leq 2n. \quad (1.3.3)$$

Alternatively, if the original equations (3.1) arose from a Lagrangian, they can also be converted into a first-order set by passing to a Hamiltonian formulation. See Sections 1.5 and 1.6.

Now hear the pronouncements of mathematicians. They provide definitive results for what is called the *Cauchy* (or initial value) problem:³⁹

Theorem 1.3.1. *Consider any set of m first-order differential equations of the form*

$$\dot{y}_j = f_j(y_1, \dots, y_m; t), \quad j = 1, \dots, m. \quad (1.3.4)$$

Here m may be even or odd. Assume that the right sides of (3.4), which define the set of differential equations, are sufficiently well behaved. In particular, assume that the f_j and the partial derivatives $\partial f_j / \partial y_k$ exist and are continuous in the y_k and in t within some region R of the m -dimensional space y_1, \dots, y_m and for t in some interval T about a fixed value t^0 . Let (y_1^0, \dots, y_m^0) be a point in R . Then there exists a unique solution

$$y_j(t) = g_j(y_1^0, \dots, y_m^0; t^0; t), \quad j = 1, \dots, m \quad (1.3.5)$$

of (3.4) with the property

$$y_j(t^0) = g_j(y_1^0, \dots, y_m^0; t^0; t^0) = y_j^0, \quad j = 1, \dots, m. \quad (1.3.6)$$

This solution is guaranteed to exist for a finite interval of time about the point t^0 , and can be extended forward or backward in time as long as the f_j are continuous in the y_k and t , and the $y_j(t)$ remain within a region R' where the $\partial f_j / \partial y_k$ exist and are continuous in the y_k and t . Furthermore, the solution (3.5) is continuous (and bounded) in all the variables y_j^0, t^0 , and t . See Figure 3.1. The quantities y_j^0 are called initial conditions and t^0 is called the initial time. To put the matter naively, we may think of first-order differential equations as a set of “marching orders” instructing us how to move at each instant of time. Once the initial starting time t^0 and the initial starting point (the initial conditions y_j^0) for the march are specified, the whole march is completely determined.

Theorem 1.3.2. *Suppose the f_j also depend on a set of parameters $\lambda_1, \dots, \lambda_n$. Assume that all $\partial f_j / \partial \lambda_k$ are continuous. Then the solution (3.5) will also be continuous in the parameters λ_k .*

Theorem 1.3.3. *Suppose the f_j are analytic in the variables y_j , λ_k , and t . (A function is analytic in some variable if it has a convergent Taylor series expansion in that variable when all other variables are held fixed. For more detail, see Sections 38.1 and 38.2.) Then the solution (3.5) will also be analytic in the variables y_j^0 , λ_k , t^0 , and t .*

³⁹Augustin-Louis Cauchy (1789-1857), a French mathematician/physicist, was one of the first to state and actually prove theorems of calculus, and made numerous important contributions to many areas of mathematics. His collected works comprise some 27 volumes.

The proofs of these theorems may be found in most reasonably complete books on differential equations. Including possible parameter dependence, the m differential equations to be solved are of the form

$$\dot{y}_j = f_j(y_1 \cdots y_m; \lambda_1 \cdots \lambda_n; t), \quad j = 1, \dots, m, \quad (1.3.7)$$

and are equivalent to the integral equations,

$$y_j(t) = y_j^0 + \int_{t^0}^t f_j[y_1(\tau) \cdots y_m^{p-1}(\tau); \lambda_1 \cdots \lambda_n; \tau] d\tau. \quad (1.3.8)$$

(Note that these integral equations automatically incorporate the initial conditions y_j^0 .) In turn, these integral equations are usually analyzed by showing that successive *Picard* iterations y_j^p of (3.8) defined by

$$y_j^p(t) = y_j^0 + \int_{t^0}^t f_j[y_1^{p-1}(\tau) \cdots y_m^{p-1}(\tau); \lambda_1 \cdots \lambda_n; \tau] d\tau, \quad p \geq 1, \quad (1.3.9)$$

converge to g_j as $p \rightarrow \infty$, and that the limit has the stated properties.⁴⁰

We should also mention that Theorem 3.1 can be proved under weaker conditions than the existence of various partial derivatives. For example, *Peano* proved existence under the assumption of simple continuity of the f_j in t and the y_j (however, as illustrated in the Exercises for this section, in this case there are examples for which uniqueness fails); simple continuity in t and *Lipschitz* continuity in the y_j are sufficient for both existence and uniqueness.⁴¹ Usually, however, the results we have stated are adequate.

Next a few words about the content of the theorems themselves. Theorem 3.1, when applied to the second-order equations (3.1), says that these equations have a unique solution providing we specify the initial coordinates

$$q_j(t^0) = q_j^0$$

and the initial “velocities”

$$\dot{q}_j(t^0) = \dot{q}_j^0.$$

[Alternatively, in a Hamiltonian formulation, these equations have a unique solution providing we specify the initial coordinates q_j^0 (as before) and the initial momenta p_j^0 .] Again we call these quantities, when taken together, a set of initial conditions. Thus, in general there is a unique trajectory for each set of initial conditions, and each trajectory varies continuously with the initial conditions, their time of imposition t^0 , and the time t . Needless to say, this continuity is in accord with our physical intuition of motion. However, the fact that initial coordinates and velocities *alone* are enough to completely specify a trajectory, i.e. that the physical equations of motion (3.1) are of second order, is not at all obvious. Or, put another way, it is not obvious that all effects of past history are in fact subsumed in a knowledge of present positions and velocities. Rather, this fact should be regarded as one of the greater discoveries of our ancestors.

⁴⁰Picard was a son-in-law of Hermite.

⁴¹Hadamard, a student of Picard, defined a problem to be *well posed* if a solution exists, is unique, and depends continuously on initial conditions and parameters. Thus, the assumptions of Theorems 3.1 through 3.3 assure that the problem of computing trajectories is well posed.

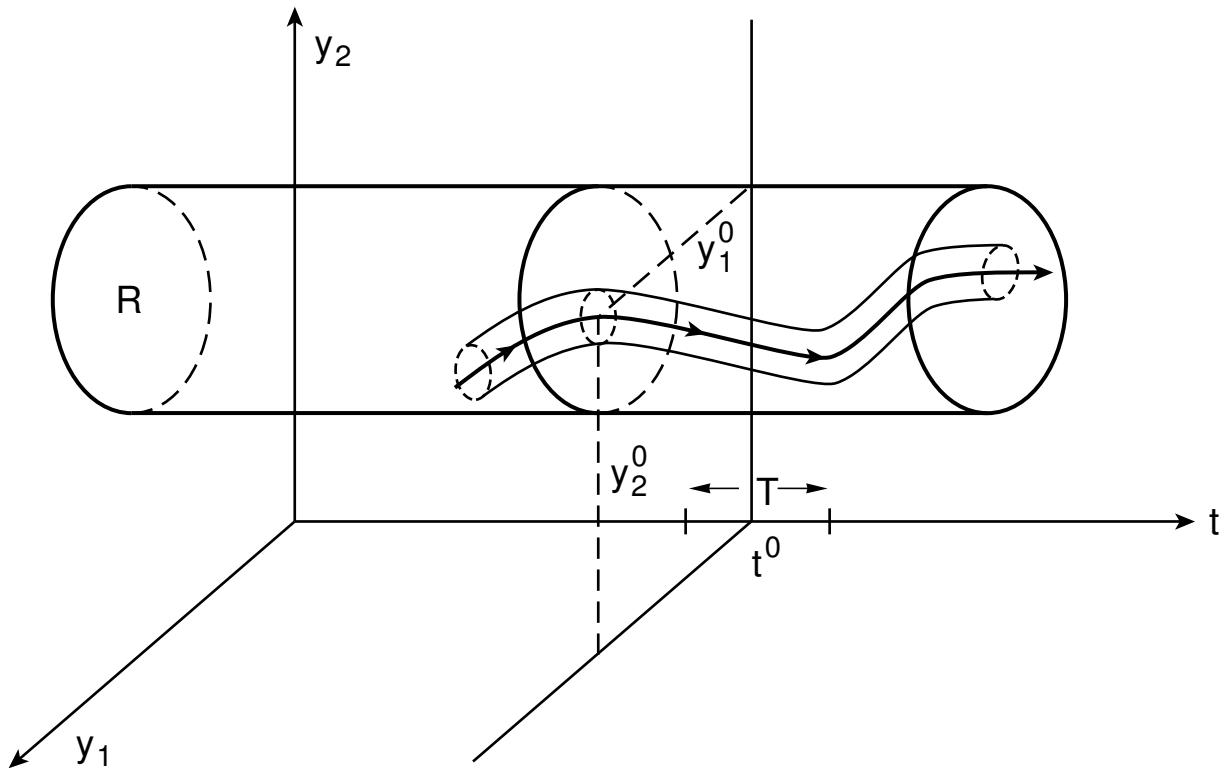


Figure 1.3.1: An illustration of Theorem 3.1 in the case that “ y ” space is two dimensional. The solution \mathbf{y} exists, is unique, and is continuous in t as long as it remains within the large cylinder of base R where \mathbf{f} is continuous and the $\partial\mathbf{f}/\partial y_j$ are continuous. If the point \mathbf{y}^0 is varied slightly, the solution also changes only slightly so that nearby solutions form a bundle.

Theorem 3.2, and particularly Theorem 3.3, are often of practical computational use. First, parameters often occur either quite naturally or can be introduced into problems of physical interest. Consider the motion, for example, of the sun-earth-moon system. There the mass ratios $\lambda_1 = M_{\text{moon}}/M_{\text{sun}}$ and $\lambda_2 = M_{\text{earth}}/M_{\text{sun}}$ appear in a natural way. Their smallness suggests the possibility of making a power series expansion of the equations of motion in terms of λ_1 and λ_2 , and then solving the resulting equations term by term. The success of such a perturbation technique is intimately related to the contents of Theorem 3.3. The use of perturbative power series was first systematically studied by Poincaré. In fact, Theorem 3.3 is often called *Poincaré's holomorphic lemma* or its results are referred to as *Poincaré analyticity*.⁴² Second, as will be seen later, it is often useful to expand a solution as a power series in the initial conditions. Finally, analyticity in t , or at least the existence of several derivatives in t , is supposed in carrying out numerical integration. See Chapter 2.

We also note that the conditions for Theorem 3.3 can be relaxed. Suppose the f_j are analytic in the y_j and the λ_k , but only have n derivatives in t . Remarkably, the final conditions will still be analytic functions of the initial conditions and the parameters, and will have $n+1$ derivatives in t . If the f_j are analytic in the y_j and the λ_k , but are only continuous in t , then the final conditions will still be analytic functions of the initial conditions and the parameters, and will have first derivatives in t . If the f_j are analytic in the y_j and the λ_k , but are only piece-wise continuous in t , then the final conditions will still be analytic functions of the initial conditions and the parameters, and will be piece-wise (first) differentiable in t . Finally, as it stands, the notation (3.6) indicates that the initial conditions are assumed to be independent of any parameters. All conclusions concerning analyticity continue to hold if the initial conditions are allowed to depend on parameters providing this dependence is analytic.

As an application of these relaxed conditions, suppose the time axis is broken up into a finite number of intervals and that the f_j are analytic in the y_j , the λ_k , and at least continuous in t for each interval. Then the final conditions will be piece-wise differentiable in t and will still be analytic functions of the initial conditions and the parameters. In the context of Accelerator Physics, where some coordinate related to path length plays the role of time, this situation arises in the idealization that an accelerator is treated as a sequence of discrete beam-line elements with a separate Hamiltonian, and therefore a separate transfer map, for each element. See Subsection 2.4 and Sections 4 and 6. Each such transfer map will be analytic in the initial conditions and parameters, and their product will then also be analytic in these quantities.

Finally, we remark that Poincaré's holomorphic lemma has important applications outside of Classical Mechanics. It is used in advanced Quantum Mechanics, for example, to show that solutions to the Schrödinger equation are analytic in energy, angular momentum, and coupling constant. This analyticity is in turn used to suggest that various processes involving elementary particles at high energies obey certain integral conditions called dispersion relations.

⁴²The terms *analytic* and *holomorphic* are commonly used interchangeably, particularly in the context of several complex variables. (The definitions of analytic and holomorphic are different, but can be proven to be mathematically equivalent. See Sections 38.1 and 38.2.) Poincaré derived his analyticity results on a case-by-case basis as needed using *Cauchy's method of majorants*.

Exercises

1.3.1. Consider the differential equation

$$t^3 \dot{y} = 2y$$

with the initial condition $y(0) = 0$. Show that it has *two* solutions: $y(t) = e^{-(1/t)^2}$, $y(0) = 0$; and $y(t) = 0$ for $t \leq 0$, $y(t) = e^{-(1/t)^2}$ for $t > 0$. Does this lack of uniqueness violate Theorem 3.1? Are there even more solutions?

1.3.2. Consider the differential equation

$$\dot{y} = -(1 - y)^{1/2}$$

with the initial condition $y(0) = 1$. Show that it has the *two* solutions

$$y(t) \equiv 1 \text{ and } y(t) = 1 - t^2/4.$$

What causes this lack of uniqueness?

1.3.3. Consider the differential equation

$$\dot{y} = (1 - y)^{-1}$$

with the initial condition $y(0) = 0$. Find the solution and show that it cannot be extended arbitrarily far forward in time. In view of Theorem 3.1, what went wrong?

1.3.4. Consider the growth of a crystal in a supersaturated solution. Let V be the volume of the crystal and A its surface area. We assume the growth rate is proportional to the surface area, that is,

$$\dot{V} = k_1 A$$

where k_1 is some constant. But for a regular geometric figure there is a definite relation between A and V of the form

$$A = k_2 V^{2/3}.$$

For example, $k_2 = (36\pi)^{1/3}$ for a sphere and $k_2 = 6$ for a cube. Thus, for a regular figure we have a growth law of the form

$$\dot{V} = k V^{2/3}.$$

Show that with the initial condition $V(0) = 0$, one has the *family* of solutions

$$\begin{aligned} V(t) &= 0 , \quad 0 \leq t \leq \tau \\ &= [(k/3)(t - \tau)]^3 , \quad t \geq \tau \end{aligned}$$

for any positive τ . What causes this lack of uniqueness mathematically? Physically, τ is the time that elapses before random fluctuations form a “seed” which initiates crystal growth.

1.3.5. Consider one-dimensional motion with position coordinate x . Let $f(x)$ be a position dependent but time independent force defined by the rule

$$f(x) = 0 \text{ for } x \leq 0; \quad f(x) = +12x^{1/2} \text{ for } x \geq 0. \quad (1.3.10)$$

Note that $f(x)$ is continuous and satisfies $f(x) \geq 0$. Consider the equation of motion

$$\ddot{x} = f(x) \quad (1.3.11)$$

with the initial conditions

$$x(0) = \dot{x}(0) = 0. \quad (1.3.12)$$

Let c be any constant satisfying $c \geq 0$. Verify that (3.11) with the initial conditions (3.12) has the solution

$$x(t) = 0 \text{ for } t \leq c \text{ and } x(t) = (t - c)^4 \text{ for } t \geq c. \quad (1.3.13)$$

That is, verify that both (3.11) and (3.12) are satisfied. Note that $x(t)$ is continuous. How many continuous derivatives does it have? Why is the solution not unique? Are there still more solutions? What are the solutions to (3.11) for other initial conditions?

1.3.6. In computing and managing the trajectory of a space craft, one is obliged to use tracking data that inevitably contain at least some small errors. Also various parameters, such as anomalies in the gravitational field, the mass of the space ship, and the impulses provided by various rockets and thrusters, are not exactly known. Comment on the effect of these errors in view of Theorems 3.1 through 3.3.

1.3.7. Consider a set of differential equations of the form (3.4), and assume that the existence and uniqueness conditions of Theorem 3.1 are met. Show that no two different trajectories in (\mathbf{y}, t) space can ever join or intersect in finite time. Suppose the quantities f_j are independent of the time t . Then the set of differential equations is called *autonomous*. Show that in this case no trajectory in \mathbf{y} space can *cross* itself in finite time. (We say that a trajectory crosses itself if the two tangent lines to the two portions of the trajectory at the point of intersection have a finite angle between them.) Show that if a trajectory does intersect itself in finite time, it must join itself smoothly to form a periodic trajectory.

1.4 Transfer Maps Produced by Differential Equations

Suppose we rewrite the set of first-order differential equations (3.4) in the more compact vector form

$$\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}; t). \quad (1.4.1)$$

Then, according to Theorem 3.1 and again using vector notation, their solution can be written in the form

$$\mathbf{y}(t) = \mathbf{g}(\mathbf{y}^0; t^0; t). \quad (1.4.2)$$

That is, the quantities $\mathbf{y}(t)$ at any time t are uniquely specified by the initial quantities \mathbf{y}^0 given at the initial time t^0 .

We capitalize on this fact by introducing a slightly different notation. First, use t^i instead of t^0 to denote the *initial* time, and similarly use \mathbf{y}^i to denote initial conditions by writing

$$\mathbf{y}^i = \mathbf{y}^0 = \mathbf{y}(t^i). \quad (1.4.3)$$

Next, let t^f be some *final* time, and define final conditions \mathbf{y}^f by writing

$$\mathbf{y}^f = \mathbf{y}(t^f). \quad (1.4.4)$$

Then, with this notation, (4.2) can be rewritten in the form

$$\mathbf{y}^f = \mathbf{g}(\mathbf{y}^i; t^i; t^f). \quad (1.4.5)$$

We now view (4.5) as a map that sends the initial conditions \mathbf{y}^i to the final conditions \mathbf{y}^f . This map will be called the *transfer map* between the times t^i and t^f , and will often be denoted by the symbol \mathcal{M} . What we have learned is that a set of first-order differential equations of the form (4.1) can be integrated to produce a transfer map \mathcal{M} . We express the fact that \mathcal{M} sends \mathbf{y}^i to \mathbf{y}^f in symbols by writing the equation

$$\mathbf{y}^f = \mathcal{M}\mathbf{y}^i, \quad (1.4.6)$$

and illustrate this relation by the picture shown in Figure 4.1. Finally, as noted earlier, \mathcal{M} is always invertible: Given \mathbf{y}^f , t^f , and t^i , we can always march (integrate) backward in time to the moment t^i and thereby find the initial conditions \mathbf{y}^i .



Figure 1.4.1: The transfer map \mathcal{M} sends the initial conditions \mathbf{y}^i to the final conditions \mathbf{y}^f .

1.4.1 Map for Simple Harmonic Oscillator

To fix these ideas more clearly in the mind, we consider three examples. The first is a one-dimensional harmonic oscillator described by the Hamiltonian

$$H = p^2/(2m) + (k/2)q^2. \quad (1.4.7)$$

In this case the equations of motion are

$$\begin{aligned} \dot{q} &= \partial H / \partial p = p/m, \\ \dot{p} &= -\partial H / \partial q = -kq. \end{aligned} \quad (1.4.8)$$

(See Section 1.5 for a review of Hamilton's equations of motion.) These equations can be solved easily enough. However, for future use, it is convenient to make the (canonical) change of variables

$$\begin{aligned} Q &= (km)^{1/4}q, \\ P &= (km)^{-1/4}p. \end{aligned} \quad (1.4.9)$$

In these new variables the equations of motion become

$$\begin{aligned} \dot{Q} &= \omega P, \\ \dot{P} &= -\omega Q, \end{aligned} \quad (1.4.10)$$

where

$$\omega = \sqrt{(k/m)}. \quad (1.4.11)$$

It is easily verified that the equations of motion (4.10) are produced by the new Hamiltonian K given by the relation

$$K = (\omega/2)(P^2 + Q^2). \quad (1.4.12)$$

The equations (4.10) are easily integrated to give the transfer map \mathcal{M} described by the relations

$$\begin{aligned} Q^f &= Q^i \cos[\omega(t^f - t^i)] + P^i \sin[\omega(t^f - t^i)], \\ P^f &= -Q^i \sin[\omega(t^f - t^i)] + P^i \cos[\omega(t^f - t^i)]. \end{aligned} \quad (1.4.13)$$

We see that for this example the transfer map is a linear relation between the initial and final conditions and (in the Q, P variables) simply consists of a (clockwise) rotation in phase space by the angle $[\omega(t^f - t^i)]$.

In view of the assertion (2.43), the map described by (4.13) can also be written formally as

$$\mathcal{M} = \exp\{-(t^f - t^i) : (\omega/2)[(P^i)^2 + (Q^i)^2] :\}. \quad (1.4.14)$$

Note that this claim is consistent with (2.37) and (2.38).

1.4.2 Maps for Monomial Hamiltonians

The second example of a transfer map is somewhat more complicated, and leads to a nonlinear relation between initial and final conditions. It too will be useful in the future. Consider, for the case of a two-dimensional phase space, the monomial Hamiltonian

$$H = \lambda q^r p^s. \quad (1.4.15)$$

Here λ is a parameter, and r and s are integers. The Hamiltonian (4.15) produces the equations of motion

$$\dot{q} = \lambda s q^r p^{s-1}, \quad (1.4.16)$$

$$\dot{p} = -\lambda r q^{r-1} p^s. \quad (1.4.17)$$

Since H has no explicit time dependence, we conclude that H must be a constant of motion. If you doubt this, see (5.14) in the next section. Let us solve (4.15) for p . Doing so gives the result

$$p = (H/\lambda)^{\frac{1}{s}} q^{-\frac{r}{s}}. \quad (1.4.18)$$

Next substitute (4.18) into (4.16) to get the relation

$$\dot{q} = \lambda s(H/\lambda)^{\frac{s-1}{s}} q^{\frac{r}{s}}. \quad (1.4.19)$$

Assume for the moment that $r \neq s$. Then (4.19) can be integrated immediately to give the result

$$(q^f)^{\frac{s-r}{s}} - (q^i)^{\frac{s-r}{s}} = \lambda(s-r)(t^f - t^i)(H/\lambda)^{\frac{s-1}{s}}. \quad (1.4.20)$$

Also, since H is a constant of motion, we may write

$$H = \lambda(q^i)^r(p^i)^s. \quad (1.4.21)$$

Equations (4.20) and (4.21) can now be combined and solved for q^f in terms of q^i and p^i . Finally, (4.17) can be integrated in a similar manner. The net result is the transfer map relations

$$q^f = q^i[1 + \lambda(s-r)(t^f - t^i)(q^i)^{r-1}(p^i)^{s-1}]^{\frac{s}{s-r}}, \quad (1.4.22)$$

$$p^f = p^i[1 + \lambda(s-r)(t^f - t^i)(q^i)^{r-1}(p^i)^{s-1}]^{\frac{r}{r-s}}, \quad (1.4.23)$$

when $r \neq s$.

The equations of motion for the case $r = s$ can also be solved. In this case (4.18) can be integrated in terms of logarithms. Also, (4.17) can be integrated similarly. The net result is the transfer map relations

$$q^f = q^i \exp[\lambda r(t^f - t^i)(q^i p^i)^{r-1}], \quad (1.4.24)$$

$$p^f = p^i \exp[-\lambda r(t^f - t^i)(q^i p^i)^{r-1}], \quad (1.4.25)$$

when $r = s$.

Note that the relations (4.22) through (4.25) are indeed nonlinear. The transfer maps for monomial Hamiltonians in higher dimensional phase spaces can also be found exactly. See Exercise 4.3. Also, we remark that the relations (4.22) and (4.23) can become *singular* in finite time. That is, the solutions to the equations of motion (4.16) and (4.17) cannot always be extended arbitrarily far forward and backward in time. See Exercise 4.4.

Again because of (2.43), the maps described by (4.22) through (4.25) can formally be written as

$$\mathcal{M} = \exp\{-(t^f - t^i) : \lambda(q^i)^r(p^i)^s :\}. \quad (1.4.26)$$

And summation of the exponential series (4.26), when acting on the initial conditions, will produce the maps (4.22) through (4.25).

1.4.3 Stroboscopic Maps and Duffing Equation Example

For a last example of a transfer map produced by a differential equation, we will begin a study of the behavior of a periodically driven damped *nonlinear* oscillator described by the equation of motion

$$\ddot{x} + a\dot{x} + bx + cx^3 = d \cos(\Omega t + \psi). \quad (1.4.27)$$

This equation, or sometimes a variant with x^3 replaced by x^2 , is commonly called *Duffing's equation*. Here ψ is an arbitrary phase factor that is often set to zero. For our purposes it is more convenient to set

$$\psi = \pi/2. \quad (1.4.28)$$

Evidently any particular choice of ψ simply results in a shift of the origin in time, and this shift has no physical consequence since the left side of (4.27) is independent of time.

We assume $b, c > 0$, which is the case of a positive hard spring restoring force.⁴³ We make these assumptions because we want the Duffing oscillator to behave like an ordinary harmonic oscillator when the amplitude is small, and we want the motion to be bounded away from infinity when the amplitude is large. Then, by a suitable choice of time and length scales that introduces new variables q and τ , the equation of motion can be brought to the form

$$\ddot{q} + 2\beta\dot{q} + q + q^3 = -\epsilon \sin \omega\tau, \quad (1.4.29)$$

where now a dot denotes $d/d\tau$ and we have made use of (4.28). See Exercise 4.10. In this form it is evident that there are 3 free parameters: β , ϵ , and ω .

Unlike the previous examples, this problem is dissipative (assuming $\beta > 0$) and time dependent. There is, however, the simplifying feature that the driving force is *periodic* with period

$$T = 2\pi/\omega. \quad (1.4.30)$$

Let us convert (4.29) into a pair of first-order equations by making the definition

$$p = \dot{q}, \quad (1.4.31)$$

with the result

$$\begin{aligned} \dot{q} &= p, \\ \dot{p} &= -2\beta p - q - q^3 - \epsilon \sin \omega\tau. \end{aligned} \quad (1.4.32)$$

Let q_0, p_0 denote initial conditions at $\tau = 0$, and let q_1, p_1 be the final conditions resulting from integrating the pair (4.32) one full period to the time $\tau = T$. Let \mathcal{M} denote the transfer map that relates q_1, p_1 to q_0, p_0 . Then, using the definition (2.39) and the notation (4.6), we may write

$$z_1 = \mathcal{M}z_0. \quad (1.4.33)$$

Suppose we now integrate for a second full period to find q_2, p_2 . Since the right side of (4.32) is periodic, the rules for integrating from $\tau = T$ to $\tau = 2T$ are the same as the rules for integrating from $\tau = 0$ to $\tau = T$. Therefore we may write

$$z_2 = \mathcal{M}z_1 = \mathcal{M}^2z_0, \quad (1.4.34)$$

and in general

$$z_{n+1} = \mathcal{M}z_n = \mathcal{M}^{n+1}z_0. \quad (1.4.35)$$

⁴³Other authors consider other cases, particularly the ‘double well’ case $b < 0$ and $c > 0$.

We may regard the quantities z_n as the result of viewing the motion of the Duffing oscillator by the light provided by a stroboscope that flashes at the times⁴⁴

$$\tau^n = nT. \quad (1.4.36)$$

Because of the periodicity of the right side of the equations of motion, the rule for sending z_n to z_{n+1} over the intervals between successive flashes is always the same, namely \mathcal{M} . For these reasons \mathcal{M} is called a *stroboscopic map*. Despite the explicit time dependence in the equations of motion, because of periodicity we have been able to describe the long-term motion by the repeated application of a single fixed map. A moment's reflection shows that what we have done here for the Duffing oscillator is quite general. The behavior of any periodically (not necessarily sinusoidally) driven system can be described by a stroboscopic map.⁴⁵

It follows from (4.35) that the long-term behavior of the driven Duffing oscillator is equivalent to the behavior of the Duffing stroboscopic map \mathcal{M} under repeated iteration. As we have seen from the examples of Section 1.2, the iteration of maps generally leads to enormous complications. Correspondingly, the driven Duffing oscillator displays an enormously rich behavior that varies widely with the parameter values β, ϵ, ω . This richness is typical of the long-term behavior of damped driven nonlinear systems. Indeed, without editorial restraint, the detailed study of any one of them could fill this entire book. However, rich as it is, the behavior of the driven Duffing oscillator, since it is governed by relatively few *attracting* (due to the presence of damping) fixed points, is trivial compared to that of most nonlinear *Hamiltonian* systems where fixed points are numerous and none are attracting.

Because even providing an overview of what can happen under repeated iteration of the stroboscopic Duffing map requires considerable work, at least an entire chapter is required for this purpose. Such an overview is provided in Chapter 28 where the subject is studied numerically and Section 29.12 where the behavior of polynomial approximations to the stroboscopic Duffing map is explored. See also Sections 10.12.7 and 10.12.8 and Appendix S.4.

Exercises

1.4.1. Verify equations (4.8) through (4.13) and all assertions made about them.

⁴⁴Note that, with the choice (4.28) for ψ , the driving term described by the right side of (4.29) vanishes at the stroboscopic times τ^n .

⁴⁵Consider a set of n second-order differential equations of the form (3.1) with the further assumption that the h_j do not depend on the time. We will say that such a set of equations (which is equivalent to a set of $2n$ autonomous first-order differential equations) describes a system having n *autonomous* degrees of freedom. Suppose next that the h_j do depend on the time, and in an *arbitrary* way. By choosing a new independent variable, it is possible to convert such a set of second-order differential equations into $(2n + 2)$ first-order autonomous differential equations. (See Exercise 6.5 for a discussion of how this can be done in the Hamiltonian case.) Thus, when t is present in the equations (3.1), we may say that these $2n$ nonautonomous equations describe a system having $(n + 1)$ autonomous degrees of freedom. As the discussion of this section shows, the case where the h_j depend on the time in a *periodic* way lies somewhere in between. Such systems are sometimes said to have $(n + 1/2)$ autonomous degrees of freedom. Thus, the Duffing oscillator may be said to have $3/2 = 1$ and $\frac{1}{2}$ degrees of freedom.

1.4.2. Verify equations (4.16) through (4.25). Suppose $s = 0$ and $r \neq 0$. Show that in this case

$$\begin{aligned} q^f &= q^i, \\ p^f &= p^i - \lambda r(t^f - t^i)(q^i)^{r-1}. \end{aligned} \quad (1.4.37)$$

Suppose $s \neq 0$ and $r = 0$. Show that in this case

$$\begin{aligned} q^f &= q^i + \lambda s(t^f - t^i)(p^i)^{s-1}, \\ p^f &= p^i. \end{aligned} \quad (1.4.38)$$

1.4.3. Consider, for the case of a four-dimensional phase space, the monomial Hamiltonian

$$H = \lambda q_1^{r_1} p_1^{s_1} q_2^{r_2} p_2^{s_2}. \quad (1.4.39)$$

Define “sub” Hamiltonians H_1 and H_2 by the relations

$$H_j = q_j^{r_j} p_j^{s_j}, \quad j = 1 \text{ and } 2. \quad (1.4.40)$$

With these definitions (4.43) can be rewritten in the form

$$H = \lambda H_1 H_2. \quad (1.4.41)$$

Show that both H_1 and H_2 are constants (in fact, integrals) of motion. [Hint: If you are having trouble, use (7.4) and (7.7) of Section 1.7.] Show that the equations of motion generated by H can be integrated and (when $r_j \neq s_j$) have solutions of the form

$$q_1^f = q_1^i [1 + \lambda(s_1 - r_1)(t^f - t^i)(q_2^i)^{r_2}(p_2^i)^{s_2}(q_1^i)^{r_1-1}(p_1^i)^{s_1-1}]^{\frac{s_1}{s_1-r_1}}, \text{ etc.} \quad (1.4.42)$$

Find complete results for all q_j^f , p_j^f and for all cases of the exponents r_j , s_j .

1.4.4. Consider the solution to (4.16) and (4.17) as given by (4.22) and (4.23) for the case $r = 1$ and $s = 4$. Show that the solution has a branch point in t at a finite time. Find other integer values of r , s for which the solution (4.22) and (4.23) has singularities for finite time. Conversely, given any neighborhood of the origin in the initial conditions q^i and p^i , show that (for suitable r , s values) the solution (4.22) and (4.23), when viewed as a function of the initial conditions, has singularities in this neighborhood for t finite (and real) providing t is sufficiently large. In view of Theorems 3.1 and 3.3, what is going wrong?

1.4.5. From the general discussion of transfer maps it is clear that non-Hamiltonian systems also can be described in terms of maps. All that is required is that the set of differential equations be written in the first-order form (4.1). Consider the one-dimensional motion of an object moving vertically and subject to gravity and viscous drag. Newton’s equation of motion for such an object can be written in the form

$$m\ddot{z} = -mg - \gamma\dot{z}. \quad (1.4.43)$$

Here m is the mass of the particle, g is the acceleration due to gravity, and γ (with $\gamma > 0$) is some measure of the viscous drag. Convert (4.47) into a first-order set of differential equations, and find the associated transfer map.

1.4.6. Let \mathcal{M} be a map of an m -dimensional space into itself as in (4.6). What happens to the final conditions when small changes are made in the initial conditions? From calculus we have the differential relation

$$dy_j^f = \sum_k (\partial y_j^f / \partial y_k^i) dy_k^i, \quad (1.4.44)$$

which can be written in the form

$$dy_j^f = \sum_k M_{jk}(\mathbf{y}^i) dy_k^i \quad (1.4.45)$$

where $M(\mathbf{y}^i)$ is the $m \times m$ matrix

$$M_{jk}(\mathbf{y}^i) = \partial y_j^f / \partial y_k^i. \quad (1.4.46)$$

This matrix is called the *Jacobian* matrix of \mathcal{M} . According to (4.81) it describes how small changes in the initial conditions \mathbf{y}^i produce small changes in the final conditions \mathbf{y}^f . Note that generally the Jacobian matrix depends on the initial conditions, and therefore we write $M(\mathbf{y}^i)$.

In the case that \mathcal{M} is a transfer map arising from a differential equation as in (4.1), the associated Jacobian matrix can be found by integrating the *variational* equations derived from (4.1). Here, as before, we assume \mathbf{y} has m components. Let \mathbf{y}^i be a set of initial conditions and let $\mathbf{y}^d(t)$ be the trajectory (sometimes called the *design* trajectory) that has these initial conditions,

$$\mathbf{y}^d(t^i) = \mathbf{y}^i. \quad (1.4.47)$$

Because it is a trajectory, it satisfies the differential equation

$$\dot{\mathbf{y}}^d = \mathbf{f}(\mathbf{y}^d; t). \quad (1.4.48)$$

Next consider nearby trajectories of the form

$$\mathbf{y}(t) = \mathbf{y}^d(t) + \epsilon \boldsymbol{\eta}(t) \quad (1.4.49)$$

where ϵ is small. Insertion of (4.49) into (4.1) gives the equation

$$\dot{\mathbf{y}}^d + \epsilon \dot{\boldsymbol{\eta}} = \mathbf{f}(\mathbf{y}^d + \epsilon \boldsymbol{\eta}; t). \quad (1.4.50)$$

Now take components of both sides of (4.50) and expand in powers of ϵ to find the relation

$$\dot{y}_j^d + \epsilon \dot{\eta}_j = f_j(\mathbf{y}^d; t) + \sum_k [(\partial f_j / \partial y_k) \Big|_{\mathbf{y}=\mathbf{y}^d} \epsilon \eta_k + O(\epsilon^2)]. \quad (1.4.51)$$

Define the $m \times m$ matrix $A(t)$ by the rule

$$A_{jk}(t) = (\partial f_j / \partial y_k) \Big|_{\mathbf{y}=\mathbf{y}^d}. \quad (1.4.52)$$

Use (4.48), (4.51), and (4.52) and equate powers of ϵ to show that $\boldsymbol{\eta}$ satisfies the set of equations

$$\dot{\boldsymbol{\eta}} = A(t) \boldsymbol{\eta}. \quad (1.4.53)$$

These are the *variational equations* associated with (4.1) around the trajectory \mathbf{y}^d .⁴⁶ Note that there are m such (usually coupled) equations because $\boldsymbol{\eta}$ is m dimensional, and that they are *linear* even if (4.1) is nonlinear.

Let $L(t)$ be the $m \times m$ matrix defined by the *matrix* differential equation (a collection of m^2 ordinary differential equations)

$$\dot{L} = A(t)L \quad (1.4.54)$$

with the initial condition

$$L(t^i) = I \quad (1.4.55)$$

where I denotes the $m \times m$ identity matrix. Show that the solution to (4.53) with the initial condition $\boldsymbol{\eta}^i$ is given by the prescription

$$\boldsymbol{\eta}(t) = L(t)\boldsymbol{\eta}^i. \quad (1.4.56)$$

Show that the desired Jacobian matrix is given in terms of $L(t)$ by the relation

$$M = L(t^f). \quad (1.4.57)$$

The solution of the differential equations (4.48) for the design trajectory, which is required to determine A using (4.52), generally requires numerical integrator. Solution of the variational equations (4.53), or their matrix counterpart (4.54), even though they are linear, also generally requires numerical integration because they are coupled and A is usually time dependent. However, assuming A is known, it is possible to calculate the *determinant* of M analytically. Use (4.54) to write the Taylor expansion

$$\begin{aligned} L(t + dt) &= L(t) + \dot{L}(t)dt + O[(dt)^2] \\ &= L(t) + dtA(t)L(t) + O[(dt)^2] \\ &= [I + dtA(t)]L(t) + O[(dt)^2]. \end{aligned} \quad (1.4.58)$$

Take determinants of both sides of (4.58) to get the result

$$\begin{aligned} \det[L(t + dt)] &= \det\{[I + dtA(t)][L(t)]\} + O[(dt)^2] \\ &= \{\det[I + dtA(t)]\}\{\det[L(t)]\} + O[(dt)^2] \\ &= \{1 + dt \operatorname{tr}[A(t)]\}\{\det[L(t)]\} + O[(dt)^2]. \end{aligned} \quad (1.4.59)$$

Here use has been made of (3.7.132). Show that (4.59) produces the differential equation

$$(d/dt) \det[L(t)] = \{\operatorname{tr}[A(t)]\}\{\det[L(t)]\} \quad (1.4.60)$$

and, in view of (4.55), that this equation has the explicit solution

$$\det[L(t)] = \exp\left\{\int_{t^i}^t dt' \operatorname{tr}[A(t')]\right\}. \quad (1.4.61)$$

⁴⁶We could more accurately call them the first-variation equations or lowest-order variational equations. For what we call the *complete* variational equations, see Section 10.12.

In particular, there is the result

$$\det(M) = \exp\left\{\int_{t^i}^{t^f} dt \operatorname{tr}[A(t)]\right\}. \quad (1.4.62)$$

Subsequently, this result will be related, in the context of Hamiltonian dynamics, to what is called *Liouville's theorem*. The result itself, in the context of linear differential equations, which is what the variational equations are, is sometimes called the *Abel-Liouville-Jacobi-Ostrogradski formula*.

From this result show that the determinant of the Jacobian matrix associated with any transfer map arising from a *real* differential equation must satisfy the condition

$$\det(M) > 0. \quad (1.4.63)$$

Geometrically, this condition means that \mathcal{M} preserves *orientation*. For example, in the case $m = 3$, the \mathcal{M} arising from any real differential equation cannot send a right-handed triad into a left-handed triad. Comment: There is also a simpler but more subtle topological argument that leads to the result (4.63). Since a transfer map arising from integrating a differential equation evolves in a continuous way starting from the identity map, it can be written as a product of several transfer maps, all of which are near the identity map. See Section 6.4.1. Since each of these maps is near the identity, by continuity the determinant of the Jacobian matrix of each must be positive. But, by the chain rule, the Jacobian matrix of a product of maps must be the product of the Jacobian matrices of the individual factors. Finally, the determinant of a product of matrices is the product of the determinants of the individual factors.

The determinant of the Jacobian matrix also has further geometrical significance. For the purpose of this exercise, let us refer to the m -dimensional space we have been considering as *variable* space. This variable space need not be phase space because the dimension may be odd, and even if m is even the equations of motion need not be Hamiltonian in form and the coordinates may not necessarily come in canonically conjugate pairs. The equations of motion and the coordinates can be completely general.

Consider a particular trajectory with initial conditions given by (4.47) and also all other trajectories whose initial conditions lie within a small volume dV^i about the initial conditions for the particular trajectory. Then at some final time t^f the final conditions for these trajectories will lie within a small volume dV^f about the final conditions for the particular trajectory. From standard advanced calculus lore the initial and final volumes are related by the equation

$$dV^f = \{\det[M(\mathbf{y}^i)]\}dV^i. \quad (1.4.64)$$

Thus, the determinant of M specifies the evolution of volume elements in variable space.

1.4.7. For the case of the complex logistic map in the form (2.112), write

$$w = u + iv \quad (1.4.65)$$

and show that the Jacobian matrix is given by the relation

$$M(w_n) = 2 \begin{pmatrix} u_n & -v_n \\ v_n & u_n \end{pmatrix}. \quad (1.4.66)$$

See Exercise 4.6. Thus for this map

$$\det[M(w_n)] = 4(u_n^2 + v_n^2) = 4|w_n|^2, \quad (1.4.67)$$

and the map preserves orientation except at the origin. Verify that the map is not invertible in the neighborhood of the origin. Consider any map of the form

$$w_{n+1} = f(w_n) \quad (1.4.68)$$

where f is an analytic function. Show that

$$\det[M(w_n)] = |f'(w_n)|^2. \quad (1.4.69)$$

Thus, all analytic maps are orientation preserving.

1.4.8. Consider the Hénon map in the product form (2.23). Compute the Jacobian matrix for each factor. See Exercise 4.6. Verify that the Jacobian matrix for each factor has determinant one and therefore, by the chain rule, the determinant of the Jacobian matrix for the full map also has determinant one. It follows, as will be described in detail later, that the Hénon map is *area preserving*.

1.4.9. Let $\delta_{per}(t)$ denote the 2π periodic delta function defined by the relation

$$\delta_{per}(t) = \sum_{n=-\infty}^{\infty} \delta(t + 2n\pi). \quad (1.4.70)$$

Show that the map $\mathcal{M}(\theta)$ given by (2.50) is the stroboscopic map resulting from integrating from $t^i = 0$ to $t^f = 2\pi$ the motion arising from the 2π periodic Hamiltonian

$$H = [\theta/(4\pi)][p^2 + q^2] - \delta_{per}(t - \pi)q^3. \quad (1.4.71)$$

1.4.10. Choose appropriate time and length scales by writing $x = \lambda q$ and $t = \sigma\tau$ to convert (4.27) into (4.29).

1.5 Lagrangian and Hamiltonian Equations

It is a remarkable discovery that all the known fundamental dynamical laws of Nature are expressible in Lagrangian or Hamiltonian form, and therefore also in variational form. Indeed, as Euler wrote in his (and the first by any author) publication on variational calculus,

Because the shape of the whole universe is most perfect and, in fact, designed by the wisest Creator, nothing in all of the world will occur in which no maximum or minimum rule is somehow shining forth.

Since the construction of the entire universe is absolutely perfect and is due to a Creator with infinite knowledge, nothing exists in the world which does not exhibit some property of maximum or minimum. Therefore, there cannot be any doubt whatsoever about the possibility that all the effects are determined by their final aims with the help of the maxima method, in the same way in which they are also determined by the initial causes.

The last five sections of this chapter are devoted to needed aspects of Lagrangian and Hamiltonian dynamics and the Theory of Special Relativity.

Usual Lagrangian L for Charged Particle Motion in Electromagnetic Fields

Since much of this book concerns the motion of charged particles in electromagnetic fields, we recall that the usual Lagrangian for the motion of a particle of mass m and charge q in an electromagnetic field is given by the expression

$$L(\mathbf{r}, \mathbf{v}, t) = -mc^2(1 - v^2/c^2)^{1/2} - q\psi(\mathbf{r}, t) + q\mathbf{v} \cdot \mathbf{A}(\mathbf{r}, t). \quad (1.5.1)$$

Here ψ and \mathbf{A} are the scalar and vector potentials defined in such a way that the electromagnetic fields \mathbf{E} and \mathbf{B} are given by the standard relations

$$\mathbf{B} = \nabla \times \mathbf{A},$$

$$\mathbf{E} = -\nabla\psi - \partial\mathbf{A}/\partial t. \quad (1.5.2)$$

[For the Hamiltonian H associated with L see (5.49).] We note that this formulation ignores spin, radiation reaction (synchrotron radiation), and quantum effects. It is also not manifestly Lorentz invariant, and therefore we might be suspect of its validity in a relativistic context. (It is assumed that the reader already has some background in Special Relativity.) However in Section 7 we will find a connection between this Lagrangian L and a manifestly Lorentz invariant Lagrangian L_R which shows that L is relativistically correct even though not manifestly Lorentz invariant. For further discussion of the Lorentz group and related material, see Chapter 28.

Poincaré, in a 1905 paper, coined the terms *Lorentz transformation* and *Lorentz group*. Hendrik Lorentz (1853-1928) was a Dutch physicist who made many contributions to Physics including the discovery and theoretical explanation of the Zeeman effect for which he and Zeeman jointly shared the 1902 Nobel Prize in Physics. For a video of Lorentz's funeral procession, which included Einstein, see <https://www.youtube.com/watch?v=H2VtrJD0xJk>. Pieter Zeeman (1865-1943) was a student and subsequent colleague of Lorentz. Hendrik Lorentz is not to be confused with the Danish physicist Ludvig Lorenz (1829-1891) for whom the Lorenz gauge/condition is named or with the meteorologist Edward Norton Lorenz (1917-2008) who was a pioneer in chaos theory.

1.5.1 The Nonsingular Case

Lagrange's equations of motion for a system having n degrees of freedom are

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0, \quad (1.5.3)$$

where $(q_1 \cdots q_n)$ is any set of generalized coordinates. [Note that in general L is a function of the q_i , \dot{q}_i , and t ; $L = L(q, \dot{q}, t)$.] According to Section 1.3, what we ultimately need are equations of the form (3.1). By the chain rule there are the relations

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_j} = \frac{\partial^2 L}{\partial t \partial \dot{q}_j} + \sum_i \left[\frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j} \ddot{q}_i + \frac{\partial^2 L}{\partial q_i \partial \dot{q}_j} \dot{q}_i \right] \quad (1.5.4)$$

so that Lagrange's equations can also be written in the form

$$\sum_i \frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j} \ddot{q}_i = \frac{\partial L}{\partial q_j} - \frac{\partial^2 L}{\partial t \partial \dot{q}_j} - \sum_i \frac{\partial^2 L}{\partial q_i \partial \dot{q}_j} \dot{q}_i. \quad (1.5.5)$$

The quantity $[\partial^2 L / \partial \dot{q}_i \partial \dot{q}_j]$ is called the *Hessian* (matrix) of L .⁴⁷ In order to solve the relations (5.5) for the \ddot{q}_i to obtain equations of the form (3.1), the Hessian of L must be invertible/nonsingular and therefore must satisfy the condition

$$\det(\partial^2 L / \partial \dot{q}_i \partial \dot{q}_j) \neq 0. \quad (1.5.6)$$

We call this the *nonsingular* or *regular* case; and, when (5.6) fails to hold, we call this the *singular* case.⁴⁸

The momentum p_i canonically conjugate to the variable q_i is defined by the relation

$$p_i = p_i(q, \dot{q}, t) = \partial L / \partial \dot{q}_i, \quad (1.5.7)$$

and the Hamiltonian H associated with the Lagrangian L is defined by the *Legendre* transformation

$$H(q, p, t) = \sum_i p_i \dot{q}_i - L(q, \dot{q}, t). \quad (1.5.8)$$

(For a study of Legendre transformations, see Exercise 6.2.9. They are an application of the theory of *gradient* maps.) Note that as it stands, and in view of (5.7), the right side of (5.8) is a function of the variables q, \dot{q}, t . However the left side describes H as a function of the variables q, p, t . That is, the variables \dot{q} are to be eliminated in terms of the p 's. According to the *inverse function theorem*, this is possible if and only if the determinant of the associated *Jacobian* matrix is nonzero,

$$\det(\partial p_i / \partial \dot{q}_j) \neq 0. \quad (1.5.9)$$

From (5.7) there is the relation

$$\partial p_i / \partial \dot{q}_j = \partial^2 L / \partial \dot{q}_j \partial \dot{q}_i = \partial^2 L / \partial \dot{q}_i \partial \dot{q}_j. \quad (1.5.10)$$

Here we have used the equality of mixed partial derivatives.⁴⁹ Thus, the conditions (5.6) and (5.9) are the same.

Hamilton's equations of motion for the $2n$ canonical variables $(q_1 \cdots q_n)$ and $(p_1 \cdots p_n)$ are given in terms of the Hamiltonian $H(q, p, t)$ by the rules

$$\dot{q}_i = \partial H / \partial p_i \quad , \quad \dot{p}_i = -\partial H / \partial q_i. \quad (1.5.11)$$

⁴⁷Otto Hesse (1811-1874).

⁴⁸Abraham and Marsden call the nonsingular case *hyperregular* if the map $q, \dot{q}, t \leftrightarrow q, p, t$ is a diffeomorphism; that is, it is a differentiable map with a differentiable inverse. See (5.6) through (5.10). (For our purposes we are happy to assume differentiability, or even analyticity.) They call the singular case *degenerate*. Some other authors call the singular case *irregular*.

⁴⁹The Clairaut-Schwarz-Young theorem.

There is also the additional relation

$$\partial H/\partial t = -\partial L/\partial t. \quad (1.5.12)$$

For later use, it is convenient to append yet one more equation to the set (5.11) and (5.12). Consider the total time rate of change of the Hamiltonian H along a trajectory in q, p space. Using the chain rule, one finds the result

$$dH/dt = \partial H/\partial t + \sum_i [(\partial H/\partial q_i)\dot{q}_i + (\partial H/\partial p_i)\dot{p}_i]. \quad (1.5.13)$$

However, the quantity under the summation sign vanishes because of (5.11). It follows that the Hamiltonian has the special property

$$dH/dt = \partial H/\partial t = -\partial L/\partial t. \quad (1.5.14)$$

Suppose that H (or L) does not depend explicitly on the time ($\partial H/\partial t = 0$ or $\partial L/\partial t = 0$). A system that does not explicitly depend on the independent variable (the time) is called *autonomous*. We see from (5.14) that if a Hamiltonian system is autonomous, then the Hamiltonian H must be a *constant* of motion, and conversely. Moreover, because it has no explicit time dependence, such an H is also an *integral* of motion. For a discussion of constants and integrals of motion see Section 5.2.

1.5.2 A Common Singular Case

We end this section by noting that there is a fairly frequently encountered case in which (5.6) and (5.9) fail to hold, namely when L is *homogeneous* of degree *one* in the \dot{q}_i ,

$$L(q, \lambda\dot{q}, t) = \lambda L(q, \dot{q}, t). \quad (1.5.15)$$

See, for examples, Exercises 5.15, 6.5, 6.9, and 6.16. In this case the p_i are homogeneous of degree *zero* in the \dot{q}_i and, according to Euler's relation, there will be the result

$$\sum_j (\partial p_i / \partial \dot{q}_j) \dot{q}_j = 0. \quad (1.5.16)$$

See Exercise 5.12. The quantities $(\partial p_i / \partial \dot{q}_j)$ may be viewed as the entries in a matrix, and the quantities \dot{q}_j may be viewed as the entries in a vector. Since (5.16) must hold for any value of the \dot{q}_j , we conclude that the matrix $(\partial p_i / \partial \dot{q}_j)$ has a nonzero vector as an eigenvector with eigenvalue 0.⁵⁰ Since the determinant of a matrix equals the product of its eigenvalues, it follows that in this case

$$\det(\partial p_i / \partial \dot{q}_j) = \det(\partial^2 L / \partial \dot{q}_i \partial \dot{q}_j) = 0. \quad (1.5.17)$$

⁵⁰Note that the $(\partial p_i / \partial \dot{q}_j)$ may depend on the \dot{q}_k . However if at least one $\dot{q}_k \neq 0$, there is a *nonzero* vector for which (5.16) holds, and therefore this vector is an eigenvector with eigenvalue 0.

(Remarkably, although the $(\partial p_i / \partial \dot{q}_j)$ may depend on the \dot{q}_k , in this case their determinant does not!) Moreover, since L is assumed homogeneous of degree 1 in the \dot{q}_i , Euler's relation also gives the result

$$\sum_i p_i \dot{q}_i = \sum_i (\partial L / \partial \dot{q}_i) \dot{q}_i = L, \quad (1.5.18)$$

and hence, according to (5.8), the Hamiltonian associated with L vanishes identically.

Finally, suppose \mathcal{A} is the action functional associated with L and that L does not explicitly depend on the time,

$$\mathcal{A}[q(t)] = \int_{t^1}^{t^2} L(q, dq/dt) dt. \quad (1.5.19)$$

Let $\tau(t)$ be any monotonic function of t so that we may also write $t = t(\tau)$. Here we view τ as a parameter. Given a path $q(t)$, we will define a related path $Q(\tau)$ by the rule

$$Q_i(\tau) = q_i(t(\tau)). \quad (1.5.20)$$

Assign an action to any such path using the same functional (5.19),

$$\mathcal{A}[Q(\tau)] = \int_{\tau^1}^{\tau^2} L(Q, dQ/d\tau) d\tau. \quad (1.5.21)$$

By the chain rule we have the relations

$$d\tau = (d\tau/dt)dt, \quad (1.5.22)$$

$$dQ_i/d\tau = (dq_i/dt)(dt/d\tau). \quad (1.5.23)$$

Therefore, upon changing integration variables, there is the result

$$\mathcal{A}[Q(\tau)] = \int_{t^1}^{t^2} L\{q, (dq/dt)(dt/d\tau)\}(d\tau/dt) dt. \quad (1.5.24)$$

Under the assumption that L is homogeneous of degree one in the \dot{q}_i , there is also the relation

$$L\{q, (dq/dt)(dt/d\tau)\} = L(q, dq/dt)(dt/d\tau). \quad (1.5.25)$$

See (5.15). Inserting (5.25) into (5.24) gives the final result

$$\mathcal{A}[Q(\tau)] = \int_{t^1}^{t^2} L(q, dq/dt)(dt/d\tau)(d\tau/dt) dt = \int_{t^1}^{t^2} L(q, dq/dt) dt = \mathcal{A}[q(t)]. \quad (1.5.26)$$

We have learned that in this case $\mathcal{A}[Q(\tau)]$ is independent of the parameterization employed. That is, there are an infinite number of paths $Q(\tau)$, corresponding to the infinite number of parameterizations $t(\tau)$, all of which have the same action. This independence implies that we should *not* expect to find a unique solution that extremizes \mathcal{A} since any reparameterization also gives a solution.

Is all lost when L is homogeneous of degree one in the \dot{q}_i ? The answer is *no*. What we may expect in this case is that additional information beyond Hamilton's principle (or Lagrange's equations) will be required to specify a trajectory. Some further information has to be provided about the parameterization. Again see, for examples, Exercises 5.15, 6.5, 6.9, and 6.16.

Exercises

1.5.1. For the Lagrangian (5.1), show that the *canonical* momenta in Cartesian coordinates are given by the equation

$$\mathbf{p}^{\text{can}} = m\mathbf{v}/(1 - v^2/c^2)^{1/2} + q\mathbf{A}. \quad (1.5.27)$$

Here we have used the superscript *can* to emphasize that we are deriving the *canonical* momenta. Note that the first term in (5.27) is just the relativistic *mechanical* momentum,

$$\mathbf{p}^{\text{mech}} = m\mathbf{v}/(1 - v^2/c^2)^{1/2} = \gamma m\mathbf{v} \quad (1.5.28)$$

where γ is the standard relativistic factor

$$\gamma = 1/(1 - v^2/c^2)^{1/2}. \quad (1.5.29)$$

Consequently, the relation (5.27) can also be written in the forms

$$\mathbf{p}^{\text{can}} = \mathbf{p}^{\text{mech}} + q\mathbf{A} \quad \text{and} \quad \mathbf{p}^{\text{mech}} = \mathbf{p}^{\text{can}} - q\mathbf{A}. \quad (1.5.30)$$

1.5.2. The purpose of this exercise is to derive and study the equations of motion associated with the Langrangian L given by (5.1). Begin by reviewing Exercise 5.1. Let \mathbf{p}^{mech} denote the *mechanical* momentum given by (5.28). In the case that the generalized coordinates are taken to be the usual Cartesian coordinates, verify that Lagrange's equations for the Lagrangian (5.1) produce for the mechanical momentum the equation of motion

$$\dot{\mathbf{p}}^{\text{mech}} = d\mathbf{p}^{\text{mech}}/dt = \mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \quad (1.5.31)$$

Here \mathbf{F} is the Lorentz force.

For reasons that will become clear shortly, let us calculate the quantity $(d\gamma/dt)$. Rewrite (5.28) in the form

$$\mathbf{v} = \mathbf{p}^{\text{mech}}/(\gamma m). \quad (1.5.32)$$

Verify that squaring and inverting both sides of (5.29), and use of (5.32), produce the chain of relations

$$1/\gamma^2 = 1 - v^2/c^2 = 1 - (\mathbf{p}^{\text{mech}} \cdot \mathbf{p}^{\text{mech}})/(\gamma mc)^2, \quad (1.5.33)$$

from which it follows that

$$\gamma^2 = 1 + (\mathbf{p}^{\text{mech}} \cdot \mathbf{p}^{\text{mech}})/(mc)^2. \quad (1.5.34)$$

Next differentiate both sides of (5.34) to find that

$$\gamma(d\gamma/dt) = [1/(mc)^2](\mathbf{p}^{\text{mech}} \cdot \dot{\mathbf{p}}^{\text{mech}}) = [\gamma/(mc^2)](\mathbf{v} \cdot \dot{\mathbf{p}}^{\text{mech}}), \quad (1.5.35)$$

from which it follows that

$$d\gamma/dt = [1/(mc^2)](\mathbf{v} \cdot \dot{\mathbf{p}}^{\text{mech}}). \quad (1.5.36)$$

Now use (5.31) to show that

$$\mathbf{v} \cdot \dot{\mathbf{p}}^{\text{mech}} = \mathbf{v} \cdot \mathbf{F} = q\mathbf{v} \cdot \mathbf{E}, \quad (1.5.37)$$

and thereby verify that

$$d\gamma/dt = [1/(mc^2)](\mathbf{v} \cdot \mathbf{F}) = [q/(mc^2)](\mathbf{v} \cdot \mathbf{E}). \quad (1.5.38)$$

Define the relativistic energy \mathcal{E} by the rule

$$\mathcal{E} = \gamma mc^2. \quad (1.5.39)$$

Show from (5.36) that it obeys the equation of motion

$$d\mathcal{E}/dt = \mathbf{v} \cdot \mathbf{F} = q(\mathbf{v} \cdot \mathbf{E}). \quad (1.5.40)$$

Note that $\mathbf{v} \cdot \mathbf{F}$ is simply the rate at which work is being done by the Lorentz force. Indeed, verify that (5.40) is equivalent to the differential relation

$$d\mathcal{E} = \mathbf{F} \cdot d\mathbf{r} = q(\mathbf{E} \cdot d\mathbf{r}). \quad (1.5.41)$$

As a particle moves, its change in energy equals the work done by the Lorentz force, more specifically by the *electric* part of the Lorentz force.

Verify from (5.34) and (5.39) that there is the relation

$$\mathcal{E}^2 = m^2 c^4 + (\mathbf{p}^{\text{mech}} \cdot \mathbf{p}^{\text{mech}}) c^2. \quad (1.5.42)$$

Show that (5.40) also follows from (5.31) and (5.42).

Solve (5.32) and (5.34) for \mathbf{v} to find the relation

$$\dot{\mathbf{r}} = \mathbf{v} = \mathbf{p}^{\text{mech}} / (m^2 + \mathbf{p}^{\text{mech}} \cdot \mathbf{p}^{\text{mech}}/c^2)^{1/2}. \quad (1.5.43)$$

Show that (5.31) can be rewritten in the form

$$\dot{\mathbf{p}}^{\text{mech}} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) = q\{\mathbf{E} + [\mathbf{p}^{\text{mech}} / (m^2 + \mathbf{p}^{\text{mech}} \cdot \mathbf{p}^{\text{mech}}/c^2)^{1/2}] \times \mathbf{B}\}. \quad (1.5.44)$$

Taken together, (5.43) and (5.44) provide equations of motion for the quantities \mathbf{r} and \mathbf{p}^{mech} in terms of \mathbf{r} , \mathbf{p}^{mech} , and t . Note that these equations only involve the fields \mathbf{E} and \mathbf{B} , and not the vector and scalar potentials \mathbf{A} and ψ . They are therefore gauge independent.

Suppose we seek equations of motion for the quantities $(\mathbf{r}; \mathbf{v})$ with t taken to be the independent variable. That is, what are desired are equations for the quantities $\ddot{\mathbf{r}}$ in terms of the variables \mathbf{r} , \mathbf{v} , and t . Verify that differentiating (5.32) yields the result

$$\ddot{\mathbf{r}} = \dot{\mathbf{v}} = \dot{\mathbf{p}}^{\text{mech}} / (m\gamma) - \mathbf{p}^{\text{mech}}[1/(m\gamma^2)](d\gamma/dt) = \dot{\mathbf{p}}^{\text{mech}} / (m\gamma) - (\mathbf{v}/\gamma)(d\gamma/dt). \quad (1.5.45)$$

For the first term on the far right side of (5.45), namely the term involving $\dot{\mathbf{p}}^{\text{mech}}$, we will use (5.31). For the second term involving the $d\gamma/dt$ factor we will use (5.38). Verify that use of (5.31) and (5.38) in (5.45) yields, in the form desired, the result

$$\ddot{\mathbf{r}} = [q/(\gamma m)](\mathbf{E} + \mathbf{v} \times \mathbf{B}) - [q/(\gamma m c^2)]\mathbf{v}(\mathbf{v} \cdot \mathbf{E}). \quad (1.5.46)$$

Equivalently, there is the coupled pair of first-order equations

$$\dot{\mathbf{r}} = \mathbf{v}, \quad (1.5.47)$$

$$\dot{\mathbf{v}} = \ddot{\mathbf{r}} = [q/(\gamma m)](\mathbf{E} + \mathbf{v} \times \mathbf{B}) - [q/(\gamma m c^2)]\mathbf{v}(\mathbf{v} \cdot \mathbf{E}). \quad (1.5.48)$$

1.5.3. Show that the Hamiltonian associated with the Lagrangian (5.1) is given in Cartesian coordinates by the expression

$$H = [m^2c^4 + c^2(\mathbf{p}^{\text{can}} - q\mathbf{A}) \cdot (\mathbf{p}^{\text{can}} - q\mathbf{A})]^{1/2} + q\psi = [m^2c^4 + c^2(\mathbf{p}^{\text{can}} - q\mathbf{A})^2]^{1/2} + q\psi. \quad (1.5.49)$$

Verify, using (5.27) through (5.30), that there is the result

$$H = [m^2c^4 + c^2(\mathbf{p}^{\text{mech}} \cdot \mathbf{p}^{\text{mech}})]^{1/2} + q\psi = \gamma mc^2 + q\psi. \quad (1.5.50)$$

Here we have used the superscripts *can* and *mech* to emphasize the distinction between *canonical* and *mechanical* momenta. We also apologize for continuing to use the symbol q to denote particle charge, which is not to be confused with the use of q to denote a canonical position variable.

1.5.4. Let x, y, z denote the usual Cartesian coordinates. In the x, z plane introduce polar coordinates ρ, ϕ by the relations

$$\begin{aligned} x &= \rho \cos \phi, \\ z &= \rho \sin \phi. \end{aligned} \quad (1.5.51)$$

View the triplet ρ, y, ϕ as a cylindrical coordinate system, and let $\mathbf{e}_\rho, \mathbf{e}_y, \mathbf{e}_\phi$ be the associated right-handed orthonormal triad. See Figure 5.1. (Note that this choice of cylindrical coordinates differs from the usual choice ρ, ϕ, z .) The purpose of this exercise is to find the canonical momenta and the Hamiltonian associated with the Lagrangian (5.1) when the cylindrical coordinates ρ, y, ϕ are used as generalized coordinates.

Verify that there are the relations

$$\mathbf{r} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z = \rho \cos \phi \mathbf{e}_x + y\mathbf{e}_y + \rho \sin \phi \mathbf{e}_z, \quad (1.5.52)$$

and

$$\begin{aligned} \mathbf{e}_\rho &= \cos \phi \mathbf{e}_x + \sin \phi \mathbf{e}_z, \\ \mathbf{e}_\phi &= -\sin \phi \mathbf{e}_x + \cos \phi \mathbf{e}_z, \end{aligned} \quad (1.5.53)$$

so that there is also the relation

$$\mathbf{r} = \rho\mathbf{e}_\rho + y\mathbf{e}_y. \quad (1.5.54)$$

Note that the directions of \mathbf{e}_ρ and \mathbf{e}_ϕ depend on ϕ , and hence on x and z . For example, the pair \mathbf{e}_ρ and \mathbf{e}_ϕ appearing in Figure 5.1 are shown pointing in the direction they would have at the x, z location where they are displayed. Verify that $\mathbf{e}_\rho, \mathbf{e}_y, \mathbf{e}_\phi$ do indeed form a right-handed orthonormal triad.

Answer: It is easily verified from (5.52) and (5.53) that \mathbf{e}_ρ and \mathbf{e}_ϕ satisfy the equations

$$\mathbf{e}_\rho = \frac{\partial \mathbf{r}}{\partial \rho} / \left\| \frac{\partial \mathbf{r}}{\partial \rho} \right\|, \quad (1.5.55)$$

$$\mathbf{e}_\phi = \frac{\partial \mathbf{r}}{\partial \phi} / \left\| \frac{\partial \mathbf{r}}{\partial \phi} \right\|,$$

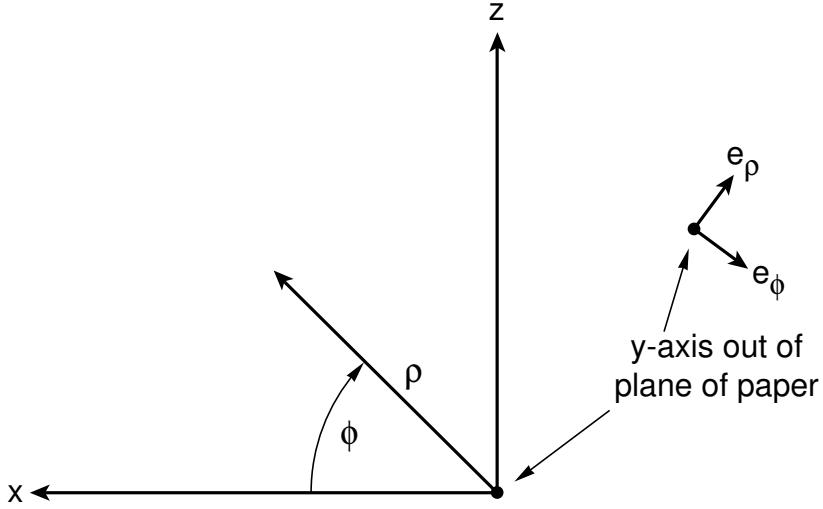


Figure 1.5.1: Illustration of the ρ, y, ϕ cylindrical coordinate system and a sample unit-vector pair e_ρ and e_ϕ .

and therefore are properly defined. Moreover, it is easily checked that e_ρ, e_y , and e_ϕ are orthonormal and satisfy the relation

$$e_\rho \times e_y = e_\phi. \quad (1.5.56)$$

They therefore form a right-handed triad.

Verify that it also follows from (5.54) or the second part of (5.52) that

$$d\mathbf{r} \cdot d\mathbf{r} = (d\rho)^2 + (dy)^2 + \rho^2(d\phi)^2. \quad (1.5.57)$$

Consequently, the line element squared can be written in the standard form

$$d\mathbf{r} \cdot d\mathbf{r} = h_1^2(dq_1)^2 + h_2^2(dq_2)^2 + h_3^2(dq_3)^2, \quad (1.5.58)$$

where

$$h_1 = 1, \quad h_2 = 1, \quad h_3 = \rho, \quad (1.5.59)$$

and

$$q_1 = \rho, \quad q_2 = y, \quad q_3 = \phi. \quad (1.5.60)$$

Correspondingly, the unit vectors are numbered in the order

$$\mathbf{e}_1 = \mathbf{e}_\rho, \quad \mathbf{e}_2 = \mathbf{e}_y, \quad \mathbf{e}_3 = \mathbf{e}_\phi. \quad (1.5.61)$$

With the above prescription, the curl of an arbitrary vector \mathbf{A} is given by the relation

$$(\nabla \times \mathbf{A})_1 = \frac{1}{h_2 h_3} \left[\frac{\partial(h_3 A_3)}{\partial q_2} - \frac{\partial(h_2 A_2)}{\partial q_3} \right], \quad (1.5.62)$$

and the relations obtained from it by cyclic permutations of the coordinate indices. Here the components A_i of \mathbf{A} are defined by the relations

$$A_i = \mathbf{e}_i \cdot \mathbf{A}. \quad (1.5.63)$$

Verify that in terms of the coordinates (5.51) there are the relations

$$\dot{x} = \dot{\rho} \cos \phi - \rho \dot{\phi} \sin \phi, \quad (1.5.64)$$

$$\dot{z} = \dot{\rho} \sin \phi + \rho \dot{\phi} \cos \phi.$$

Show from (5.51) and (5.64) that consequently there are the relations

$$\begin{aligned} \mathbf{v} &= d\mathbf{r}/dt = \dot{x}\mathbf{e}_x + \dot{y}\mathbf{e}_y + \dot{z}\mathbf{e}_z \\ &= \dot{\rho}(\cos \phi \mathbf{e}_x + \sin \phi \mathbf{e}_z) + \dot{y}\mathbf{e}_y + \rho \dot{\phi}(\cos \phi \mathbf{e}_z - \sin \phi \mathbf{e}_x) \\ &= \dot{\rho}\mathbf{e}_\rho + \dot{y}\mathbf{e}_y + \rho \dot{\phi}\mathbf{e}_\phi, \end{aligned} \quad (1.5.65)$$

$$v^2 = \dot{x}^2 + \dot{y}^2 + \dot{z}^2 = \dot{\rho}^2 + \dot{y}^2 + \rho^2 \dot{\phi}^2, \quad (1.5.66)$$

$$\mathbf{v} \cdot \mathbf{A} = \dot{x}A_x + \dot{y}A_y + \dot{z}A_z = \dot{\rho}A_\rho + \dot{y}A_y + \rho \dot{\phi}A_\phi, \quad (1.5.67)$$

where

$$A_\rho = \cos \phi A_x + \sin \phi A_z = \mathbf{e}_\rho \cdot \mathbf{A}, \quad (1.5.68)$$

$$A_\phi = -\sin \phi A_x + \cos \phi A_z = \mathbf{e}_\phi \cdot \mathbf{A}.$$

Using these results, show that the Lagrangian (5.1) can also be written in the form

$$L = -mc^2[1 - (\dot{\rho}^2 + \dot{y}^2 + \rho^2 \dot{\phi}^2)/c^2]^{1/2} - q\psi + q(\dot{\rho}A_\rho + \dot{y}A_y + \rho \dot{\phi}A_\phi). \quad (1.5.69)$$

The Hamiltonian H corresponding to the Lagrangian L given by (5.69) can now be found by the usual procedure. Show that for the conjugate momenta there are the results

$$\begin{aligned} p_\rho &= \frac{\partial L}{\partial \dot{\rho}} = \frac{m\dot{\rho}}{\sqrt{1 - (\dot{\rho}^2 + \dot{y}^2 + \rho^2 \dot{\phi}^2)/c^2}} + qA_\rho, \\ p_y &= \frac{\partial L}{\partial \dot{y}} = \frac{m\dot{y}}{\sqrt{1 - (\dot{\rho}^2 + \dot{y}^2 + \rho^2 \dot{\phi}^2)/c^2}} + qA_y, \\ p_\phi &= \frac{\partial L}{\partial \dot{\phi}} = \frac{m\rho^2 \dot{\phi}}{\sqrt{1 - (\dot{\rho}^2 + \dot{y}^2 + \rho^2 \dot{\phi}^2)/c^2}} + q\rho A_\phi. \end{aligned} \quad (1.5.70)$$

Finally, verify that H is given by the relation

$$\begin{aligned} H &= \dot{\rho}p_\rho + \dot{y}p_y + \dot{\phi}p_\phi - L \\ &= \{m^2 c^4 + c^2[(p_\rho - qA_\rho)^2 + (p_y - qA_y)^2 + (p_\phi/\rho - qA_\phi)^2]\}^{1/2} + q\psi. \end{aligned} \quad (1.5.71)$$

Here is a cautionary note: Let \mathbf{p} be the momentum vector as defined by (5.27). Then, from (5.65) and (5.70), verify that there are the results

$$p_\rho = \mathbf{p} \cdot \mathbf{e}_\rho, \quad (1.5.72)$$

$$p_y = \mathbf{p} \cdot \mathbf{e}_y, \quad (1.5.73)$$

but

$$p_\phi = \rho \mathbf{p} \cdot \mathbf{e}_\phi \neq \mathbf{p} \cdot \mathbf{e}_\phi. \quad (1.5.74)$$

1.5.5. Show that a uniform electric field in the z direction can be derived from the scalar and vector potentials

$$\psi = 0, \quad (1.5.75)$$

$$\mathbf{A} = -Ete_z.$$

1.5.6. Show that a uniform electric field in the z direction can be derived from the scalar and vector potentials

$$\psi = -Ez, \quad (1.5.76)$$

$$\mathbf{A} = 0.$$

1.5.7. Show that a uniform vertical magnetic field $\mathbf{B} = Be_y$, such as that produced by an idealized (normal) dipole bending magnet, can be derived from the scalar and vector potentials

$$\psi = 0, \quad (1.5.77)$$

$$\mathbf{A} = -Bxe_z.$$

Assuming the magnet has iron pole faces, sketch the pole faces and windings required to produce such a field, and label the pole faces N and S . Also sketch the magnetic field lines and the directions the current must flow in the windings.

1.5.8. Show that when cylindrical coordinates ρ, y, ϕ are used, a uniform magnetic field in the y direction can be derived from the scalar and vector potentials

$$\psi = 0, \quad (1.5.78)$$

$$\mathbf{A} = -(\rho/2)Be_\phi.$$

Answer: See Figure 5.1. From (5.62) one has the results

$$B_\rho = (\nabla \times \mathbf{A})_\rho = \frac{\partial A_\phi}{\partial y} - \frac{1}{\rho} \frac{\partial A_y}{\partial \phi} = 0, \quad (1.5.79)$$

$$B_y = (\nabla \times \mathbf{A})_y = \frac{1}{\rho} \frac{\partial A_\rho}{\partial \phi} - \frac{1}{\rho} \frac{\partial}{\partial \rho} (\rho A_\phi) = \frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\frac{\rho^2 B}{2} \right) = B,$$

$$B_\phi = (\nabla \times \mathbf{A})_\phi = \frac{\partial A_y}{\partial \rho} - \frac{\partial A_\rho}{\partial y} = 0.$$

1.5.9. Review Exercises 5.1 and 5.2. Suppose there is a *uniform* static magnetic field given by

$$\mathbf{B} = Be_y \text{ with } B > 0, \quad (1.5.80)$$

and no electric field. Consider charged-particle motion in this simple case. Show that a possible trajectory is uniform motion on a circle of radius ρ in the $y = 0$ plane, and show that $p = \|\mathbf{p}^{\text{mech}}\| = \|\gamma m\mathbf{v}\|$, the magnitude of the mechanical momentum given by (5.28), is related to B and ρ by the equation

$$B\rho = p/|q|. \quad (1.5.81)$$

The product $B\rho$ is called the *magnetic rigidity*. In Accelerator Physics it is common to characterize the mechanical momentum of a particle by its equivalent magnetic rigidity.

Show that, in the case of uniform circular motion, the circle is traced out with angular velocity ω given by the relation

$$\omega = |q|B/(\gamma m). \quad (1.5.82)$$

The quantity ω , particularly in the nonrelativistic limit $\gamma \simeq 1$, is called the *cyclotron frequency*.

1.5.10. Show that a magnetic quadrupole field with midplane ($\pm y$) symmetry can be derived from the scalar and vector potentials

$$\psi = 0, \quad (1.5.83)$$

$$\mathbf{A} = -(Q/2)(x^2 - y^2)\mathbf{e}_z.$$

Answer:

$$B_x = Qy, \quad (1.5.84)$$

$$B_y = Qx, \quad (1.5.85)$$

$$B_z = 0. \quad (1.5.86)$$

Assuming the quadrupole magnet has iron pole faces, sketch the pole faces and windings required to produce such a field, and label the pole faces N and S . Also sketch the magnetic field lines and the directions the current must flow in the windings.

1.5.11. Show that a magnetic sextupole field with midplane ($\pm y$) symmetry can be derived from the scalar and vector potentials

$$\psi = 0, \quad (1.5.87)$$

$$\mathbf{A} = -(S/3)(x^3 - 3xy^2)\mathbf{e}_z.$$

Assuming the sextupole magnet has iron pole faces, sketch the pole faces and windings required to produce such a field, and label the pole faces N and S . Also sketch the magnetic field lines and the directions the current must flow in the windings.

1.5.12. Let f be a function of the ℓ variables z_1, \dots, z_ℓ . The function f is said to be *homogeneous* of degree m if it satisfies the relation

$$f(\lambda z) = \lambda^m f(z) \text{ (for } \lambda > 0\text{).} \quad (1.5.88)$$

Evidently homogeneous polynomials provide examples of homogeneous functions. However, a function need not be polynomial to be homogeneous. Verify, for example, that the function

$$f = (ax^2 + bxy + cy^2)^{1/2} \quad (1.5.89)$$

is homogeneous of degree 1. Show that if f is homogeneous of degree m , then the functions $(\partial f / \partial z_j)$ are homogeneous of degree $(m - 1)$. Show that if f is homogeneous of degree m , then it satisfies *Euler's relation*

$$\sum_a z_a (\partial f / \partial z_a) = mf, \quad (1.5.90)$$

and conversely.

1.5.13. Given a Lagrangian L , one can find the associated Hamiltonian H by a Legendre transformation provided (5.6) is satisfied. Consider the inverse question. Given H , show that one can find an associated L using the *inverse* Legendre transformation provided by rewriting (5.8) in the form

$$L(q, \dot{q}, t) = \sum_i p_i \dot{q}_i - H(q, p, t) \quad (1.5.91)$$

with the proviso that

$$\dot{q}_i = \partial H / \partial p_i. \quad (1.5.92)$$

Show, as required, that the variables p can be eliminated in terms of the \dot{q} 's, provided

$$\det(\partial \dot{q}_i / \partial p_j) \neq 0. \quad (1.5.93)$$

Verify that

$$\partial \dot{q}_i / \partial p_j = \partial^2 H / \partial p_i \partial p_j \quad (1.5.94)$$

so that (5.93) is equivalent to the condition

$$\det(\partial^2 H / \partial p_i \partial p_j) \neq 0. \quad (1.5.95)$$

In analogy to (5.6), when (5.95) holds we will call this the nonsingular Hamiltonian case.

Suppose, as is true in these kinds of calculations, that the variables q are held fixed. Show that then, by the chain rule, there is the differential relation

$$d\dot{q}_i = \sum_j (\partial \dot{q}_i / \partial p_j) dp_j \quad (1.5.96)$$

which can be written in the matrix-vector form

$$d\dot{\mathbf{q}} = T d\mathbf{p} \quad (1.5.97)$$

where T is the matrix

$$T_{ij} = (\partial \dot{q}_i / \partial p_j). \quad (1.5.98)$$

Show, in view of (5.93), that (5.97) may be solved for the $d\mathbf{p}$ to yield the relation

$$d\mathbf{p} = T^{-1} d\dot{\mathbf{q}}. \quad (1.5.99)$$

Argue, on the other hand, that there is the relation

$$d\mathbf{p} = U d\dot{\mathbf{q}} \quad (1.5.100)$$

where U is the matrix

$$U_{ij} = (\partial p_i / \partial \dot{q}_j). \quad (1.5.101)$$

Verify that comparison of (5.99) and (5.100) gives the result

$$U = T^{-1}. \quad (1.5.102)$$

Finally, show that there is the two-directional logical implication

$$\det(\partial^2 L / \partial \dot{q}_i \partial \dot{q}_j) \neq 0 \Leftrightarrow \det(\partial^2 H / \partial p_i \partial p_j) \neq 0. \quad (1.5.103)$$

Thus, if a Legendre transformation can be made in one direction, it can also be made in the reverse direction. The nonsingular Lagrangian case leads to the nonsingular Hamiltonian case, and vice versa.

1.5.14. Review Subsection 5.2 and Exercise 5.13. Show that if the Hamiltonian $H(q, p, t)$ is homogeneous of degree one in the p_i , then (5.95) fails to hold, and we are dealing with the singular Hamiltonian case. Make an analysis of this case similar to that which was done for the Lagrangian case of Subsection 5.2. Show, in particular, that the Lagrangian associated with H vanishes identically.

1.5.15. Review Exercises 5.3 and 5.13. Show, by making an inverse Legendre transformation, that the Lagrangian associated with the Hamiltonian (5.49) is the Lagrangian (5.1).

1.5.16. Let $x(\tau), y(\tau)$ be a parameterized path in two-dimensional space. Let \mathcal{A} be the *distance* functional defined by

$$\mathcal{A} = \int ds = \int (ds/d\tau)d\tau = \int (\dot{x}^2 + \dot{y}^2)^{1/2}d\tau \quad (1.5.104)$$

where

$$(ds)^2 = (dx)^2 + (dy)^2 \quad (1.5.105)$$

and a dot denotes $d/d\tau$. Specifically, consider all paths for $\tau \in [0, 1]$ with the end points

$$x(0) = y(0) = 0, \quad (1.5.106)$$

$$x(1) = y(1) = 1. \quad (1.5.107)$$

Then we may write

$$\mathcal{A} = \int_0^1 L(\dot{x}, \dot{y})d\tau \quad (1.5.108)$$

with

$$L = (\dot{x}^2 + \dot{y}^2)^{1/2}. \quad (1.5.109)$$

Verify that L is homogeneous of degree one in the quantities \dot{x}, \dot{y} , and verify by direct calculation that (5.6) fails. Visualize the paths $x(\tau), y(\tau)$ as curves in the three-dimensional x, y, τ space. Show that there are an *infinity* of curves (corresponding to different parameterizations) that extremize \mathcal{A} . Show that all of these curves, when projected onto the x, y plane, fall on the straight line joining $(0, 0)$ to $(1, 1)$. Specifically, consider all curves of the form

$$x(\tau) = \tau + f(\tau), \quad (1.5.110)$$

$$y(\tau) = \tau + f(\tau), \quad (1.5.111)$$

where f is any function satisfying

$$f(0) = f(1) = 0, \quad (1.5.112)$$

$$|f'(\tau)| \leq 1. \quad (1.5.113)$$

Show that all these curves extremize \mathcal{A} . Show that for all these curves \mathcal{A} has the value

$$\mathcal{A} = \sqrt{2}. \quad (1.5.114)$$

1.5.17. Review Exercise 5.16. Instead of using the parameterization $x(\tau), y(\tau)$, simply write

$$y = y(x), \quad (1.5.115)$$

which is equivalent to taking the coordinate x to be the parameter,

$$\tau = x, \quad (1.5.116)$$

and thereby providing information about the parameterization.

Verify that in this case the distance functional takes the form

$$\mathcal{A} = \int ds = \int (ds/dx)dx = \int [1 + (y')^2]^{1/2} dx \quad (1.5.117)$$

where a prime denotes d/dx . Specifically, consider all paths $y(x)$ with the end points

$$y(0) = 0, \quad (1.5.118)$$

$$y(1) = 1. \quad (1.5.119)$$

Show that now we may write

$$\mathcal{A} = \int_0^1 L(y')dx \quad (1.5.120)$$

with

$$L = [1 + (y')^2]^{1/2}. \quad (1.5.121)$$

Verify that this L is *not* homogeneous of degree one in the quantity y' . Show that

$$p_y = \partial L / \partial y' = y' / [1 + (y')^2]^{1/2}, \quad (1.5.122)$$

and verify that this relation can be solved for y' in terms of p_y to give the result

$$y' = p_y / (1 - p_y^2)^{1/2}. \quad (1.5.123)$$

Therefore, we are dealing with the nonsingular case. Verify that, in fact,

$$\partial^2 L / (\partial y')^2 \neq 0 \quad (1.5.124)$$

so that (5.6) holds. Show that the Hamiltonian associated with the Lagrangian (5.121) is given by the relation

$$H = -(1 - p_y^2)^{1/2}. \quad (1.5.125)$$

Show that the solution to Lagrange's (or Hamilton's) equations in this case takes the form

$$y(x) = a + bx \quad (1.5.126)$$

where a and b are constants to be determined by the end-point conditions (5.118) and (5.119). Show that imposition of the end-point conditions yields the *unique* solution

$$y(x) = x, \quad (1.5.127)$$

the straight line (and therefore the shortest path) between the end points. Verify that for this path \mathcal{A} has the value (5.114).

1.5.18. Exercises 5.16 and 5.17 treated a simple example of finding *geodesics*, the shortest paths between two points, in terms of the distance functional. It involved the Lagrangians (5.109) and (5.121). Consider as before the parameterized path $x(\tau), y(\tau)$, but now employ instead the Lagrangian

$$\hat{L} = (1/2)(\dot{x}^2 + \dot{y}^2) \quad (1.5.128)$$

and seek to extremize what is called the *energy* functional $\hat{\mathcal{A}}$ defined by

$$\hat{\mathcal{A}} = \int \hat{L} d\tau. \quad (1.5.129)$$

The solution to this goal is an example of an *affine* geodesic. For a further description of geodesics and affine geodesics, see Exercise 6.16.

Verify that the Hessian of \hat{L} is invertible so that we are dealing with the nonsingular case. Show that the Lagrange equations associated with \hat{L} have, for the end-point conditions (5.106) and (5.107), the *unique* solution

$$x(\tau) = \tau, \quad (1.5.130)$$

$$y(\tau) = \tau. \quad (1.5.131)$$

Note that the extremizing path is again the straight line between the end points. Show that for this path $\hat{\mathcal{A}} = 1$.

1.5.19. This problem concerns fluid flow in two dimensions and its relation to Hamiltonian dynamics. Consider a fluid flowing in two dimensions and let $v_x(x, y, t)$ and $v_y(x, y, t)$ be the components of the velocity \mathbf{v} of a small portion of the fluid at the point with coordinates x and y . (It is assumed that there is no motion/velocity in the z direction and that \mathbf{v} does not depend on z .) We are interested in the solutions to the coupled pair of differential equations

$$\dot{x} = v_x(x, y, t), \quad (1.5.132)$$

$$\dot{y} = v_y(x, y, t). \quad (1.5.133)$$

Moreover assume that the flow is divergence free (which follows from the assumption that the fluid density remains constant, i.e., the flow is incompressible) so that

$$\nabla \cdot \mathbf{v} = \partial_x v_x + \partial_y v_y = 0. \quad (1.5.134)$$

Define an associated two-dimensional vector field $\mathbf{u}(x, y, t)$ by the rule

$$u_x = -v_y, \quad (1.5.135)$$

$$u_y = v_x. \quad (1.5.136)$$

Verify that (5.134) through (5.136) imply the relation

$$\partial_x u_y = \partial_x v_x = -\partial_y v_y = \partial_y u_x. \quad (1.5.137)$$

That is, the differential form associated with the vector field $\mathbf{u}(x, y, t)$ is closed. Consequently there is a function $\psi(x, y, t)$ defined by

$$\psi(x, y, t) = \int^{x,y} [u_x(x', y', t)dx' + u_y(x', y', t)dy'] \quad (1.5.138)$$

such that

$$u_x = \partial_x \psi, \quad (1.5.139)$$

$$u_y = \partial_y \psi. \quad (1.5.140)$$

See Exercise 6.1.1.

Verify that the results obtained so far can be combined to yield the differential equation pair

$$\dot{x} = \partial_y \psi, \quad (1.5.141)$$

$$\dot{y} = -\partial_x \psi. \quad (1.5.142)$$

Evidently these are Hamilton's equations with ψ playing the role of the Hamiltonian and x and y playing the roles of q and p .

In the case that \mathbf{v} is time independent, \mathbf{u} and therefore ψ will have no explicit time dependence. Then, because ψ is a Hamiltonian, there will be the relation

$$\psi\{x(t), y(t)\} = \text{constant} \quad (1.5.143)$$

on any solution of the pair (5.132) and (5.133). Call the pair $x(t)$ and $y(t)$ a *flow line*. According to (5.143), lines of constant ψ (*level lines* of ψ) are flow lines. For this reason (and the fact that Lagrange first arrived at this result in 1781) ψ is called a (Lagrange) *stream function*.

It is also possible to set up a stream function in three dimensions in the case of axial symmetry. The result is called a *Stokes* (1819-1903) stream function. Let ρ, ϕ, z be the usual choice of cylindrical coordinates with associated unit vectors $\mathbf{e}_\rho, \mathbf{e}_\phi, \mathbf{e}_z$. See (15.2.12) through (15.2.14), (15.2.20) through (15.2.25), and Exercise 15.2.2. Suppose the fluid velocity has only \mathbf{e}_ρ and \mathbf{e}_z components and does not depend on ϕ ,

$$\mathbf{v}(\rho, z, t) = v_\rho(\rho, z, t)\mathbf{e}_\rho + v_z(\rho, z, t)\mathbf{e}_z. \quad (1.5.144)$$

Recall that in general there is the relation

$$\mathbf{v} = d\mathbf{r}/dt = \dot{\rho}\mathbf{e}_\rho + \rho\dot{\phi}\mathbf{e}_\phi + \dot{z}\mathbf{e}_z \quad (1.5.145)$$

so that we are then interested in the solutions to the coupled pair

$$\dot{\rho} = v_\rho(\rho, z, t), \quad (1.5.146)$$

$$\dot{z} = v_z(\rho, z, t). \quad (1.5.147)$$

Again assume the flow is divergence free so that

$$\nabla \cdot \mathbf{v} = (1/\rho)\partial_\rho(\rho v_\rho) + \partial_z v_z = 0. \quad (1.5.148)$$

Multiply the last two pieces of (5.148) by ρ to get the result

$$\partial_\rho(\rho v_\rho) + \partial_z(\rho v_z) = 0. \quad (1.5.149)$$

Define a vector $\mathbf{u}(\rho, z, t)$ by the rule

$$u_\rho = -(\rho v_z), \quad (1.5.150)$$

$$u_z = (\rho v_\rho). \quad (1.5.151)$$

Verify from (5.149) through (5.151) that there is the relation

$$\partial_\rho u_z = \partial_\rho(\rho v_\rho) = -\partial_z(\rho v_z) = \partial_z u_\rho. \quad (1.5.152)$$

That is, the differential form associated with the vector field $\mathbf{u}(\rho, z, t)$ is closed. Consequently there is a function $\psi_S(\rho, z, t)$ defined by

$$\psi_S(\rho, z, t) = \int^{\rho, z} [u_\rho(\rho', z', t)d\rho' + u_z(\rho', z', t)dz'] \quad (1.5.153)$$

such that

$$u_\rho = \partial_\rho \psi_S, \quad (1.5.154)$$

$$u_z = \partial_z \psi_S. \quad (1.5.155)$$

Here we have added a subscript S to ψ to distinguish it from Lagrange's stream function and to honor Stokes.

Verify that the results obtained so far for the case of axial symmetry can be combined to yield the differential equation pair

$$\dot{\rho} = (1/\rho)\partial_z \psi_S, \quad (1.5.156)$$

$$\dot{z} = -(1/\rho)\partial_\rho \psi_S. \quad (1.5.157)$$

Because of the $(1/\rho)$ factor these are not Hamilton's equations. Nevertheless we will be able to draw from them similar conclusions.

In the case that \mathbf{v} is time independent, \mathbf{u} and therefore ψ_S will have no explicit time dependence. For this case let us compute the change in ψ_S along a flow line. Verify that so doing yields, with the aid of (5.156) and (5.157), the result

$$d\psi_S/dt = \partial_\rho \psi_S \dot{\rho} + \partial_z \psi_S \dot{z} = -\rho \dot{z} \dot{\rho} + \rho \dot{\rho} \dot{z} = 0. \quad (1.5.158)$$

Thus ψ_S , which is called the Stokes stream function, has the property that lines of constant ψ_S (level lines of ψ_S) are flow lines.

1.6 Hamilton's Equations with a Coordinate as an Independent Variable

In the usual Hamiltonian formulation (as in the usual Lagrangian formulation) the time t plays the distinguished role of an *independent* variable, and all the q 's and p 's are *dependent* variables. That is, the canonical variables are viewed as functions $q(t), p(t)$ of the independent variable t .

In some cases, it is more convenient to take some coordinate to be the independent variable rather than the time. So doing may facilitate the use of maps. For example, consider the passage of a collection of particles through a rectangular magnet such as is shown in Figures 6.1 and 6.2. In such a situation, particles with different initial conditions will require different times to pass through the magnet. If the quantities of interest are primarily the locations and momenta of the particles as they leave the exit face of the magnet, then it would clearly be more convenient to use some coordinate that measures the progress of a particle through the magnet as an independent variable. With such a choice, the relation between entering coordinates and momenta and exiting coordinates and momenta could be treated as a transfer map.

In the case of a magnet with parallel faces as shown in Figures 6.1 and 6.2, a convenient independent variable would be the z coordinate. In the case of a wedge magnet as shown in Figure 6.3, a convenient independent variable would be the angle ϕ of a cylindrical coordinate triad ρ, y, ϕ . See Exercise 5.4.

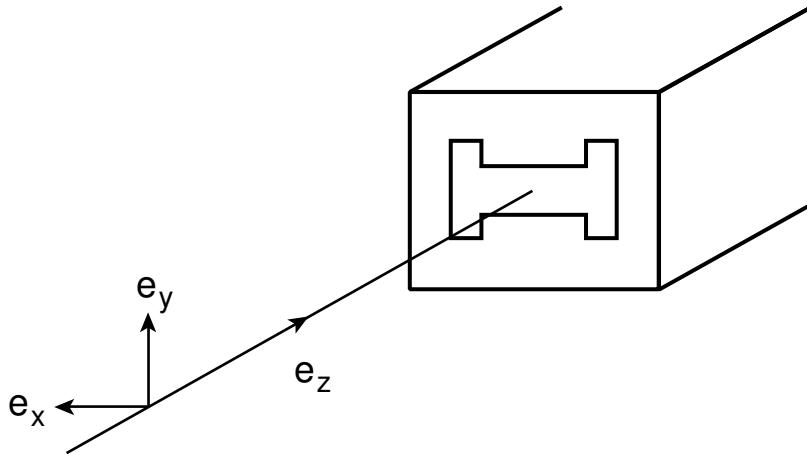


Figure 1.6.1: Typical choice of a Cartesian coordinate system for the description of charged-particle trajectories in a magnet.

Suppose some coordinate is indeed chosen to be the independent variable. Is it then still possible to have a Hamiltonian (or Lagrangian) formulation of the equations of motion? The answer in general is *yes* as is shown by the following theorem:

Theorem 1.6.1. *Suppose $H(q, p, t)$ is a Hamiltonian for a system having n degrees of freedom. Suppose further that $\dot{q}_1 = \partial H / \partial p_1 \neq 0$ for some interval of time T in some region R of the phase space described by the $2n$ variables (q_1, \dots, q_n) and (p_1, \dots, p_n) . Then in*

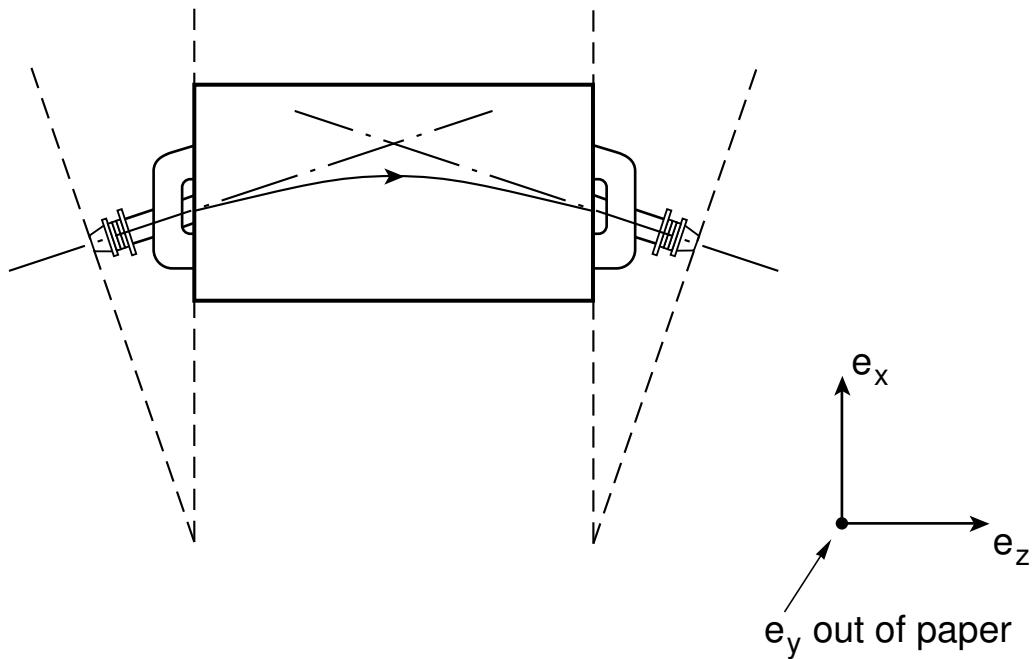


Figure 1.6.2: Top view of a particle trajectory in a rectangular magnet.

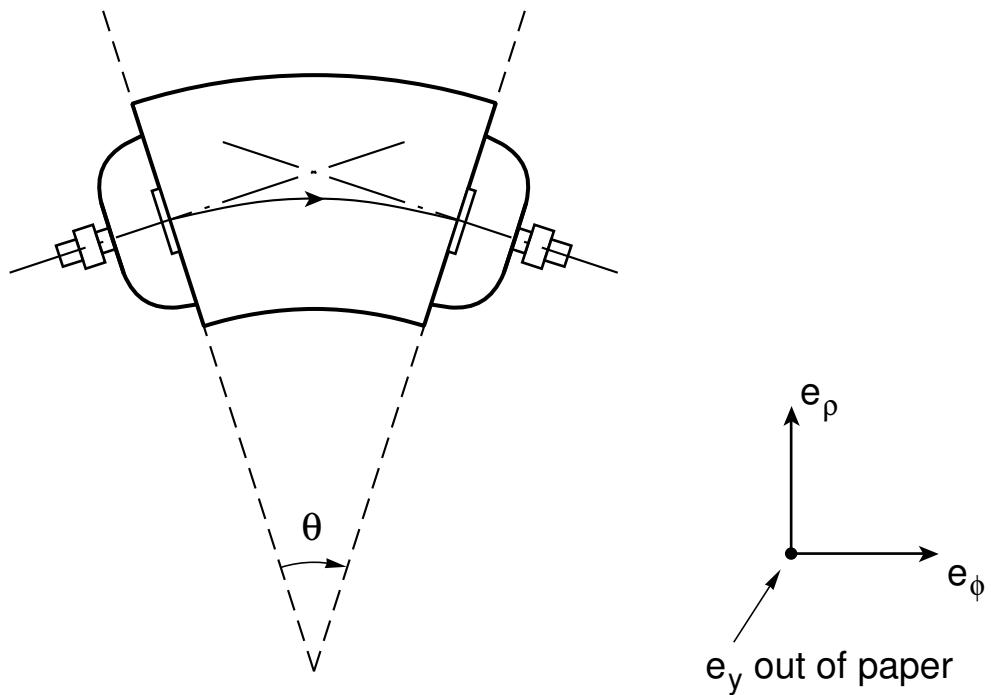


Figure 1.6.3: Top view of a particle trajectory in a wedge magnet. The trajectory is conveniently described using the cylindrical coordinates ρ, y, ϕ . See Figure 5.1.

this region and time interval, q_1 can be introduced as an independent variable in place of the time t . Moreover, the equations of motion with q_1 as an independent variable can be obtained from a Hamiltonian that will be called K .

Proof. Consider the $2n - 2$ quantities (q_2, \dots, q_n) and (p_2, \dots, p_n) . They obey Hamilton's equations of motion

$$\begin{aligned}\dot{q}_i &= \partial H / \partial p_i, \quad i = 2, \dots, n, \\ \dot{p}_i &= -\partial H / \partial q_i, \quad i = 2, \dots, n.\end{aligned}\tag{1.6.1}$$

Denote total derivatives with respect to q_1 by a prime. Then, applying the chain rule to equations (6.1), one finds the relations

$$\begin{aligned}q'_i &= dq_i/dq_1 = (dq_i/dt)(dt/dq_1) = (\partial H / \partial p_i)(\partial H / \partial p_1)^{-1}, \\ p'_i &= dp_i/dq_1 = (dp_i/dt)(dt/dq_1) = -(\partial H / \partial q_i)(\partial H / \partial p_1)^{-1}.\end{aligned}\tag{1.6.2}$$

To these $2n - 2$ relations it is convenient to add two more. First, suppose the time t is added to the list of *coordinates* as a *dependent* variable. Then one immediately has the relation

$$t' = dt/dq_1 = (dq_1/dt)^{-1} = (\partial H / \partial p_1)^{-1}.\tag{1.6.3}$$

Second, suppose the quantity p_t defined by writing $p_t = -H$ is formally added to the list of momenta. Then, using (5.11) and (5.14), one finds the relation

$$p'_t = dp_t/dq_1 = (dp_t/dt)(dt/dq_1) = -(\partial H / \partial t)(\partial H / \partial p_1)^{-1}.\tag{1.6.4}$$

Equations (6.2) through (6.4) are the desired equations of motion for the $2n$ variables (t, q_2, \dots, q_n) and (p_t, p_2, \dots, p_n) with q_1 as an independent variable. What remains to be shown is that the quantities on the right sides of these equations can be obtained by applying the standard rules to some Hamiltonian K .

Look once again at the defining relation for p_t ,

$$p_t = -H(q, p, t).\tag{1.6.5}$$

Suppose that this relation is solved for p_1 to give a relation of the form

$$p_1 = -K(t, q_2, \dots, q_n; p_t, p_2, \dots, p_n; q_1).\tag{1.6.6}$$

Such an inversion is possible according to the inverse function theorem because $\partial H / \partial p_1 \neq 0$ by assumption. Then, as the notation is intended to suggest, K is the desired new Hamiltonian.

To see that this is so, compute the total differential of (6.5) to find the relation

$$dp_t = -(\partial H / \partial t)dt - \sum_i (\partial H / \partial q_i)dq_i - \sum_i (\partial H / \partial p_i)dp_i.\tag{1.6.7}$$

Now solve (6.7) for dp_1 to get the relation

$$dp_1 = \left(\frac{\partial H}{\partial p_1} \right)^{-1} \left[-dp_t - (\partial H / \partial t)dt - \sum_i (\partial H / \partial q_i)dq_i - \sum_{i \neq 1} (\partial H / \partial p_i)dp_i \right].\tag{1.6.8}$$

Also, compute the total differential of (6.6) to find the relation

$$dp_1 = -(\partial K/\partial p_t)dp_t - (\partial K/\partial t)dt - \sum_i (\partial K/\partial q_i)dq_i - \sum_{i \neq 1} (\partial K/\partial p_i)dp_i. \quad (1.6.9)$$

Upon comparing (6.8) and (6.9), and looking at equations (6.1–6.4), one obtains the advertised result:

$$\begin{aligned} \partial K/\partial p_t &= (\partial H/\partial p_1)^{-1} = t', \\ \partial K/\partial p_i &= (\partial H/\partial p_i)(\partial H/\partial p_1)^{-1} = q'_i, \quad i = 2, \dots, n, \\ \partial K/\partial t &= (\partial H/\partial t)(\partial H/\partial p_1)^{-1} = -p'_t, \\ \partial K/\partial q_i &= (\partial H/\partial q_i)(\partial H/\partial p_1)^{-1} = -p'_i, \quad i = 2, \dots, n. \end{aligned} \quad (1.6.10)$$

That is, the indicated partial derivates of K do indeed produce the required right sides of equations (6.2) through (6.4). Note that according to equations (6.10), the quantity p_t may be viewed as the momentum canonically conjugate to the time t . \square

How might one have guessed that (6.6) gives the desired Hamiltonian? One way is to employ (a modified) Hamilton's principle. According to this principle, the *action* \mathcal{A} associated with a path in phase space should be defined by the relation

$$\mathcal{A} = \int dt \left(\sum_{i=1}^n p_i \dot{q}_i - H \right) = \int \left(\sum_{i=1}^n p_i dq_i - H dt \right); \quad (1.6.11)$$

and the equations of motion (5.11) through (5.14) follow from requiring that \mathcal{A} be an extremum,

$$\delta \mathcal{A} = 0, \quad (1.6.12)$$

and use of the calculus of variations. Now introduce the notation

$$q_{n+1} = t, \quad p_{n+1} = -H = p_t. \quad (1.6.13)$$

With this notation the action (6.11) takes the symmetrical form

$$\mathcal{A} = \int \sum_{i=1}^{n+1} p_i dq_i. \quad (1.6.14)$$

In this form it is evident that we may regard any of the p_i as being related to some Hamiltonian. Suppose we choose p_1 , and then write (6.6). When this is done, \mathcal{A} takes the form

$$\begin{aligned} \mathcal{A} &= \int \sum_{i=2}^{n+1} p_i dq_i - K dq_1 = \int dq_1 \left[\sum_{i=2}^{n+1} p_i (dq_i/dq_1) - K \right] \\ &= \int dq_1 \left(\sum_{i=2}^{n+1} p_i q'_i - K \right). \end{aligned} \quad (1.6.15)$$

Since the requirement (6.12) is intrinsic in nature and therefore coordinate independent, it must also hold when \mathcal{A} is written in the form (6.15). (An extremum is an extremum independent of parameterization.) But then use of (6.12), and application of the calculus of variations to (6.15), give (6.10).

Exercises

1.6.1. Find the Hamiltonian K corresponding to the Hamiltonian H given by (5.49) when the z coordinate is taken to be the independent variable. Assume that $\dot{z} > 0$ for the trajectories in question. Answer:

$$K = -[(p_t + q\psi)^2/c^2 - m^2c^2 - (p_x - qA_x)^2 - (p_y - qA_y)^2]^{1/2} - qA_z. \quad (1.6.16)$$

Here the quantities p_x and p_y denote *canonical* momenta. Note that according to (6.5), p_t is usually negative. Show, using (5.50), that

$$p_t = -[m^2c^4 + c^2(\mathbf{p}^{\text{mech}} \cdot \mathbf{p}^{\text{mech}})]^{1/2} - q\psi = -\gamma mc^2 - q\psi. \quad (1.6.17)$$

1.6.2. Find the Hamiltonian K corresponding to the Hamiltonian H given by (5.71) when the coordinate ϕ is taken to be the independent variable. Assume that $\dot{\phi} > 0$ for trajectories of interest. Answer:

$$K = -\rho[(p_t + q\psi)^2/c^2 - m^2c^2 - (p_\rho - qA_\rho)^2 - (p_y - qA_y)^2]^{1/2} - q\rho A_\phi. \quad (1.6.18)$$

Here the quantities p_ρ and p_y denote *canonical* momenta. Verify that (6.17) continues to hold.

1.6.3. The derivation of (6.10) based on the modified Hamilton's principle, Equations (6.11) through (6.15), is a bit heuristic. Make the derivation more precise by indicating exactly what changes of variables are being made; what the limits of integration are in (6.11), (6.14), and (6.15); what (6.12) means; etc. Begin your discussion by reviewing exactly how (5.11) and (5.14) follow from (6.11) and (6.12). Hint: To derive (5.14) from Hamilton's principle, consider variations in t as well as those in the q_i and p_i . That is, introduce a new independent variable τ such that the dependent variables are parameterized in the form $q_i(\tau)$, $p_i(\tau)$, $t(\tau)$.

1.6.4. How might one have guessed that p_t should be defined as in (6.5)? According to Hamilton's principle stated in Lagrangian terms, the action \mathcal{A} associated with a path in configuration space is given by the relation

$$\mathcal{A} = \int_{t^1}^{t^2} L(q, \dot{q}, t) dt. \quad (1.6.19)$$

Suppose we introduce a new independent variable τ such that the time t and the other dependent variables are parameterized in the form $t(\tau)$, $q_i(\tau)$. Then, using a prime to denote differentiation with respect to τ , we have the relation

$$dt = (dt/d\tau)d\tau = t'd\tau, \quad (1.6.20)$$

$$\dot{q}_i = dq_i/dt = (dq_i/d\tau)(d\tau/dt) = q'_i/t'. \quad (1.6.21)$$

Correspondingly, the action (6.18) takes the form

$$\mathcal{A} = \int L dt = \int Lt' d\tau = \int [L(q, q'/t', t)t'] d\tau, \quad (1.6.22)$$

and we see that in terms of τ there is an *effective* Lagrangian L^{eff} given by the expression

$$L^{\text{eff}}(q, t; q', t') = L(q, q'/t', t)t'. \quad (1.6.23)$$

Justify this assertion by treating all the necessary details. (See the analogous case of Exercise 6.3.) Following the standard procedure (5.7), the momentum p_t canonically conjugate to the variable t is defined by the relation

$$p_t = \partial L^{\text{eff}} / \partial t'. \quad (1.6.24)$$

By using (6.23) and (6.24) and the chain rule show that

$$p_t = L - \sum_i p_i \dot{q}_i = -H. \quad (1.6.25)$$

The Lagrange equation for the t coordinate is

$$\frac{d}{d\tau} \frac{\partial L^{\text{eff}}}{\partial t'} - \frac{\partial L^{\text{eff}}}{\partial t} = 0. \quad (1.6.26)$$

Show that p_t is conserved if L (and therefore L^{eff}) does not explicitly contain the time t . In view of (6.25), we may say that energy (the Hamiltonian) is conserved if the time is an *ignorable* coordinate. Use (5.14) to obtain the same result.

1.6.5. Review Exercise 6.4. Suppose we wish to find the Hamiltonian H^{eff} associated with L^{eff} . To do so we must first compute all the conjugate momenta p_i^{eff} . Using (6.23), show that

$$p_i^{\text{eff}} = \partial L^{\text{eff}} / \partial q'_i = \partial L / \partial \dot{q}_i = p_i. \quad (1.6.27)$$

Next, following the rule (5.8), find the result

$$\begin{aligned} H^{\text{eff}} &= p_t t' + \sum_i p_i^{\text{eff}} q'_i - L^{\text{eff}} = p_t t' + \sum_i p_i \dot{q}_i t' - Lt' \\ &= t'(p_t + H). \end{aligned} \quad (1.6.28)$$

At this stage, two complications arise: First, in view of (6.25) and (6.27), it is evident that p_t does not depend on t' , and therefore the Jacobian determinant (5.9) vanishes. Verify this assertion. Second, because of (6.25), we see from (6.28) that H^{eff} vanishes identically.

These complications should not surprise us. Review Subsection 5.2. Show that L^{eff} as given by (6.23) is homogeneous of degree one in the velocities and does not explicitly depend on τ . Therefore, these complications must occur.

What to do? Some further information has to be provided about the parameterization. Suppose we make the dependence of t on τ a bit more explicit by writing a relation of the form

$$d\tau = f(q, p, t)dt \quad (1.6.29)$$

where f is a function to be specified. Then, by the chain rule, we have the relations

$$q'_j = (dq_j/dt)(dt/d\tau) = (1/f)(\partial H/\partial p_j),$$

$$\begin{aligned} t' &= (1/f), \\ p'_j &= (dp_j/dt)(dt/d\tau) = -(1/f)(\partial H/\partial q_j), \\ H' &= (dH/dt)(dt/d\tau) = (1/f)(\partial H/\partial t). \end{aligned} \quad (1.6.30)$$

In the last of these equations use has been made of (5.14). Is there a Hamiltonian that will produce these equations?

There is. Inspired by (6.28), *define* an effective Hamiltonian \bar{H}^{eff} [on the *extended* $(2n+2)$ -dimensional phase space consisting of the variables q_1, q_2, \dots, q_n, t and $p_1, p_2 \dots p_n, p_t$] by writing

$$\bar{H}^{\text{eff}}(q, t; p, p_t) = (1/f)(p_t + H) \quad (1.6.31)$$

where the relation (6.25) is to be ignored (but soon recovered as a special case).⁵¹ Then, taking partial derivatives, we find the results

$$\begin{aligned} \partial \bar{H}^{\text{eff}} / \partial p_j &= (1/f)(\partial H / \partial p_j) + (p_t + H)[\partial(1/f) / \partial p_j] \\ &= q'_j + (p_t + H)[\partial(1/f) / \partial p_j], \\ \partial \bar{H}^{\text{eff}} / \partial p_t &= (1/f) = t', \\ \partial \bar{H}^{\text{eff}} / \partial q_j &= (1/f)(\partial H / \partial q_j) + (p_t + H)[\partial(1/f) / \partial q_j] \\ &= -p'_j + (p_t + H)[\partial(1/f) / \partial q_j], \\ \partial \bar{H}^{\text{eff}} / \partial t &= (1/f)(\partial H / \partial t) + (p_t + H)[\partial(1/f) / \partial t] \\ &= H' + (p_t + H)[\partial(1/f) / \partial t]. \end{aligned} \quad (1.6.32)$$

Here, we have also used (6.30). Next observe that \bar{H}^{eff} does not depend on the independent variable τ , and therefore must be *constant* on each trajectory it generates. Consider those trajectories on which $\bar{H}^{\text{eff}} = 0$. Then for those trajectories (6.25) holds and the relations (6.32) take the form

$$\begin{aligned} q'_j &= \partial \bar{H}^{\text{eff}} / \partial p_j, \\ t' &= \partial \bar{H}^{\text{eff}} / \partial p_t, \\ p'_j &= -\partial \bar{H}^{\text{eff}} / \partial q_j, \\ p'_t &= -\partial \bar{H}^{\text{eff}} / \partial t. \end{aligned} \quad (1.6.33)$$

Thus, a *special* class of trajectories generated by \bar{H}^{eff} , namely those for which $\bar{H}^{\text{eff}} = 0$, gives $q(\tau)$, $t(\tau)$, $p(\tau)$, and $p_t(\tau)$.

A particularly simple case is to set $f = 1$ so that

$$t' = (1/f) = 1. \quad (1.6.34)$$

In this case find the result

$$\bar{H}^{\text{eff}}(q, t; p, p_t) = p_t + H(q, p, t). \quad (1.6.35)$$

⁵¹The transformation (6.31) is sometimes called a *Poincaré transformation*, is useful for *regularization*, but should not be confused with the Poincaré transformations of Relativity Theory. See Chapter 28.

Show, by one of Hamilton's equations, that there is now the relation

$$t' = \partial \bar{H}^{\text{eff}} / \partial p_t = 1, \quad (1.6.36)$$

which is consistent with the requirement (6.34).

Suppose, for any choice of f , we consider trajectories in q, t, p, p_t space generated by \bar{H}^{eff} for which the initial conditions happen to satisfy the relation

$$p_t = -H \quad (1.6.37)$$

at some initial value of τ . Then $\bar{H}^{\text{eff}} = 0$ at this value of τ . But since \bar{H}^{eff} is constant on trajectories, (6.37) must then hold all along such trajectories.

One moral of this exercise is that a nonautonomous Hamiltonian system can always be converted into an autonomous one in an extended phase space (with the two additional phase-space variables t, p_t) by use of (6.35) or, more generally, (6.31). Another is that the time t can be replaced by a new independent variable τ , while remaining within a Hamiltonian framework, such that a relation of the form (6.29) holds. Such a replacement may be useful for *regularization*. See Subsection 2.7.4 and the regularization references at the end of Chapter 2.

1.6.6. Read Exercise 6.4. Let K be the Hamiltonian defined by (6.6). By reversing the Legendre transformation that relates a Lagrangian and a Hamiltonian, see (5.8), show that the Lagrangian L_K associated with K is given by the relation

$$L_K = p_t t' + \sum_{i=2}^n p_i q'_i - K = (\dot{q}_1)^{-1} L. \quad (1.6.38)$$

Suppose we rewrite (6.19) in the form

$$\mathcal{A} = \int L dt = \int L(dt/dq_1)dq_1 = \int L^{\text{eff}} dq_1, \quad (1.6.39)$$

with

$$L^{\text{eff}} = L(dt/dq_1). \quad (1.6.40)$$

Verify the relation

$$L_K = L^{\text{eff}}. \quad (1.6.41)$$

Conversely show that, starting with the Lagrangian L_K defined by (6.40), one arrives at the Hamiltonian K defined by (6.6).

1.7 Manifestly Lorentz Invariant Formulation of Equations of Motion

The expression (5.1) for the Lagrangian L and the expression (5.49) for the Hamiltonian H lead to the correct equations of motion for charged-particles moving in electromagnetic fields. Review Exercises 5.2 and 5.3. However they do not seem particularly aesthetically

pleasing because they contain a square root and because they are not manifestly Lorentz invariant. The purpose of this section is to explore other possible Lagrangians, and to show that the particular forms of the Lagrangian (5.1) and the Hamiltonian (5.49) come about because of a decision to treat time as an independent variable, and the spatial coordinates as dependent variables. We will also find other interesting results along the way.

1.7.1 Relativistic Preliminaries

We assume the reader has some prior knowledge of the Theory of Special Relativity. However, partially to establish a standard nomenclature, we will review some of the basic concepts of the theory. For example, we will take special care to deal properly with “down” (covariant) and “up” (contravariant) indices.⁵² Also, some of the exposition will consist of statements to be verified by the reader. Finally, detailed information about the Lorentz and related groups is presented in Chapter 28.

In the spirit of relativity, and following the insight of Hermann Minkowski (1864-1909), it is reasonable to try to treat space and time on a similar footing. Suppose the world line of a particle through space-time is parameterized in terms of some parameter τ by specifying four functions $x^\mu(\tau)$ that, taken together, form a vector with four contravariant components x^μ . We adopt the convention that the first three components of x^μ are the spatial coordinates of the particle, and the fourth (with a factor of c) is its temporal coordinate. Specifically, we write $\mu = 1, 2, 3, 4$ with $x^4 = ct$. That is, we write

$$x^\mu = (x, y, z, ct) = (\mathbf{r}, ct). \quad (1.7.1)$$

We will define associated covariant quantities x_μ shortly. See (7.9).

Introduction of a Metric Tensor and associated Infinitesimal Space-Time Interval

Central to the Theory of Special Relativity is the concept of a metric tensor $g_{\mu\nu}$ and an associated infinitesimal space-time interval ds^2 . In Cartesian coordinates and for *flat* (not *curved*) space-time, only the diagonal entries of g are nonzero, and we take them to have the *constant* values

$$\begin{aligned} g_{11} &= g_{22} = g_{33} = -1, \\ g_{44} &= 1. \end{aligned} \quad (1.7.2)$$

Correspondingly, the associated space-time *infinitesimal interval* ds^2 is taken to be given by the relation

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = -(dx^1)^2 - (dx^2)^2 - (dx^3)^2 + (dx^4)^2 = -(d\mathbf{r})^2 + c^2(dt)^2. \quad (1.7.3)$$

⁵²Is there an easy way to remember the association between down/up and covariant/contravariant? Here is one way: The third letter from the left of the word covariant is a v, which may be viewed as the tip of a *downward* pointing arrow. Correspondingly, covariant components have down indices. And the third letter from the left in the word contravariant is an n, which, with somewhat more imagination, may be viewed as the tip of a blunted *upward* pointing arrow. Correspondingly, contravariant components have up indices.

In writing (7.3) we have employed the usual Einstein convention that repeated indices are to be summed over. Moreover, we have used the adjective *infinitesimal* to emphasize that ds^2 is small, and also because in Subsection 8.3 we will define a related *net interval* $I(*, *)$ that is related to ds^2 but need not be small. Finally we remark that the notation ds^2 appearing in (7.3), although universally employed, can be misleading since, depending on circumstances, ds^2 can be negative, zero, or positive, and therefore is not necessarily the square of anything. But note that $ds^2 > 0$ for time-like displacements. Space-time endowed with the metric (7.2) is sometimes called *Minkowski* space.⁵³

For future use we also define quantities $g^{\mu\nu}$ by the rule

$$g^{\mu\nu} = (g^{-1})_{\mu\nu}. \quad (1.7.4)$$

For the choice (7.2) use of (7.4) immediately yields the relation

$$g^{\mu\nu} = g_{\mu\nu}. \quad (1.7.5)$$

We also note that if the quantities $g^{\mu\nu}$ are viewed as the entries of a 4×4 matrix g , then g has the properties

$$g^T = g, \quad \det(g) = -1, \quad g^2 = I \Leftrightarrow g^{-1} = g. \quad (1.7.6)$$

The metric tensor can be used to raise and lower indices. That is, “up” index quantities can be defined given “down” index quantities, and vice versa. For example there are the definitions/relations

$$\begin{aligned} x_\mu &= g_{\mu\nu}x^\nu, \quad x^\mu = g^{\mu\nu}x_\nu, \\ w_\mu x^\mu &= g_{\mu\nu}w^\nu x^\mu = w^\nu g_{\nu\mu}x^\mu = w^\nu x_\nu = w^\mu x_\mu, \end{aligned} \quad (1.7.7)$$

$$g_\mu{}^\sigma = g_{\mu\nu}g^{\nu\sigma} = \delta_\mu^\sigma. \quad (1.7.8)$$

In particular, x_μ has the entries

$$x_\mu = (-x, -y, -z, ct) = (-\mathbf{r}, ct). \quad (1.7.9)$$

Indices on *tensors* with multiple indices can be raised and lowered in an analogous fashion. See, for example, (7.36). Also, as is easily checked using (7.7), the operations of raising and lowering are inverses of each other. For example suppose some “down” index, on some tensor having a down index, is raised to become an “up” index, and is then subsequently lowered. The net result is the original tensor. Finally, we note that these raising and lowering rules, *per se*, have nothing to do with Special Relativity, but rather are a convenient bookkeeping device.

⁵³Many authors adopt the convention $\mu = 0, 1, 2, 3$ with $x^0 = ct$ and the remaining x^μ being the spatial coordinates. Accordingly, they would write $x^\mu = (ct, x, y, z) = (ct, \mathbf{r})$ and $g = \text{diag}(1, -1, -1, -1)$. The relation (7.3) between its far left and far right sides holds in either case.

Introduction of Lorentz Transformations

The second central ingredient for the Theory of Special Relativity is the concept of a Lorentz transformation. It is a transformation, often assumed to be a *linear* transformation, of space-time into itself with special properties.⁵⁴ Let Λ be a linear transformation described by a 4×4 matrix which, under the assumption of linearity, sends space-time points x to the points \bar{x} by the rule

$$\bar{x} = \Lambda x \Leftrightarrow \bar{x}^\mu = \sum_\nu \Lambda^{\mu\nu} x^\nu. \quad (1.7.10)$$

Then by the assumption of linearity, which implies that Λ does not depend on x , there is the relation

$$d\bar{x} = \Lambda dx \Leftrightarrow d\bar{x}^\mu = \sum_\nu \Lambda^{\mu\nu} dx^\nu. \quad (1.7.11)$$

Let us examine the effect of Λ on the infinitesimal space-time interval ds^2 . Evidently, employing (7.3), it becomes the transformed interval $(ds^2)^{\text{tran}}$ given by

$$(ds^2)^{\text{tran}} = g_{\mu\nu} d\bar{x}^\mu d\bar{x}^\nu. \quad (1.7.12)$$

Observe that ds^2 can be written in the vector-matrix inner product form

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = g^{\mu\nu} dx^\mu dx^\nu = (dx, gdx). \quad (1.7.13)$$

Similarly, we may write

$$(ds^2)^{\text{tran}} = g_{\mu\nu} d\bar{x}^\mu d\bar{x}^\nu = (d\bar{x}, gd\bar{x}). \quad (1.7.14)$$

Next employ (7.11) in (7.14) to find

$$(ds^2)^{\text{tran}} = (d\bar{x}, gd\bar{x}) = (\Lambda dx, g\Lambda dx) = (dx, \Lambda^T g \Lambda dx). \quad (1.7.15)$$

We are ready for the master step that, we will see, specifies Λ : we require that

$$(ds^2)^{\text{tran}} = ds^2 \Leftrightarrow (dx, \Lambda^T g \Lambda dx) = (dx, gdx). \quad (1.7.16)$$

That is, the infinitesimal space-time interval should be *invariant* under a Lorentz transformation.

To extract/explicate the consequences of this requirement, define a matrix Δ by the rule

$$\Delta = \Lambda^T g \Lambda - g. \quad (1.7.17)$$

Note/verify for future use that because g is symmetric, see (7.6), so is Δ ,

$$\Delta^T = \Delta. \quad (1.7.18)$$

With the definition (7.17), the requirement (7.16) is equivalent to the requirement

$$(dx, \Delta dx) = 0. \quad (1.7.19)$$

⁵⁴The assumption of linearity is not necessary. With the use of affine geodesics in Minkowski space and a related definition of what is meant by a Lorentz transformation, it can be shown to be linear. See Subsection 8.4.

Next, to exploit (7.19), employ a variant of a mathematical move called *polarization*. Suppose u and v are any two space-time points and let x be the sum

$$x = u + v. \quad (1.7.20)$$

[Note that in writing (7.20) we have assumed that space-time has a vector-space structure.] For x given by (7.20) there is the differential result

$$dx = du + dv \quad (1.7.21)$$

and (7.19) becomes

$$\begin{aligned} 0 &= ([du + dv], \Delta[du + dv]) = \\ &= (du, \Delta du) + [(du, \Delta dv) + (dv, \Delta du)] + (dv, \Delta dv). \end{aligned} \quad (1.7.22)$$

Observe that the first and last terms on the right side of (7.22) vanish because of (7.19). Also, by (7.18), there is the relation

$$(dv, \Delta du) = (\Delta du, dv) = (du, \Delta^T dv) = (du, \Delta dv). \quad (1.7.23)$$

Taking these results into account, (7.22) becomes the relation

$$2(du, \Delta dv) = 0, \quad (1.7.24)$$

from which it follows, since du and dv are arbitrary, that

$$\Delta = 0. \quad (1.7.25)$$

Correspondingly, Λ must satisfy the key relation

$$\Lambda^T g \Lambda = g. \quad (1.7.26)$$

Subsequently we will learn that matrices Λ that satisfy (7.26) exist, and form a 6-dimensional group called the Lorentz group. Surprisingly, we will also learn that the Lorentz group is closely related to $Sp(2, \mathbb{C})$, the group of 2×2 *symplectic* matrices with complex entries. We may say that Special Relativity has a symplectic flavor.

As a first step in verifying that Lorentz transformations form a group, and for further use in Subsection 8.5, let us momentarily pause to verify that Lorentz transformation matrices Λ are invertible. Form the determinant of both sides (7.26) to find that

$$\det(\Lambda^T g \Lambda) = \det(g) \Leftrightarrow \det(\Lambda^T) \det(g) \det(\Lambda) = \det(g) \Leftrightarrow [\det(\Lambda)]^2 = 1 \Leftrightarrow \det(\Lambda) = \pm 1. \quad (1.7.27)$$

Here we have used (7.6) and the fact that $\det(M^T) = \det(M)$ for any matrix M .

In summary, and to return to the main discussion, transformation of a space-time point x with contravariant components x^ν into a related point \bar{x} is defined to be a Lorentz transformation if x and \bar{x} are connected by a relation of the form

$$\bar{x} = \Lambda x \Leftrightarrow \bar{x}^\mu = \sum_\nu \Lambda^{\mu\nu} x^\nu \quad (1.7.28)$$

with Λ being a matrix that satisfies (7.26). Moreover, if some 4-component entity transforms in the manner (7.28), we say that it is a *4-vector*. By their transformation properties ye shall know them.

Let us turn matters around. Suppose w and x are two 4-vectors. Define a “relativistic” dot product between them by the rule

$$w \cdot x = (w, gx) = \sum_{\mu\nu} w^\mu g^{\mu\nu} x^\nu = \sum_{\mu\nu} w^\mu g_{\mu\nu} x^\nu = w^\mu x_\mu = w_\nu x^\nu = \sum_{\mu\nu} w_\mu g^{\mu\nu} x_\nu. \quad (1.7.29)$$

Here we have also used (7.5). Next assume that w and x are acted upon by a Lorentz transformation described by Λ to become \bar{w} and \bar{x} . Verify that

$$\bar{w} \cdot \bar{x} = (\Lambda w, g\Lambda x) = (w, \Lambda^T g \Lambda x) = (w, gx) = w \cdot x. \quad (1.7.30)$$

Here we have used (7.26). Evidently, Lorentz transformations preserve the relativistic 4-vector dot product; and that is their essential property since the relations (7.26) and (7.30) are logically equivalent. A quantity that is unchanged under a Lorentz transformation is called a *scalar*.

With the experience we have now accumulated, we are able to re-examine the effect of a Lorentz transformation on the infinitesimal space-time interval ds^2 . Denote the *transformed* space-time interval by $(ds^2)^{\text{tran}}$. Verify that

$$(ds^2)^{\text{tran}} = g_{\mu\nu} d\bar{x}^\mu d\bar{x}^\nu = d\bar{x} \cdot d\bar{x} = dx \cdot dx = ds^2. \quad (1.7.31)$$

Here we have used the infinitesimal version of (7.30). We again see that Lorentz transformations preserve the infinitesimal space-time interval, now as a consequence of (7.26).

More about World Lines

We are now prepared to continue our study of world lines. Let $(x')^\mu$ denote the four quantities defined by the equations

$$(x')^\mu = dx^\mu / d\tau. \quad (1.7.32)$$

Under the assumption that the parameterization is unchanged by a Lorentz transformation, $(x')^\mu$ is evidently also a 4-vector, which will be called the 4-velocity. The 3-velocity \mathbf{v} of a particle is given by the ratio $\mathbf{v} = (dr/d\tau)/(dt/d\tau)$. Since the speed of a massive particle must be less than c , $\|\mathbf{v}\| < c$, verify that the 4-velocity must satisfy the condition

$$x' \cdot x' = (x')^\mu (x')^\nu g_{\mu\nu} > 0. \quad (1.7.33)$$

Introduction of the 4-potential A^μ , the tensor $F^{\mu\nu}$, and the fields \mathbf{E} and \mathbf{B}

To continue our construction of a manifestly Lorentz invariant treatment of charged-particle motion, define a 4-potential A^μ with contravariant entries

$$A^\mu = (A_x, A_y, A_z, \psi/c) = (\mathbf{A}, \psi/c). \quad (1.7.34)$$

We will soon need the *antisymmetric* tensor $F^{\mu\nu}$ and its lowered counterpart $F_{\mu\nu}$ defined by the relations

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu, \quad (1.7.35)$$

$$F_{\mu\nu} = g_{\mu\sigma}g_{\nu\tau}F^{\sigma\tau} = g_{\mu\sigma}g_{\nu\tau}(\partial^\sigma A^\tau - \partial^\tau A^\sigma) = \partial_\mu A_\nu - \partial_\nu A_\mu. \quad (1.7.36)$$

Here we have used the notation

$$\begin{aligned}\partial^\mu &= \partial/\partial x_\mu = (-\partial/\partial x, -\partial/\partial y, -\partial/\partial z, c^{-1}\partial/\partial t) = (-\nabla, c^{-1}\partial/\partial t), \\ \partial_\mu &= \partial/\partial x^\mu = (\partial/\partial x, \partial/\partial y, \partial/\partial z, c^{-1}\partial/\partial t) = (\nabla, c^{-1}\partial/\partial t),\end{aligned}\quad (1.7.37)$$

which reminds us, for example, that the derivative of a scalar with respect to a covariant variable yields a contravariant result, and vice versa. [See (8.159) in Subsection 8.5.] Verify using (7.35) and (5.2) that the entries of $F^{\mu\nu}$ are the components of \mathbf{B} and \mathbf{E}/c arranged in the form

$$F^{\mu\nu} = \begin{pmatrix} 0 & -B_z & B_y & E_x/c \\ B_z & 0 & -B_x & E_y/c \\ -B_y & B_x & 0 & E_z/c \\ -E_x/c & -E_y/c & -E_z/c & 0 \end{pmatrix}, \quad (1.7.38)$$

so that $F^{12} = -B_z$, etc. And use of (7.36) gives

$$F_{\mu\nu} = \begin{pmatrix} 0 & -B_z & B_y & -E_x/c \\ B_z & 0 & -B_x & -E_y/c \\ -B_y & B_x & 0 & -E_z/c \\ E_x/c & E_y/c & E_z/c & 0 \end{pmatrix}, \quad (1.7.39)$$

so that $F_{12} = -B_z$, etc. Evidently the elements of $F_{\mu\nu}$ are obtained from those of $F^{\mu\nu}$ by making the substitution $\mathbf{E} \rightarrow -\mathbf{E}$.

1.7.2 A Relativistic Lagrangian L_R and Associated Relativistic Hamiltonian H_R

Consider the *relativistic* Lagrangian L_R defined by the relation

$$L_R = (1/2)mc(x')^\mu(x')^\nu g_{\mu\nu} + q(x')^\mu A^\nu g_{\mu\nu}. \quad (1.7.40)$$

It has the pleasing property that it is algebraically simple and treats space and time on a similar footing. In particular, L_R is evidently a scalar. That is, it is invariant under Lorentz transformations. Finally, we will see that use of L_R produces the expected/correct equations of motion for charged particles moving in electromagnetic \mathbf{E} and \mathbf{B} fields.⁵⁵

We will now find the equations of motion that L_R produces, and will also find the associated Hamiltonian H_R and the equations of motion it produces.

⁵⁵The quantity $(x')^\mu A^\nu g_{\mu\nu}$ is a scalar under Lorentz transformations providing the 4-potential A^ν actually transforms as a 4-vector. This can be shown to be the case if the \mathbf{E} and \mathbf{B} fields described by A^ν arise from an external current j_{ext}^ν that vanishes sufficiently rapidly at infinity. But in some cases, such as that of an electromagnetic plane wave or wave packet, the associated 4-potential A^ν is sourceless and does not transform like a 4-vector under Lorentz transformations. Instead the new 4-potential is the Lorentz transformation of the old (as if it were a 4-vector) plus a gauge transformation term. However, the additional gauge transformation term, when combined with the term arising from $(x')^\mu g_{\mu\nu}$, forms a total τ derivative. As discussed in standard Classical Mechanics texts, such total derivatives, when added to the Lagrangian, have no effect on the equations of motion. Moreover, they do not contribute to the variation of the action \mathcal{A} associated with the Lagrangian when the path is varied with fixed end points. They may therefore be dropped. Thus, Lorentz invariance is again restored, even in those cases in which the 4-potential A^ν does

Equations of Motion produced by L_R

With τ as the independent variable, Lagrange's equations of motion when applied to L_R read

$$(d/d\tau)[\partial L_R/\partial(x')^\mu] = \partial L_R/\partial x^\mu. \quad (1.7.41)$$

- a) Show that the *canonical* momenta p_μ are given by the relation

$$p_\mu = \partial L_R/\partial(x')^\mu = mc(x')_\mu + qA_\mu, \quad (1.7.42)$$

which can also be written in the form

$$p_\mu = p_\mu^{\text{mech}} + qA_\mu \quad (1.7.43)$$

where the *mechanical* momenta are given by

$$p_\mu^{\text{mech}} = mc(x')_\mu. \quad (1.7.44)$$

Note again that the derivative of a scalar (in this case L_R) with respect to a contravariant ("up" index) variable yields a covariant ("down" index) result.⁵⁶ Verify that there are the results

$$p^\mu = mc(x')^\mu + qA^\mu = (p^{\text{mech}})^\mu + qA^\mu \quad (1.7.45)$$

where

$$(p^{\text{mech}})^\mu = mc(x')^\mu. \quad (1.7.46)$$

- b) Verify that

$$\partial^2 L_R/\partial(x')^\mu \partial(x')^\nu = mc g_{\mu\nu}, \quad (1.7.47)$$

and therefore (5.6) is satisfied if $m \neq 0$.

- c) Show that differentiating and rearranging both sides of (7.45) produces the relation

$$d(p^{\text{mech}})^\mu/d\tau = dp^\mu/d\tau - q(dA^\mu/d\tau), \quad (1.7.48)$$

and verify by the chain rule that

$$q(dA^\mu/d\tau) = q \sum_\nu (\partial A^\mu/\partial x^\nu)(dx^\nu/d\tau). \quad (1.7.49)$$

not transform as a 4-vector.

A similar discussion of Lorentz invariance is required in the case of the Lagrangian (classical or quantal) for the combined system of electromagnetic fields and charged particles. In this case, the charge conservation relation $\partial_\nu j^\nu = 0$ again allows conversion of possibly non Lorentz invariant terms into total derivatives that may be dropped. Note that in the single particle case, the quantities $(x')^\nu$ may be viewed as being proportional to the single-particle current 4-vector. Thus, the single particle case is a special instance of the general case.

⁵⁶We also remark that, according to (7.46), the mechanical momentum $(p^{\text{mech}})^\mu$ transforms like a 4-vector under Lorentz transformations because $(x')^\mu$ transforms like a 4-vector. From (7.45) we see that the canonical momentum also transforms like a 4-vector to the extent that the 4-potential does so. If a gauge transformation is also involved in the transformation of the 4-potential, then this same additional term appears in the transformation of the canonical momentum. According to ? this additional term may be viewed as the result of a symplectic map. Finally, we remark that a Lorentz transformation is itself a symplectic map. See ?.

Combine (7.48) and (7.49) to find

$$d(p^{\text{mech}})^{\mu}/d\tau = dp^{\mu}/d\tau - q \sum_{\nu} (\partial A^{\mu}/\partial x^{\nu})(dx^{\nu}/d\tau). \quad (1.7.50)$$

Observe that

$$\sum_{\nu} (\partial A^{\mu}/\partial x^{\nu})(dx^{\nu}/d\tau) = \sum_{\nu} \partial_{\nu} A^{\mu}(dx^{\nu}/d\tau) = \sum_{\nu} \partial^{\nu} A^{\mu}(dx_{\nu}/d\tau). \quad (1.7.51)$$

Here we have used (7.37) and (7.7). Therefore we may rewrite (7.50) in the form

$$d(p^{\text{mech}})^{\mu}/d\tau = dp^{\mu}/d\tau - q \sum_{\nu} \partial^{\nu} A^{\mu}(dx_{\nu}/d\tau). \quad (1.7.52)$$

- d) Show from Lagrange's equations, see (7.41), that the canonical momenta obey the equations of motion

$$p'_{\mu} = dp_{\mu}/d\tau = (d/d\tau)[\partial L_R/\partial(x')^{\mu}] = \partial L_R/\partial x^{\mu} = q \sum_{\nu} (x')^{\nu}(\partial A_{\nu}/\partial x^{\mu}). \quad (1.7.53)$$

Observe that by index manipulation, see (7.7) and (7.37), we may write

$$\sum_{\nu} (x')^{\nu}(\partial A_{\nu}/\partial x^{\mu}) = \sum_{\nu} (x')_{\nu}(\partial A^{\nu}/\partial x^{\mu}) = \sum_{\nu} \partial^{\mu} A^{\nu}(dx_{\nu}/d\tau). \quad (1.7.54)$$

Therefore (7.53) can be written in the form

$$dp_{\mu}/d\tau = q \sum_{\nu} \partial^{\mu} A^{\nu}(dx_{\nu}/d\tau). \quad (1.7.55)$$

- e) Let us work out the consequences of Lagrange's equations of motion, with τ as an independent variable, for the Lagrangian L_R . See (7.41). Combine (7.52) and (7.55) to find that

$$\begin{aligned} d(p^{\text{mech}})^{\mu}/d\tau &= q \sum_{\nu} \partial^{\mu} A^{\nu}(dx_{\nu}/d\tau) - q \sum_{\nu} (\partial^{\nu} A^{\mu})(dx_{\nu}/d\tau) \\ &= q \sum_{\nu} (\partial^{\mu} A^{\nu} - \partial^{\nu} A^{\mu})(dx_{\nu}/d\tau). \end{aligned} \quad (1.7.56)$$

We may now invoke (7.35) in (7.56) to convert it, with the aid of (7.44), to the final forms

$$d(p^{\text{mech}})^{\mu}/d\tau = q F^{\mu\nu}(dx_{\nu}/d\tau) \Leftrightarrow d(p^{\text{mech}})^{\mu}/d\tau = [q/(mc)] F^{\mu\nu} (p^{\text{mech}})_{\nu}. \quad (1.7.57)$$

There is a second equation involving the quantities x^{μ} . Differentiate both sides of (7.46) to find

$$(mc)d^2 x^{\mu}/d\tau^2 = d(p^{\text{mech}})^{\mu}/d\tau, \quad (1.7.58)$$

and combine (7.57) and (7.58) to find that

$$d^2x^\mu/d\tau^2 = [q/(mc)]F^{\mu\nu}(dx_\nu/d\tau). \quad (1.7.59)$$

The equations of motion, when written in the forms (7.57) and (7.59), are manifestly Lorentz invariant.⁵⁷ Indeed, this is an ideal opportunity to reiterate the meaning of Lorentz invariance: Lorentz invariance, as embodied by (7.59), states that if the world line $x^\mu(\tau)$ is a solution of the equations of motion, then so is its Lorentz transformed world line $\bar{x}^\mu(\tau)$ provided $F^{\mu\nu}$ is replaced by $\bar{F}^{\mu\nu}$ where $\bar{F}^{\mu\nu}$ is the tensor composed of $\bar{\mathbf{E}}$ and $\bar{\mathbf{B}}$, the Lorentz transformed electric and magnetic fields.⁵⁸ We note that this happy circumstance comes about because, as we have already seen, the variation of the action \mathcal{A} associated with L_R is *unchanged* by a Lorentz transformation. Therefore if $x^\mu(\tau)$ with its specified endpoints extremizes \mathcal{A} , so will $\bar{x}^\mu(\tau)$ with its transformed end points. Finally, we observe that the equations of motion (7.57 and (7.59) do not involve the vector potential, but only the fields \mathbf{E} and \mathbf{B} . They are therefore gauge independent. Moreover, since the equations of motion do not in fact involve the 4-potential, its transformation properties are irrelevant.

- f) Suppose we wish to use (7.59) to compute a world line (trajectory). Verify that inverting (7.46) yields the relations

$$dx^\mu/d\tau = [1/(mc)](p^{\text{mech}})^\mu. \quad (1.7.60)$$

Use (7.60) to rewrite (7.57) in the form

$$d(p^{\text{mech}})^\mu/d\tau = [q/(mc)]F^{\mu\nu}g_{\nu\sigma}(p^{\text{mech}})^\sigma. \quad (1.7.61)$$

Verify that taken together the relations (7.60) and (7.61) constitute a (coupled) set of first-order differential equations for the variables x^μ and $(p^{\text{mech}})^\mu$.

- g) Again suppose we wish to use (7.59) to compute a world line. Introduce auxiliary variables u^μ by the rule

$$u^\mu = dx^\mu/d\tau. \quad (1.7.62)$$

Verify that (7.59) and (7.62) can be rewritten in the form

$$dx^\mu/d\tau = u^\mu, \quad (1.7.63)$$

$$du^\mu/d\tau = [q/(mc)]F^{\mu\nu}g_{\nu\sigma}u^\sigma \quad (1.7.64)$$

to yield a (coupled) set of first-order differential equations for the variables x^μ and u^μ .

⁵⁷Some authors would say instead that the equations of motion (7.57) and (7.59) are *covariant*. We prefer not to use such terminology because we wish to reserve the use of the term *covariant*, and the complementary term *contravariant*, to refer to the “down” and “up” index components of vectors and tensors. Perhaps even better would be to say that the equations of motion (7.57) and (7.59) are *form* invariant; they have the same form in every inertial frame.

⁵⁸For a discussion how \mathbf{E} and \mathbf{B} transform, see Chapter 28.

h) Show that the equation of motion (7.57) has the constant and integral of motion

$$(p^{\text{mech}})^\mu p_\mu^{\text{mech}} = p^{\text{mech}} \cdot p^{\text{mech}} = \text{const}, \quad (1.7.65)$$

and the equation of motion (7.59) has the constant and integral of motion

$$(x')^\mu (x')_\mu = x' \cdot x' = \text{const}. \quad (1.7.66)$$

[Hint: For (7.65) use the second version of (7.57) and the antisymmetry of $F^{\mu\nu}$. For (7.66) use (7.59) and again the antisymmetry of $F^{\mu\nu}$.] Whatever values these quantities have for some initial value of τ , they retain these same values for all values of τ .

Associated Relativistic Hamiltonian H_R

i) Define the associated relativistic Hamiltonian H_R by the rule

$$\begin{aligned} H_R &= \left\{ \sum_\mu [\partial L_R / \partial (x')^\mu] (x')^\mu \right\} - L_R \\ &= \left\{ \sum_\mu p_\mu (x')^\mu \right\} - L_R. \end{aligned} \quad (1.7.67)$$

Show that H_R is given by the relation

$$\begin{aligned} H_R &= (1/2)mc(x')^\mu (x')^\nu g_{\mu\nu} = [1/(2mc)](p^\mu - qA^\mu)(p^\nu - qA^\nu)g_{\mu\nu} \\ &= [1/(2mc)](p_\mu - qA_\mu)(p_\nu - qA_\nu)g^{\mu\nu} \\ &= [1/(2mc)](p^\mu - qA^\mu)(p_\mu - qA_\mu). \end{aligned} \quad (1.7.68)$$

Note that H_R , like L_R , is Lorentz invariant.

j) If H_R is viewed as a function of the variables x^μ , p_μ , and τ , it has the total differential

$$dH_R = \left\{ \sum_\mu (\partial H_R / \partial x^\mu) dx^\mu + (\partial H_R / \partial p_\mu) dp_\mu \right\} + (\partial H_R / \partial \tau) d\tau. \quad (1.7.69)$$

On the other hand, if it is viewed as a function of the variables x^μ , $(x')^\mu$, and τ , H_R has [using (7.67)] the total differential

$$\begin{aligned} dH_R &= \left\{ \sum_\mu [p_\mu - \partial L_R / \partial (x')^\mu] d(x')^\mu + (x')^\mu dp_\mu - (\partial L_R / \partial x^\mu) dx^\mu \right\} - (\partial L_R / \partial \tau) d\tau \\ &= \left\{ \sum_\mu (x')^\mu dp_\mu - (p')_\mu dx^\mu \right\} - (\partial L_R / \partial \tau) d\tau. \end{aligned} \quad (1.7.70)$$

Here we have also used (7.42) and (7.53). By comparing (7.68) and (7.69), deduce the equations of motion.

$$(x')^\mu = \partial H_R / \partial p_\mu, \quad (1.7.71)$$

$$(p')_\mu = -\partial H_R / \partial x^\mu, \quad (1.7.72)$$

$$\partial H_R / \partial \tau = -\partial L_R / \partial \tau. \quad (1.7.73)$$

Equations of Motion produced by H_R

- k) Let us check that use of the Lorentz invariant Hamiltonian H_R given by (7.68), and the associated equations of motion (7.71) through (7.73), reproduces some previous results. Verify that use of (7.71) yields (7.46). Also work out the consequences of (7.72) and compare your results with those produced by use of (7.58). Show that (7.59) is a consequence of the Hamiltonian equations (7.71) and (7.72).
- l) Verify that L_R as given by (7.40) does not depend explicitly on τ ,

$$\partial L_R / \partial \tau = 0. \quad (1.7.74)$$

It follows, see (5.14), that

$$dH_R/d\tau = \partial H_R / \partial \tau = -\partial L_R / \partial \tau = 0. \quad (1.7.75)$$

That is, H_R is a constant and integral of motion and therefore the quantity $[ds^2/(d\tau)^2]$ defined by

$$ds^2/(d\tau)^2 = g_{\mu\nu}(x')^\mu(x')^\nu = (x')^\mu(x')_\mu = x' \cdot x' \quad (1.7.76)$$

is a constant and an integral of motion,

$$ds^2/(d\tau)^2 = \text{const.} \quad (1.7.77)$$

Note that this result agrees with (7.66).

- m) Suppose we restrict our attention to those solutions that satisfy the relation

$$x' \cdot x' = \lambda \quad (1.7.78)$$

where λ is a constant that can have any value including negative and zero values as well as positive values. Show that for these solutions $(p^{\text{mech}})^\mu$, as given by (7.46), satisfies the mass-shell condition

$$p_\mu^{\text{mech}}(p^{\text{mech}})^\mu = \lambda m^2 c^2. \quad (1.7.79)$$

Thus there are solutions for which the quantity $p_\mu^{\text{mech}}(p^{\text{mech}})^\mu$ can have any value including negative and zero values as well as positive values. Show that for these solutions H_R has the values

$$H_R = \lambda(mc/2). \quad (1.7.80)$$

Thus H_R can also have any value including negative and zero values as well as positive values.

- n) Suppose we restrict our attention to those solutions that satisfy the relation

$$x' \cdot x' = \lambda = 1. \quad (1.7.81)$$

Show that for these solutions the particle has mass m ,

$$p_\mu^{\text{mech}}(p^{\text{mech}})^\mu = p^{\text{mech}} \cdot p^{\text{mech}} = m^2 c^2, \quad (1.7.82)$$

and H_R has the value

$$H_R = mc/2. \quad (1.7.83)$$

- o) For those solutions that satisfy (7.81) verify that $ds^2 > 0$ and therefore we may select, in accord with (7.3), (7.76), and (7.81), a parameterization such that

$$ds/d\tau = 1. \quad (1.7.84)$$

Show that these solutions satisfy the equations

$$d(p^{\text{mech}})^\mu/ds = qF^{\mu\nu}(dx_\nu/ds), \quad (1.7.85)$$

$$d^2x^\mu/ds^2 = [q/(mc)]F^{\mu\nu}(dx_\nu/ds). \quad (1.7.86)$$

- p) Again restrict attention to those solutions that satisfy (7.81). Show that for these solutions there is the result

$$\dot{x}' = dx/d\tau = (dx/d\tau)(d\tau/ds) = dx/ds = (dx/dt)(dt/ds) = \dot{x}(dt/ds). \quad (1.7.87)$$

Verify from (7.1) that

$$\dot{x}^\mu = dx^\mu/dt = (d\mathbf{r}/dt, c) = (\mathbf{v}, c). \quad (1.7.88)$$

Also verify, starting with (7.43), that there is the relation

$$ds/dt = c(1 - v^2/c^2)^{1/2} = c/\gamma, \quad (1.7.89)$$

and therefore

$$dt/ds = \gamma/c. \quad (1.7.90)$$

Recall the definition (5.29). Conclude that

$$(x')^\mu = (\gamma/c)(\mathbf{v}, c), \quad (1.7.91)$$

and therefore, by the definition (7.46), there is the relation

$$(p^{\text{mech}})^\mu = mc(x')^\mu = (m\gamma)(\mathbf{v}, c) = (\mathbf{p}^{\text{mech}}, \mathcal{E}/c) \quad (1.7.92)$$

with

$$\mathbf{p}^{\text{mech}} = \gamma m \mathbf{v} = (p_x^{\text{mech}}, p_y^{\text{mech}}, p_z^{\text{mech}}) \quad (1.7.93)$$

and

$$\mathcal{E} = \gamma mc^2. \quad (1.7.94)$$

Recall (5.28) and (5.39). Verify that combining (7.82) and (7.92) yields the relation

$$\mathcal{E}^2 = m^2c^4 + (\mathbf{p}^{\text{mech}} \cdot \mathbf{p}^{\text{mech}})c^2. \quad (1.7.95)$$

Verify also that

$$p^4 = (p^{\text{mech}})^4 + qA^4 = \mathcal{E}/c + q\psi/c = (1/c)(\gamma mc^2 + q\psi) = -(1/c)p_t. \quad (1.7.96)$$

Recall Exercise 6.1.

q) Show for the solutions of the equations of motion that satisfy (7.81) there is the relation

$$p^\mu = \{\mathbf{p}^{\text{can}}, -(1/c)p_t\} \quad (1.7.97)$$

with

$$p_t = -q\psi - \gamma mc^2 = -q\psi - \mathcal{E} \quad (1.7.98)$$

and

$$\mathbf{p}^{\text{can}} = \mathbf{p}^{\text{mech}} + q\mathbf{A} = \gamma m\mathbf{v} + q\mathbf{A}. \quad (1.7.99)$$

r) Multiply both sides of (7.85) by ds/dt to find the intermediate result

$$[d(p^{\text{mech}})^\mu/ds](ds/dt) = qF^{\mu\nu}(dx_\nu/ds)(ds/dt). \quad (1.7.100)$$

Verify the relations

$$[d(p^{\text{mech}})^\mu/ds](ds/dt) = d(p^{\text{mech}})^\mu/dt, \quad (1.7.101)$$

$$(dx_\nu/ds)(ds/dt) = dx_\nu/dt, \quad (1.7.102)$$

and conclude that (7.85) can be rewritten in the form

$$d(p^{\text{mech}})^\mu/dt = qF^{\mu\nu}(dx_\nu/dt). \quad (1.7.103)$$

Verify using (1.79) that

$$dx_\nu/dt = (-\mathbf{v}, c). \quad (1.7.104)$$

Use this result to show that (7.100) yields and is equivalent to the relations

$$d\mathbf{p}^{\text{mech}}/dt = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}), \quad (1.7.105)$$

$$d\mathcal{E}/dt = q\mathbf{v} \cdot \mathbf{E}. \quad (1.7.106)$$

Recall (5.31) and (5.40).

1.7.3 Relation between L_R and H_R and L and H

With the preparation provided by Subsection 7.2, the purpose of this section is to relate the Lagrangian L_R given by (7.40) and the Hamiltonian H_R given by (7.68) to the Lagrangian L given by (5.1) and the Hamiltonian H given by (5.49).

a) Review the machinery of Section 6. Start with the Hamiltonian H_R given by (7.68), for which τ is the independent variable, and make the definition

$$p_\tau = -H_R. \quad (1.7.107)$$

Verify that (7.107), with H_R given by (7.68), can be rewritten in the form

$$\begin{aligned} (p_4 - qA_4)^2 &= (p^4 - qA^4)^2 = [-2mc p_\tau + (\mathbf{p}^{\text{can}} - q\mathbf{A}) \cdot (\mathbf{p}^{\text{can}} - q\mathbf{A})] \\ &= [-2mc p_\tau + (\mathbf{p}^{\text{can}} - q\mathbf{A})^2], \end{aligned} \quad (1.7.108)$$

from which it follows that

$$p_4 - qA_4 = \pm[-2mc p_\tau + (\mathbf{p}^{\text{can}} - q\mathbf{A})^2]^{1/2}. \quad (1.7.109)$$

Here we have made the definition

$$\mathbf{p}^{\text{can}} = \mathbf{p}^{\text{mech}} + q\mathbf{A}. \quad (1.7.110)$$

Observe that (7.43) and (7.44) can be combined and rewritten in the form

$$p_4 - qA_4 = p_4^{\text{mech}} = mc(x')_4 = mc^2(dt/d\tau). \quad (1.7.111)$$

Require that the parameterization of the world line be such that $dt/d\tau > 0$. Verify that, upon taking into account this requirement, (7.109) can be rewritten in the form

$$p_4 = qA_4 + [-2mc p_\tau + (\mathbf{p}^{\text{can}} - q\mathbf{A})^2]^{1/2}. \quad (1.7.112)$$

- b) Let K be the new Hamiltonian for which x^4 is the independent variable. Recall that x^4 and p_4 are canonically conjugate. See also Exercise 7.6 for further discussion of this point. Verify that there is the result

$$K(\mathbf{r}, \tau, \mathbf{p}^{\text{can}}, p_\tau; x^4) = -p_4 = -qA_4 - [-2mc p_\tau + (\mathbf{p}^{\text{can}} - q\mathbf{A})^2]^{1/2}. \quad (1.7.113)$$

- c) Note that H_R and hence K are, in fact, independent of τ . Therefore p_τ is a constant of motion. Relate this constant to equation (7.80). That is, verify the relation

$$p_\tau = -\lambda(mc/2). \quad (1.7.114)$$

- d) Since K is independent of τ , and p_τ is a constant, suppose attention is restricted to the remaining variables in K . Moreover, let us assign to p_τ the value it has for trajectories of interest, namely those with $\lambda = 1$. That is, we restrict our attention to the case where

$$p_\tau = -mc/2. \quad (1.7.115)$$

Verify that there is then the result

$$K(\mathbf{r}, \tau, \mathbf{p}^{\text{can}}, -mc/2; x^4) = -qA_4 - [m^2c^2 + (\mathbf{p}^{\text{can}} - q\mathbf{A})^2]^{1/2}. \quad (1.7.116)$$

Upon comparing (5.49) and (7.116), verify that there must be the relation

$$K(\mathbf{r}, \tau, \mathbf{p}^{\text{can}}, -mc/2; x^4) = -(1/c)H. \quad (1.7.117)$$

- e) Does (7.117) agree with what we already know? Suppose K , as given by (7.116), is used to produce equations of motion. Then, in view of (7.70) and (7.71), show that we expect the results

$$(1/c)(dx^\mu/dt) = dx^\mu/dx^4 = \partial K/\partial p_\mu \text{ for } \mu = 1, 2, 3; \quad (1.7.118)$$

$$(1/c)(dp_\mu/dt) = dp_\mu/dx^4 = -\partial K/\partial x^\mu \text{ for } \mu = 1, 2, 3. \quad (1.7.119)$$

But, there are the relations

$$p_\mu = -p^\mu \text{ for } \mu = 1, 2, 3. \quad (1.7.120)$$

Verify that, consequently, (7.118) and (7.119) can be rewritten in the form

$$(1/c)(dx^\mu/dt) = -\partial K/\partial p^\mu \text{ for } \mu = 1, 2, 3; \quad (1.7.121)$$

$$(1/c)(dp^\mu/dt) = +\partial K/\partial x^\mu \text{ for } \mu = 1, 2, 3. \quad (1.7.122)$$

But, as we already know, we wish to have the relations

$$(1/c)(dx^\mu/dt) = (1/c)(\partial H/\partial p^\mu) \text{ for } \mu = 1, 2, 3; \quad (1.7.123)$$

$$(1/c)(dp^\mu/dt) = -(1/c)(\partial H/\partial x^\mu) \text{ for } \mu = 1, 2, 3. \quad (1.7.124)$$

Verify that (7.121) through (7.124) are consistent with (7.117).

Let us summarize our results. In Exercise 6.7 you showed that use of the manifestly Lorentz invariant Lagrangian L_R given by (7.40) leads to the manifestly Lorentz invariant Hamiltonian H_R given by (7.68). Subsequently, in this exercise you showed that deciding to treat the time as the independent variable, and restricting attention to the variables x^μ and p^μ (with $\mu = 1, 2, 3$), leads from H_R to the Hamiltonian K given by (7.113) and then to the Hamiltonian H given by (5.49). Finally, see Exercise 5.13, by an inverse Legendre transformation the Hamiltonian H yields the Lagrangian L given by (5.1).

1.7.4 An Alternate Relativistic Lagrangian L_A ?

Review Exercise 6.7. Some texts claim that the equations of motion for relativistic charged-particle motion can also be derived from the action functional $\mathcal{A}[x]$ given by

$$\mathcal{A}[x] = \int mc ds + \int q g_{\mu\nu} A^\mu dx^\nu \quad (1.7.125)$$

with ds^2 given by (7.3). Since this \mathcal{A} can also be written in the form

$$\mathcal{A}[x] = \int [mc(ds/d\tau) + q g_{\mu\nu} A^\mu(x')^\nu] d\tau \quad (1.7.126)$$

where τ parameterizes the world line $x^\mu(\tau)$, show that the use of this action is equivalent to using the *Alternate* Lagrangian L_A given by the rule

$$L_A = mc[g_{\mu\nu}(x')^\mu(x')^\nu]^{1/2} + q(x')^\mu A^\nu g_{\mu\nu}. \quad (1.7.127)$$

Evidently this Lagrangian, like L_R as given (7.40), is also invariant under Lorentz transformations provided the parameterization is Lorentz invariant.

- a) Show that L_A is homogeneous of degree one in the velocities and does not explicitly depend on τ . Verify directly that $\mathcal{A}[x]$ is *independent* of the parameterization employed. This independence implies that we should not expect to find a unique solution that extremizes \mathcal{A} since any reparametrization also gives a solution. See the discussion at the end of Subsection 5.2. Consequently, as expected, additional information will be required. By contrast, show that the action $\mathcal{A}_R[x]$ associated with the Lagrangian L_R given by (7.40) is *not* parameterization independent.
- b) Show that for the Lagrangian (7.127) the *canonical* momenta p_μ^{can} are given by the relations

$$p_\mu^{\text{can}} = \partial L_A / \partial (x')^\mu = p_\mu^{\text{mech}} + q A_\mu \quad (1.7.128)$$

where

$$p_\mu^{\text{mech}} = mc(x')_\mu / (x' \cdot x')^{1/2}. \quad (1.7.129)$$

Here, consistent with (7.33), the parameterization and the sign of the square root are selected in such a way that both $(x')^4$ and $(p^{\text{mech}})^4$ are positive. Show that both p^{mech} and p^{can} are independent of the choice of parameterization τ . Verify that the quantities p_μ^{mech} comprise a 4-vector, and that there is the Lorentz invariant relation

$$p_\mu^{\text{mech}} (p^{\text{mech}})^\mu = m^2 c^2. \quad (1.7.130)$$

- c) Show that Lagrange's equations of motion when applied to L_A yield the result

$$d(p^{\text{mech}})^\mu / d\tau = q F^{\mu\nu} (dx_\nu / d\tau). \quad (1.7.131)$$

The equations of motion, when written in the form (7.131), are manifestly Lorentz invariant. However note that, while superficially similar, (7.131) is not the same as (7.57) because the definitions of p^{mech} in (7.46) and (7.129) are not the same. Show that the form of the equations of motion (7.131) is unchanged if the world-line parameterization is changed. Show that the equations of motion (7.131) preserve the relation (7.130).

- d) Verify that, as it stands, L_A as given by (7.127) is not a very promising Lagrangian because it has the property

$$\det[\partial^2 L_A / \partial (x')^\mu \partial (x')^\nu] = 0. \quad (1.7.132)$$

That is, the requirement (5.6) is violated. [Compare (7.132) with the analogous result in Exercise 6.7.] Also, because L_A is homogeneous of degree one in the variables $(x')^\mu$, it satisfies the relation

$$\sum_\mu [\partial L_A / \partial (x')^\mu] (x')^\mu = L_A \quad (1.7.133)$$

See Exercise 5.12. Consequently, verify that the Hamiltonian associated with L_A , call it H_A , vanishes identically! By contrast, the Lagrangian L_R given by (7.70) satisfies (5.6), has a well-defined Hamiltonian counterpart H_R , and also automatically provides the supplementary condition (7.77).

- e) In point of fact the equations (7.131), in the absence of further information, do *not* provide equations of motion in the form (3.1) as is required in order to specify trajectories. To see this, compute $d(p^{\text{mech}})^{\mu}/d\tau$ using the chain rule,

$$d(p^{\text{mech}})^{\mu}/d\tau = \sum_{\nu} [\partial(p^{\text{mech}})^{\mu}/\partial(x')^{\nu}] (x'')^{\nu}. \quad (1.7.134)$$

Show that

$$\det[\partial(p^{\text{mech}})^{\mu}/\partial(x')^{\nu}] = 0 \quad (1.7.135)$$

so that (7.131) and (7.134) *cannot* be solved for the $(x'')^{\nu}$ to produce equations of motion of the form (3.1). Hint: Either verify (7.135) directly by brute force using (7.129) or, more elegantly and following the discussion in Subsection 5.2, show that each $(p^{\text{mech}})^{\mu}$ is a homogeneous function of degree zero. It then follows from Euler's relation, see Exercise 5.12, that there is the result

$$\sum_{\nu} [(x')^{\nu}] [\partial(p^{\text{mech}})^{\mu}/\partial(x')^{\nu}] = 0. \quad (1.7.136)$$

This result shows that the matrix $[\partial(p^{\text{mech}})^{\mu}/\partial(x')^{\nu}]$ has the (generally nonzero) vector x' as an eigenvector with eigenvalue zero. Therefore (7.135) must hold.

- f) Nevertheless, as we will see, the equations of motion provided by L_A give satisfactory results when supplemented by additional information. As might be expected, what is required is some information about how the parameter τ is to be selected. Suppose, for example, that τ is selected in such a way that

$$x^4 = c\tau \Leftrightarrow t = \tau. \quad (1.7.137)$$

(Alternatively, we may proceed as in Exercise 6.10.) Note that this parameterization is not Lorentz invariant. However, since both p^{mech} and p^{can} do not depend on the choice of parameter, they continue to be 4-vectors. With the parameter choice (7.137) there is the additional information

$$(x')^4 = c, \quad (1.7.138)$$

and the equations of motion (7.131) take the form

$$d(p^{\text{mech}})^{\mu}/dt = qF^{\mu\nu}(dx_{\nu}/dt). \quad (1.7.139)$$

Show that if (7.138) holds, then there is the relation

$$x' \cdot x' = c^2(1 - v^2/c^2) = c^2/\gamma^2, \quad (1.7.140)$$

and therefore there is the relation

$$(x' \cdot x')^{1/2} = c/\gamma. \quad (1.7.141)$$

Consequently, verify using (7.129) and (7.141) that $(p^{\text{mech}})^{\mu}$ now takes the form

$$(p^{\text{mech}})^{\mu} = (p_x, p_y, p_z, \mathcal{E}/c) = (\mathbf{p}, \mathcal{E}/c) \quad (1.7.142)$$

with the relativistic momentum \mathbf{p} given by the relation

$$\mathbf{p} = \gamma m \mathbf{v} \quad (1.7.143)$$

and the relativistic energy \mathcal{E} given by the relation

$$\mathcal{E} = \gamma mc^2. \quad (1.7.144)$$

Show that, when written out in component form, the equations of motion (7.139) become

$$d\mathbf{p}/dt = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}), \quad (1.7.145)$$

$$d\mathcal{E}/dt = q\mathbf{v} \cdot \mathbf{E}. \quad (1.7.146)$$

Recall (5.31) and (5.40).

Let us summarize our results so far. We have seen that, because it is homogeneous of degree 1 in the variables $(x')^\mu$, the Lagrangian L_A given by (7.127) has no Hamiltonian counterpart. It is therefore of limited interest if we wish, as we do in this book, to exploit the symplectic symmetries associated with Hamiltonian formulations. However, with the aid of additional information specifying the parameterization, it is possible to obtain the equations of motion (7.145) and (7.146).

We have explored some consequences of using the parameterization (7.137). Another attractive parameterization possibility (which is also Lorentz invariant) is to select τ in such a way that there is the relation

$$d\tau = +(ds^2)^{1/2}, \text{ which we commonly casually write as } d\tau = ds, \quad (1.7.147)$$

with ds^2 given by (7.3). Here we assume that all displacements are positive time like so that $dt > 0$ and $ds^2 > 0$. [This assumption makes the square root in (7.147) well defined and the magnitude $\|\mathbf{v}\|$ of the 3-velocity less than c . See (7.33).] That is, again speaking casually, we stipulate/say that the world line is parameterized by the *space-time* path length. Show that in this case there is the additional (Lorentz invariant) information

$$(x') \cdot (x') = 1 \text{ for all } \tau \quad (1.7.148)$$

so that now

$$(p^{\text{mech}})^\mu = mc(x')^\mu, \quad (1.7.149)$$

and the equations of motion (7.131) take the (manifestly Lorentz invariant) form

$$d(p^{\text{mech}})^\mu/ds = qF^{\mu\nu}(dx_\nu/ds),$$

$$d^2x^\mu/ds^2 = [q/(mc)]F^{\mu\nu}(dx_\nu/ds) = [q/(mc)]F^{\mu\nu}g_{\nu\sigma}(dx^\sigma/ds). \quad (1.7.150)$$

Verify that these equations of motion preserve the conditions (7.130) and (7.148). Moreover, observe that they are of the desired form (3.1).

Since the equations of motion (7.150) agree with those given by (7.85) and (7.86), verify that one can use them to derive the remaining results in items p through r in Subsection 7.2.

Exercises

1.7.1. Review Exercises 6.7 and 6.8. Starting with the Hamiltonian H_R , as given by (6.77) and for which τ is the independent variable, find a new Hamiltonian (call it K) for which $x^3 = z$ is the independent variable. Use (7.96) and show that it is correct to make the identification $p_t = -p^4 c = -p_4 c$. Compare your result with (6.16).

1.7.2. Exercise 5.2 determined the equations of motion for the *mechanical* variables \mathbf{r} and \mathbf{p} with the time t as the independent variable. See (5.43) and (5.44). The purpose of this exercise is to determine the equations of motion for mechanical variables when some coordinate is taken to be the independent variable. Specifically, suppose that the coordinate z is taken to be the independent variable. Introduce the notation

$$D = [(p_t + q\psi)^2/c^2 - m^2c^2 - (p_x - qA_x)^2 - (p_y - qA_y)^2]^{1/2}. \quad (1.7.151)$$

From (6.16) derive the equations of motion

$$x' = \partial K / \partial p_x = (p_x - qA_x)/D, \quad (1.7.152)$$

$$y' = \partial K / \partial p_y = (p_y - qA_y)/D, \quad (1.7.153)$$

$$t' = \partial K / \partial p_t = -(1/c^2)(p_t + q\psi)/D, \quad (1.7.154)$$

$$\begin{aligned} p'_x &= -\partial K / \partial x \\ &= q[(p_x - qA_x)(\partial A_x / \partial x) + (p_y - qA_y)(\partial A_y / \partial x) + (1/c^2)(p_t + q\psi)(\partial \psi / \partial x)]/D \\ &\quad + q\partial A_z / \partial x, \end{aligned} \quad (1.7.155)$$

$$\begin{aligned} p'_y &= -\partial K / \partial y \\ &= q[(p_x - qA_x)(\partial A_x / \partial y) + (p_y - qA_y)(\partial A_y / \partial y) + (1/c^2)(p_t + q\psi)(\partial \psi / \partial y)]/D \\ &\quad + q\partial A_z / \partial y, \end{aligned} \quad (1.7.156)$$

$$\begin{aligned} p'_t &= -\partial K / \partial t \\ &= q[(p_x - qA_x)(\partial A_x / \partial t) + (p_y - qA_y)(\partial A_y / \partial t) + (1/c^2)(p_t + q\psi)(\partial \psi / \partial t)]/D \\ &\quad + q\partial A_z / \partial t, \end{aligned} \quad (1.7.157)$$

where a prime denotes d/dz . Next employ (7.152) through (7.154) in (7.155) through (7.157) to find the results

$$p'_x = q[x'(\partial A_x / \partial x) + y'(\partial A_y / \partial x) - t'(\partial \psi / \partial x)] + q\partial A_z / \partial x, \quad (1.7.158)$$

$$p'_y = q[x'(\partial A_x / \partial y) + y'(\partial A_y / \partial y) - t'(\partial \psi / \partial y)] + q\partial A_z / \partial y, \quad (1.7.159)$$

$$p'_t = q[x'(\partial A_x / \partial t) + y'(\partial A_y / \partial t) - t'(\partial \psi / \partial t)] + q\partial A_z / \partial t. \quad (1.7.160)$$

The relations (7.151) through (7.160) involve canonical momenta. Since we are interested in employing mechanical variables, introduce the notation

$$\tilde{p}_x = p_x - qA_x, \quad (1.7.161)$$

$$\tilde{p}_y = p_y - qA_y, \quad (1.7.162)$$

$$\tilde{p}_t = p_t + q\psi. \quad (1.7.163)$$

From (5.30) we see that \tilde{p}_x and \tilde{p}_y are mechanical momenta, and from (7.98) we conclude that

$$\tilde{p}_t = p_t + q\psi = -\gamma mc^2 = -\mathcal{E} = -E^{\text{mech}}. \quad (1.7.164)$$

(See also Exercise 7.10.) In terms of these variables the equations of motion (7.152) through (7.154) for the coordinates take the form

$$x' = \tilde{p}_x/\tilde{D}, \quad (1.7.165)$$

$$y' = \tilde{p}_y/\tilde{D}, \quad (1.7.166)$$

$$t' = -(1/c^2)\tilde{p}_t/\tilde{D}, \quad (1.7.167)$$

where

$$\tilde{D} = [\tilde{p}_t^2/c^2 - m^2c^2 - \tilde{p}_x^2 - \tilde{p}_y^2]^{1/2}. \quad (1.7.168)$$

The remaining task is to find the equations of motion for the mechanical momenta. Differentiate and apply the chain rule to (7.161) through (7.163) to find the results

$$\begin{aligned} \tilde{p}'_x &= p'_x - qA'_x \\ &= p'_x - q[(\partial A_x/\partial x)x' + (\partial A_x/\partial y)y' + (\partial A_x/\partial z)z' + (\partial A_x/\partial t)t'], \end{aligned} \quad (1.7.169)$$

$$\begin{aligned} \tilde{p}'_y &= p'_y - qA'_y \\ &= p'_y - q[(\partial A_y/\partial x)x' + (\partial A_y/\partial y)y' + (\partial A_y/\partial z)z' + (\partial A_y/\partial t)t'], \end{aligned} \quad (1.7.170)$$

$$\begin{aligned} \tilde{p}'_t &= p'_t + q\psi' \\ &= p'_t + q[(\partial\psi/\partial x)x' + (\partial\psi/\partial y)y' + (\partial\psi/\partial z)z' + (\partial\psi/\partial t)t']. \end{aligned} \quad (1.7.171)$$

Now combine (7.158) and (7.169) to obtain the result

$$\begin{aligned} \tilde{p}'_x &= p'_x - qA'_x \\ &= q[x'(\partial A_x/\partial x) + y'(\partial A_y/\partial x) - t'(\partial\psi/\partial x)] + q\partial A_z/\partial x \\ &\quad - q[(\partial A_x/\partial x)x' + (\partial A_x/\partial y)y' + (\partial A_x/\partial z)z' + (\partial A_x/\partial t)t'] \\ &= q[y'(\partial A_y/\partial x) - \partial A_x/\partial y] + (\partial A_z/\partial x - \partial A_x/\partial z) \\ &\quad - qt'[(\partial\psi/\partial x) + (\partial A_x/\partial t)] \\ &= q[y'B_z - B_y] + qt'E_x. \end{aligned} \quad (1.7.172)$$

Here we have used (5.2). Similarly, verify that

$$\tilde{p}'_y = q[B_x - x'B_z] + qt'E_y. \quad (1.7.173)$$

Next, combine (7.160) and (7.171) to find the result

$$\begin{aligned}
 \tilde{p}'_t &= p'_t + q\psi' \\
 &= q[x'(\partial A_x/\partial t) + y'(\partial A_y/\partial t) - t'(\partial \psi/\partial t)] + q\partial A_z/\partial t \\
 &\quad + q[(\partial \psi/\partial x)x' + (\partial \psi/\partial y)y' + (\partial \psi/\partial z) + (\partial \psi/\partial t)t'] \\
 &= q[x'(\partial \psi/\partial x + \partial A_x/\partial t) + y'(\partial \psi/\partial x + \partial A_x/\partial t) + (\partial \psi/\partial z + \partial A_z/\partial t)] \\
 &= -q[x'E_x + y'E_y + E_z].
 \end{aligned} \tag{1.7.174}$$

Verify that the relations (7.172) through (7.174) are what one would expect in view of (7.105) and (7.106).

There is one final step. We would like the right sides of (7.172) through (7.174) to involve only the coordinates and mechanical momenta, and not the quantities x' , x' , and t' . This can be accomplished with the aid of (7.165) through (7.168). Show that the net results are equations of motion for the mechanical momenta in the form

$$\tilde{p}'_x = q[(\tilde{p}_y/\tilde{D})B_z - B_y] - q[(1/c^2)\tilde{p}_t/\tilde{D}]E_x, \tag{1.7.175}$$

$$\tilde{p}'_y = q[B_x - (\tilde{p}_x/\tilde{D})B_z] - q[(1/c^2)\tilde{p}_t/\tilde{D}]E_y, \tag{1.7.176}$$

$$\tilde{p}'_t = -q[(\tilde{p}_x E_x + \tilde{p}_y E_y)/\tilde{D} + E_z]. \tag{1.7.177}$$

Taken together, the relations (7.165) through (7.168) and (7.175) through (7.177) provide equations of motion in mechanical variables when z is taken to be the independent variable. That is, the dependent variables are $(x, y, t; \tilde{p}_x, \tilde{p}_y, \tilde{p}_t)$, and z is the independent variable. Note that these equations of motion, like their similar counterparts in Exercises 5.2, 6.7, 6.9, and 6.10, involve only the fields \mathbf{E} and \mathbf{B} and make no reference to the vector and scalar potentials \mathbf{A} and ψ .

1.7.3. Review Exercise 6.12. It formulated equations of motion for the dependent variables $(x, y, t; \tilde{p}_x, \tilde{p}_y, \tilde{p}_t)$ with z taken to be the independent variable. Your task for this exercise is to formulate equations of motion for the dependent variables $(x, y, t; x', y', t')$ with z taken to be the independent variable. What are desired are equations for the quantities (x'', y'', t'') in terms of the variables $(x, y, t; x', y', t')$ and z . Here a prime denotes (d/dz) .

1.7.4. Consider charged-particle motion in the case of a *static* magnetic field $\mathbf{B}(\mathbf{r})$ and no electric field. (Note that, according to Maxwell's equations, there must be an electric field if \mathbf{B} is not static.) Show from (5.40) that in this case the energy \mathcal{E} is constant and, by (5.39), γ is constant. Next show from (5.48) that the equations of motion take the form

$$m^* d^2\mathbf{r}/dt^2 = q(\mathbf{v} \times \mathbf{B}) \tag{1.7.178}$$

where

$$m^* = \gamma m. \tag{1.7.179}$$

Thus, in the case of a static magnetic field and no electric field, the only difference between relativistic and nonrelativistic motion is that m must be replaced by m^* . To be more precise, suppose m^* (with $m^* \geq m$) and hence γ are specified numbers. The equations of motion

(7.178) have solutions for any set $(\mathbf{r}^{\text{in}}, \mathbf{v}^{\text{in}})$ of initial conditions. Those solutions for which \mathbf{v}^{in} satisfies

$$[1 - (\mathbf{v}^{\text{in}}/c) \cdot (\mathbf{v}^{\text{in}}/c)]^{-1/2} = m^*/m = \gamma \quad (1.7.180)$$

will also be solutions of the relativistic equations of motion.

Review Exercise 5.9. Show that the results in that exercise are consistent with the results of this exercise.

1.7.5. Review Exercise 6.14. Again view m^* as a specified number. Show that the equations of motion (7.178) follow from the “nonrelativistic” Lagrangian

$$L = (m^*/2)\mathbf{v} \cdot \mathbf{v} + q\mathbf{v} \cdot \mathbf{A}(\mathbf{r}). \quad (1.7.181)$$

Show that the canonical momentum \mathbf{p} is given by the relation

$$\mathbf{p} = m^*\mathbf{v} + q\mathbf{A}, \quad (1.7.182)$$

and that the Hamiltonian H associated with L is given by

$$H = (\mathbf{p} - q\mathbf{A}) \cdot (\mathbf{p} - q\mathbf{A})/(2m^*). \quad (1.7.183)$$

Finally show that for trajectories of physical interest, namely those that satisfy (7.180), H has the constant value

$$H = (1/2)mc^2(\gamma^2 - 1)/\gamma = (1/2)mc^2[(m^*/m) - (m/m^*)] = (1/2)m^*v^2. \quad (1.7.184)$$

1.8 Something About Riemannian Manifolds

Roughly speaking, an n -dimensional *manifold* is a set that locally at each point looks like an n -dimensional space with local coordinates x^1, \dots, x^n . When equipped with a (possibly position dependent) metric tensor $g(x)$, a manifold becomes a *Riemannian* manifold. Two-dimensional Euclidean space is a simple example of a Riemannian manifold for which the metric tensor is constant and, when viewed as a matrix, is equal to the identity matrix.

Now consider a general Riemannian manifold with local coordinates x^i and metric tensor $g(x)$. We assume that g is invertible. This manifold is called *proper* Riemannian if g is positive definite, and *pseudo* Riemannian if g is not positive (or negative) definite. Thus for example, according to (7.2), space-time in the theory of special relativity (Minkowski space) is a pseudo Riemannian manifold.

1.8.1 Geodesics and Affine Geodesics

This subsection describes/defines geodesics and affine geodesics. As background, review Exercises 5.16 through 5.18. They treat the problem of finding shortest paths in two-dimensional Euclidean space. Now consider a general Riemannian manifold with local coordinates x^i and metric tensor $g(x)$. Let y and z be any two nearby points in the manifold, and consider all paths $x(\tau)$ joining y and z such that

$$x(0) = y,$$

$$x(1) = z. \quad (1.8.1)$$

Let a dot denote $(d/d\tau)$. If the manifold is proper Riemannian, we may define a *distance* functional $D[x]$ by the rule

$$D[x] = \int_0^1 d\tau \left[\sum_{ij} g_{ij}(x) \dot{x}^i \dot{x}^j \right]^{1/2}. \quad (1.8.2)$$

If the manifold is either proper or pseudo Riemannian, we may define an “*energy*” functional $E[x]$ by the rule

$$E[x] = (1/2) \int_0^1 d\tau \sum_{ij} g_{ij}(x) \dot{x}^i \dot{x}^j. \quad (1.8.3)$$

A path that extremizes D is called a *geodesic*, and a path that extremizes E is called an *affine geodesic*.⁵⁹ Note that the functional $D[x]$ may not be defined for all paths in a pseudo-Riemannian space because in that case the argument of the square root appearing in (8.2) may be negative. Correspondingly, geodesics do not necessarily exist between all y, z pairs in a pseudo-Riemannian space. By contrast, the functional $E[x]$ is well defined for all paths in both the proper and pseudo-Riemannian cases. (Note that in this simplified discussion we have assumed that the topology of the manifold is that of Euclidean space since we have assumed global coordinates in defining D and/or E . A more general discussion would involve the use of overlapping local coordinate patches.)

Is there a Relation between Geodesics and Affine Geodesics?

Geodesics

- a) Let us begin with the geodesic case. Verify that the functional $D[x]$ does not depend on the parameterization of x . That is, one may replace $x(\tau)$ by $x(\sigma(\tau))$ where $\sigma(\tau)$ is any function satisfying

$$\begin{aligned} \sigma(0) &= 0, \\ \sigma(1) &= 1. \end{aligned} \quad (1.8.4)$$

Therefore, as described in Subsection 5.2 and illustrated in Exercises 5.16, 5.17, 6.5, and 6.9, there will eventually be a need for further information.

- b) The condition for a geodesic is $\delta D = 0$. Verify, by the standard calculus of variations, that this condition is equivalent to Lagrange’s equations for the Lagrangian L_D given by

$$L_D = (g_{ij} \dot{x}^i \dot{x}^j)^{1/2} = [g_{ij} (dx^i/d\tau)(dx^j/d\tau)]^{1/2} = [g_{ij} (dx^i)(dx^j)]^{1/2}/d\tau = ds/d\tau. \quad (1.8.5)$$

⁵⁹The value of $D[x]$ for a geodesic is called the shortest distance between its end points. The value of $E[x]$ for an affine geodesic between its end points is called its “Energy”. Thus, in the context of geodesics, shortest distance and Energy are analogous concepts. Presumably the term “Energy” is employed because L_E , see (8.31), resembles the expression $(1/2)m\mathbf{v} \cdot \mathbf{v}$ for the kinetic energy of a particle in Lagrangian mechanics. The term “Action” probably would have been better since, for a free particle, $\int[(1/2)m\mathbf{v} \cdot \mathbf{v}]d\tau$ is commonly called Action. Strictly speaking, in Physics Energy is not an attribute of paths, but Action is.

Here, and in what follows, we again employ the Einstein summation convention. Verify that L_D has the (unpromising) property

$$\det(\partial^2 L_D / \partial \dot{x}^i \partial \dot{x}^j) = 0, \quad (1.8.6)$$

and that this property arises from the fact that L_D is homogeneous of degree one in the \dot{x}^i , which is why $D[x]$ is parameterization independent. Also verify that the Hamiltonian H_D associated with L_D vanishes identically!

c) Nevertheless, let us push on. As a first step, verify that

$$\partial L_D / \partial \dot{x}^i = g_{ij} \dot{x}^j / (g_{k\ell} \dot{x}^k \dot{x}^\ell)^{1/2} = g_{ij} \dot{x}^j / (ds/d\tau). \quad (1.8.7)$$

Next verify the relations

$$\frac{d}{d\tau} \left(\frac{\partial L_D}{\partial \dot{x}^i} \right) = \left[\frac{d(g_{ij} \dot{x}^j)}{d\tau} \right] \left(\frac{ds}{d\tau} \right)^{-1} + (g_{ij} \dot{x}^j) \frac{d(ds/d\tau)^{-1}}{d\tau}, \quad (1.8.8)$$

$$d(g_{ij} \dot{x}^j) / d\tau = g_{ij} \ddot{x}^j + (\partial g_{ij} / \partial x^k) \dot{x}^j \dot{x}^k, \quad (1.8.9)$$

$$\frac{d(ds/d\tau)^{-1}}{d\tau} = - \left(\frac{ds}{d\tau} \right)^{-2} \frac{d^2 s}{d\tau^2}, \quad (1.8.10)$$

$$\begin{aligned} \partial L_D / \partial x^i &= (1/2)(g_{jk} \dot{x}^j \dot{x}^k)^{-1/2} (\partial g_{jk} / \partial x^i) \dot{x}^j \dot{x}^k \\ &= (1/2)(ds/d\tau)^{-1} (\partial g_{jk} / \partial x^i) \dot{x}^j \dot{x}^k. \end{aligned} \quad (1.8.11)$$

Also verify the identity

$$\partial g_{ij} / \partial x^k = (1/2)(\partial g_{ij} / \partial x^k + \partial g_{ik} / \partial x^j) + (1/2)(\partial g_{ij} / \partial x^k - \partial g_{ik} / \partial x^j), \quad (1.8.12)$$

which decomposes $(\partial g_{ij} / \partial x^k)$ into symmetric and antisymmetric parts under the interchange of j and k . Note that only the symmetric part contributes to the sum $(\partial g_{ij} / \partial x^k) \dot{x}^j \dot{x}^k$ that occurs in (8.9) and (8.11). Thus, verify that Lagrange's equations (5.3) for L_D produce the relations

$$\begin{aligned} g_{ij} \ddot{x}^j (ds/d\tau)^{-1} + (ds/d\tau)^{-1} (1/2)(\partial g_{ij} / \partial x^k + \partial g_{ik} / \partial x^j) \dot{x}^j \dot{x}^k \\ - (g_{ij} \dot{x}_j) (ds/d\tau)^{-2} (d^2 s / d\tau^2) = (1/2)(ds/d\tau)^{-1} (\partial g_{jk} / \partial x^i) \dot{x}^j \dot{x}^k. \end{aligned} \quad (1.8.13)$$

d) Next multiply both sides of (8.13) by $(ds/d\tau)$ and group terms to get the result

$$\begin{aligned} g_{ij} \ddot{x}^j &= g_{ij} \dot{x}^j (ds/d\tau)^{-1} (d^2 s / d\tau^2) + (1/2)\{(\partial g_{jk} / \partial x^i) \\ &\quad - [(\partial g_{ij} / \partial x^k) + (\partial g_{ik} / \partial x^j)]\} \dot{x}^j \dot{x}^k. \end{aligned} \quad (1.8.14)$$

Since g is invertible, it appears that we may solve (8.14) for the \ddot{x}^j . Indeed, multiply both sides of (8.14) by $g^{\ell i}$, where $g^{\ell i}$ is defined by the rule

$$g^{\ell i} = (g^{-1})_{\ell i}, \quad (1.8.15)$$

and sum over i to get the intermediate results

$$g^{\ell i} g_{ij} \ddot{x}^j = g^{\ell i} g_{ij} \dot{x}^j (ds/d\tau)^{-1} (d^2 s/d\tau^2) - \Gamma_{jk}^\ell \dot{x}^j \dot{x}^k. \quad (1.8.16)$$

Here the Γ_{jk}^ℓ are the *Christoffel* symbols/coefficients defined in terms of the metric tensor by the rule

$$\Gamma_{jk}^\ell = (1/2) g^{\ell i} \{ [(\partial g_{ij}/\partial x^k) + (\partial g_{ik}/\partial x^j)] - (\partial g_{jk}/\partial x^i) \}. \quad (1.8.17)$$

Note that they are symmetric under the interchange of the two lower indices. Verify that carrying out the indicated sums in (8.16) and using (8.15) yield the final results

$$\ddot{x}^\ell = \dot{x}^\ell (ds/d\tau)^{-1} (d^2 s/d\tau^2) - \Gamma_{jk}^\ell \dot{x}^j \dot{x}^k. \quad (1.8.18)$$

- e) Have we, contrary to (8.6), succeeded in solving for the \ddot{x}^j ? The answer is *no*, because in general the quantity $(d^2 s/d\tau^2)$ also involves the \dot{x}^j . What is needed is some information about the parameterization. One possibility, also discussed in Exercise 5.17, is to take one of the x^j as the parameter. Another, more democratic, approach is to select the parameterization in such a way that

$$d^2 s/d\tau^2 = 0. \quad (1.8.19)$$

Verify that (8.19) implies relations of the form

$$\begin{aligned} ds/d\tau &= \text{const} = b, \\ s &= a + b\tau, \end{aligned} \quad (1.8.20)$$

where a and b are constants. Moreover, it is natural to set $a = 0$ so that $s = b\tau$ and $s = 0$ when $\tau = 0$. Finally, since both sides of (8.18) are homogeneous of degree 2 in τ when (8.19) holds, verify that we may as well set $b = 1$ so that

$$s = \tau. \quad (1.8.21)$$

When this is done, verify that the equations (8.18) for a geodesic become

$$d^2 x^\ell / ds^2 + \Gamma_{jk}^\ell (dx^j/ds)(dx^k/ds) = 0, \quad (1.8.22)$$

and on this geodesic, according to (8.21), there is the relation

$$(ds/d\tau)^2 = g_{ij} \dot{x}^i \dot{x}^j = 1. \quad (1.8.23)$$

- f) There is a consistency check that may dispel any lingering doubts about the correctness of what we have done. Suppose we solve the equations

$$d^2 x^\ell / d\tau^2 + \Gamma_{jk}^\ell (dx^j/d\tau)(dx^k/d\tau) = 0. \quad (1.8.24)$$

What can be said about $L_D = ds/d\tau$ for such solutions? Verify, by undoing some of the previous steps, that (8.24) is equivalent to the relation

$$g_{ij} \ddot{x}^j + [(\partial g_{ij}/\partial x^k) - (1/2)(\partial g_{jk}/\partial x^i)] \dot{x}^j \dot{x}^k = 0. \quad (1.8.25)$$

Next, verify that multiplying (8.25) by \dot{x}^i and summing over i yields the result

$$g_{ij}\dot{x}^i\ddot{x}^j + (1/2)(\partial g_{ij}/\partial x_k)\dot{x}^i\dot{x}^j\dot{x}^k = 0. \quad (1.8.26)$$

According to (8.5) there is the relation

$$(L_D)^2 = g_{ij}\dot{x}^i\dot{x}^j. \quad (1.8.27)$$

Verify that (8.27) implies the relation

$$L_D(dL_D/d\tau) = g_{ij}\dot{x}^i\ddot{x}^j + (1/2)(\partial g_{ij}/\partial x^k)\dot{x}^i\dot{x}^j\dot{x}^k, \quad (1.8.28)$$

and that (8.26) and (8.28), when combined, yield the relation

$$dL_D/d\tau = 0. \quad (1.8.29)$$

Therefore, the relation

$$ds/d\tau = \text{const} \quad (1.8.30)$$

is a consequence of (8.24), and hence is consistent with (8.24).

Affine Geodesics

- g) Having discussed geodesics at some length, let us now turn to affine geodesics. The condition for an affine geodesic is $\delta E = 0$. Verify that, unlike $D[x]$, the functional $E[x]$ does depend on parameterization. Evidently the Lagrangian L_E for an affine geodesic is given by

$$L_E = (1/2)g_{ij}\dot{x}^i\dot{x}^j. \quad (1.8.31)$$

Verify that in this case

$$\det(\partial^2 L_E/\partial\dot{x}^i\partial\dot{x}^j) \neq 0. \quad (1.8.32)$$

Verify that

$$p_i = \partial L_E/\partial\dot{x}^i = g_{ij}\dot{x}^j = \dot{x}_i, \quad (1.8.33)$$

and that the Hamiltonian H_E associated with L_E is given by

$$H_E = p_i\dot{x}^i - L_E = (1/2)g_{ij}\dot{x}^i\dot{x}^j = (1/2)g^{ij}p_ip_j. \quad (1.8.34)$$

Verify that H_E is a constant of motion, and hence

$$H_E = (1/2)g_{ij}\dot{x}^i\dot{x}^j = (1/2)(ds/d\tau)^2 = \text{const}. \quad (1.8.35)$$

Verify that

$$dp_i/d\tau = g_{ij}\ddot{x}^j + (\partial g_{ij}/\partial x^k)\dot{x}^j\dot{x}^k, \quad (1.8.36)$$

$$\partial L_E/\partial x^i = (1/2)(\partial g_{jk}/\partial x^i)\dot{x}^j\dot{x}^k, \quad (1.8.37)$$

and hence Lagrange's equations of motion yield the relations

$$g_{ij}\ddot{x}^j + [(\partial g_{ij}/\partial x^k) - (1/2)(\partial g_{jk}/\partial x^i)]\dot{x}^j\dot{x}^k = 0. \quad (1.8.38)$$

Verify that these relations can be solved for the \ddot{x}^j to yield the results

$$\ddot{x}^\ell + \Gamma_{jk}^\ell\dot{x}^j\dot{x}^k = 0. \quad (1.8.39)$$

Summary of Results

We have demonstrated that an affine geodesic satisfies (8.35) and (8.39). Comparison of (8.24) and 8.39 shows that a geodesic, when it exists and is parameterized to satisfy $\tau = s/b$, is also an affine geodesic. Conversely, an affine geodesic always exists, is always automatically parameterized to satisfy (8.35), and yields a geodesic parameterized to satisfy $\tau = s/b$ when such exists. Thus, there is no loss of generality in working with affine geodesics, and they have the advantage of being defined even when the metric is not positive definite.

Closing Remarks

There is yet one more set of remarks of interest. Let x^0 be some point and consider some affine geodesic through x^0 parameterized in such a way that $x(\tau) = x^0$ when $\tau = 0$. Let us see what can be said about the quantity $(1/2)g_{ij}\dot{x}^i\dot{x}^j$ at this point. Since the metric tensor $g_{ij}(x^0)$ at this point, when regarded as a matrix, is a real symmetric matrix, it can be diagonalized by a similarity transformation employing a real orthogonal matrix. Moreover, all its eigenvalues will be real. Next, by proper scaling of the coordinates, g_{ij} at this point can be brought to a diagonal form where each of its eigenvalues are either $+1$, 0 , or -1 ; and the numbers of eigenvalues of each kind are invariants.⁶⁰ Since we have assumed that g_{ij} is invertible, we will exclude from our discussion the case where any of the eigenvalues vanish. Then, in the case that g_{ij} is positive definite, all the eigenvalues (after diagonalization and suitable coordinate scaling) may be taken to be $+1$. Correspondingly, the value of $(1/2)g_{ij}\dot{x}^i\dot{x}^j$ at this point will be positive, and (after suitable rescaling of the parameter τ) we may confine our attention to the case for which it has the value $1/2$.⁶¹ Similarly, if g_{ij} is negative definite, all the eigenvalues may be taken to be -1 . In this case the value occurring $(1/2)g_{ij}\dot{x}^i\dot{x}^j$ at this point will be negative, and we may confine our attention to the case for which it has the value $-1/2$. Finally, in the pseudo-Riemannian case, some of the eigenvalues will be positive and some will be negative.⁶² In this circumstance we may confine our attention to three classes of cases: those for which $(1/2)g_{ij}\dot{x}^i\dot{x}^j$ has the value $+1/2$, those for which it has the value 0 , and those for which it has the value $-1/2$.

All these considerations apply to the possible values of $(1/2)g_{ij}\dot{x}^i\dot{x}^j$ at the point x^0 for affine geodesics through the point x^0 . But now, according to (8.35), the value of $(1/2)g_{ij}\dot{x}^i\dot{x}^j$ remains constant *all along* an affine geodesic. Therefore in the positive definite case we may restrict our attention to affine geodesics for which the constant appearing on the right side of (8.35) has the value $1/2$; and in the negative definite case we may restrict our attention to affine geodesics for which the constant has the value $-1/2$. Finally, in the pseudo Riemannian case, we may restrict our attention to those affine geodesics for which the constant has the values $1/2$, 0 , and $-1/2$.

⁶⁰This result is called *Sylvester's law of inertia* for quadratic forms.

⁶¹Note that (8.39) is invariant under rescaling of τ .

⁶²For example, in the case (7.2), one of the eigenvalues is $+1$ and three are -1 .

1.8.2 Affine Geodesics in Minkowski Space

Review Subsection 8.1 which described geodesics and affine geodesics. Our task in this subsection is to find affine geodesics in Minkowski space and to study their properties.

Let y and z be any two points in Minkowski space. Let $x(\tau)$ be a parameterized path in Minkowski space with parameter τ ranging between an *initial* value τ^{in} and a *final* value τ^{fin} ,

$$\tau \in [\tau^{in}, \tau^{fin}], \quad (1.8.40)$$

and stipulate that $x(\tau)$ connects y and z by requiring that

$$x(\tau^{in}) = y \quad (1.8.41)$$

and

$$x(\tau^{fin}) = z. \quad (1.8.42)$$

For the case of affine geodesics in Minkowski space we will employ the Lagrangian (8.31) and the metric (7.2) to find

$$L_E = (1/2) \sum_{ij} g_{ij} \dot{x}^i \dot{x}^j = (1/2) [-(\dot{x}^1)^2 - (\dot{x}^2)^2 - (\dot{x}^3)^2 + (\dot{x}^4)^2]. \quad (1.8.43)$$

For the energy functional $E[x]$ (and employing the parameterization just described) we will write

$$E[x] = \int_{\tau^{in}}^{\tau^{fin}} L_E(\tau) d\tau = (1/2) \int_{\tau^{in}}^{\tau^{fin}} d\tau \sum_{ij} g_{ij} \dot{x}^i \dot{x}^j. \quad (1.8.44)$$

Verify that use of Lagrange's equations and the Lagrangian (8.43) yields for the *affine geodesic*, which we will call $x_{ag}(\tau)$, the "equations of motion"

$$\ddot{x}_{ag}^i = 0. \quad (1.8.45)$$

[Alternatively, verify (as is to be expected) that for the Minkowski metric use of (8.17) and (8.39) also yields (8.45).] Evidently each $x_{ag}^i(\tau)$ must be a sum of constant and linear functions of τ . Therefore we may write the vector relations

$$x_{ag}(\tau) = a + b\tau, \quad (1.8.46)$$

and the boundary conditions (8.41) and (8.42) yield the relations

$$a + b\tau^{in} = y \quad (1.8.47)$$

and

$$a + b\tau^{fin} = z. \quad (1.8.48)$$

Verify that the solution to the pair (8.47) and (8.48) is

$$a = (y\tau^{fin} - z\tau^{in}) / (\tau^{fin} - \tau^{in}) \quad (1.8.49)$$

and

$$b = (z - y) / (\tau^{fin} - \tau^{in}). \quad (1.8.50)$$

Consequently, (8.46) becomes

$$x_{ag}(\tau) = (y\tau^{fin} - z\tau^{in})/(\tau^{fin} - \tau^{in}) + \tau(z - y)/(\tau^{fin} - \tau^{in}). \quad (1.8.51)$$

The affine geodesic joining any two points y and z in Minkowski space is the *straight line* between y and z .

So far we have viewed the problem of finding an affine geodesic as a *two-point boundary value* problem. Namely, solve the affine geodesic differential equation (8.45) with the two boundary conditions (8.41) and (8.42). It is known from general differential equation theory that two-point boundary value problems do not always have solutions. Alternatively, we may consider the affine geodesic problem as an *initial value* problem. Namely, solve (8.45) with the initial values given by (8.41) and

$$(dx/d\tau)|_{\tau=\tau^{in}} = b. \quad (1.8.52)$$

Unlike two-point boundary value problems, initial value problems, as described in Subsection 1.3, generally have solutions. The difficulty with two-point boundary value problems is that, given y and τ^{fin} , there may be no value of b for which the second boundary condition (8.42) can be satisfied. Fortunately, as we have already seen, that must not be the case for affine geodesics in Minkowski space.

To continue, let us now solve the initial value problem. In this case the affine geodesic differential equation (8.45) with the initial conditions (8.41) and (8.52) has the solution

$$x_{ag}(\tau) = y + (\tau - \tau^{in})b, \quad (1.8.53)$$

and one hopes to satisfy the second boundary-condition relation

$$z = x_{ag}(\tau^{fin}) = y + (\tau^{fin} - \tau^{in})b. \quad (1.8.54)$$

Verify that indeed there is a value of b which satisfies (8.54) and that this value is given by (8.50). Finally, verify that when this value for b is employed in (8.53), the result is (8.51).

Let E be the value of the Energy functional $E[x]$ for the affine geodesic given by (8.51). By almost slight if not abuse of hand, we write

$$\sum_{ij} g_{ij} \dot{x}^i \dot{x}^j = \sum_{ij} g_{ij} (dx^i/d\tau)(dx^j/d\tau) = (d\tau)^{-2} \sum_{ij} g_{ij} (dx^i)(dx^j) = ds^2/(d\tau)^2. \quad (1.8.55)$$

Correspondingly, with no further abuse of notation, we may write

$$\begin{aligned} E = E[x_{ag}] &= \int_{\tau^{in}}^{\tau^{fin}} L_E(\tau) d\tau = (1/2) \int_{\tau^{in}}^{\tau^{fin}} d\tau \sum_{ij} g_{ij} \dot{x}_{ag}^i \dot{x}_{ag}^j \\ &= (1/2) \int_{\tau^{in}}^{\tau^{fin}} d\tau [ds^2/(d\tau)^2]|_{x=x_{ag}}, \end{aligned} \quad (1.8.56)$$

thereby indicating that E involves use of the infinitesimal interval ds^2 in an integration process.

What remains is to evaluate E , which in this case is easier than might have been imagined. Verify using (8.51) that, for points on the affine geodesic in Minkowski space, $\dot{x}_{ag}(\tau)$ has the *constant* value

$$\dot{x}_{ag} = (z - y)/(\tau^{fin} - \tau^{in}), \quad (1.8.57)$$

and therefore L_E , when evaluated on the affine geodesic, has the *constant* value

$$L_E(\tau) = (1/2) \sum_{ij} g_{ij} \dot{x}_{ag}^i \dot{x}_{ag}^j = (1/2)(\tau^{fin} - \tau^{in})^{-2} \sum_{ij} g_{ij}(z - y)^i(z - y)^j. \quad (1.8.58)$$

Correspondingly, show that the energy functional (8.3), when evaluated on the affine geodesic joining y and z , has the value

$$E = \int_{\tau^{in}}^{\tau^{fin}} d\tau L_E = (1/2)(\tau^{fin} - \tau^{in})^{-1} \sum_{ij} g_{ij}(z - y)^i(z - y)^j. \quad (1.8.59)$$

The quantity $I(u, v)$ defined for any pair of points u, v by the rule

$$I(u, v) = \sum_{ij} g_{ij}(v - u)^i(v - u)^j \quad (1.8.60)$$

will be called the *net interval* between u and v . Evidently it has the symmetry property

$$I(u, v) = I(v, u). \quad (1.8.61)$$

With the aid of the definition (8.60) we have for (8.59) the result

$$E = (1/2)(\tau^{fin} - \tau^{in})^{-1} I(y, z). \quad (1.8.62)$$

1.8.3 Relation Between Infinitesimal Interval ds^2 and Net Interval I

One of the aims of this subsection is to relate the net interval $I(u, v)$ and the infinitesimal interval ds^2 . Upon comparing (8.56) and (8.62) we see that, for the affine geodesic joining the points y and z in Minkowski space, there is the result

$$\int_{\tau^{in}}^{\tau^{fin}} d\tau [ds^2/(d\tau)^2]|_{x=x_{ag}} = (\tau^{fin} - \tau^{in})^{-1} I(y, z) \quad (1.8.63)$$

where the integral appearing on the left side is to be carried out over the affine geodesic connecting y and z . The net and infinitesimal intervals I and ds^2 are related by integration.

To continue on, consider a curve in some space and suppose this curve is viewed as a collection of pieces. In the case of Euclidean space the *length* of a curve is the sum of the lengths of its pieces. And if the curve is a geodesic, then each piece will also be a geodesic connecting its ends. In the case of Minkowski space there is a unique affine geodesic connecting any two points y and z in Minkowski space. Are there analogous relations between the interval $I(y, z)$ for the whole affine geodesic and the intervals for pieces of the affine geodesic?

Imagine, for example, that the affine geodesic joining y and z is viewed as consisting of two pieces. Let τ^{int} be some *intermediate* parameter value between τ^{in} and τ^{fin} ,

$$\tau^{int} \in (\tau^{in}, \tau^{fin}). \quad (1.8.64)$$

Let x_{int} be the corresponding intermediate point on the affine geodesic joining y and z . That is, define x_{int} to be given by (6.244) evaluated at $\tau = \tau^{int}$,

$$x_{int} = x_{ag}(\tau^{int}). \quad (1.8.65)$$

Show, using (6.234), (6.235), (6.239), and (6.243), that then there are the relations

$$\begin{aligned} x_{int} - y &= x_{ag}(\tau^{int}) - y = x_{ag}(\tau^{int}) - x_{ag}(\tau^{in}) = b(\tau^{int} - \tau^{in}) \\ &= [(\tau^{int} - \tau^{in})/(\tau^{fin} - \tau^{in})](z - y) \end{aligned} \quad (1.8.66)$$

and

$$\begin{aligned} z - x_{int} &= z - x_{ag}(\tau^{int}) = x_{ag}(\tau^{fin}) - x_{ag}(\tau^{int}) = b(\tau^{fin} - \tau^{int}) \\ &= [(\tau^{fin} - \tau^{int})/(\tau^{fin} - \tau^{in})](z - y). \end{aligned} \quad (1.8.67)$$

Verify, using (6.259) and (6.260), that there are the relations

$$I(y, x_{int}) = [(\tau^{int} - \tau^{in})/(\tau^{fin} - \tau^{in})]^2 I(y, z) \quad (1.8.68)$$

and

$$I(x_{int}, z) = [(\tau^{fin} - \tau^{int})/(\tau^{fin} - \tau^{in})]^2 I(y, z). \quad (1.8.69)$$

Evidently both $I(y, x_{int})$ and $I(x_{int}, z)$ are fractions of $I(y, z)$, but these fractions do not sum to one, nor to any number independent of τ^{int} . Their sum is

$$\begin{aligned} &[(\tau^{int} - \tau^{in})/(\tau^{fin} - \tau^{in})]^2 + [(\tau^{fin} - \tau^{int})/(\tau^{fin} - \tau^{in})]^2 \\ &= [(\tau^{fin})^2 + (\tau^{in})^2 - 2\tau^{int}(\tau^{fin} + \tau^{in}) + 2(\tau^{int})^2]/[\tau^{fin} - \tau^{in}]^2. \end{aligned} \quad (1.8.70)$$

Verify that they sum to one only in the cases $\tau^{int} = \tau^{in}$ and $\tau^{int} = \tau^{fin}$.

We have seen that the *interval* for Minkowskian geometry affine geodesics, unlike curve *length* for Euclidean geometry, is *not* additive. However, we will soon see that there are suitably weighted intervals that do have admirable additive properties. We begin by observing that the *energy* functional is additive. Start by rewriting (6.249), using (6.251), in the form

$$\begin{aligned} E_{\tau^{in}, \tau^{fin}} = E &= \int_{\tau^{in}}^{\tau^{fin}} d\tau L_E = (1/2)(\tau^{fin} - \tau^{in})^{-2} \sum_{ij} g_{ij}(z - y)^i(z - y)^j \int_{\tau^{in}}^{\tau^{fin}} d\tau \\ &= (\tau^{fin} - \tau^{in})(1/2)(\tau^{fin} - \tau^{in})^{-2} I(y, z) \\ &= (\tau^{fin} - \tau^{in})^{-1}(1/2)I(y, z). \end{aligned} \quad (1.8.71)$$

Here the notation $E_{\tau^{in}, \tau^{fin}}$ is meant to signify the value of the energy functional for the *full* affine geodesic joining y and z . Similarly define $E_{\tau^{in}, \tau^{int}}$ and $E_{\tau^{int}, \tau^{fin}}$ by the rules

$$\begin{aligned} E_{\tau^{in}, \tau^{int}} &= \int_{\tau^{in}}^{\tau^{int}} d\tau L_E = (1/2)(\tau^{fin} - \tau^{in})^{-2} \sum_{ij} g_{ij}(z - y)^i(z - y)^j \int_{\tau^{in}}^{\tau^{int}} d\tau \\ &= (\tau^{int} - \tau^{in})(1/2)(\tau^{fin} - \tau^{in})^{-2} I(y, z) \end{aligned} \quad (1.8.72)$$

and

$$\begin{aligned} E_{\tau^{int}, \tau^{fin}} &= \int_{\tau^{int}}^{\tau^{fin}} d\tau L_E = (1/2)(\tau^{fin} - \tau^{in})^{-2} \sum_{ij} g_{ij}(z-y)^i(z-y)^j \int_{\tau^{int}}^{\tau^{fin}} d\tau \\ &= (\tau^{fin} - \tau^{int})(1/2)(\tau^{fin} - \tau^{in})^{-2} I(y, z). \end{aligned} \quad (1.8.73)$$

They are the values of the energy functional for the pieces of the affine geodesics corresponding to $\tau \in (\tau^{in}, \tau^{int})$ and $\tau \in (\tau^{int}, \tau^{fin})$, respectively. Check that, essentially by construction, they satisfy the relation

$$E_{\tau^{in}, \tau^{int}} + E_{\tau^{int}, \tau^{fin}} = E_{\tau^{in}, \tau^{fin}}, \quad (1.8.74)$$

thereby illustrating that the energy functional is additive when an affine geodesic is broken into two pieces.

There is one last possible defect in our presentation. Namely the quantities $E_{\tau^{in}, \tau^{int}}$ and $E_{\tau^{int}, \tau^{fin}}$ for the two pieces, as given in (6.265) and (6.266), appear to depend on y and z since they involve $I(y, z)$ on their right sides. Show that this defect can be removed by employing (6.261) and (6.262) in (6.265) and (6.266) to rewrite them in the forms

$$\begin{aligned} E_{\tau^{in}, \tau^{int}} &= (\tau^{int} - \tau^{in})(1/2)(\tau^{fin} - \tau^{in})^{-2} I(y, z) \\ &= (1/2)(\tau^{int} - \tau^{in})^{-1} I(y, x_{int}) \end{aligned} \quad (1.8.75)$$

and

$$\begin{aligned} E_{\tau^{int}, \tau^{fin}} &= (\tau^{fin} - \tau^{int})(1/2)(\tau^{fin} - \tau^{in})^{-2} I(y, z) \\ &= (1/2)(\tau^{fin} - \tau^{int})^{-1} I(x_{int}, z). \end{aligned} \quad (1.8.76)$$

Using (6.264), (6.268), and (6.269) verify that the sum rule (6.267) becomes

$$(\tau^{int} - \tau^{in})^{-1} I(y, x_{int}) + (\tau^{fin} - \tau^{int})^{-1} I(x_{int}, z) = (\tau^{fin} - \tau^{in})^{-1} I(y, z). \quad (1.8.77)$$

Note that each term in (6.270) depends *only* on quantities associated with a particular piece of affine geodesic. You have shown (in the case of two pieces) that intervals for pieces of affine geodesics obey what we may call *weighted* sum rules.

We expect analogous results may hold when the full affine geodesic joining y and z is broken into many pieces. Let us explore one way of doing so and its consequences. Suppose the parameter range $[\tau^{in}, \tau^{fin}]$ is broken into N pieces of *equal size*/“duration” Δ . (For a treatment of the more general case where the pieces may have different durations, see Chapter 28.) When the pieces have equal duration, define Δ by writing

$$\Delta = (\tau^{fin} - \tau^{in})/N. \quad (1.8.78)$$

Also, define intermediate parameter values τ_n by the rule

$$\tau_n = \tau^{in} + n\Delta \quad (1.8.79)$$

so that

$$\tau_0 = \tau^{in}, \tau_1 = \tau^{in} + \Delta, \dots \tau_N = \tau^{in} + N\Delta = \tau^{fin} \quad (1.8.80)$$

and

$$\tau_{n+1} - \tau_n = \Delta. \quad (1.8.81)$$

And define points x_n on the affine geodesic by writing

$$x_n = x_{ag}(\tau_n) \quad (1.8.82)$$

so that

$$\begin{aligned} x_0 &= x_{ag}(\tau_0) = x_{ag}(\tau^{in}) = y, \quad x_1 = x_{ag}(\tau_1) = x_{ag}(\tau^{in} + \Delta), \quad \dots \\ x_N &= x_{ag}(\tau_N) = x_{ag}(\tau^{fin}) = z. \end{aligned} \quad (1.8.83)$$

With the aid of the notation just introduced, conjecture that the N -piece version of (6.270) should read

$$\begin{aligned} &(\tau_1 - \tau^{in})^{-1} I(y, x_1) + (\tau_2 - \tau_1)^{-1} I(x_1, x_2) + \dots + (\tau^{fin} - \tau_{N-1})^{-1} I(x_{N-1}, z) \\ &= (\tau^{fin} - \tau^{in})^{-1} I(y, z). \end{aligned} \quad (1.8.84)$$

Let us check. Note that

$$\begin{aligned} x_{n+1} - x_n &= x_{ag}(\tau_{n+1}) - x_{ag}(\tau_n) = (a + b\tau_{n+1}) - (a + b\tau_n) = b(\tau_{n+1} - \tau_n) \\ &= b\Delta = \Delta(z - y)/(\tau^{fin} - \tau^{in}) = \Delta(z - y)/(N\Delta) = (z - y)/N. \end{aligned} \quad (1.8.85)$$

Here we have used (6.239), (6.243), (6.271), and (6.275). Next verify that

$$I(x_n, x_{n+1}) = I(y, z)/N^2 \quad (1.8.86)$$

and

$$(\tau_{n+1} - \tau_n)^{-1} I(x_n, x_{n+1}) = (1/\Delta) I(y, z)/N^2 = (1/N) (\tau^{fin} - \tau^{in})^{-1} I(y, z). \quad (1.8.87)$$

Finally, since all the terms on the left side of (6.277) have the same value (6.280) and there are N of them, we see that (6.277) is indeed correct.

Now suppose N is large so that Δ is small and the differences $(x_{n+1} - x_n)$ are small. Since in our application all the differences $(x_{n+1} - x_n)$ are also the *same*, see (6.278), introduce the not entirely awkward notation

$$(x_{n+1} - x_n) = dx. \quad (1.8.88)$$

Then for the interval $I(x_{n+1}, x_n)$ we may write, again in somewhat awkward but also usual notation, the relation

$$I(x_n, x_{n+1}) = \sum_{ij} g_{ij} (x_{n+1} - x_n)^i (x_{n+1} - x_n)^j = \sum_{ij} g_{ij} (dx)^i (dx)^j = ds^2. \quad (1.8.89)$$

Correspondingly, there is the result

$$(\tau_{n+1} - \tau_n)^{-1} I(x_n, x_{n+1}) = N(\tau^{fin} - \tau^{in})^{-1} ds^2. \quad (1.8.90)$$

Therefore we may also write

$$\begin{aligned} & (\tau_1 - \tau^{in})^{-1} I(y, x_1) + (\tau_2 - \tau_1)^{-1} I(x_1, x_2) + \cdots + (\tau^{fin} - \tau_{N-1})^{-1} I(x_{N-1}, z) \\ &= N(\tau_{n+1} - \tau_n)^{-1} I(x_n, x_{n+1}) = (\tau^{fin} - \tau^{in})^{-1} N^2 ds^2, \end{aligned} \quad (1.8.91)$$

from which it follows that, in the limit of large N and (quadratically) vanishing ds^2 , there is the relation

$$N(Nds^2) = N^2 ds^2 = I(y, z). \quad (1.8.92)$$

Compare (6.277) and (6.284).

Thus, the interval $I(y, z)$ may be viewed as the result of summing the N quantities Nds^2 associated with the N pieces into which the affine geodesic has been broken.⁶³ By breaking the affine geodesic into an ever larger number of pieces and summing the results from each piece, we have performed a species of *integration*, along an affine geodesic path, that converts ds^2 into $I(y, z)$. Recall (6.256).

Conversely, going the other way, we find the reverse relation

$$I(x, x + dx) = \sum_{ij} g_{ij}(x + dx - x)^i (x + dx - x)^j = \sum_{ij} g_{ij}(dx)^i (dx)^j = ds^2. \quad (1.8.93)$$

Note that (6.281) and (6.282), taken together, is a special case of (6.286).

Before moving on to one final subtopic there are two more items to be explored. The first is an aspect of the interval function $I(*, *)$, namely the large N limit of the individual-piece contribution $(\tau_{n+1} - \tau_n)^{-1} I(x_n, x_{n+1})$. Let us begin. Since in our application all the quantities $(\tau_{n+1} - \tau_n)$ are the same, introduce the notation

$$(\tau_{n+1} - \tau_n) = \Delta = d\tau. \quad (1.8.94)$$

Then we have the approximation

$$(x_{n+1} - x_n) = [x(\tau_n + d\tau) - x(\tau_n)] = d\tau \dot{x}(\tau_n) + O[(d\tau)^2]. \quad (1.8.95)$$

Correspondingly show, with the aid of (6.254) and (6.288), that there is the result

$$I(x_n, x_{n+1}) = \sum_{ij} g_{ij}(x_{n+1} - x_n)^i (x_{n+1} - x_n)^j = (d\tau)^2 \sum_{ij} g_{ij}(\dot{x})^i (\dot{x})^j + O[(d\tau)^3], \quad (1.8.96)$$

and therefore, see (6.224), there is also the relation

$$(\tau_{n+1} - \tau_n)^{-1} I(x_n, x_{n+1}) = (d\tau) 2L_E + O[(d\tau)^2]. \quad (1.8.97)$$

Finally, show that summing all the N (equal) quantities that appear on the left side of (6.277) and letting N approach infinity produces the result

$$\begin{aligned} & (\tau_1 - \tau^{in})^{-1} I(y, x_1) + (\tau_2 - \tau_1)^{-1} I(x_1, x_2) + \cdots + (\tau^{fin} - \tau_{N-1})^{-1} I(x_{N-1}, z) \\ &= N(\tau_{n+1} - \tau_n)^{-1} I(x_n, x_{n+1}) = N(d\tau) 2L_E + NO[(d\tau)^2] \rightarrow 2 \int_{\tau^{in}}^{\tau^{fin}} L_E d\tau = 2E. \end{aligned} \quad (1.8.98)$$

⁶³Note that the quantity Nds^2 vanishes *linearly* as N goes to infinity.

Verify that (6.256), (6.264), (6.277), and (6.291) are consistent.

Here is the second item: So far, in our study of what happens when an affine geodesic is broken into multiple pieces, we have assumed that the parameter values τ_n were *equally* spaced. See (6.272) through (6.274). What happens if we relax this assumption? Suppose we assume only that

$$\tau_{n+1} > \tau_n \quad (1.8.99)$$

so that

$$\tau^{in} < \tau_1 < \tau_2 \cdots \tau_{N-1} < \tau^{fin}. \quad (1.8.100)$$

For the corresponding points x_n on the affine geodesic we will still retain the rule (6.275). What then can be said about $I(x_n, x_{n+1})$? Verify that

$$x_{n+1} - x_n = x_{ag}(\tau_{n+1}) - x_{ag}(\tau_n) = b(\tau_{n+1} - \tau_n) = [(\tau_{n+1} - \tau_n)/(\tau^{fin} - \tau^{in})](z - y). \quad (1.8.101)$$

Consequently, find that

$$\begin{aligned} I(x_n, x_{n+1}) &= \sum_{ij} g_{ij}(x_{n+1} - x_n)^i (x_{n+1} - x_n)^j \\ &= [(\tau_{n+1} - \tau_n)/(\tau^{fin} - \tau^{in})]^2 \sum_{ij} g_{ij}(z - y)^i (z - y)^j \\ &= [(\tau_{n+1} - \tau_n)/(\tau^{fin} - \tau^{in})]^2 I(y, z). \end{aligned} \quad (1.8.102)$$

Correspondingly, verify that

$$(\tau_{n+1} - \tau_n)^{-1} I(x_n, x_{n+1}) = (\tau_{n+1} - \tau_n)(\tau^{fin} - \tau^{in})^{-2} I(y, z). \quad (1.8.103)$$

With the aid of (6.296), show that the sum rule (6.277) still holds even when the τ_n are not equally spaced. All that is required is that the affine geodesic be broken into N pieces which, for computational convenience, we have taken to be contiguous.

At this point we may also talk about the Energy associated with a piece of an affine geodesic. Let $E_{\tau_n, \tau_{n+1}}$ be the Energy associated with that portion of the affine geodesic $x_{ag}(\tau)$ for which $\tau \in [\tau_n, \tau_{n+1}]$. Consistent with our earlier discussion/notation, it is defined by the integral

$$E_{\tau_n, \tau_{n+1}} = \int_{\tau_n}^{\tau_{n+1}} L_E(\tau) d\tau. \quad (1.8.104)$$

In our application we also know that on the affine geodesic L_E has the constant value given by (6.251). Verify that (6.251) can be rewritten in the form

$$L_E = (1/2)(\tau^{fin} - \tau^{in})^{-2} I(y, z). \quad (1.8.105)$$

It follows that

$$E_{\tau_n, \tau_{n+1}} = \int_{\tau_n}^{\tau_{n+1}} L_E(\tau) d\tau = L_E \int_{\tau_n}^{\tau_{n+1}} d\tau = (1/2)(\tau_{n+1} - \tau_n)(\tau^{fin} - \tau^{in})^{-2} I(y, z). \quad (1.8.106)$$

Verify that combining (6.296) and (6.299) yields the result

$$2E_{\tau_n, \tau_{n+1}} = (\tau_{n+1} - \tau_n)^{-1} I(x_n, x_{n+1}). \quad (1.8.107)$$

Finally, verify that combining (6.264) and (6.300) converts the sum rule (6.277) into the relation

$$E_{\tau^{in}, \tau_1} + E_{\tau_1, \tau_2} + \cdots + E_{\tau_{N-1}, \tau^{fin}} = E_{\tau^{in}, \tau^{fin}} = E. \quad (1.8.108)$$

As expected, the energy for the full geodesic connecting y and z is the sum of the energies of its pieces. But recall that an earlier footnote in Exercise 6.16 commented that what, in the context of affine geodesics, is called Energy could better be called Action. Thus, what would be better to say is that the Action for the full affine geodesic connecting points y and z in Minkowski space is the sum of the Actions of its pieces.

Your last chore is to verify some statements about affine geodesics in Minkowski space and Poincaré and Lorentz transformations. Let w be any point in Minkowski space, and suppose f is a function defined on Minkowski space that sends the point w to the point \bar{w} , also in Minkowski space,

$$\bar{w} = f(w). \quad (1.8.109)$$

Next let y and z be any pair of points in Minkowski space, and let \bar{y} and \bar{z} be their images under the action of f ,

$$\bar{y} = f(y) \text{ and } \bar{z} = f(z). \quad (1.8.110)$$

Now make the requirement that

$$I(\bar{y}, \bar{z}) = I(y, z) \quad (1.8.111)$$

for all pairs y, z . Then it can be shown that f must be of the form

$$\bar{w} = f(w) = d + \Lambda w \quad (1.8.112)$$

where d is some fixed four-component vector and Λ is some fixed 4×4 matrix. That is, f consists of a linear transformation described by Λ and a translation described by d .⁶⁴ See Exercise 7.3.26. (The converse of this assertion is treated in Exercise 6.2.6.) Transformations of the form (6.305) are called *Poincaré* transformations. In the special case $d = 0$ they are called *Lorentz* transformations.

We will now have the pleasure of examining the effect of Poincaré transformations on affine geodesics. We have already shown that the affine geodesic $x(\tau)$ joining y and z is given by (6.244), which we repeat below:

$$x_{ag}(\tau) = (y\tau^{fin} - z\tau^{in})/(\tau^{fin} - \tau^{in}) + \tau(z - y)/(\tau^{fin} - \tau^{in}). \quad (1.8.113)$$

Verify that similarly the affine geodesic, call it $\hat{x}_{ag}(\tau)$, joining \bar{y} and \bar{z} is given by

$$\hat{x}_{ag}(\tau) = (\bar{y}\tau^{fin} - \bar{z}\tau^{in})/(\tau^{fin} - \tau^{in}) + \tau(\bar{z} - \bar{y})/(\tau^{fin} - \tau^{in}). \quad (1.8.114)$$

⁶⁴Transformations that consist of a linear transformation followed by a translation are called *affine* transformations.

Verify that employing (6.303) and (6.305) in (6.307) yields the result

$$\begin{aligned}
 \hat{x}_{ag}(\tau) &= (\bar{y}\tau^{fin} - \bar{z}\tau^{in})/(\tau^{fin} - \tau^{in}) + \tau(\bar{z} - \bar{y})/(\tau^{fin} - \tau^{in}) \\
 &= [(d + \Lambda y)\tau^{fin} - (d + \Lambda z)\tau^{in}]/(\tau^{fin} - \tau^{in}) \\
 &\quad + \tau[(d + \Lambda z) - (d + \Lambda y)]/(\tau^{fin} - \tau^{in}) \\
 &= d + [\Lambda(y\tau^{fin} - z\tau^{in})]/(\tau^{fin} - \tau^{in}) \\
 &\quad + \tau[\Lambda(z - y)]/(\tau^{fin} - \tau^{in}) \\
 &= d + \Lambda x_{ag}(\tau) \\
 &= \bar{x}_{ag}(\tau).
 \end{aligned} \tag{1.8.115}$$

You have shown that the affine geodesic joining the transformed points \bar{y} and \bar{z} is the result of applying the Poincaré transformation under consideration to the affine geodesic joining the original points y and z . Poincaré transformations map, point by point, affine geodesics into affine geodesics.

What have we learned in this exercise? Below is a list.

- Affine geodesics in Minkowski space based on the Minkowski metric g given by (6.45) are *straight* lines. There is a unique affine geodesic that connects any pair of points y, z . See (6.244).
- Suppose we define the *duration* of an affine geodesic, or piece of an affine geodesic, to be the difference of the parameter (τ) values at its endpoints. For example, the duration of a full affine geodesic is $\tau^{fin} - \tau^{in}$. And, if an affine geodesic is broken into N pieces, then the duration of the n^{th} piece is $\tau_{n+1} - \tau_n$. By definition, duration is *additive*. For example, there is the relation

$$(\tau_2 - \tau_1) + (\tau_3 - \tau_2) = (\tau_3 - \tau_1). \tag{1.8.116}$$

- A key property of any two points in Minkowski space is the *net interval* I defined by (6.253). The net interval I and the *infinitesimal interval* ds^2 are connected by the integral and differential relations (6.256) and (6.286).
- Poincaré transformations are defined to be mappings of Minkowski space into itself that preserve the net interval between all pairs of points, and can be *proved* to be linear/affine. See Exercise 7.3.26.
- Poincaré transformations map affine geodesics into affine geodesics. See (6.308).
- The *Action* (Energy) of a full affine geodesic, or any piece of an affine geodesic, is defined to be the ratio

$$\text{Action} = (1/2)(\text{Net Interval})/\text{Duration}. \tag{1.8.117}$$

See (6.255) and (6.300). Action is additive. See (6.277) and (6.301).

1.8.4 Proof that Lorentz Transformations Must Be Linear

Also review Exercise 6.2.6 which explored transformations of Minkowski space that preserved ds^2 under the assumption that they were *affine*, that is were *linear* transformations combined with translations. The purpose of this exercise is to show that the assumption of linearity/“affinity” is unnecessary.

Let y be any point in Minkowski space, and suppose f is a function defined on Minkowski space that sends the point y to the point y' , also in Minkowski space,

$$y' = f(y). \quad (1.8.118)$$

Suppose that f has the property that

$$I(y', z') = I(y, z) \quad (1.8.119)$$

for all pairs of points y, z in Minkowski space. Our goal is to show/prove, without any additional assumptions of linearity/affinity, that f must be of the form

$$y' = f(y) = d + \Lambda y \quad (1.8.120)$$

where d is some fixed four-component vector and Λ is some fixed 4×4 matrix. (The converse of this assertion has already been treated in Exercise 6.2.6.)

Begin by defining d to be the image of the origin,

$$d = f(0), \quad (1.8.121)$$

and define a new transformation $h(y)$ by the rule

$$h(y) = f(y) - d. \quad (1.8.122)$$

Verify that h maps the origin into itself,

$$h(0) = 0, \quad (1.8.123)$$

and also satisfies

$$I\{h(y), h(z)\} = I(y, z). \quad (1.8.124)$$

Verify that explicit evaluation of both sides of (3.150) gives the result

$$h(y) \cdot h(y) - 2h(y) \cdot h(z) + h(z) \cdot h(z) = y \cdot y - 2y \cdot z + z \cdot z. \quad (1.8.125)$$

Show that setting $z = 0$ in (3.151) gives the result

$$h(y) \cdot h(y) = y \cdot y. \quad (1.8.126)$$

Show that combining (3.151) and (3.152) yields the result

$$h(y) \cdot h(z) = y \cdot z. \quad (1.8.127)$$

Let e^1 to e^4 be the points/vectors

$$e^1 = (1000), e^2 = (0100), \text{ etc.} \quad (1.8.128)$$

Define points/vectors c^j by the rule

$$c^j = h(e^j). \quad (1.8.129)$$

Show that using (3.153) and (3.155) gives the result

$$c^i \cdot c^j = h(e^i) \cdot h(e^j) = e^i \cdot e^j = g^{ij}. \quad (1.8.130)$$

Prove, therefore, that the vectors c^j are linearly independent and can be used as a basis set. Now define a matrix Λ by the rule

$$\Lambda e^j = c^j, \quad (1.8.131)$$

and show that this definition results in the explicit relation

$$\Lambda^{ij} = (e^i, \Lambda e^j) = (e^i, c^j) \quad (1.8.132)$$

where $(*, *)$ denotes the usual/ordinary scalar product.

Let y be an arbitrary point having the expansion

$$y = \sum_j \eta^j e^j, \quad (1.8.133)$$

and set

$$y'' = h(y). \quad (1.8.134)$$

Show, since the c^j form a basis, that one may write

$$y'' = \sum_j g^{jj} \{c^j \cdot y''\} c^j. \quad (1.8.135)$$

Using (3.153), (3.155), (3.159), and (3.160), verify that

$$c^j \cdot y'' = h(e^j) \cdot h(y) = e^j \cdot y = g^{jj} \eta^j. \quad (1.8.136)$$

Show that combining this information with (3.157), (3.160), and (3.161) yields the result

$$h(y) = y'' = \sum_j (g^{jj})^2 \eta^j c^j = \sum_j \eta^j \Lambda e^j = \Lambda y. \quad (1.8.137)$$

Finally, verify that going back to (3.148) gives the advertised result

$$f(y) = h(y) + d = \Lambda y + d. \quad (1.8.138)$$

1.8.5 Vector and Tensor Transformation Properties

The purpose of this subsection is to study vector and tensor transformation properties. Recall that under a Lorentz transformation the contravariant components of a four vector transform according to the rule (6.310), which we repeat below:

$$\bar{x}^\alpha = \sum_\mu \Lambda^{\alpha\mu} x^\mu. \quad (1.8.139)$$

See Exercise 6.18. The covariant components of the same four vector are given by the relation

$$x_\mu = \sum_\nu g_{\mu\nu} x^\nu. \quad (1.8.140)$$

See (6.49). Your first task is to determine how the covariant components transform under the same Lorentz transformation.

Begin with some preparatory steps. Observe that (6.348) implies the differential relations

$$d\bar{x}^\alpha = \sum_\mu \Lambda^{\alpha\mu} dx^\mu. \quad (1.8.141)$$

If we view the \bar{x}^α as functions of the x^μ there are also the differential relations

$$d\bar{x}^\alpha = \sum_\mu (\partial \bar{x}^\alpha / \partial x^\mu) dx^\mu. \quad (1.8.142)$$

Upon comparison of (6.350) and (6.351) we see that there are the partial differential relations

$$\partial \bar{x}^\alpha / \partial x^\mu = \Lambda^{\alpha\mu}. \quad (1.8.143)$$

Consequently (6.348) can also be written in the form

$$\bar{x}^\alpha = \sum_\mu (\partial \bar{x}^\alpha / \partial x^\mu) x^\mu. \quad (1.8.144)$$

Next show that (6.350) can be inverted. Multiply both sides of (6.350) by $(\Lambda^{-1})^{\nu\alpha}$ and sum over α to find the result

$$\begin{aligned} \sum_\alpha (\Lambda^{-1})^{\nu\alpha} d\bar{x}^\alpha &= \sum_\alpha (\Lambda^{-1})^{\nu\alpha} \sum_\mu \Lambda^{\alpha\mu} dx^\mu = \sum_\mu \sum_\alpha (\Lambda^{-1})^{\nu\alpha} \Lambda^{\alpha\mu} dx^\mu \\ &= \sum_\mu (\Lambda^{-1}\Lambda)^{\nu\mu} dx^\mu = \sum_\mu (I)^{\nu\mu} dx^\mu = dx^\nu. \end{aligned} \quad (1.8.145)$$

(That Λ^{-1} exists follows from Exercise 7.3.27.) If we view the x^ν as functions of the \bar{x}^α there are also the differential relations

$$dx^\nu = \sum_\alpha (\partial x^\nu / \partial \bar{x}^\alpha) d\bar{x}^\alpha. \quad (1.8.146)$$

Upon comparison of (6.354) and (6.355) we see that there are the partial differential relations

$$\partial x^\nu / \partial \bar{x}^\alpha = (\Lambda^{-1})^{\nu\alpha}. \quad (1.8.147)$$

We are now ready to proceed with the first task. From (6.48) through (6.50) there are the relations

$$\bar{x}_\alpha = \sum_\beta g_{\alpha\beta} \bar{x}^\beta = \sum_\beta g^{\alpha\beta} \bar{x}^\beta, \quad (1.8.148)$$

$$x^\mu = \sum_\nu g^{\mu\nu} x_\nu. \quad (1.8.149)$$

Use these relations and a relation of the form (6.348) to show that

$$\bar{x}_\alpha = \sum_\beta g^{\alpha\beta} \sum_\mu \Lambda^{\beta\mu} \sum_\nu g^{\mu\nu} x_\nu, \quad (1.8.150)$$

which can be rewritten in the matrix form

$$\bar{x}_\alpha = \sum_\beta g^{\alpha\beta} \sum_\mu \Lambda^{\beta\mu} \sum_\nu g^{\mu\nu} x_\nu = \sum_\nu (g\Lambda g)^{\alpha\nu} x_\nu. \quad (1.8.151)$$

Next verify that the Minkowski metric g as given by (6.45) or (6.48), when viewed as a matrix, satisfies the relation

$$g^2 = I \quad (1.8.152)$$

where I is the 4×4 identity matrix. See (7.6). Show from this result and (6.312) that there is the relation

$$g\Lambda g = (\Lambda^T)^{-1} \quad (1.8.153)$$

so that (6.360) can also be rewritten in the form

$$\bar{x}_\alpha = \sum_\nu K^{\alpha\nu} x_\nu \quad (1.8.154)$$

where

$$K = (\Lambda^T)^{-1} = (\Lambda^{-1})^T. \quad (1.8.155)$$

Show for comparison that (6.348), upon a suitable relabeling of indices, takes the form

$$\bar{x}^\alpha = \sum_\nu \Lambda^{\alpha\nu} x^\nu. \quad (1.8.156)$$

Evidently (6.363) and (6.365) take the same form with K given in terms of Λ by the relation (6.364). At this point we remark that it can be shown that K is a Lorentz transformation matrix if Λ is, and vice versa. See Exercise 6.2.6. For a further discussion of the relation (6.364), see Exercise 3.7.37 and (3.7.241).

You have completed the first task. You have shown that the covariant components of a four vector transform according to the rule given by (6.363) and (6.364). But now

somewhat more can be said. Show that (6.363), (6.364), and (6.356) can be combined to give the relation

$$\bar{x}_\alpha = \sum_\nu K^{\alpha\nu} x_\nu = \sum_\nu (K^T)^{\nu\alpha} x_\nu = \sum_\nu (\Lambda^{-1})^{\nu\alpha} x_\nu = \sum_\nu (\partial x^\nu / \partial \bar{x}^\alpha) x_\nu. \quad (1.8.157)$$

You have shown that the covariant components of a four vector transform according to the rule

$$\bar{x}_\alpha = \sum_\nu (\partial x^\nu / \partial \bar{x}^\alpha) x_\nu. \quad (1.8.158)$$

Show for comparison that (6.353), upon a suitable relabeling of indices, takes the form

$$\bar{x}^\alpha = \sum_\nu (\partial \bar{x}^\alpha / \partial x^\nu) x^\nu. \quad (1.8.159)$$

Evidently, (6.367) and (6.368) are related by the symbol interchange $\partial x^\nu \leftrightarrow \partial \bar{x}^\alpha$. Note also that comparison of (6.362) and (6.367) gives the relation

$$\partial x^\nu / \partial \bar{x}^\alpha = K^{\alpha\nu}. \quad (1.8.160)$$

It should be compared with (6.352), which we rewrite in the form

$$\partial \bar{x}^\alpha / \partial x^\nu = \Lambda^{\alpha\nu}. \quad (1.8.161)$$

Note that again there is the symbol interchange $\partial x^\nu \leftrightarrow \partial \bar{x}^\alpha$.

Earlier in our discussion we observed that the metric tensor g is symmetric, $g_{\mu\nu} = g_{\nu\mu}$. See (7.6). And since for any invertible matrix G there is the result $(G^T)^{-1} = (G^{-1})^T$, it follows from (6.48) that $g^{\mu\nu} = g^{\nu\mu}$. We will need these results for what follows.

Your second task is to apply what you have learned about the transformation properties of four vectors to the case of general tensors. To begin, suppose there are quantities B^μ which obey the four-vector contravariant transformation rule

$$\bar{B}^\alpha = \sum_\mu \Lambda^{\alpha\mu} B^\mu = \sum_\mu (\partial \bar{x}^\alpha / \partial x^\mu) B^\mu, \quad (1.8.162)$$

and suppose there are quantities C_μ which obey the four-vector covariant transformation rule

$$\bar{C}_\alpha = \sum_\mu K^{\alpha\mu} C_\mu = \sum_\mu (\partial x^\mu / \partial \bar{x}^\alpha) C_\mu. \quad (1.8.163)$$

Using (6.48) through (6.51) and the fact the metric tensor is symmetric, verify immediately the results

$$\sum_{\alpha\beta} g_{\alpha\beta} B^\alpha C^\beta = \sum_{\alpha\beta} g^{\alpha\beta} B_\alpha C_\beta = \sum_\alpha B^\alpha C_\alpha = \sum_\alpha B_\alpha C^\alpha, \quad (1.8.164)$$

that these results are independent of the transformation rules, and that analogous results hold for the \bar{B} and \bar{C} components. Next verify that there is the more subtle result

$$\begin{aligned} \sum_\alpha \bar{B}^\alpha \bar{C}_\alpha &= \sum_{\alpha\mu\nu} \Lambda^{\alpha\mu} K^{\alpha\nu} B^\mu C_\nu = \sum_{\alpha\mu\nu} (\Lambda^T)^{\mu\alpha} K^{\alpha\nu} B^\mu C_\nu = \\ &= \sum_{\mu\nu} (\Lambda^T K)^{\mu\nu} B^\mu C_\nu = \sum_{\mu\nu} I^{\mu\nu} B^\mu C_\nu = \sum_\nu B^\nu C_\nu. \end{aligned} \quad (1.8.165)$$

[Here we have used the dummy index principle to replace (6.372) by the equivalent statement $\bar{C}_\alpha = \sum_\nu K^{\alpha\nu} C_\nu = \sum_\nu (\partial x^\nu / \partial \bar{x}^\alpha) C_\nu$.] That is, the quantity $\sum_\alpha \bar{B}^\alpha \bar{C}_\alpha$ is *invariant* (has the value $\sum_\nu B^\nu C_\nu$) no matter what the Lorentz transformation Λ may be. Indeed, Λ need not even be a Lorentz transformation. Evidently the invariance relation (6.374) holds for any (but nonsingular) matrix Λ in any number of dimensions. Finally, the invariance relation (6.374) holds for *all* (invertible) *maps* \mathcal{M} , including possibly *nonlinear* maps, that send quantities x^ν to quantities \bar{x}^α because the \bar{B}^α and \bar{C}_α can also be defined in terms of the B^μ and C_μ using only partial derivatives of \mathcal{M} and its inverse. See the far right sides of (6.371) and (6.372). Therefore, although our discussion began in the context of Special Relativity, the results we have found may also be applicable in other contexts.

The invariance principle that the four-vector contravariant and covariant transformation properties compensate each other so that (6.374) holds can be extended to general tensors. For example, let $T_{\bullet\bullet\sigma\bullet}^{\mu\nu\bullet\tau}$ be a quantity that depends on the contravariant indices $\mu\nu\tau$ and the covariant index σ . Here, to keep track of index positions, we have placed \bullet symbols below contravariant indices and above covariant indices to indicate empty spaces where these indices would go if they were lowered or raised, respectively. The quantities $T_{\bullet\bullet\sigma\bullet}^{\mu\nu\bullet\tau}$ are said to comprise a (mixed rank 4) tensor if they transform according to the rule

$$\bar{T}_{\bullet\bullet\gamma\bullet}^{\alpha\beta\bullet\delta} = \sum_{\mu\nu\sigma\tau} \Lambda^{\alpha\mu} \Lambda^{\beta\nu} K^{\gamma\sigma} \Lambda^{\delta\tau} T_{\bullet\bullet\sigma\bullet}^{\mu\nu\bullet\tau}. \quad (1.8.166)$$

That is, Λ matrices are used for contravariant indices and K matrices are used for covariant indices.⁶⁵ Now pick a pair of indices associated with T , one being contravariant and one being covariant. For example, the pair could be the first contravariant index (which in this example would be μ or α) and the only covariant index (which in this example would be the third index and therefore would be σ or γ). Form the rank 2 objects $S_{\bullet\bullet}^{\nu\tau}$ and $\bar{S}_{\bullet\bullet}^{\beta\delta}$ by the rules

$$S_{\bullet\bullet}^{\nu\tau} = \sum_{\theta} T_{\bullet\bullet\theta\bullet}^{\theta\nu\bullet\tau}, \quad (1.8.167)$$

$$\bar{S}_{\bullet\bullet}^{\beta\delta} = \sum_{\theta} \bar{T}_{\bullet\bullet\theta\bullet}^{\theta\beta\bullet\delta}. \quad (1.8.168)$$

We characterize these tensors as being of *rank 2* because they only have two free indices (indices that have not been summed over). Verify that employing (6.375) in (6.377) and

⁶⁵Some authors write relations such as (6.365) in the form $\bar{x}^\alpha = \sum_\nu \Lambda_{\bullet\nu}^{\alpha\bullet} x^\nu$ and would also use both contravariant and covariant indices on the Λ and K appearing in expressions such as (6.375). Although doing so appears to neatly marry indices, we do not think such notation is a good idea because it makes Λ and K look like tensors, which they are not. They are transformation coefficients.

then using (6.376) yields the result

$$\begin{aligned}
\bar{S}_{\bullet\bullet}^{\beta\delta} &= \sum_{\theta} \bar{T}_{\bullet\bullet\theta\bullet}^{\theta\beta\bullet\delta} = \sum_{\theta\mu\nu\sigma\tau} \Lambda^{\theta\mu} \Lambda^{\beta\nu} K^{\theta\sigma} \Lambda^{\delta\tau} T_{\bullet\bullet\sigma\bullet}^{\mu\nu\bullet\tau} \\
&= \sum_{\theta\mu\nu\sigma\tau} (\Lambda^T)^{\mu\theta} \Lambda^{\beta\nu} K^{\theta\sigma} \Lambda^{\delta\tau} T_{\bullet\bullet\sigma\bullet}^{\mu\nu\bullet\tau} = \sum_{\theta\mu\nu\sigma\tau} \Lambda^{\beta\nu} (\Lambda^T)^{\mu\theta} K^{\theta\sigma} \Lambda^{\delta\tau} T_{\bullet\bullet\sigma\bullet}^{\mu\nu\bullet\tau} \\
&= \sum_{\mu\nu\sigma\tau} \Lambda^{\beta\nu} (\Lambda^T K)^{\mu\sigma} \Lambda^{\delta\tau} T_{\bullet\bullet\sigma\bullet}^{\mu\nu\bullet\tau} = \sum_{\mu\nu\sigma\tau} \Lambda^{\beta\nu} (I)^{\mu\sigma} \Lambda^{\delta\tau} T_{\bullet\bullet\sigma\bullet}^{\mu\nu\bullet\tau} \\
&= \sum_{\mu\nu\tau} \Lambda^{\beta\nu} \Lambda^{\delta\tau} T_{\bullet\bullet\mu\bullet}^{\mu\nu\bullet\tau} = \sum_{\nu\tau} \Lambda^{\beta\nu} \Lambda^{\delta\tau} \sum_{\mu} T_{\bullet\bullet\mu\bullet}^{\mu\nu\bullet\tau} \\
&= \sum_{\nu\tau} \Lambda^{\beta\nu} \Lambda^{\delta\tau} S_{\bullet\bullet}^{\nu\tau}.
\end{aligned} \tag{1.8.169}$$

You have shown that the quantities $S_{\bullet\bullet}^{\nu\tau}$ comprise a second rank contravariant tensor. The process you have executed is called *contraction*. Evidently contraction can be carried out as often as desired or possible, thereby yielding tensors of successively lower ranks by decrements of 2, until only free contravariant or free covariant indices remain (depending on which were more abundant initially) or no free indices remain if contravariant and covariant indices were equally abundant initially. Also, if there are multiple ways of choosing contravariant and covariant pairs (as there are in this example), the net result generally depends on the choice(s) of pairs. Finally, to put our findings another way, we may say that the operations of tensor transformation and tensor contraction *commute*. That is, we may first contract one or some index pairs and then transform using the remaining indices, or we may first transform using all indices and then contract. Both operation orders yield the same result.

We have seen that contravariant and covariant components of vectors and tensors are characterized by their transformation properties. Our last task in this exercise is to apply this concept to the relations (6.55). Consider the differential operators $\partial/\partial\bar{x}^\alpha$ and $\partial/\partial x^\beta$. According to the chain rule they are related by the equation

$$\partial/\partial\bar{x}^\alpha = \sum_{\beta} (\partial x^\beta / \partial\bar{x}^\alpha) (\partial/\partial x^\beta). \tag{1.8.170}$$

As in (6.55), make the definitions and index assignments/placements

$$\bar{\partial}_\alpha = \partial/\partial\bar{x}^\alpha, \tag{1.8.171}$$

$$\partial_\beta = \partial/\partial x^\beta. \tag{1.8.172}$$

Also verify that, by relabeling indices, (6.369) can be rewritten in the form

$$K^{\alpha\beta} = \partial x^\beta / \partial\bar{x}^\alpha. \tag{1.8.173}$$

Finally, using (6.380), (6.381), and (6.379), verify that (6.379) can be rewritten in the form

$$\bar{\partial}_\alpha = \sum_{\beta} K^{\alpha\beta} \partial_\beta. \tag{1.8.174}$$

Observe, by comparing (6.363) and (6.383), that (6.383) is the expected transformation rule for *covariant* components, and therefore the index placements in (6.380) and (6.381) are correct.

1.9 Definition of Poisson Bracket

Life is good for only two things: to discover/do mathematics and to teach mathematics.

Siméon-Denis Poisson

In subsequent chapters we will learn that Hamiltonian dynamics can be placed in a Lie-algebraic context. Key to this placement is the *Poisson bracket*.⁶⁶ In this section we will review its definition and some of its properties.

Let $H(q, p, t)$ be the Hamiltonian for some dynamical system and let f be any *dynamical variable*. That is, let $f(q, p, t)$ be any function of the phase-space variables q, p and the time t . Consider the problem of computing the *total* time rate of change of f along a trajectory generated by H . According to the chain rule, this derivative is given by the expression

$$df/dt = \partial f/\partial t + \sum_i \{(\partial f/\partial q_i)\dot{q}_i + (\partial f/\partial p_i)\dot{p}_i\}. \quad (1.9.1)$$

However, the \dot{q} 's and \dot{p} 's are given by Hamilton's equations of motion (5.11). Consequently, the expression for df/dt can also be written in the form

$$df/dt = \partial f/\partial t + \sum_i \{(\partial f/\partial q_i)(\partial H/\partial p_i) - (\partial f/\partial p_i)(\partial H/\partial q_i)\}. \quad (1.9.2)$$

The second quantity appearing on the right side of (7.2) occurs so often that it is given a special symbol and a special name in honor of Poisson. Let f and g be any two functions of the variables q, p, t . Then the Poisson bracket of f and g , denoted by the symbol $[f, g]$, is another function defined by the equation

$$[f, g] = \sum_i \{(\partial f/\partial q_i)(\partial g/\partial p_i) - (\partial f/\partial p_i)(\partial g/\partial q_i)\}. \quad (1.9.3)$$

With this new notation, (7.2) can be written in the compact form

$$df/dt = \partial f/\partial t + [f, H]. \quad (1.9.4)$$

The Poisson bracket operation has three important and obvious properties that are easily checked from its definition:

1. Distributive property,

$$[(af + bg), h] = a[f, h] + b[g, h] \quad (1.9.5)$$

for arbitrary constants a, b .

2. Antisymmetry condition,

$$[f, g] = -[g, f]. \quad (1.9.6)$$

⁶⁶Poisson (1781-1840) was a student of Lagrange and Laplace and, at age 25, succeeded Fourier as a professor at the École Polytechnique. Poisson is the French word for fish. Sometimes, now mostly for fun and perhaps originally due to an error in translation, Poisson brackets are referred to as *fishermen's brackets*. There actually are such things, and are available in marine supply stores.

3. Derivation with respect to multiplication,

$$[f, gh] = [f, g]h + g[f, h]. \quad (1.9.7)$$

(For those unfamiliar with the term, a *derivation* is an operation that behaves like “differentiation” in the sense that it obeys a product rule analogous to the product rule for differentiating a product in ordinary calculus.) From the definition (7.3) one also easily finds the so-called *fundamental* Poisson brackets,

$$\begin{aligned} [q_i, q_j] &= 0, \\ [p_i, p_j] &= 0, \\ [q_i, p_j] &= \delta_{ij}. \end{aligned} \quad (1.9.8)$$

There is a fourth important property of the Poisson bracket called the *Jacobi* identity or condition. It is less obvious, and will be discussed in Section 5.1.

At this point it is convenient to introduce a more compact notation for the phase-space variables $(q_1 \cdots q_n), (p_1 \cdots p_n)$. To do this, introduce the $2n$ variables (z_1, \dots, z_{2n}) by the rule

$$z = (z_1, \dots, z_n; z_{n+1}, \dots, z_{2n}) = (q_1, \dots, q_n; p_1, \dots, p_n). \quad (1.9.9)$$

That is, the first n z 's are the q 's and the last n z 's are the p 's. We will also adopt the convention of using lower case latin letters near the beginning of the alphabet to denote indices that range from 1 to $2n$.

With the definition (7.9), it is easily verified that the fundamental Poisson brackets (7.8) can also be written in the form

$$[z_a, z_b] = J_{ab}. \quad (1.9.10)$$

Here J is a $2n \times 2n$ matrix defined in block form by the equation

$$J = \begin{pmatrix} \mathbf{0} & I \\ -I & \mathbf{0} \end{pmatrix}, \quad (1.9.11)$$

where each entry in J is an $n \times n$ matrix, I denotes the $n \times n$ identity matrix, and all other entries are zero. The matrix J is sometimes called the *Poisson* matrix.

Suppose functions f and g of the variables q, p, t are written more compactly as $f(z, t)$, $g(z, t)$. Then the general Poisson bracket (7.3) can be written more compactly in the form

$$[f, g] = \sum_{a,b} (\partial f / \partial z_a) J_{ab} (\partial g / \partial z_b). \quad (1.9.12)$$

Suppose further that the $2n$ quantities $(\partial f / \partial z_a)$ are viewed as the components of a vector conveniently written as $\partial_z f$, and similarly for the quantities $(\partial g / \partial z_b)$. Then the right side of (7.12) can be viewed as a combination of two vectors and a matrix that can be written even more compactly using matrix and scalar product notation,

$$[f, g] = (\partial_z f, J \partial_z g). \quad (1.9.13)$$

Exercises

1.9.1. Verify the relations (7.5) through (7.7).

1.9.2. Derive (5.14) using (7.4) and (7.6). Show that if H does not explicitly depend on time, then it is a constant of motion and an integral of motion. See Section 5.2 for the definitions of constants and integrals of motion.

1.9.3. Verify (7.8) and (7.10).

1.9.4. Verify (7.12) and (7.13).

1.9.5. Review Exercise 6.7. Recall the Lorentz invariant Hamiltonian H_R given by (6.77) and the associated equations of motion (6.80) through (6.82). The purpose of this exercise is to study Poisson brackets in the context of a manifestly Lorentz invariant Hamiltonian formulation of the equations of motion.

- a) Using the x^μ and p_μ as phase-space variables, suppose $f(x^\mu, p_\mu, \tau)$ is any dynamical variable. Repeat the steps (7.1) through (7.4) to show that in this case the Poisson bracket should be defined by the rule

$$[f, g] = \sum_{\mu} [(\partial f / \partial x^\mu)(\partial g / \partial p_\mu) - (\partial f / \partial p_\mu)(\partial g / \partial x^\mu)]. \quad (1.9.14)$$

As a consequence of this rule, show that

$$[x^\mu, x^\nu] = 0, \quad [p_\mu, p_\nu] = 0, \quad [x^\mu, p_\nu] = \delta_\nu^\mu \quad (1.9.15)$$

where δ_ν^μ is defined, as expected, by the equations

$$\begin{aligned} \delta_\nu^\mu &= 0 \text{ for } \mu \neq \nu, \\ &= 1 \text{ for } \mu = \nu. \end{aligned} \quad (1.9.16)$$

Thus, the x^μ and p_ν are canonically conjugate variables. Next, show that

$$[x^\mu, p^\nu] = g^{\mu\nu}. \quad (1.9.17)$$

- b) Also show, based on (6.54), (6.59), (6.61), and (7.14), that in the presence of an electromagnetic field there are the Poisson bracket relations

$$[p_\mu^{\text{mech}}, p_\nu^{\text{mech}}] = qF_{\mu\nu} \quad \text{and} \quad [(p^{\text{mech}})^\mu, (p^{\text{mech}})^\nu] = qF^{\mu\nu}. \quad (1.9.18)$$

As a special case of (7.18), show that there are the relations

$$[p_x^{\text{mech}}, p_y^{\text{mech}}] = [(p^{\text{mech}})^1, (p^{\text{mech}})^2] = qF^{12} = -qB_z, \text{ etc.} \quad (1.9.19)$$

We see that the *mechanical* momenta are *not* canonical variables because, unlike the corresponding relation in (7.15), the right sides in (7.18) are nonzero. It follows that the equations of motion (6.67) and (6.69), although first order, are *not* canonical

because they involve mechanical momenta. That is, these equations of motion do not arise from any Hamiltonian. Similarly, show that converting the second-order set of equations (6.68) or (6.95) into an associated first-order set using the method of Section 1.3 yields noncanonical equations. Thus these equations of motion are not particularly useful if one wishes to exploit the symplectic (canonical) symmetry associated with Hamiltonian systems. See Exercise 6.4.11.

- c) From (6.42) there is the relation

$$t = x^4/c. \quad (1.9.20)$$

Define p_t by the rule

$$p_t = -cp^4 = -cp_4. \quad (1.9.21)$$

Then, from (7.15) with $\mu = \nu = 4$, show that there is the result

$$[t, p_t] = [x^4/c, -cp_4] = [x^4, -p_4] = -1. \quad (1.9.22)$$

Also, again from (7.15), show that there are the results

$$[x^\mu, p^\nu] = -[x^\mu, p_\nu] = -\delta^{\mu\nu} \text{ for } \mu, \nu = 1 \cdots 3 \quad (1.9.23)$$

where, as again expected, $\delta^{\mu\nu}$ is defined for $\mu, \nu = 1 \cdots 3$ by the equations

$$\begin{aligned} \delta^{\mu\nu} &= 0 \text{ for } \mu \neq \nu, \\ &= 1 \text{ for } \mu = \nu. \end{aligned} \quad (1.9.24)$$

Evidently the quantities x^μ, p^ν for $\mu, \nu = 1 \cdots 3$ and t, p_t behave like canonical variables save for an annoying/alarming minus sign. We would, in fact, like to use the variables x^μ, p^ν for $\mu, \nu = 1 \cdots 3$ because then all indices are up so that we do not have to distinguish between up and down indices, and can eventually even forget about their position. Also, up index quantities are directly related to variables of interest without any additional minus signs. Contrast, for example, (6.42) and (6.52).

- d) What to do? Suppose we define a new Hamiltonian \hat{H}_R by the rule

$$\hat{H}_R = H_R/(-1) = -H_R. \quad (1.9.25)$$

With this definition in mind, check that the equations of motion (6.80) and (6.81) yield for the variables x^μ, p^ν for $\mu, \nu = 1 \cdots 3$ and t, p_t the results

$$(x')^\mu = \partial H_R / \partial p_\mu = -\partial H_R / \partial p^\mu = \partial \hat{H}_R / \partial p^\mu \text{ for } \mu = 1 \cdots 3, \quad (1.9.26)$$

$$t' = (1/c)(x')^4 = (1/c)\partial H_R / \partial p_4 = (1/c)\partial H_R / \partial p^4 = -\partial H_R / \partial p_t = \partial \hat{H}_R / \partial p_t; \quad (1.9.27)$$

$$(p')^\mu = -(p')_\mu = \partial H_R / \partial x^\mu = -\partial \hat{H}_R / \partial x^\mu \text{ for } \mu = 1 \cdots 3, \quad (1.9.28)$$

$$(p_t)' = -c(p')_4 = c\partial H_R / \partial x_4 = \partial H_R / \partial t = -\partial \hat{H}_R / \partial t. \quad (1.9.29)$$

Upon examining the far left and far right sides of (7.26) through (7.29) verify that, if we agree to use the Hamiltonian \hat{H}_R instead of H_R , then we may *redefine* the

fundamental Poisson brackets for the variables x^μ, p^ν (with $\mu, \nu = 1 \cdots 3$) and t, p_t to be the standard ones:

$$[x^\mu, t] = 0 \text{ for } \mu = 1 \cdots 3, \quad (1.9.30)$$

$$[x^\mu, x^\nu] = 0 \text{ for } \mu, \nu = 1 \cdots 3; \quad (1.9.31)$$

$$[p^\mu, p_t] = 0 \text{ for } \mu = 1 \cdots 3, \quad (1.9.32)$$

$$[p^\mu, p^\nu] = 0 \text{ for } \mu, \nu = 1 \cdots 3; \quad (1.9.33)$$

$$[t, p^\mu] = 0 \text{ for } \mu = 1 \cdots 3, \quad (1.9.34)$$

$$[x^\mu, p_t] = 0 \text{ for } \mu = 1 \cdots 3, \quad (1.9.35)$$

$$[t, p_t] = 1, \quad (1.9.36)$$

$$[x^\mu, p^\nu] = \delta^{\mu\nu} \text{ for } \mu, \nu = 1 \cdots 3. \quad (1.9.37)$$

Note that the relation between H and K given by (6.126) contains a minus sign just like the relation (7.25) between \hat{H}_R and H_R . We also observe that the replacement of -1 by 1 in the Poisson bracket rules and the replacement of H_R by $\hat{H}_R = -H_R$ in the equations of motion is a special case of what we may call a *scaling* transformation. See Subsection 13.1.5.

- e) Suppose there is no electromagnetic field so that all the components A^μ vanish. For the identifications (6.42), (6.101), and (6.102) show that (7.36) and (7.37) imply the relations

$$[x, p_x] = [y, p_y] = [z, p_z] = [t, -\mathcal{E}] = 1. \quad (1.9.38)$$

Observe that these relations are consistent with (7.8) and (6.105).

1.9.6. Suppose we employ the Hamiltonian H given by (5.49). Note that in this case there is no mention of up and down index quantities. There are simply the components of the vectors \mathbf{r} and \mathbf{p}^{can} . That is, there are the dynamical variables $(x, y, z; p_x^{\text{can}}, p_y^{\text{can}}, p_z^{\text{can}})$, and the time t is treated as the independent variable. For this Hamiltonian follow the recipe of Section 1.7 to define Poisson brackets. Show that doing so yields the result that all Poisson brackets involving only components of \mathbf{r} and \mathbf{p}^{can} vanish save for

$$[x, p_x^{\text{can}}] = [y, p_y^{\text{can}}] = [z, p_z^{\text{can}}] = 1. \quad (1.9.39)$$

Show that for this definition of the Poisson bracket there are the results

$$[p_x^{\text{mech}}, p_y^{\text{mech}}] = qB_z, \text{ etc.} \quad (1.9.40)$$

Note that (7.19) and (7.40) differ by a sign. This difference occurs because the definition of the Poisson bracket depends on what Hamiltonian is being employed.

1.9.7. Suppose we employ the Hamiltonian K given by (6.16). In this case the dynamical variables are $(x, y, t; p_x, p_y, p_t)$ and z is the independent variable. Note that, although not indicated by our imprecise notation, the quantities (p_x, p_y, p_t) are canonical and not mechanical momenta. For this Hamiltonian follow the recipe of Section 1.7 to define Poisson

brackets. Show that doing so yields the result that all poisson brackets among the dynamical variables $(x, y, t; p_x, p_y, p_t)$ vanish save for

$$[x, p_x] = [y, p_y] = [t, p_t] = 1. \quad (1.9.41)$$

Show that for this definition of the Poisson bracket there is the result

$$[p_x^{\text{mech}}, p_y^{\text{mech}}] = qB_z. \quad (1.9.42)$$

Note that (7.19) and (7.42) differ by a sign. This difference occurs because the definition of the Poisson bracket depends on what Hamiltonian is being employed

1.9.8. Suppose that

$$\psi = 0 \quad (1.9.43)$$

in (5.1), (5.49), and (6.16) so that K takes the form

$$K = -[(p_t/c)^2 - m^2c^2 - (p_x - qA_x)^2 - (p_y - qA_y)^2]^{1/2} - qA_z. \quad (1.9.44)$$

Show from (5.27) through (5.30), (6.5), and (7.43) that there are the relations

$$[(\mathbf{p}^{\text{mech}})^2]^{1/2} = p^{\text{mech}} = \gamma mv = \gamma\beta mc, \quad (1.9.45)$$

$$p_t = -[m^2c^4 + (\mathbf{p}^{\text{mech}}c)^2]^{1/2} = -\gamma mc^2, \quad (1.9.46)$$

$$p_t^2 = m^2c^4 + (\mathbf{p}^{\text{mech}}c)^2, \quad (1.9.47)$$

$$v = c[1 - (mc^2/p_t)^2]^{1/2}. \quad (1.9.48)$$

Here β and γ are the usual relativistic factors,

$$\beta = v/c, \quad (1.9.49)$$

$$\gamma = (1 - \beta^2)^{-1/2}. \quad (1.9.50)$$

The quantity p_t obeys the equation of motion

$$dp_t/dz = -\partial K/\partial t. \quad (1.9.51)$$

See (6.10). Therefore if \mathbf{A} is time independent (which amounts to the case of motion in a static magnetic field), there are the relations

$$\partial K/\partial t = 0, \quad (1.9.52)$$

$$p_t = \text{constant}. \quad (1.9.53)$$

From (7.46) and (7.48) through (7.50) show that in this case β , γ , and v are also constants of motion.

In Accelerator Physics, when studying orbits in a magnetic field, it is common to introduce the quantity δ by the definition

$$p^{\text{mech}} = (1 + \delta)p_0^{\text{mech}} \quad (1.9.54)$$

where $p_0^{\text{mech}} = \|\mathbf{p}_0^{\text{mech}}\|$ is the magnitude of some reference or *design* mechanical momentum and $p^{\text{mech}} = \|\mathbf{p}^{\text{mech}}\|$ is the magnitude of the actual mechanical momentum. The quantity δ is called the *momentum deviation*. By combining (7.47) and (7.54) show that there are the relations

$$p_t^2 = m^2 c^4 + (1 + \delta)^2 (p_0^{\text{mech}} c)^2, \quad (1.9.55)$$

$$\delta = [(p_t^2 - m^2 c^4)^{1/2} / (p_0^{\text{mech}} c)] - 1. \quad (1.9.56)$$

Consider the quantity ℓ defined by

$$\ell = (p_0^{\text{mech}} c) [1 - (mc^2/p_t)^2]^{1/2} t. \quad (1.9.57)$$

Show from (7.48) that ℓ can also be written in the form

$$\ell = (p_0^{\text{mech}}) v t. \quad (1.9.58)$$

Evidently, if v is constant (which will be the case for motion in a static magnetic field), the quantity ℓ is proportional to *path length* with proportionality constant p_0^{mech} . Note that the quantity ℓ is still defined by (7.57) in the time-dependent case, but then it has no such simple physical interpretation. Show, however, that in the extreme relativistic limit $-p_t \gg mc^2$ where $v \simeq c$ there is the relation

$$\ell \simeq (p_0^{\text{mech}}) c t \quad (1.9.59)$$

so that in this limit the interpretation of ℓ as being proportional to path length is regained even in the time-dependent case.

Show, starting from the known Poisson bracket relation

$$[t, p_t] = 1, \quad (1.9.60)$$

that there is the relation

$$[\delta, \ell] = 1. \quad (1.9.61)$$

Also show that there are the relations

$$[x, \delta] = [y, \delta] = [p_x, \delta] = [p_y, \delta] = 0,$$

$$[x, \ell] = [y, \ell] = [p_x, \ell] = [p_y, \ell] = 0. \quad (1.9.62)$$

Thus, δ and ℓ are *canonically conjugate* with δ being “coordinate like” and ℓ being “momentum like”. See (7.8). We may therefore view the quantities $x, p_x; y, p_y; \delta, \ell$ as a set of canonical coordinates obtained from the set $x, p_x; y, p_y; t, p_t$ by a canonical transformation. (Recall that a canonical transformation is a transformation that preserves the fundamental Poisson brackets. See Section 6.1.2.)

Show that there are the inverse relations

$$p_t = -[m^2 c^4 + (1 + \delta)^2 (p_0^{\text{mech}} c)^2]^{1/2}, \quad (1.9.63)$$

$$t = [\ell / (p_0^{\text{mech}} c)] \{1 - m^2 c^4 / [m^2 c^4 + (1 + \delta)^2 (p_0^{\text{mech}} c)^2]\}^{-1/2}. \quad (1.9.64)$$

If a canonical transformation does not depend explicitly on the independent variable (the quantity z in the case), then the new Hamiltonian \bar{K} equals the old Hamiltonian K expressed in terms of the new variables,

$$\bar{K}\{x, p_x, y, p_y, \delta, \ell; z\} = K\{x, p_x, y, p_y, t(\delta, \ell), p_t(\delta, \ell); z\}. \quad (1.9.65)$$

(See Appendix D.) Show, using (7.55) and (7.65), that

$$\bar{K} = -[(1 + \delta)^2(p_0^{\text{mech}})^2 - (p_x - q\bar{A}_x)^2 - (p_y - q\bar{A}_y)^2]^{1/2} - q\bar{A}_z \quad (1.9.66)$$

where

$$\bar{\mathbf{A}}\{\mathbf{r}, \delta, \ell\} = \mathbf{A}\{\mathbf{r}, t(\delta, \ell)\}. \quad (1.9.67)$$

If all is well, there should be the relation

$$d\ell/dz = [\ell, \bar{K}] = -\partial\bar{K}/\partial\delta. \quad (1.9.68)$$

See (7.4). Show, from (6.10) and (7.65), that the right side of (7.68) is given by the relation

$$\partial\bar{K}/\partial\delta = (\partial K/\partial t)(\partial t/\partial\delta) + (\partial K/\partial p_t)(\partial p_t/\partial\delta) = -(dp_t/dz)(\partial t/\partial\delta) + (dt/dz)(\partial p_t/\partial\delta). \quad (1.9.69)$$

Evaluate the partial derivatives on the right side of (7.69) using (7.63) and (7.64) to find the results

$$(\partial t/\partial\delta) = -(p_0^{\text{mech}})t/(mc\beta\gamma^3), \quad (1.9.70)$$

$$(\partial p_t/\partial\delta) = -(p_0^{\text{mech}})v. \quad (1.9.71)$$

Evaluate the left side of (7.68) using (7.57), and verify that (7.68) is correct. Similarly, verify that

$$d\delta/dz = [\delta, \bar{K}]. \quad (1.9.72)$$

Sometimes it is convenient to introduce scaled variables P_x , P_y , and $\hat{\ell}$ by the rules

$$P_x = p_x/p_0^{\text{mech}}, \quad (1.9.73)$$

$$P_y = p_y/p_0^{\text{mech}}, \quad (1.9.74)$$

$$\hat{\ell} = \ell/p_0^{\text{mech}} = c[1 - (mc^2/p_t)^2]^{1/2}t = vt. \quad (1.9.75)$$

See Section 13.1.5. Note that P_x and P_y are dimensionless. Also now, when v is constant, $\hat{\ell}$ is the path length. If we now regard the pairs x, P_x ; y, P_y ; and $\delta, \hat{\ell}$ as canonically conjugate, their evolution will be governed by the Hamiltonian \hat{K} given by

$$\hat{K} = (1/p_0^{\text{mech}})\bar{K} = -[(1 + \delta)^2 - (P_x - q\hat{A}_x)^2 - (P_y - q\hat{A}_y)^2]^{1/2} - q\hat{A}_z \quad (1.9.76)$$

where

$$\hat{\mathbf{A}}\{\mathbf{r}, \delta, \hat{\ell}\} = (1/p_0^{\text{mech}})\mathbf{A}\{\mathbf{r}, t(\delta, \hat{\ell})\}. \quad (1.9.77)$$

(Again see Appendix D.)

1.9.9. Review Exercise 7.6. It treated the Cartesian-coordinate Hamiltonian (6.16). Show that the cylindrical-coordinate Hamiltonian (6.18) can be treated analogously. Conclude that in this respect there is nothing special about the use of Cartesian coordinates.

1.9.10. Review Exercise 5.1 that related mechanical and canonical momentum. Show that the *mechanical* energy E^{mech} is given by the relation

$$E^{\text{mech}} = \gamma mc^2 = [m^2 c^4 + c^2(\mathbf{p}^{\text{mech}})^2]^{1/2} = [m^2 c^4 + c^2(\mathbf{p} - q\mathbf{A})^2]^{1/2}. \quad (1.9.78)$$

Review Exercise 5.3. Using the definition (6.5), show that

$$p_t = -E^{\text{mech}} - q\psi. \quad (1.9.79)$$

Make the definition

$$p_t^{\text{mech}} = -E^{\text{mech}}, \quad (1.9.80)$$

in which case

$$p_t = p_t^{\text{mech}} - q\psi = -\gamma mc^2 - q\psi, \quad (1.9.81)$$

which is a relation analogous to those in Exercise 5.1. Also compare the above results with (6.59), those of Exercise 6.11, and (7.21). Note, using (6.45) and (6.53), that there are the relations

$$A^4 = A_4 = \psi/c. \quad (1.9.82)$$

Bibliography

Maps, Map Iteration, Chaos, and Fractals

Entering the words *dynamical systems* or *chaos* or *fractal* into the *Amazon* search window produces overwhelming lists of books on these subjects. Also, Google Dynamical Systems-Scholarpedia, and Encyclopedia Dynamical Systems-Scholarpedia, and see the Web site <http://www.scholarpedia.org>.

- [1] P. Cvitanović, R. Artuso, R. Mainieri, G. Tanner, G. Vattay, N. Whelan, and A. Wirzba, *Chaos: Classical and Quantum*, (2016). See the Web site <http://chaosbook.org/chapters/ChaosBook.pdf>. See also the Web site <http://chaosbook.org>.
- [2] R. M. May, “Simple Mathematical Models with very Complicated Dynamics”, *Nature*, **261**, 459 (1976).
- [3] B. B. Mandelbrot, *The Fractal Geometry of Nature*, W.H. Freeman (1983).
- [4] B. B. Mandelbrot, *Fractals and Chaos: The Mandelbrot Set and Beyond*, (Springer, 2004).
- [5] H. G. Schuster and W. Just, *Deterministic Chaos, An Introduction*, (Wiley-VCH, 2005).
- [6] H.-O. Peitgen and D. Saupe, Eds., *The Science of Fractal Images*, Springer-Verlag (1988).
- [7] H.-O. Peitgen, H. Jürgens, D. Saupe, *Chaos and Fractals: New Frontiers of Science*, Springer-Verlag (1992).
- [8] G. Velo and A. S. Wightman, edit., *Regular and Chaotic Motions in Dynamic Systems*, Plenum Press (1985).
- [9] D. Arrowsmith and C. Place, *An Introduction to Dynamical Systems*, Cambridge University Press (1990).
- [10] D. Arrowsmith and C. Place, *Dynamical Systems: differential equations, maps, and chaotic behavior*, Chapman & Hall (1992).
- [11] M. F. Barnsley, *Fractals Everywhere*, Academic Press (1993).
- [12] K. Falconer, *The Geometry of Fractal Sets*, Cambridge University Press (1985).

- [13] D. R. Hofstadter, *Mathematical Themes: Questing for the Essence of Mind and Pattern*, chapter 16, (Basic Books, 1996).
- [14] G. L. Baker and J. P. Gollub, *Chaotic Dynamics: an introduction*, 2nd ed., (Cambridge University Press, 1996).
- [15] A. J. Lichtenberg and M.A. Lieberman, *Regular and Chaotic Dynamics*, 2nd ed., (Springer, 1992).
- [16] F.C. Moon, *Chaotic and Fractal Dynamics: An Introduction for Applied Scientists and Engineers*, (Wiley, 1992).
- [17] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, (Addison-Wesley, 1989).
- [18] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, (Benjamin/Cummings, 1986).
- [19] R. L. Devaney and L. Keen, edit., *Chaos and Fractals: The Mathematics Behind the Computer Graphics*, Proceedings of the Symposia in Applied Mathematics, Vol. 39, (AMS, 1989).
- [20] R. L. Devaney, *A First Course in Chaotic Dynamical Systems*, (Perseus, 1992).
- [21] M. Hénon, Quarterly of Applied Mathematics **27**, 291 (1969).
- [22] J. Moser, *On Quadratic Symplectic Mappings*, Math. Zeitschrift **216**, 417-430 (1994).
- [23] M. Lyubich, *The Quadratic Family as a Qualitatively Solvable Model of Chaos*, Notices of the American Mathematical Society **47**, p. 1042 (2000).
- [24] R. A. Holmgren, *A First Course in Discrete Dynamics*, Springer (1996).
- [25] G. Contopoulos, *Order and Chaos in Dynamical Astronomy*, Springer (2004).
- [26] E. Ott, *Chaos in Dynamical Systems*, Cambridge (2002).
- [27] D. Ruelle, *Elements of Differentiable Dynamics and Bifurcation Theory*, Academic Press (1989).
- [28] D. Ruelle, Edit., *Turbulence, Strange Attractors, and Chaos*, World Scientific (1995).
- [29] D. Ruelle, “What Is a Strange Attractor?”, *Notices of the American Mathematical Society* **53**, p. 764, August 2006.
- [30] Y. Ueda, *The Road to Chaos*, Aerial Press (1992).
- [31] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*, Cambridge University Press (1999).

- [32] B. Hasselblatt and A. Katok, *A First Course in Dynamics: with a Panorama of Recent Developments*, Cambridge University Press (2003).
- [33] B. Hasselblatt and A. Katok, Edit., *Handbook of Dynamical Systems*, Vol 1A, Elsevier (2002).
- [34] B. Hasselblatt and A. Katok, Edit., *Handbook of Dynamical Systems*, Vol 1B, Elsevier (2006).
- [35] B. Fiedler, Edit., *Handbook of Dynamical Systems*, Vol 2, Elsevier (2002).
- [36] H. Broer, B. Hasselblatt, and F. Takens, Edit., *Handbook of Dynamical Systems*, Vol 3, Elsevier (2010).
- [37] M. Brin and G. Stuck, *Introduction to Dynamical Systems*, Cambridge University Press (2002).
- [38] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*, Perseus Books/WestviewPress (1994 and 2015).
- [39] S. S. Abdullaev, *Construction of Mappings for Hamiltonian Systems and Their Applications*, Springer (2006).
- [40] G. Zaslavsky, R. Sagdeev, D. Usikov, and A. Chernikov, *Weak chaos and quasi-regular patterns*, Cambridge University Press (1991).
- [41] G. Zaslavsky, *Physics of Chaos in Hamiltonian Systems*, Imperial College Press (1998).
- [42] G. Zaslavsky, *Hamiltonian Chaos and Fractional Dynamics*, Oxford University Press (2005).
- [43] J. Sprott, *Elegant Chaos: Algebraically Simple Chaotic Flows*, World Scientific (2010).
- [44] T. Tél and M. Gruiz, *Chaotic Dynamics: An Introduction Based on Classical Mechanics*, Cambridge University Press (2006).

Maps in the Complex Domain

- [45] H.-O. Peitgen and P.H. Richter, *The Beauty of Fractals, Images of Complex Dynamical Systems*, Springer-Verlag (1986).
- [46] A. Douady and J. Hubbard, *Étude dynamique des polynômes complex*, Publications mathématiques d'Orsay, Université de Paris-Sud (1984).
- [47] J. Hubbard, “The Hénon mapping in the complex domain”, published in *Chaotic Dynamics and Fractals*, M. Barnsley and S. Demko, Eds., p. 101, Academic Press (1986).

- [48] J. H. Hubbard and R. W. Oberste-Vorth, *Hénon Mappings in the Complex Domain I: The Global Topology of Dynamical Space*, Institut Des Hautes Étude Scientifiques Publications Mathématiques 79 (1994); *II: Projective and Inductive Limits of Polynomials, in Real and Complex Dynamical Systems*, B. Branner and P. Hjorth, eds., NATO ASI Series C: Mathematical and Physical Sciences Vol. 464 (Kluwer, Amsterdam 1995).
- [49] J. Hubbard, *Bulletin of the American Mathematical Society* **38**, p. 495 (2001).
- [50] H. Kriete, ed., *Progress in holomorphic dynamics*, Addison Wesley Longman (1998).
- [51] S. Heinemann, “Julia sets for endomorphisms of C^n ”, *Ergodic Theory and Dynamical Systems* **16** p. 1275 (1996).
- [52] K. Schmidt, *Dynamical Systems of Algebraic Origin*, Birkhäuser (1995).
- [53] J. Smillie and G. T. Buzzard, “Complex Dynamics in Several Variables”, published in *Flavors of Geometry*, S. Levy, Ed., Cambridge (1997).
- [54] E. Bedford and J. Smillie, “Polynomial diffeomorphisms of C^2 : currents, equilibrium measure and hyperbolicity”, *Invent. Math.* **103**, 69 (1991).
- [55] E. Bedford, M. Lyubich, and J. Smillie, “Polynomial diffeomorphisms of C^2 , IV. The measure of maximal entropy and laminar currents”, *Invent. Math.* **112**, 77 (1993).
- [56] S. Friedland and and J. Milnor, “Dynamical properties of plane polynomial automorphisms”, *Ergodic Theory Dyn. Syst.* **9**, p. 67 (1989).
- [57] J. Milnor, *Dynamics in One Complex Variable*, Third Edition, Princeton University Press (2006).
- [58] C. T. McMullen, *Complex Dynamics and Renormalization*, Annals of Mathematics Studies Number 135, Princeton University Press (1994).
- [59] L. Carleson and T.W. Gamelin, *Complex Dynamics*, Springer-Verlag (1993).
- [60] J. E. Fornaess, *Dynamics in Several Complex Variables*, Conference Board of the Mathematical Sciences Regional Conference Series in Mathematics Number 87, American Mathematical Society (1996).
- [61] J. E. Fornaess and N. Sibony, “Complex Dynamics in Higher Dimension”, in *Several Complex Variables*, M. Schneider and Y.-T. Siu, Edit., Cambridge University Press (1999).
- [62] B. Branner and P. Hjorth, edits., *Real and Complex Dynamical Systems*, Kluwer Academic Publishers (1995).
- [63] A. F. Beardon, *Iteration of Rational Functions*, Springer-Verlag (1991).
- [64] N. Steinmetz, *Rational iteration, complex analytic dynamical systems*, de Gruyter (1993).

- [65] R. L. Devaney, edit., *Complex Dynamical Systems: The Mathematics Behind the Mandelbrot and Julia Sets*, Proceedings of Symposia in Applied Mathematics, Vol. 49, American Mathematical Society (1994).
- [66] V. Dolotin and A. Morozov, *The Universal Mandelbrot Set: Beginning of the Story*, World Scientific (2006).
- [67] Mandelbrot set Web sites. Any search engine will find several Web sites devoted to the Mandelbrot set. Search under fractal, Julia, and Mandelbrot. Two such sites are listed below:
<http://aleph0.clarku.edu/~djoyce/julia/explorer.html>
<http://math.bu.edu/DYSYS/>

Universality and Renormalization

- [68] M. J. Feigenbaum, “Quantitative universality for a class of non-linear transformations”, *J. Statist. Phys.* **19**, 25 (1978).
- [69] O. E. Lanford III, “A Shorter Proof of the Existence of the Feigenbaum Fixed Point”, *Communications in Mathematical Physics* **96**, 521 (1984).
- [70] P. Cvitanović, Ed., *Universality in Chaos*, Second Edition, Adam Hilger (1989).
- [71] P. Collet and J.-P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems*, Birkhäuser (1980).
- [72] R. S. MacKay, *Renormalization in Area Preserving Maps*, Princeton University Ph.D. Thesis (1982).
- [73] J. M. Greene, R. S. MacKay, F. Vivaldi, and M.J. Feigenbaum, “Universal Behaviour of Area-Preserving Maps”, *Physica* **3D**, 468 (1981).
- [74] T. C. Bountis, “Period-Doubling Bifurcations and Universality in Conservative Systems”, *Physica* **3D**, 577 (1981).

Differential Equations and Dynamical Systems

- [75] SIAM Dynamical Systems Web Site : <http://www.siam.org/activity/ds/>
- [76] G. D. Birkhoff, *Dynamical Systems*, American Mathematical Society (1966).
- [77] E. R. Scheinerman, *Invitation to Dynamical Systems*, Dover (2012).
- [78] S. Sternberg, *Dynamical Systems*, Dover (2010).
- [79] Garrett Birkhoff and G-C. Rota, *Ordinary Differential Equations*, 4th Ed., Wiley (1989).
- [80] A. Andronov and C. Chaikin, *Theory of Oscillations*, Princeton University Press (1949).

- [81] Francis J. Murray and Kenneth S. Miller, *Existence Theorems for Ordinary Differential Equations*, (New York University Press and Interscience Publishing Co. 1954).
- [82] Y. Ilyashenko and S. Yakovenko, *Lectures on Analytic Differential Equations*, American Mathematical Society (2008).
- [83] W. Walter, *Ordinary Differential Equations*, Springer (1998).
- [84] H. Zoladek, *The Monodromy Group*, Birkhäuser Verlag (2006).
- [85] Fred Brauer and John A. Nohel, *The Qualitative Theory of Ordinary Differential Equations*, Benjamin (1969).
- [86] Zhang Zhi-fen, Ding Tong-ren, Huang Wen-zao, Dong Zhen-xi, *The Qualitative Theory of Differential Equations*, American Mathematical Society (1992).
- [87] Earl A. Coddington and Norman Levinson, *Theory of Ordinary Differential Equations*, McGraw-Hill (1955).
- [88] Philip Hartman, *Ordinary Differential Equations*, Birkhäuser (1982).
- [89] I. G. Petrovski, *Ordinary Differential Equations*, Prentice-Hall (1966).
- [90] V. I. Arnold, *Ordinary Differential Equations*, third edition Springer Verlag (1992).
- [91] V. I. Arnold, *Geometrical Methods in the Theory of Ordinary Differential Equations*, Springer Verlag (1983).
- [92] E. Hille, *Ordinary Differential Equations in the Complex Domain*, John Wiley (1976).
- [93] E. Hille, *Lectures on Ordinary Differential Equations*, Addison-Wesley (1969).
- [94] J. H. Hubbard and B. H. West, *Differential Equations: a Dynamical Systems Approach*, Springer (1971).
- [95] J. Hale and H. Kocak, *Dynamics and Bifurcations*, Springer Verlag (1991).
- [96] M. W. Hirsch, S. Smale, and R. L. Devaney, *Differential Equations, Dynamical Systems, and an Introduction to Chaos*, Elsevier (2013).
- [97] P. Blanchard, R. L. Devaney, and G.R. Hall, *Differential Equations*, Brooks/Cole (2002).
- [98] S. Lefschetz, *Differential Equations: Geometrical Theory*, 2nd Edition, Interscience (1963).
- [99] J. Meiss, *Differential Dynamical Systems*, SIAM (2007).
- [100] L. Perko, *Differential Equations and Dynamical Systems*, third edition, (Springer 2002).

- [101] C. Chicone, *Ordinary Differential Equations with Applications*, second edition, Springer-Verlag (2006).
- [102] N. Minorsky, *Nonlinear Oscillations*, Krieger Publishing Company, New York (1974).
- [103] P. Hagedorn, *Non-linear Oscillations*, Clarendon Press, Oxford (1982).
- [104] V. M. Starzhinskii, *Applied Methods in the Theory of Nonlinear Oscillations*, Mir Publishers, Moscow (1980).
- [105] J. K. Hale, *Ordinary Differential Equations*, Wiley-Interscience (1969).
- [106] R. E. Bellman, *Stability Theory of Differential Equations*, Dover (1969).
- [107] R. E. Bellman, *Methods of Nonlinear Analysis*, Vols. 1 and 2, Academic Press (1970).
- [108] F. Verhulst, *Nonlinear Differential Equations and Dynamical Systems*, Springer-Verlag (1990).
- [109] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer-Verlag (1990).
- [110] J. Guckenheimer, J. Moser, and S. Newhouse, *Dynamical Systems: C.I.M.E, Lectures*, Birkhäuser (1983).
- [111] R. H. Rand and D. Armbruster, *Perturbation Methods, Bifurcation Theory, and Computer Algebra*, Springer-Verlag (1987).
- [112] R. H. Rand, *Topics in Nonlinear Dynamics with Computer Algebra*, Gordan and Breach (1994).
- [113] J. M. T. Thompson and H. B. Stewart, *Nonlinear Dynamics and Chaos*, Second Edition, John Wiley (2002).
- [114] R. C. Hilborn, *Chaos and Nonlinear Dynamics: An Introduction for Scientists and Engineers*, Second Edition, Oxford University Press (2006).
- [115] R. C. Robinson, *Dynamical Systems: Stability, Symbolic Dynamics, and Chaos*, CRC Press (1995).
- [116] R. C. Robinson, *An Introduction to Dynamical Systems: Continuous and Discrete*, Pearson Prentice Hall (2004).
- [117] H. Stephani, *Differential Equations: Their solution using symmetries*, M. Maccallum, Edit., Cambridge University Press (1989).
- [118] P. J. Olver, *Applications of Lie Groups to Differential Equations*, Springer-Verlag (1993).
- [119] R. Hermann, *Lie-Theoretic ODE Numerical Analysis, Mechanics, and Differential Systems*, Math Sci Press (1994).

- [120] N. Ibragimov, *Transformation Groups and Lie Algebras*, World Scientific (2013).
 Existence, Uniqueness, Differentiability, and Analyticity Theorems
- [121] E. Hille, *Ordinary Differential Equations in the Complex Domain*, John Wiley (1976).
- [122] E. Hille, *Lectures on Ordinary Differential Equations*, Addison-Wesley (1969).
- [123] Earl A. Coddington and Norman Levinson, *Theory of Ordinary Differential Equations*, McGraw-Hill (1955).
- [124] Francis J. Murray and Kenneth S. Miller, *Existence Theorems for Ordinary Differential Equations*, New York University Press and Interscience Publishing Co. (1954).
- [125] W. Walter, *Ordinary Differential Equations*, Springer (1998).
- [126] Y. Il'yashenko and S. Yakovenko, *Lectures on Analytic Differential Equations*, American Mathematical Society (2008).
- [127] Zhang Zhi-fen, Ding Tong-ren, Huang Wen-zao, Dong Zhen-xi, *The Qualitative Theory of Differential Equations*, American Mathematical Society (1992).
- [128] Philip Hartman, *Ordinary Differential Equations*, Birkhäuser (1982).
- [129] H. Amann, *Ordinary Differential Equations: An Introduction to Nonlinear Analysis*, Walter de Gruyter (1990).
- [130] S-N. Chow and J. Hale, *Methods of Bifurcation Theory*, Springer-Verlag (1982).
- [131] R. Agarwal and V. Lakshmikantham, *Uniqueness and Nonuniqueness Criteria for Ordinary Differential Equations*, World Scientific (1993).
- [132] M. Irwin, *Smooth Dynamical Systems*, Academic Press (1980).
- [133] V. Nemytskii and V. Stepanov, *Qualitative Theory of Differential Equations*, Princeton University Press (1960).
- [134] W. Hurewicz, *Lectures on Ordinary Differential Equations*, M.I.T. Press (1963), Dover (1990).
- [135] L. Piccinini, G. Stampacchia, and G. Vidossich, *Ordinary Differential Equations in R^n : Problems and Methods*, Springer-Verlag (1984).
- Solution for Monomial Hamiltonian
- [136] F. J. Testa, *J. Math Phys.* **14**, p. 1097 (1973).
- [137] P. J. Channell, *Explicit Integration of Kick Hamiltonians in Three Degrees of Freedom*, Accelerator Theory Note AT-6:ATN-86-6, Los Alamos National Laboratory (1986).
- [138] I. Gjaja, *Particle Accelerators*, vol. 43 (3), pp. 133-144 (1994).

- [139] L. Michelotti, *Comment on the exact evaluation of symplectic maps*, Fermilab preprint (1992).

Original Sources and Histories

- [140] I. Newton, *The Principia*, Translated by B. Cohen and A. Whitman, University of California Press (1999).
- [141] S. Chandrasekhar, *Newton's Principia for the Common Reader*, Clarendon Press (1995).
- [142] C. Pask, *Magnificent Principia: Exploring Isaac Newton's Masterpiece*, Prometheus Books (2013).
- [143] Jed Z. Buchwald and Mordechai Feingold, *Newton and the Origin of Civilization*, Princeton University Press (2013).
- [144] Rob Iliffe, *Priest of Nature: The Religious Worlds of Isaac Newton*, Oxford University Press (2017).
- [145] G. Alexanderson and L. Klosinski, “The Newton-Leibniz Controversy”, *Bulletin of the American Mathematical Society* **53**, p. 295 (2016).
- [146] G. W. Leibniz, *Von dem Verhängnisse*, in *Hauptschriften zur Grundlegung der Philosophie*, Vol II, pp. 129-134 (Ernst Cassirer, Leipzig 1906), cited by P. Cvitanović et al. in reference 1 above.
- [147] R. J. Boscovich, *A Theory of Natural Philosophy*, reprinted by MIT Press, p. 141 (1966).
- [148] P. S. Laplace, *A Philosophical Essay on Probabilities*, Chapter II, On or Concerning Probability, Dover (1951) and Springer-Verlag (1995).
- [149] W. R. Hamilton, *Trans. R. Irish Acad.* **15**, 69 (1828); **16**, 1 (1830); **16**, 93 (1831); **17**, 1 (1837). Reprinted in *The Mathematical Papers of Sir W. R. Hamilton*, Vol. I, *Geometrical Optics*, A. W. Conway and J. L. Synge, eds., Cambridge U. Press, Cambridge (1931). All the mathematical papers of Hamilton may be found at <https://www.maths.tcd.ie/pub/HistMath/People/Hamilton/Papers.html>
- [150] H. Poincaré, *New Methods of Celestial Mechanics*, Parts 1, 2, and 3. (Originally published as *Les Méthodes nouvelles de la Méchanique céleste*.) American Institute of Physics *History of Modern Physics and Astronomy*, Volume 13, D. L. Goroff, Edit., American Institute of Physics (1993). See page 145 of the Editor’s Introduction for a statement of Poincaré’s holomorphic lemma, and Sections §20 through §27 for Poincaré’s proof. The Editor’s Introduction also provides a very useful summary of Poincaré’s contributions to Dynamical Systems and his legacy.
- [151] Lizhen Ji and Athanase Papadopoulos, Edit., *Sophus Lie and Felix Klein: The Erlangen Program and Its Impact in Mathematics and Physics*, European Mathematical Society and American Mathematical Society (2015).

- [152] G. Duffing, *Erzwungene Schwingungen bei veränderlicher Eigenfrequenz und ihre technische Bedeutung*, Braunschweig, Druck und Verlag von Friedr. Vieweg und Sohn (1918).
- [153] E. D. Courant and H. Snyder, “Theory of the Alternating-Gradient Synchrotron”, *Annals Phys.* **3**, p. 1 (1958).
- [154] S. Penner, “Calculations of Properties of Magnetic Deflection Systems”, *Rev. of Sci. Instr.* **32**, p. 150 (1961).
- [155] K. L. Brown, *TRANSPORT*, SLAC-75, revision 3 (1975); K.L. Brown, F. Rothacker, D. C. Carey, and Ch. Iselin, *TRANSPORT*, SLAC-91, revision 2 (1977).
- [156] G. Hori, “Theory of general perturbations with unspecified canonical variables”, *Publications of the Astronomical Society of Japan* **18**, p. 287 (1966).
- [157] A. Deprit, *Celest. Mech.* **1**, 12 (1969).
- [158] J. Barrow-Green, *Poincaré and the Three Body Problem*, American Mathematical Society (1997).
- [159] F. Browder, Edit., *The Mathematical Heritage of Henri Poincaré*, *Proceedings of Symposia in Pure Mathematics of the American Mathematical Society* **39**, Parts 1 and 2, American Mathematical Society (1983).
- [160] J. Gray, *Henri Poincaré. A Scientific Biography*, Princeton University Press (2013).
- [161] É. Charpentier, É. Ghys, and A. Lesne, Edit., *The Scientific Legacy of Poincaré*, History of Mathematics Volume 36, American & London Mathematical Societies (2010).
- [162] B. Duplantier and V. Rivasseau, eds., *Henri Poincaré, 1912-2012: Poincaré Seminar 2012*, Birkhäuser-Springer (2015).
- [163] R. Abraham and Y. Ueda, Edit., *The Chaos Avant-Garde: Memories of the Early Days of Chaos Theory*, World Scientific (2000).
- [164] D. Nolte, “The tangled tale of phase space”, *Physics Today* **64**(4), 33 (April 2010).
- [165] A. Motter and D. Campbell, “Chaos at fifty”, *Physics Today* **66**(5), 27 (May 2013).
- Solar System Stability; Singularities in the Newtonian N-Body Problem
- [166] E. Belbruno, *Fly Me to the Moon, an Insider’s Guide to the New Science of Space Travel*, Princeton University Press (2007).
- [167] E. Belbruno, *Capture Dynamics and Chaotic Motions in Celestial Mechanics: with Applications to the Construction of Low Energy Transfers*, Princeton University Press (2004).
- [168] F. Diacu and P. Holmes, *Celestial Encounters: The Origins of Chaos and Stability*, Princeton University Press (1996).

- [169] Z. Xia, “The existence of noncollision singularities in Newtonian systems”, *Annals of Mathematics*, **135**, pp. 411-468 (1992).
- [170] D. Saari and Z. Xia, “Off to Infinity in Finite Time”, *Notices of the AMS* **42**, 538-546 (1995).
- [171] D. Saari and Z. Xia, “Singularities in the Newtonian N -Body Problem”, (1996). <http://citeserx.ist.psu.edu/viewdoc/download?doi=10.1.1.24.1325&rep=rep1&type=pdf>
- [172] H. Poincaré, *New Methods of Celestial Mechanics*, Parts 1, 2, and 3. (Originally published as Les Méthodes nouvelles de la Méchanique céleste.) American Institute of Physics *History of Modern Physics and Astronomy*, Volume 13, D. L. Goroff, Edit., American Institute of Physics (1993). See the Editor’s Introduction for a discussion of the stability of the solar system.
- [173] Y. Kozai, Edit., *The Stability of the Solar System and of Small Stellar Systems*, D. Reidel (1974).
- [174] M. Suvakov and V. Dmitrasinovic, “Three Classes of Newtonian Three-Body Planar Periodic Orbits”, *Phys. Rev. Lett.* **110**, 114301 (2013). See also the Web site <http://suki.ipb.ac.rs/3body/>.
- Classical/Celestial/Galactic Mechanics, KAM Theory, and Nonlinear Dynamics
- [175] G. Gallavotti, *The Elements of Mechanics*, Springer-Verlag (1983). See also the Web site <http://141.108.10.74/pagine/deposito/2007/elements.pdf>.
- [176] W. S. Koon, M. W. Lo, J. E. Marsden, S. D. Ross, *Dynamical Systems, the Three-Body Problem, and Space Mission Design*, 12 October 2022 version, available at the Web site https://ross.aoe.vt.edu/books/Ross_3BodyProblem_Book_2022.pdf. Also, see the Web sites <https://ross.aoe.vt.edu/books/> and <https://ross.aoe.vt.edu>.
- [177] A. Brizard, *An Introduction to Lagrangian Mechanics*, World Scientific (2008).
- [178] H. Goldstein, *Classical Mechanics*, Addison-Wesley (1980).
- [179] L. D. Landau and E. M. Lifshitz, *Mechanics*, Addison-Wesley (1969).
- [180] R. Abraham and J. Marsden, *Foundations of Mechanics*, American Mathematical Society (2008).
- [181] R. Abraham and C. Shaw, *Dynamics: the Geometry of Behavior*, 4 Vols., Aeriel Press (1984).
- [182] V. I. Arnold, *Mathematical Methods of Classical Mechanics*, Second Edition, Springer-Verlag (1989).

- [183] V. I. Arnold, V.V. Kozlov, and A. I. Neishtadt, *Mathematical Aspects of Classical and Celestial Mechanics*, Third Edition, Springer Verlag (2006).
- [184] V. I. Arnold, Ya. G. Sinai, et al., edit, *Dynamical Systems I* through *Dynamical Systems IX*, Volumes from the Encyclopedia of Mathematical Sciences, Springer Verlag (1995).
- [185] V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics*, Benjamin (1968).
- [186] H. Cabral and F. Diacu, Edit, *Classical and Celestial Mechanics*, Princeton University Press (2002).
- [187] J. Danby, *Fundamentals of Celestial Mechanics*, Macmillan (1962).
- [188] H. S. Dumas, K. R. Meyer, and D. S. Schmidt, *Hamiltonian Dynamical Systems: History, Theory, and Applications*, Springer Verlag (1995).
- [189] H. S. Dumas, *The KAM Story: A Friendly Introduction to the Content, History, and Significance of Classical Kolmogorov-Arnold-Moser Theory*, World Scientific (2014).
- [190] G. Benettin, I. Galgani, A. Giorgilli, J.-M. Strelcyn, “A Proof of Kolmogorov’s Theorem on Invariant Tori Using Canonical Transformations Defined by the Lie Method”, *Il Nuovo Cimento* **79 B**, 201 (1984).
- [191] H. Broer and M. Sevryuk, “KAM Theory: Periodicity in dynamical systems”, see the Web link <https://www.semanticscholar.org/paper/KAM-Theory--Quasi-periodicity-in-Dynamical-Systems-Broer-Sevryuk/eaa966f12b4b497b82bb64e34a8481c2fb1b8bb6>. Also Google the authors’ names individually to find additional references.
- [192] K. Meyer, G. Hall, and D. Offin, *Introduction to Hamiltonian Dynamical Systems and the N-Body Problem*, Second Edition, Springer (2009).
- [193] J. Moser, “Lectures on Hamiltonian Systems”, *Mem. Am. Math. Soc.* **81**, 1-60 (1968).
- [194] C. Hayashi, *Nonlinear Oscillations in Physical Systems*, McGraw-Hill (1964).
- [195] E. J. Saletan and A.H. Cromer, *Theoretical Mechanics*, John Wiley (1971).
- [196] J. V. Jose and E. J. Saletan, *Classical Dynamics: A Contemporary Approach*, Cambridge University Press (1998).
- [197] C. L. Siegel and J. K. Moser, *Lectures on Celestial Mechanics*, Springer-Verlag (1995).
- [198] J. K. Moser, *Stable and Random Motions in Dynamical Systems: with Special Emphasis on Celestial Mechanics*, Princeton University Press (1973).
- [199] J. K. Moser and E. J. Zehnder, *Notes on Dynamical Systems*, American Mathematical Society (2005).

- [200] G. Benettin, J. Henrard, S. Kuksin, and A. Giorgilli, *Hamiltonian Dynamics - Theory and Applications*, Springer-Verlag (2005).
- [201] G. J. Sussman, J. Wisdom, and M.E. Mayer, *Structure and Interpretation of Classical Mechanics*, Second Edition, MIT Press (2014).
- [202] R. Talman, *Geometric Mechanics*, John Wiley (2000).
- [203] J. E. Marsden and T. S. Ratiu, *Introduction to Mechanics and Symmetry*, Springer Verlag (1999).
- [204] L. Michelotti, *Intermediate Classical Dynamics with Applications to Beam Physics*, John Wiley (1995).
- [205] J. L. McCauley, *Classical Mechanics*, Cambridge (1997).
- [206] John R. Taylor, *Classical Mechanics*, University Science Books (2005).
- [207] R. L. Devaney and Z.H. Nitecki, *Classical Mechanics and Dynamical Systems*, Lecture notes in pure and applied mathematics, Vol. 70, Dekker (1981).
- [208] L. A. Pars, *A Treatise on Analytical Dynamics*, Ox Bow Press (1979).
- [209] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies with an Introduction to the Problem of Three Bodies*, Cambridge University Press (1960).
- [210] J. Lopuszanski, *The Inverse Variational Problem in Classical Mechanics*, World Scientific (1999).
- [211] G. Vilasi, *Hamiltonian Dynamics*, World Scientific (2001).
- [212] J. Lowenstein, *Essentials of Hamiltonian Dynamics*, Cambridge (2012).
- [213] W. Thirring, *A Course in Mathematical Physics: 1. Classical Dynamical Systems*, Springer-Verlag (1978).
- [214] J. Binney and S. Tremaine, *Galactic Dynamics*, Princeton University Press (1987).
- [215] F. Scheck, *Mechanics: from Newton's Laws to Deterministic Chaos*, Springer-Verlag (2005).
- [216] S. Sternberg, *Celestial Mechanics, Parts I and II*, W. A. Benjamin (1969).
- [217] V. Szebehely, *Theory of Orbits*, Academic Press (1967).
- [218] A. Nayfeh and B. Balachandran, *Applied Nonlinear Dynamics: Analytical, Computational, and Experimental Methods*, John Wiley & Sons (1995).
- [219] A. Nayfeh, *Nonlinear Interactions: Analytical, Computational, and Experimental Methods*, John Wiley & Sons (2000).

- [220] H. Nusse and J. Yorke, *Dynamics: Numerical Explorations*, Second revised and Enlarged Edition, Springer (1998).
- [221] E. Sudarshan and M. Mukunda, *Classical Dynamics: A Modern Perspective*, Wiley (1974).
- [222] F. Gantmacher, *Lectures in Analytical Mechanics*, Mir Publishers (1975).
- [223] R. Matzner and L. Shepley, *Classical Mechanics*, Prentice Hall (1991).
- [224] G. Benettin, J. Henrard, S. Kuksin, and A. Giorgilli (Edit.), *Hamiltonian Dynamics Theory and Applications*, Lecture Notes in Mathematics 861, Springer (2005).
- [225] J.-M. Souriau, *Structure of Dynamical Systems, a Symplectic View of Physics*, Birkhäuser (1997).
- [226] D. D. Holm, *Geometric Mechanics, Part I: Dynamics and Symmetry*, Imperial College Press, World Scientific (2008).
- [227] D. D. Holm, *Geometric Mechanics, Part II: Rotating, Translating and Rolling*, Imperial College Press, World Scientific (2008).
- [228] D. D. Holm, T. Schmah, C. Stoica, and D. C. P. Ellis, *Geometric Mechanics, from Finite to Infinite Dimensions*, Oxford (2009).
- [229] W. B. Kibble and F. H. Berkshire, *Classical Mechanics*, Fifth Edition, World Scientific (2004).
- [230] M. Spivak, *Physics for Mathematicians: Mechanics I*, Publish or Perish, Inc. (2010).
- [231] M. Audin, *Hamiltonian Systems and Their Integrability*, American Mathematical Society (2008).
- [232] J. J. Morales Ruiz, *Differential Galois Theory and Non-Integrability of Hamiltonian Systems*, Springer and Birkhäuser (1999).
- [233] M. Hénon, *Generating Families in the Restricted Three-Body Problem*, Springer Verlag (1997); *Generating Families in the Restricted Three-Body Problem II. Quantitative Study of Bifurcations*, Springer Verlag (2001).
- [234] D. Heggie and P. Hut, *The Gravitational Million-Body Problem: A Multidisciplinary Approach to Star Cluster Dynamics*, Cambridge University Press (2003).
- [235] S. Aarseth, *Gravitational N-Body Simulations: Tools and Algorithms*, Cambridge University Press (2003).
- [236] M. Levi, *Classical Mechanics with Calculus of Variations and Optimal Control*, American Mathematical Society (2014).
- [237] J. Papastavridis, *Analytical Mechanics : A Comprehensive Treatise on the Dynamics of Constrained Systems*, World Scientific (2014).

- [238] Richard K. Cooper and Claudio Pellegrini, *Modern Analytic Mechanics*, Kluwer Academic (2010).
- [239] Kai S. Lam, *Fundamental Principles of Classical Mechanics : A Geometrical Perspective*, World Scientific (2014).
- [240] H. Iro, *A Modern Approach to Classical Mechanics*, Second Edition, World Scientific (2016).
- [241] L. Hand and J. Finch, *Analytical Mechanics*, Cambridge University Press (1998).
- [242] D. Tong, *Classical Dynamics*, (2004-2005). See the Web site <http://www.damtp.cam.ac.uk/user/tong/dynamics/one.pdf>.
- [243] S. Wiggins, *Chaotic Transport in Dynamical Systems*, Springer-Verlag (1992).
- [244] S. Wiggins, *Normally Hyperbolic Invariant Manifolds in Dynamical Systems*, Springer-Verlag (1994).
- [245] S. Wiggins, *Introduction to Applied Dynamical Systems and Chaos*, Springer-Verlag (2003).

Inverse and Implicit Function Theorems and Functional Analysis

- [246] R. Courant and F. John, *Introduction to Calculus and Analysis*, Vol. I, Vol. II/1, Vol. II/2, Springer-Verlag (1998, 1999, 2000).
- [247] J. E. Marsden and M. J. Hoffman, *Elementary Classical Analysis*, p. 397, W.H. Freeman (1993).
- [248] R. Abraham, J. Marsden, and T. Ratiu, *Manifolds, Tensor Analysis, and Applications*, Springer-Verlag (1988).
- [249] W. Rudin, *Principles of Mathematical Analysis*, Third Edition, McGraw-Hill (1976).
- [250] W. Rudin, *Real and Complex Analysis*, Third Edition, McGraw-Hill (1987).
- [251] A. Knapp, *Advanced Real Analysis*, Birkhäuser (2005).
- [252] B. Simon, *A Comprehensive Course in Analysis* (5-volume set), American Mathematical Society (2015).
- [253] S. G. Krantz and H. R. Parks, *The Implicit Function Theorem: History, Theory, and Applications*, Birkhäuser (2002).

Euler's Relation for Homogeneous Functions

- [254] R. Courant and F. John, *Introduction to Calculus and Analysis*, Vol. I, Vol.. II/1, Vol. II/2, Springer-Verlag (1998, 1999, 2000). See pages 119-121 of Vol. II/1.

Electromagnetism

- [255] W. R. Smythe, *Static and Dynamic Electricity*, McGraw-Hill (1939).
- [256] J. A. Stratton, *Electromagnetic Theory*, McGraw-Hill (1941).
- [257] J. D. Jackson, *Classical Electrodynamics*, John Wiley (1999).
- [258] J.D. Jackson and L.B. Okun, “Historical roots of gauge invariance”, *Reviews of Modern Physics* **73**, p. 663 (2001).
- [259] W. K. H. Panofsky and M. Phillips, *Classical Electricity and Magnetism*, Second Edition, Addison-Wesley (1962) and Dover (2005).
- [260] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, Addison-Wesley (1971).
- [261] J. Schwinger *et al.*, *Classical Electrodynamics*, Perseus Books (1998).
- [262] E. Purcell and D. Morin, *Electricity and Magnetism*, third edition, Cambridge University Press (2013).
- [263] A. Zangwill *Modern Electrodynamics*, Cambridge University Press (2013).
- [264] D. J. Griffiths, *Introduction to Electrodynamics*, Prentice Hall (1999).
- [265] W. Gibson, *The Method of Moments in Electromagnetics*, Chapman & Hall/CRC (2008).

Lorentz Invariant Formulation

- [266] M. Henneaux and C. Teitelboim, *Quantization of Gauge Systems*, Princeton University Press (1992).
- [267] J. Sipe, “New Hamiltonian for a charged particle in an applied electromagnetic field”, *Phys. Rev A* **27**, p. 615 (1983).

Extended Phase Space and Variational Calculus

- [268] C. Lanczos, *The Variational Principles of Mechanics*, fourth edition, Dover (1986).
- [269] J. Struckmeier, “Hamiltonian dynamics on the symplectic extended phase space for autonomous and non-autonomous systems”, *J. Phys. A: Math. Gen.* **38**, 1257 (2005).
- [270] J. Struckmeier, “Extended Hamilton-Lagrange Formalism and Its Application to Feynman’s Path Integral for Relativistic Quantum Physics”, *International Journal of Modern Physics E* **18**, 79 (2009).
- [271] J. Struckmeier, W. Greiner, and H. Reichau, *Extended Lagrange and Hamiltonian Formalism for Point Mechanics and Covariant Hamiltonian Field Theory*, World Scientific (2014).

- [272] H. Rund, *The Hamilton-Jacobi theory in the calculus of variations*, D. Van Nostrand (1966).
- [273] I. Gelfand and S. Fomin, *Calculus of Variations*, Prentice Hall (1963) and Dover (2000).
- [274] M. Giaquinta and S. Hildebrandt, *Calculus of Variations I: The Lagrangian Formalism*, *Calculus of Variations II: The Hamiltonian Formalism*, Springer-Verlag (2004).
- [275] M. Morse, *The Calculus of Variations in the Large*, American Mathematical Society (1934).
- [276] B. van Brunt, *The Calculus of Variations*, Springer (2006).

Singular (Degenerate) Lagrangians

- [277] H. Rund, *The Hamilton-Jacobi theory in the calculus of variations*, D. Van Nostrand (1966). See Chapter 3.
- [278] P. A. M. Dirac, “Generalized Hamiltonian dynamics” *Can. J. Math.* **2**, 129-148 (1950).
- [279] R. Abraham and J. Marsden, *Foundations of Mechanics*, American Mathematical Society (2008). See Section 3.6.

Geodesics

- [280] Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds, and Physics*, Elsevier North Holland (1987). See pages 302 and 320 through 324.
- [281] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, volume 1, second edition, Publish or Perish (1979). See Chapter 9.
- [282] L. P. Eisenhart, *Riemannian Geometry*, Princeton (1960). See Section 17.

Fluid Mechanics

- [283] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, Volume 6 of a Course of Theoretical Physics, Pergamon Press (1979).

Accelerator Physics/Charged-Particle Optics

- [284] É. Forest, *Beam Dynamics: A New Attitude and Framework*, Harwood Academic Publishers (1998).
- [285] É. Forest, *From Tracking Code to Analysis: Generalized Courant-Snyder Theory for Any Accelerator Model*, Springer Japan (2016).
- [286] A. Chao, *Lectures on Accelerator Physics*, World Scientific (2020).
- [287] A. Wolski, *Beam Dynamics in High Energy Particle Accelerators*, Imperial College Press (2014).

- [288] A. Chao, M. Tigner, H. Weise, and F. Zimmermann, Edit., *Handbook of Accelerator Physics and Engineering*, Third Edition, World Scientific (2023).
- [289] E. Wilson, *An Introduction to Particle Accelerators*, Oxford University Press (2001).
- [290] H. Wiedemann, *Particle Accelerator Physics - Basic Principles and Linear Beam Dynamics*, Springer-Verlag (1993).
- [291] H. Wiedemann, *Particle Accelerator Physics II- Nonlinear and Higher-Order Beam Dynamics*, Springer-Verlag (1995).
- [292] M. Berz, *Modern Map Methods in Particle Beam Physics*, Volume 108 of *Advances in Imaging and Electron Physics*, Academic Press (1999).
- [293] D. Edwards and M. Syphers, *An Introduction to the Physics of High Energy Accelerators*, Wiley (1993).
- [294] D. Carey, *The Optics of Charged Particle Beams*, Harwood Academic (1987).
- [295] H. Wollnik, *Optics of Charged Particles*, Academic Press (1987).
- [296] M. Conte and W. MacKay, *An Introduction to the Physics of Particle Accelerators*, Second Edition, World Scientific (2008).
- [297] M. Reiser, *Theory and Design of Charged Particle Beams*, John Wiley (1994).
- [298] P. Bryant and K. Johnsen, *The Principles of Circular Accelerators and Storage Rings*, Cambridge University Press (1993).
- [299] S. Y. Lee, *Accelerator Physics*, World Scientific (1999).
- [300] N. Dikansky and D. Pestrikov, *The Physics of Intense Beams and Storage Rings*, AIP Press (1994).
- [301] S. Bernal, *A Practical Introduction to Beam Physics and Particle Accelerators*, IOP Concise Physics (2016).

Chapter 2

Numerical Integration

Nature laughs at the difficulties of integration.

Laplace

The differential equations of motion for many systems of physical interest cannot be completely solved in terms of familiar functions. For example, there are precious few problems in Plasma Physics, Space Mechanics, or Accelerator Design that have closed-form analytical solutions. Generally, a differential equation, or a set of differential equations, should be viewed as the source of some *new* transcendental function. This fact was realized shortly after the discovery of Classical Mechanics and Differential Equations. Consequently, over the past centuries and particularly in Celestial Mechanics, considerable effort has been put into the possibility of expressing solutions not in terms of known functions, but rather in terms of *infinite* series of known functions. For example, elaborate series expansions have been worked out for the motion of the planets and their moons, and these series have been used to compute their trajectories to high precision.

The contemporary approach is somewhat different. Usually a complete knowledge of every possible “trajectory” or motion of a system is not necessary. Rather, it often suffices to have a qualitative description of the types of allowed motion supplemented by a detailed knowledge of a few representative “orbits”. Detailed knowledge of specific orbits is today most easily obtained by numerical integration using digital computers.¹ The types of allowed orbits can usually be determined best by analytical and topological methods, although even here numerical studies often precede and suggest later analytical results. Contemporary mechanics is thus an interplay between both analytical and numerical methods.

Even a survey of numerical methods is outside the scope of this text. It would require a text in itself. However, we hope that the brief discussion we are about to give will impart some of the flavor of numerical techniques, and perhaps entice the reader to explore further on his or her own. We hasten to add that numerical methods are also important outside classical mechanics, and that the techniques learned here can be applied to other situations in which ordinary differential equations arise. They also serve as a background for related methods in the numerical treatment of partial differential equations.

¹Recently, however, there has been renewed interest in series expansions with the new twist that these expansions are produced by computers programmed to perform algebraic manipulations. In some cases it is advantageous to use series expansions to transform the equations of motion, and then to integrate these transformed equations numerically.

2.1 The General Problem

2.1.1 Integrating Forward in Time

Consider a set of first-order differential equations of the form (1.3.4). For compactness of notation we shall group together the quantities $(y_1 \cdots y_m)$ and $(f_1 \cdots f_m)$, and regard them as the components of two m -dimensional vectors: \mathbf{y} and \mathbf{f} . Thus, we rewrite (1.3.4) in the form

$$\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}, t). \quad (2.1.1)$$

Suppose t^0 is some initial time and we wish to integrate *forward* to the time $t^0 + T$. Divide up the time axis into N equal steps, each of duration h , so that

$$Nh = T. \quad (2.1.2)$$

Define successive times t^n by writing²

$$t^n = t^0 + nh. \quad (2.1.3)$$

See Figure 2.1.1 below. The time step, h , is taken to be small compared to the characteristic

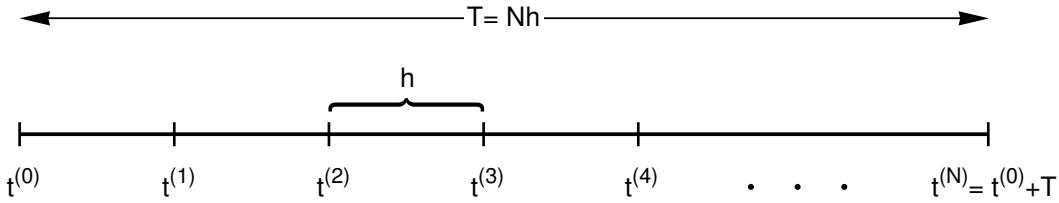


Figure 2.1.1: The Time Axis

time scale or period of the physical system we are studying. For example, in solving a pendulum problem, h should be much less than the period of oscillation. Our goal is to compute the vectors \mathbf{y}^n , where

$$\mathbf{y}^n = \mathbf{y}(t^n), \quad (2.1.4)$$

starting from the vector \mathbf{y}^0 . The vector \mathbf{y}^0 is assumed given as a set of definite numbers, i.e. the initial conditions at t^0 . To complete our notation, we make the definition

$$\mathbf{f}^n = \mathbf{f}(\mathbf{y}^n, t^n). \quad (2.1.5)$$

2.1.2 Integrating Backwards in Time

In the next several sections we will describe various methods for integrating forward in time to times later than t^0 . Suppose we instead wish to integrate backwards to times earlier than t^0 so that $T < 0$. According to Theorem 1.3.1 this should be possible. After a few moments reflection we see that this problem has already been solved if we have found how to integrate forward. To integrate backward, we simply change the sign of h . That is, once an integration method has been selected, execute it with $h < 0$.

²Warning! Here n is a superscript, not an exponent. Sometimes, however, n will be an exponent. There should be enough clues from the context for you to decide what is meant.

2.2 A Crude Solution Due to Euler

2.2.1 Procedure

Theorem 1.3.1 guarantees that the solution vectors \mathbf{y}^n exist and are uniquely specified by \mathbf{y}^0 . The question is how to find them. Proceed one step at a time! By Taylor's theorem,

$$\mathbf{y}^1 = \mathbf{y}(t^1) = \mathbf{y}(t^0 + h) = \mathbf{y}^0 + h\dot{\mathbf{y}}^0 + O(h^2) \quad (2.2.1)$$

or

$$\mathbf{y}^1 = \mathbf{y}^0 + h\mathbf{f}^0 + O(h^2). \quad (2.2.2)$$

(Here, and in what follows, we assume analyticity in t , or at least the existence of several derivatives, as guaranteed by the theorems and discussion of Section 1.3.) Since \mathbf{y}^0 and t^0 are definite numbers, \mathbf{f}^0 is explicitly computable. Let us ignore the $O(h^2)$ error in (2.2) for the moment and accept (2.2) as an exact result for \mathbf{y}^1 . Then using this \mathbf{y}^1 we can compute \mathbf{f}^1 , and from that \mathbf{y}^1 and \mathbf{f}^1 proceed in similar fashion to compute \mathbf{y}^2 and \mathbf{f}^2 , etc. In summary, we march forward step by step using the rule

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h\mathbf{f}^n. \quad (2.2.3)$$

Suppose we march to the time $t^0 + T$. This requires $N = T/h$ steps. At each step we make a *local* error of order h^2 . Consequently the *cumulative* error, barring cancellations that could only reduce it, is of order³

$$Nh^2 = Th. \quad (2.2.4)$$

We see that if the step size h is made sufficiently small and correspondingly the number of steps N sufficiently large, the error made in computing $\mathbf{y}(t^0 + T)$ using (2.3) can be made arbitrarily small.

2.2.2 Numerical Example

Consider the differential equation

$$\ddot{x} + x = 2t \quad (2.2.5)$$

with the initial conditions

$$x(0) = 0 \text{ and } \dot{x}(0) = 1. \quad (2.2.6)$$

We convert (2.5) into a first-order set by writing

$$y_1 = x, \quad y_2 = \dot{x}, \quad (2.2.7)$$

and find

$$f_1(\mathbf{y}, t) = \dot{y}_1 = y_2, \quad (2.2.8)$$

$$f_2(\mathbf{y}, t) = \dot{y}_2 = 2t - y_1. \quad (2.2.9)$$

³By “order Th ” we mean proportional to Th with a bounded but *unspecified* proportionality constant.

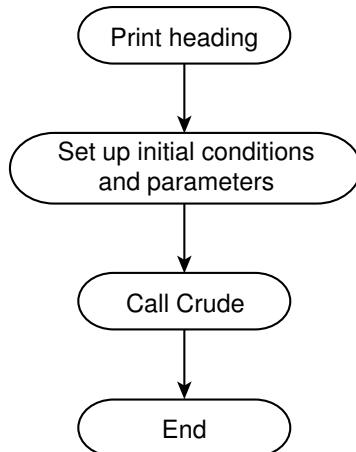
The simple computer program diagramed, listed, and annotated in Exhibit 2.1 below implements the Euler method (2.3) to integrate this set. The step size is $h = 1/10$. The differential equation we have selected is sufficiently simple that it also can be integrated analytically to give the exact result

$$y_1 = x(t) = 2t - \sin t, \quad (2.2.10)$$

$$y_2 = \dot{x}(t) = 2 - \cos t. \quad (2.2.11)$$

Note that the characteristic period of the solution is 2π so that the choice $h = 1/10$ is considerably smaller than the period as required. For comparison, both the Euler result and the exact result are tabulated.

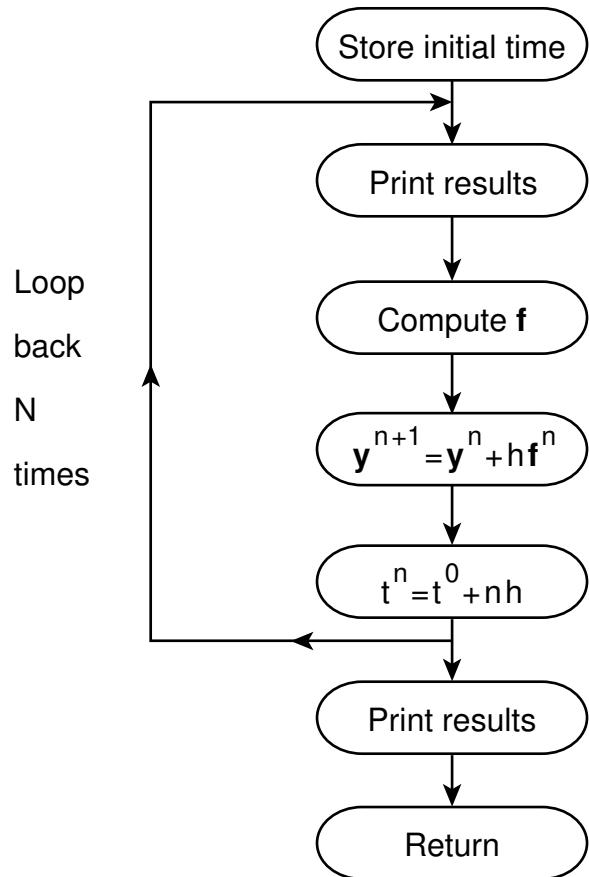
Exhibit 2.2.1: Crude Euler Integration

Block Diagram of Main Program

```

c This is the main program for illustrating the crude Euler method
c of numerical integration.
c
c Print heading.
c
      write(6,100)
100 format
     & (1h , 'time', 4x, 'y1comp', 10x, 'y2comp', 10x, 'y1true', 10x, 'y2true', '/')
c
c Set up initial conditions and parameters. n is the number of integration
c steps we wish to make.
c
      t=0.
      h=.1
      n=15
      y1=0.
      y2=1.
c
      call crude(t,h,n,y1,y2)
c
      end
  
```

Block Diagram of Integration Routine



```

c This is the crude Euler integration subroutine
c
      subroutine crude(t,h,n,y1,y2)
c
c Store initial time.
c
      tint=t
c
c Printing and integration loop.
c
      do 100 i=1,n
         call prints(t,y1,y2,y1true(t),y2true(t),0)
c
c Compute f, the right side of the differential equation.
c
         call eval(y1,y2,t,f1,f2)
c
c Make integration step and update time.
c
   
```

```
y1=y1+h*f1
y2=y2+h*f2
t=tint+float(i)*h
c
100 continue
c
c Print final results.
c
call prints(t,y1,y2,y1true(t),y2true(t),0)
c
return
end
```

Auxiliary Programs

```
c This subroutine evaluates f, the right side of the
c differential equation.
c
      subroutine eval(y1,y2,t,f1,f2)
c
      f1=y2
      f2=2.*t-y1
c
      return
      end

c
c Function for computing the exact value of y1.
c
      function y1true(t)
      y1true=2.*t-sin(t)
      return
      end

c
c Function for computing the exact value of y2.
c
      function y2true(t)
      y2true=2.-cos(t)
      return
      end

c
c Subroutine to handle printing. It need not concern the reader.
c
      Subroutine prints(t,y1,y2,y1t,y2t,iflag)
c
      if (iflag .eq. 0) then
      write(6,100) t,y1,y2,y1t,y2t
100 format (1h ,f6.4,2x,4(e14.8,2x))
      return
      endif
c
      if (iflag .ne. 0) then
      write(6,200) y1,y2
200 format (1h ,8x,2(e14.8,2x))
      return
      endif
c
      end
```

Numerical Results

time	y1comp	y2comp	y1true	y2true
0.0000	0.00000000E+00	0.10000000E+01	0.00000000E+00	0.10000000E+01
0.1000	0.10000000E+00	0.10000000E+01	0.10016658E+00	0.10049958E+01
0.2000	0.20000000E+00	0.10100000E+01	0.20133068E+00	0.10199335E+01
0.3000	0.30100000E+00	0.10300000E+01	0.30447981E+00	0.10446635E+01
0.4000	0.40400001E+00	0.10598999E+01	0.41058168E+00	0.10789391E+01
0.5000	0.50999004E+00	0.10994999E+01	0.52057445E+00	0.11224174E+01
0.6000	0.61994004E+00	0.11485009E+01	0.63535756E+00	0.11746644E+01
0.7000	0.73479015E+00	0.12065070E+01	0.75578231E+00	0.12351578E+01
0.8000	0.85544086E+00	0.12730279E+01	0.88264394E+00	0.13032933E+01
0.9000	0.98274368E+00	0.13474839E+01	0.10166732E+01	0.13783901E+01
1.0000	0.11174921E+01	0.14292095E+01	0.11585290E+01	0.14596977E+01
1.1000	0.12604131E+01	0.15174602E+01	0.13087927E+01	0.15464039E+01
1.2000	0.14121591E+01	0.16114190E+01	0.14679611E+01	0.16376423E+01
1.3000	0.15733010E+01	0.17102031E+01	0.16364419E+01	0.17325013E+01
1.4000	0.17443212E+01	0.18128730E+01	0.18145503E+01	0.18300328E+01
1.5000	0.19256085E+01	0.19184409E+01	0.20025051E+01	0.19292628E+01

We conclude that with $h = 1/10$, the Euler method integrates (2.5) over the range $t = 0$ to $t = 1.5$ with an accuracy of somewhat less than two significant figures.

Exercises

2.2.1. Consider the differential equation

$$\ddot{x} + x = 0. \quad (2.2.12)$$

a) Show that in this case Euler's method amounts to solving the set of difference equations

$$y_1^{n+1} = y_1^n + hy_2^n, \quad (2.2.13)$$

$$y_2^{n+1} = y_2^n - hy_1^n. \quad (2.2.14)$$

b) Show that the difference equations have the solution

$$\mathbf{y}^n = M^n \mathbf{y}^0 \quad (2.2.15)$$

where M is the matrix

$$M = \begin{pmatrix} 1 & h \\ -h & 1 \end{pmatrix}. \quad (2.2.16)$$

- c) Show by explicit computation that M has two linearly independent eigenvectors \mathbf{a} and \mathbf{b} with eigenvalues α and β . Expand \mathbf{y}^0 in terms of \mathbf{a} and \mathbf{b} . That is, write

$$\mathbf{y}^0 = A\mathbf{a} + B\mathbf{b} \quad (2.2.17)$$

where A and B are expansion coefficients. Show that

$$\mathbf{y}^n = \alpha^n A\mathbf{a} + \beta^n B\mathbf{b}. \quad (2.2.18)$$

- d) Study how $\mathbf{y}(t^0 + T)$, as computed by Euler's method, converges to the exact result as $h \rightarrow 0$.
- e) Show that when $h \neq 0$, the length of \mathbf{y}^n grows (exponentially) without bound as $n \rightarrow \infty$! What happens to the length of the true solution as $t \rightarrow \infty$?
- f) Make a similar analysis for the differential equation (2.5). (Hint: find a particular solution, and then use the solution of the homogeneous equation to fit the initial conditions.)

2.2.2. Consider the differential equation

$$dx/dt = A + Bx + Cx^2, \quad (2.2.19)$$

which is a variant of the logistic/Verhulst differential equation. See (1.2.114). Solve this differential equation exactly.

Show that applying Euler's method to this differential equation produces the quadratic difference equation

$$x_{n+1} = x_n + hA + hBx_n + hC(x_n)^2, \quad (2.2.20)$$

which is a quadratic map of the form (1.2.114). Compare the behavior of the solutions of the differential equation (2.19) to that of the quadratic difference equation (2.20). Consider the cases of both small and large step size h . At what value of h does chaotic behavior set in? Chaotic behavior would be a bad thing because you should have found that the solutions to (2.19) are well behaved. How small must h be to avoid period doubling? Period doubling would also be a bad thing because you should have found that the solutions to (2.19) are not periodic.

2.3 Runge-Kutta Methods

2.3.1 Introduction

Now that we have the general idea, let us see what improvements can be made. The obvious need is to improve the accuracy of the stepping formula (2.3). One procedure would be to invoke the use of the first few additional derivatives. Higher derivatives are computable, and could in principle be used. For example, differentiating (1.1) and substituting it back into its derivative gives the result

$$\ddot{y}_i = \partial f_i / \partial t + \sum_j (\partial f_i / \partial y_j) \dot{y}_j$$

or

$$\ddot{y}_i = \partial f_i / \partial t + \sum_j (\partial f_i / \partial y_j) f_j. \quad (2.3.1)$$

This procedure can be effective for differential equations whose right sides are polynomial in the y_i . However it is evident that for most systems of differential equations the expressions for the higher derivatives become quite lengthy, and their use may be a bit cumbersome. What would be delightful is a stepping procedure that only involves evaluations of \mathbf{f} .

2.3.2 Procedure

Such procedures were originally studied by *Carl Runge* (1856-1927) and *Martin Kutta* (1867-1944), and now generally bear the name Runge-Kutta.⁴ Many are available, and we shall be only able to quote a few without derivation. The general idea is to evaluate \mathbf{f} at several different points and to add the results together in such a way that \mathbf{y}^{n+1} is correctly estimated up to some error that is proportional to a large power of h , and thus quite small. (Exactly what points to use in evaluating \mathbf{f} and how to weight the results is a complicated matter. We refer the interested reader to Exercises 3.1 and 3.10 through 3.12, and then to the references.) A method called RK3, that makes *local* errors only of order h^4 , i.e. is locally correct through order h^3 , is given by

$$\mathbf{y}^{n+1} = \mathbf{y}^n + \frac{1}{6}(\mathbf{a} + 4\mathbf{b} + \mathbf{c}), \quad (2.3.2)$$

where at each step

$$\begin{aligned} \mathbf{a} &= h\mathbf{f}(\mathbf{y}^n, t^n), \\ \mathbf{b} &= h\mathbf{f}(\mathbf{y}^n + \frac{1}{2}\mathbf{a}, t^n + \frac{1}{2}h), \\ \mathbf{c} &= h\mathbf{f}(\mathbf{y}^n + 2\mathbf{b} - \mathbf{a}, t^n + h). \end{aligned} \quad (2.3.3)$$

Higher-order methods are also available. The higher the order, of course, the more work is involved. One of several fourth-order methods, and called RK4, is given by

$$\begin{aligned} \mathbf{a} &= h\mathbf{f}(\mathbf{y}^n, t^n), \\ \mathbf{b} &= h\mathbf{f}(\mathbf{y}^n + \frac{1}{2}\mathbf{a}, t^n + \frac{1}{2}h), \\ \mathbf{c} &= h\mathbf{f}(\mathbf{y}^n + \frac{1}{2}\mathbf{b}, t^n + \frac{1}{2}h), \end{aligned} \quad (2.3.4)$$

⁴Ernest Courant, who co-invented the use of matrices to approximate transfer maps as described in Section 1.1.2, is the son of the mathematician Richard Courant of Courant and Hilbert and Courant Institute fame. Ernest Courant's mother, Nerina Runge, was a daughter of Runge, and thus Ernest Courant is also a grandson of Runge. Runge was very athletic, and entertained his grandchildren at his 70th birthday by doing handstands, which Ernest Courant remembers. Runge also had a second daughter Iris who, despite the many serious obstacles facing women in her time, excelled to become a mathematician/physicist and worked with Arnold Sommerfeld. Runge also had a son Wilhelm who was an early developer of radar. There were good genes in that family. Runge was a student of Weierstrass and Kummer, and the doctoral advisor of Max Born. Wikipedia and other information about Martin Kutta is also available on the Web.

$$\begin{aligned}\mathbf{d} &= h\mathbf{f}(\mathbf{y}^n + \mathbf{c}, t^n + h), \\ \mathbf{y}^{n+1} &= \mathbf{y}^n + \frac{1}{6}(\mathbf{a} + 2\mathbf{b} + 2\mathbf{c} + \mathbf{d}).\end{aligned}\tag{2.3.5}$$

This method is locally correct through order h^4 , and makes *local* errors of order h^5 .⁵

The reader should note that when we say that either *local* or *cumulative* error is of order h^ℓ , we mean that it is proportional to h^ℓ with an unspecified constant of proportionality. Unfortunately, an analytic estimation of the proportionality constant for Runge-Kutta is very complicated. See, for example, Exercises 3.1 and 5.1.

To see the advantage of higher-order methods over (2.3), suppose we use the third-order method (3.2) to integrate from t^0 to $t^0 + T$. This time the *cumulative* error is proportional to $Nh^4 = Th^3$, which is an improvement over the earlier error by a factor of h^2 . Of course, each integration step now requires about three times as much work since \mathbf{f} must be evaluated three times for each step. But the integration error is reduced by considerably more than a factor of three. It is possible now to use a much larger step size thus actually *reducing* the total work required to remain within a specified error.

The error we have been discussing so far can in principle be made arbitrarily small by letting $h \rightarrow 0$ and $N \rightarrow \infty$. In actual practice using digital computers, the ideal is not quite realizable. This is because computers only work with a finite number of significant figures, and hence each step involves a certain unavoidable “round-off” error. If the sign of each round-off error is nearly random, their cumulative effect increases approximately as \sqrt{N} . (The IEEE hardware standard with regard to round-off procedures is designed with this goal in mind.) If the sign is systematic, their cumulative effect may grow like N . In any case, if N is made too large, not only the cost of computation increases. The total error, after reaching a certain minimum, also increases! To see how this can work out in a specific case, look over the next example and then study Figure 3.1.

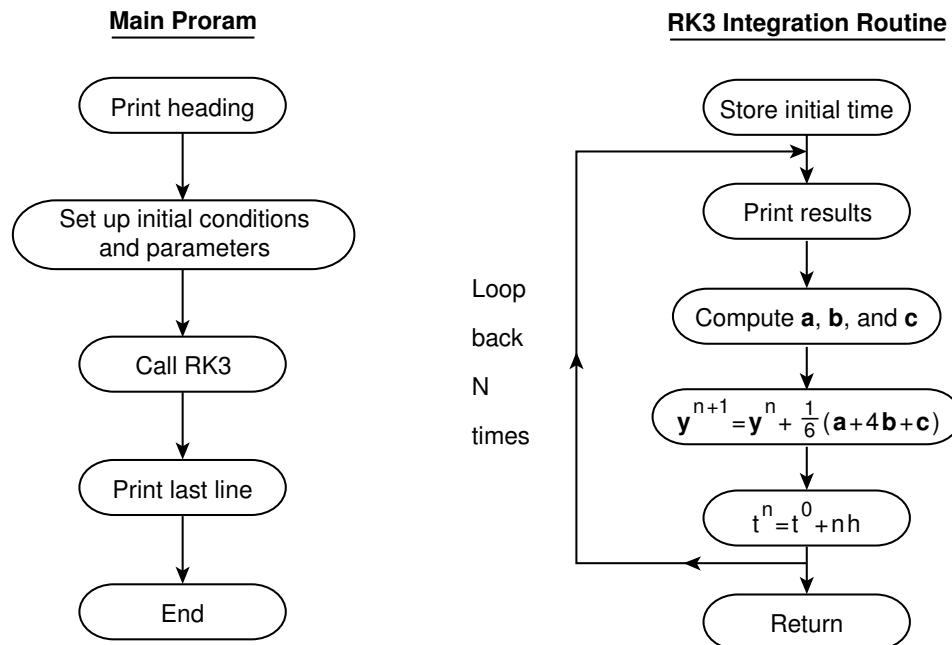
2.3.3 Numerical Example

We show below in Exhibit 3.1 a simple program that uses the third-order Runge-Kutta method (3.2) to integrate the problem of Section 2.2.2. The step size is again $h = 1/10$. We list only the main program and the subroutine RK3. The other subprograms are the same as those used in Section 2.2.2. Note that the numerical solution is now accurate to five significant figures.

In order to illustrate how the total cumulative error depends upon step size, we have also made calculations with other values of h . Figure 3.1 shows the results. Note that the cumulative error first decreases roughly as h^3 as expected, and then rises again because of round-off error.

⁵You will observe that we label a method by the order of the local accuracy. That is, an m th order method is locally correct through order h^m , and makes local errors of order h^{m+1} .

Exhibit 2.3.1: Third-Order Runge Kutta Integration



```

c This is the main program for illustrating a Runge Kutta method
c for numerical integration.
c
c Print heading.
c
      write(6,100)
100 format
     & (1h , 'time',4x,'y1comp',10x,'y2comp',10x,'y1true',10x,'y2true',/)
c
c Set up initial conditions and parameters. n is the number of integration
c steps we wish to make.
c
      t=0.
      h=.1
      n=15
      y1=0.
      y2=1.
c
      call rk3(t,h,n,y1,y2)
  
```

```
call prints(t,y1,y2,y1true(t),y2true(t),0)
c
end

c
c This is a third-order Runge Kutta integration subroutine.
c
subroutine rk3(t,h,n,y1,y2)
c
c Store initial time.
c
tint=t
c
c Printing and integration loop.
c
do 100 i=1,n
call prints(t,y1,y2,y1true(t),y2true(t),0)
c
c Set up for integration step.
c
call eval(y1,y2,t,f1,f2)
a1=h*f1
a2=h*f2
y1t=y1+a1/2.
y2t=y2+a2/2.
tt=t+h/2.
call eval(y1t,y2t,tt,f1,f2)
b1=h*f1
b2=h*f2
y1t=y1+2.*b1-a1
y2t=y2+2.*b2-a2
tt=t+h
call eval(y1t,y2t,tt,f1,f2)
c1=h*f1
c2=h*f2

c
c Make integration step and update time.
c
y1=y1+(a1+4.*b1+c1)/6.
y2=y2+(a2+4.*b2+c2)/6.
t=tint+float(i)*h
c
100 continue
c
return
end
```

Numerical Results

time	y1comp	y2comp	y1true	y2true
0.0000	0.00000000E+00	0.10000000E+01	0.00000000E+00	0.10000000E+01
0.1000	0.10016667E+00	0.10050000E+01	0.10016658E+00	0.10049958E+01
0.2000	0.20133168E+00	0.10199417E+01	0.20133068E+00	0.10199335E+01
0.3000	0.30448255E+00	0.10446757E+01	0.30447981E+00	0.10446635E+01
0.4000	0.41058692E+00	0.10789548E+01	0.41058168E+00	0.10789391E+01
0.5000	0.52058297E+00	0.11224364E+01	0.52057445E+00	0.11224174E+01
0.6000	0.63536996E+00	0.11746861E+01	0.63535756E+00	0.11746644E+01
0.7000	0.75579929E+00	0.12351816E+01	0.75578231E+00	0.12351578E+01
0.8000	0.88266593E+00	0.13033184E+01	0.88264394E+00	0.13032933E+01
0.9000	0.10167006E+01	0.13784157E+01	0.10166732E+01	0.13783901E+01
1.0000	0.11585623E+01	0.14597230E+01	0.11585290E+01	0.14596977E+01
1.1000	0.13088318E+01	0.15464280E+01	0.13087927E+01	0.15464039E+01
1.2000	0.14680060E+01	0.16376641E+01	0.14679611E+01	0.16376423E+01
1.3000	0.16364928E+01	0.17325199E+01	0.16364419E+01	0.17325013E+01
1.4000	0.18146069E+01	0.18300474E+01	0.18145503E+01	0.18300328E+01
1.5000	0.20025671E+01	0.19292722E+01	0.20025051E+01	0.19292628E+01

We close this subsection by remarking that the form of the Runge-Kutta program above was largely dictated by pedagogical considerations. A more compact version of this program using vector arrays and suitable for integrating any number of coupled equations is given in Appendix B. We commend this appendix to the reader who is considering more serious numerical work.

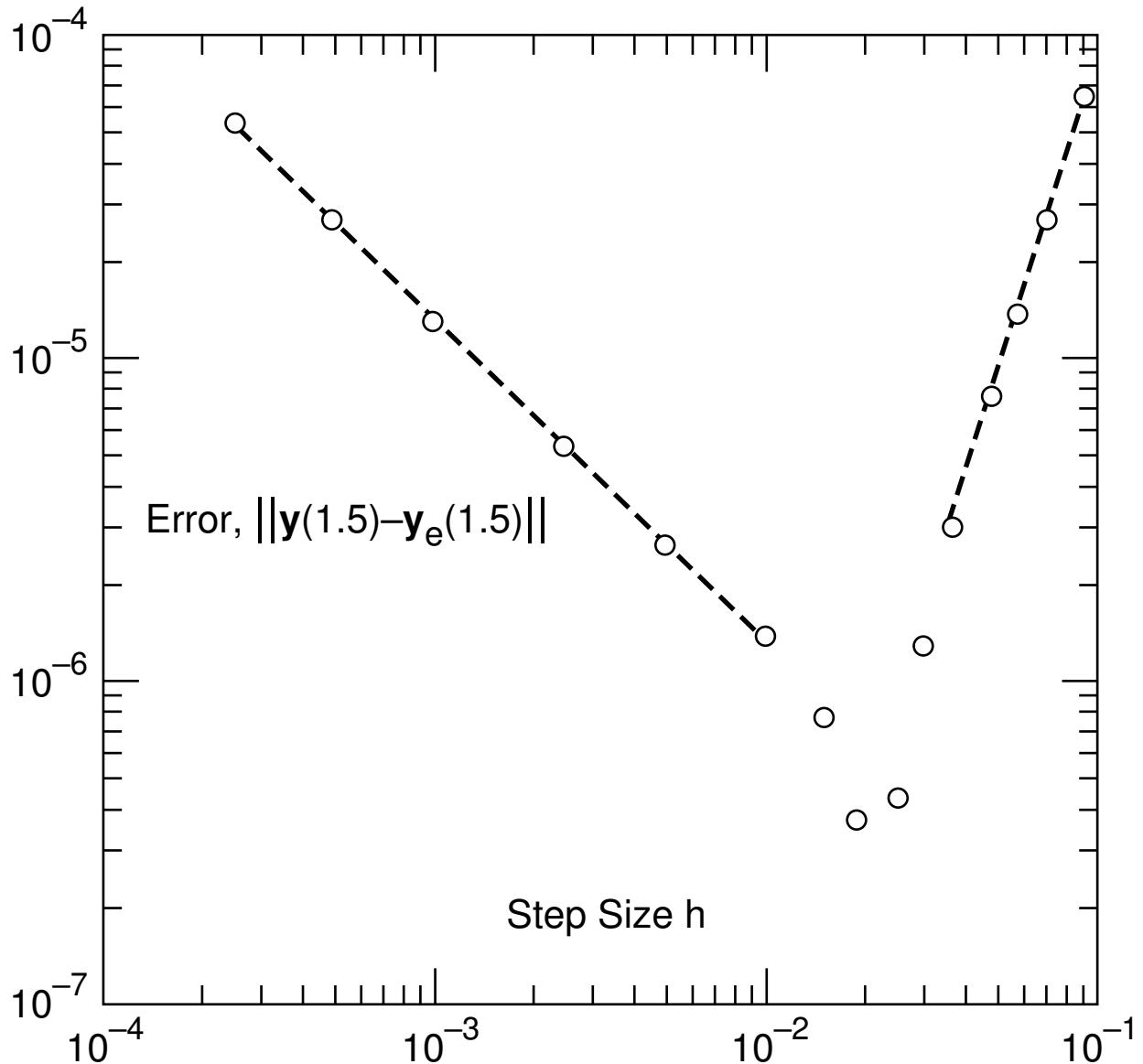


Figure 2.3.1: The result of integrating with RK3 the set (2.7) through (2.9) to $t = 1.5$ with several different step sizes to illustrate how the cumulative error depends on h . The error is measured by $\| \mathbf{y}(1.5) - \mathbf{y}_e(1.5) \|$ where \mathbf{y}_e is the exact solution. The dashed line on the right has a slope of $+3$ showing that the global truncation error at first decreases as h^3 . The dashed line on the left has a slope of -1 showing that in this example the global round-off error increases as the number of steps N . These calculations were made on a computer that had an accuracy of about $8\frac{1}{2}$ significant figures.

2.3.4 Nomenclature

Runge-Kutta methods have been studied extensively. In this subsection, as an aid to further reading, we will present briefly some of the nomenclature used to describe various Runge-Kutta concepts and methods.

Butcher Tableaux

Let b and c be s -dimensional vectors with real entries, and let a be an $s \times s$ matrix with real entries. Consider stepping formulas of the form

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h \sum_{i=1}^s b_i \mathbf{k}_i \quad (2.3.6)$$

where at each step

$$\mathbf{k}_i = \mathbf{f}(\mathbf{y}^n + h \sum_{j=1}^s a_{ij} \mathbf{k}_j, t^n + c_i h). \quad (2.3.7)$$

Observe that the integration methods RK3 and RK4 given by (3.2) through (3.5) are of this kind. The number s is called the number of *stages*. Observe that s is equal to the number of evaluations of the function \mathbf{f} required to compute the \mathbf{k}_i and thereby carry out one integration step using (3.6).

Before continuing on, it is sometimes useful to rewrite the relations (3.6) and (3.7) in a somewhat different form. At each step introduce intermediate times t_i and coordinates \mathbf{y}_i by the rules

$$t_i = t^n + c_i h, \quad (2.3.8)$$

$$\mathbf{y}_i = \mathbf{y}^n + h \sum_{j=1}^s a_{ij} \mathbf{k}_j. \quad (2.3.9)$$

With this convention (3.7) can be rewritten in the form

$$\mathbf{k}_i = \mathbf{f}(\mathbf{y}_i, t_i). \quad (2.3.10)$$

Finally we copy (3.6) and place it last,

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h \sum_{i=1}^s b_i \mathbf{k}_i. \quad (2.3.11)$$

Evidently the relations (3.8) through (3.11) are equivalent to the relations (3.6) and (3.7), but in this expanded form it is clear that the \mathbf{k}_i are the values of \mathbf{f} at the intermediate points, and the stepping rule (3.11) resembles the rule (2.3) for crude Euler except that it involves a weighted sum of these \mathbf{f} values rather than a single \mathbf{f} value.

Continue on. The problem now is to impose various conditions on the vectors b and c and the matrix a so that the integration method will be of some particular order m , and perhaps have other desirable properties. For purposes of visualization, it is convenient to arrange the vectors b and c and the matrix a in a tableau, called a *Butcher* tableau after its author, as shown below:

$$\begin{array}{c|cccc} c_1 & a_{11} & \cdots & a_{1s} \\ \vdots & \vdots & & \vdots \\ c_s & a_{s1} & \cdots & a_{ss} \\ \hline b_1 & \cdots & b_s \end{array} \quad (2.3.12)$$

The Butcher tableau for RK3 is

$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1 & -1 & 2 & 0 \\ \hline & 1/6 & 4/6 & 1/6 \end{array} . \quad (2.3.13)$$

The Butcher tableau for RK4 (often called *classic* RK4) is

$$\begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ \hline & 1/6 & 2/6 & 2/6 & 1/6 \end{array} . \quad (2.3.14)$$

There is another possible fourth-order method, also known to Kutta, sometimes called the 3/8 rule.⁶ It is given by the Butcher tableau

$$\begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 0 & 0 \\ 2/3 & -1/3 & 1 & 0 & 0 \\ 1 & 1 & -1 & 1 & 0 \\ \hline & 1/8 & 3/8 & 3/8 & 1/8 \end{array} . \quad (2.3.15)$$

Two features should be noticed about the Butcher tableaux (3.13) through (3.15): The first is that the matrix a is *strictly lower triangular*. That is, all entries on or above the diagonal vanish. This feature makes these methods *explicit*. That is, each \mathbf{k}_i is computable in terms of the \mathbf{k}_j with $j < i$. Runge-Kutta methods without this property are called *implicit*.⁷ The second feature is that the vector c is related to the matrix a by the rule

$$c_i = \sum_{j=1}^s a_{ij}. \quad (2.3.16)$$

⁶Although both classic RK4 and the 3/8 rule are fourth order (make local errors of order h^5), it can be shown that the 3/8 rule is somewhat more accurate because its local error terms proportional to h^5 have smaller coefficients. However, even though both classic RK4 and the 3/8 rule require the same number of function evaluations per step (namely, 4), the 3/8 rule is somewhat slower because its matrix a is somewhat more dense than that for RK4. Therefore, see (3.9), more additions and multiplications are required per step for the 3/8 rule than for RK4.

⁷Explicit Runge-Kutta methods are sometimes called ERK methods; and implicit Runge-Kutta methods are sometimes referred to as IRK methods. In the same spirit, if the matrix a has strictly lower triangular entries plus some nonzero diagonal entries but no entries above the diagonal, then the associated integration methods are called diagonally implicit Runge-Kutta (DIRK).

Pictorially, each c_i is the sum of the a 's in its row. This relation is called the *consistency* condition and is, for convenience, generally required of all Runge-Kutta methods. See Exercise 3.10. Exercise 3.9 briefly describes what further *order* conditions are required to achieve local accuracy through orders $m = 1$, $m = 2$, and $m = 3$.

Relation Between Number of Stages and Achievable Order

It is tempting to conjecture that with s stages it should be possible to find an explicit Runge-Kutta method whose order m satisfies $m = s$. This conjecture is true for $m = 1, 2, 3$, and 4 , but it fails for $m \geq 5$. Table 3.1 below lists the minimum s value required to achieve order m with explicit Runge-Kutta methods. As can be seen, $s \geq 6$ is needed to achieve an explicit Runge-Kutta method with $m = 5$. Thus, there are diminishing returns in going beyond order 4, which gives fourth-order methods such as RK4 a preferred status.

Table 2.3.1: Minimum Number of Stages s Required for Explicit Runge-Kutta to Achieve Order m .

m	1	2	3	4	5	6	7	8
s	1	2	3	4	6	7	9	11

With *implicit* Runge-Kutta methods it is possible for the order to even *exceed* the number of stages. Consider the one-stage method specified by the Butcher tableau

Gauss2

$$\begin{array}{c|c} 1/2 & 1/2 \\ \hline & 1 \end{array}. \quad (2.3.17)$$

It corresponds to the *implicit midpoint rule*

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h \mathbf{f}[(\mathbf{y}^n + \mathbf{y}^{n+1})/2, t^n + h/2], \quad (2.3.18)$$

which is known to be of order 2.⁸ See Exercise 3.7. It is also related to Gaussian quadrature. See Subsection T.1.3. For this reason, and because it is second order, it is given the name Gauss2.

In fact, there are implicit Runge-Kutta methods for which $m = 2s$, and this order is the best that can be hoped for with s stages.⁹ Butcher tableaux for two such methods, for the cases of two and three stages and also based on Gaussian quadrature, are given below. They have orders 4 and 6, respectively.

⁸This stepping procedure, particularly in the context of partial differential equations, is also referred to as *Crank-Nicolson*.

⁹Strictly speaking, an s -stage explicit Runge-Kutta integrator requires s function evaluations per step. Implicit Runge-Kutta methods require many more since the implicit equations involved are generally solved by multiple iteration.

Gauss4

$$\begin{array}{c|cc} 1/2 - \sqrt{3}/6 & 1/4 & 1/4 - \sqrt{3}/6 \\ 1/2 + \sqrt{3}/6 & 1/4 + \sqrt{3}/6 & 1/4 \\ \hline & 1/2 & 1/2 \end{array}, \quad (2.3.19)$$

Gauss6

$$\begin{array}{c|ccc} 1/2 - \sqrt{15}/10 & 5/36 & 2/9 - \sqrt{15}/15 & 5/36 - \sqrt{15}/30 \\ 1/2 & 5/36 + \sqrt{15}/24 & 2/9 & 5/36 - \sqrt{15}/24 \\ 1/2 + \sqrt{15}/10 & 5/36 + \sqrt{15}/30 & 2/9 + \sqrt{15}/15 & 5/36 \\ \hline & 5/18 & 8/18 & 5/18 \end{array}. \quad (2.3.20)$$

Butcher tableaux for Gauss8 and Gauss10 are also available. See the book of *Sanz-Serna* and *Calvo* listed in the Bibliography for this chapter. For further discussion of implicit Runge-Kutta methods, see Section 12.4.

Interpolation and Dense Output

There are some situations, for example when graphical output is needed, in which one desires an accurate and efficient method for finding $\mathbf{y}(t^n + \theta h)$ for any $\theta \in [0, 1]$. There are procedures that prepare, at each integration step, polynomials in θ for this purpose, and these procedures utilize the \mathbf{k} vectors computed in the course of a Runge-Kutta step. See, for example, the book of *Hairer, Nørsett, and Wanner* cited at the end of this chapter.

First Same As Last

There is one final comment worth making. It is possible to construct Runge-Kutta methods for which the Butcher tableaux take the form

$$\begin{array}{c|cccccc} 0 & 0 & 0 & \cdots & 0 & 0 \\ c_2 & a_{2,1} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ c_{s-1} & a_{s-1,1} & a_{s-1,2} & \cdots & 0 & 0 \\ 1 & b_1 & b_2 & \cdots & b_{s-1} & 0 \\ \hline & b_1 & b_2 & \cdots & b_{s-1} & 0 \end{array} \quad (2.3.21)$$

Comparison of (3.21) with (3.12) shows that we have imposed the conditions

$$a_{ij} = 0 \text{ for } j \geq i, \quad (2.3.22)$$

$$a_{sj} = b_j, \quad (2.3.23)$$

$$b_s = 0, \quad (2.3.24)$$

$$c_s = 1. \quad (2.3.25)$$

The condition (3.22) makes the associated integration method explicit. The condition (3.24) must hold if (3.22) and (3.23) are to be enforced. The condition (3.25) follows from the

consistency condition (3.16) and the desire that the method be at least of order 1. See (3.42).

What is the virtue of the condition (3.23)? Let us compute \mathbf{k}_s when the Butcher tableau has the form (3.21) and we are making the integration step from $t = t^n$ to $t = t^{n+1}$. From (3.7), (3.22), and (3.23) we find the result

$$\begin{aligned}\mathbf{k}_s|_{t=t^n} &= \mathbf{f}(\mathbf{y}^n + h \sum_{j=1}^s a_{sj} \mathbf{k}_j, t^n + c_s h) \\ &= \mathbf{f}(\mathbf{y}^n + h \sum_{j=1}^s b_j \mathbf{k}_j, t^n + h) = \mathbf{f}(\mathbf{y}^{n+1}, t^{n+1}).\end{aligned}\quad (2.3.26)$$

Here we have used (3.6). Now let us compute \mathbf{k}_1 when the Butcher tableau has the form (3.21) and we are making the integration step from $t = t^{n+1}$ to $t = t^{n+2}$. From (3.7) and (3.22) we find the result

$$\mathbf{k}_1|_{t=t^{n+1}} = \mathbf{f}(\mathbf{y}^{n+1}, t^{n+1}). \quad (2.3.27)$$

We conclude that

$$\mathbf{k}_1|_{t=t^{n+1}} = \mathbf{k}_s|_{t=t^n}, \quad (2.3.28)$$

the *first* \mathbf{k} for a successive step is the same as the *last* \mathbf{k} from the previous step. For this reason, a Butcher tableau of the form (3.21) is said to have a First Same As Last (FSAL) structure. We see that for a FSAL Runge-Kutta method, once an initial integration step has been completed, successive steps only require $s - 1$ function evaluations and are therefore the method effectively has $s - 1$ stages.¹⁰ However, the price to be paid for FSAL turns out to be a reduction in order.

For example, the Butcher tableau

	0	0	0	0	0	0	0
$\frac{1}{5}$	$\frac{1}{5}$	0	0	0	0	0	0
$\frac{3}{10}$	$\frac{3}{40}$	$\frac{9}{40}$	0	0	0	0	0
$\frac{4}{5}$	$\frac{44}{45}$	$-\frac{56}{15}$	$\frac{32}{9}$	0	0	0	0
$\frac{8}{9}$	$\frac{19372}{6561}$	$-\frac{25360}{2187}$	$\frac{64448}{6561}$	$-\frac{212}{729}$	0	0	0
1	$\frac{9017}{3168}$	$-\frac{355}{33}$	$\frac{46732}{5247}$	$\frac{49}{176}$	$-\frac{5103}{18656}$	0	0
1	$\frac{35}{384}$	0	$\frac{500}{1113}$	$\frac{125}{192}$	$-\frac{2187}{6784}$	$\frac{11}{84}$	0
	$\frac{35}{384}$	0	$\frac{500}{1113}$	$\frac{125}{192}$	$-\frac{2187}{6784}$	$\frac{11}{84}$	0

¹⁰Note also that, because $b_s = 0$, the final operation (3.6) is also already carried out in the evaluation of \mathbf{k}_s , which results in an additional savings.

describes a FSAL Runge-Kutta method that has $s = 7$ stages but acts (with respect to effort) like a $7 - 1 = 6$ stage method after the first step since 6 function evaluations are required for each subsequent step. It is a 5th order method ($m = 5$). From Table 3.1 we see that this method has the optimal order that can be achieved with a 6 stage method, and has an order that is one less than the optimal order that can be achieved with a 7 stage method. Section 2.5.1 and Appendix B describe how this method can be used as part of an embedded Runge-Kutta pair called Dormand-Prince 5(4).

Discovery/Construction of Runge-Kutta Methods

Runge-Kutta methods, particularly those of high order, are difficult to discover/construct. Exercises 3.10 through 3.12 describe some methods for doing so if it is not required that the methods be explicit. The discovery of high-order explicit methods is much harder.

Exercises

2.3.1. Consider the second-order Runge-Kutta method (sometimes called the improved Euler method or the second-order *Heun* method)

$$\mathbf{a} = h\mathbf{f}(\mathbf{y}^n, t^n), \quad (2.3.30)$$

$$\begin{aligned} \mathbf{b} &= h\mathbf{f}(\mathbf{y}^n + \mathbf{a}, t^n + h), \\ \mathbf{y}^{n+1} &= \mathbf{y}^n + \frac{1}{2}(\mathbf{a} + \mathbf{b}). \end{aligned} \quad (2.3.31)$$

Verify that the *local* truncation error is of the form $\mathbf{e}h^3 + O(h^4)$ and find a formula for \mathbf{e} . Finding error estimates for Runge-Kutta methods is not easy! Hint: Use a Taylor series to write $\mathbf{y}_{\text{true}}^{n+1} = \mathbf{y}^n + h\dot{\mathbf{y}}^n + h^2\ddot{\mathbf{y}}^n/2! + h^3\ddot{\mathbf{y}}^n/3! + O(h^4)$. Now expand the Runge-Kutta formula in a Taylor series and compare terms. You should find the result

$$\mathbf{e} = -(\ddot{\mathbf{y}} - 3 \sum \ddot{y}_i \partial \mathbf{f} / \partial y_i) / (12). \quad (2.3.32)$$

2.3.2. Review Exercise 3.1. Consider the so called *explicit midpoint rule* Runge-Kutta method

$$\mathbf{a} = h\mathbf{f}(\mathbf{y}^n, t^n), \quad (2.3.33)$$

$$\begin{aligned} \mathbf{b} &= h\mathbf{f}(\mathbf{y}^n + \mathbf{a}/2, t^n + h/2), \\ \mathbf{y}^{n+1} &= \mathbf{y}^n + \mathbf{b}. \end{aligned} \quad (2.3.34)$$

Show that this method is also second order. That is, verify that the local truncation error is of the form $\mathbf{e}h^3 + O(h^4)$. Find a formula for \mathbf{e} .

2.3.3. Show that the Euler method (2.3) is a Runge-Kutta method with Butcher tableau

$$\begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array} \quad (2.3.35)$$

2.3.4. Show that the Runge-Kutta method of Exercise 3.1 above has the Butcher tableau

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & 1/2 & 1/2 \end{array} . \quad (2.3.36)$$

Show that the Runge-Kutta method of Exercise 3.2 above has the Butcher tableau

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1/2 & 1/2 & 0 \\ \hline & 0 & 1 \end{array} . \quad (2.3.37)$$

2.3.5. Verify that (3.13) and (3.14) are the Butcher tableaux for RK3 and RK4, respectively.

2.3.6. Show that the Runge-Kutta method with Butcher tableau

$$\begin{array}{c|c} 1 & 1 \\ \hline & 1 \end{array} \quad (2.3.38)$$

describes the rule

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h\mathbf{f}(\mathbf{y}^{n+1}, t^n + h) = \mathbf{y}^n + h\mathbf{f}(\mathbf{y}^{n+1}, t^{n+1}). \quad (2.3.39)$$

This method might properly be called the implicit endpoint rule, but is more commonly called backward Euler or, simply, implicit Euler. Verify that this method has order 1 and find an estimate for the local truncation error.

2.3.7. Show that the Butcher tableau (3.17) corresponds to the implicit midpoint rule (3.18). Review Exercises 3.1 and 3.2. Verify by direct computation of Taylor series that (3.18) is of order 2, and find an estimate for the local truncation error.

2.3.8. Show that the Butcher tableau

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1/2 & 1/2 \\ \hline & 1/2 & 1/2 \end{array} . \quad (2.3.40)$$

corresponds to the Runge-Kutta formula

$$\mathbf{y}^{n+1} = \mathbf{y}^n + (h/2)[\mathbf{f}(\mathbf{y}^n, t^n) + \mathbf{f}(\mathbf{y}^{n+1}, t^n + h)]. \quad (2.3.41)$$

This method is known as the *trapezoidal rule*. What is its order? It is interesting to note that the Butcher tableaux (3.36) and (3.40) have the same b_i and c_i , but different matrix parts a .

2.3.9. Verify, for RK3 and RK4, that there are the Butcher tableau relations

Order 1:

$$\sum_i b_i = 1, \quad (2.3.42)$$

Order 2:

$$\sum_i b_i c_i = 1/2, \quad (2.3.43)$$

Order 3:

$$\sum_i b_i c_i^2 = 1/3, \quad (2.3.44)$$

$$\sum_{ij} b_i a_{ij} c_j = 1/6. \quad (2.3.45)$$

These relations are called *order conditions*.

Verify that (3.42) is necessary for a Runge-Kutta method to at least be of order 1. It can be shown that (3.42) and (3.43) are necessary for a Runge-Kutta method to at least be of order 2. Finally, all the conditions (3.42) through (3.45) are required for a Runge-Kutta method to at least be of order 3.

Verify that (3.42) holds for the Butcher tableaux (3.35) and (3.38), but (3.43) does not. Verify that (3.42) and (3.43), but not (3.44) and (3.45), hold for the Butcher tableaux (3.17), (3.36), (3.37), and (3.40).

2.3.10. Runge-Kutta methods, particularly those of high order, are difficult to discover. To simplify the problem, it is convenient to begin with the autonomous case of differential equations of the form

$$\dot{\mathbf{z}} = \mathbf{g}(\mathbf{z}), \quad (2.3.46)$$

and search for stepping formulas of the form

$$\mathbf{z}^{n+1} = \mathbf{z}^n + h \sum_{i=1}^s b_i \boldsymbol{\ell}_i \quad (2.3.47)$$

where at each step

$$\boldsymbol{\ell}_i = \mathbf{g}(\mathbf{z}^n + h \sum_{j=1}^s a_{ij} \boldsymbol{\ell}_j). \quad (2.3.48)$$

In this case there is no vector c so that the Butcher tableau takes the simpler form

$$\begin{array}{c|ccc} & a_{11} & \cdots & a_{1s} \\ \vdots & & \vdots & . \\ & a_{s1} & \cdots & a_{ss} \\ \hline & b_1 & \cdots & b_s \end{array} \quad (2.3.49)$$

Suppose that such a Runge-Kutta method of some desired order has been found for the autonomous case. We will now see that it can be parlayed into a Runge-Kutta method of the same order for the non-autonomous case (1.1).

To accomplish this feat, we will convert (1.1), which is a set of m non-autonomous equations, into a set of $m+1$ autonomous differential equations of the form (3.46). We will

then apply the method of (3.49) to these equations thereby producing an associated method for (1.1).

With reference to the set (1.1), let τ be a new independent variable and treat t as a dependent variable by adding the differential equation

$$dt/d\tau = 1 \quad (2.3.50)$$

to the set. That is, introduce a new set of $(m + 1)$ variables \mathbf{z} by the rule

$$\text{first } m \text{ components of } \mathbf{z} = \text{first } m \text{ components of } \mathbf{y}, \quad (2.3.51)$$

$$(m + 1)^{\text{th}} \text{ component of } \mathbf{z} = t; \quad (2.3.52)$$

and define an $m + 1$ -dimensional vector of functions $\mathbf{g}(\mathbf{z})$ by the rule

$$\text{first } m \text{ components of } \mathbf{g}(\mathbf{z}) = \text{first } m \text{ components of } \mathbf{f}(\mathbf{y}, t), \quad (2.3.53)$$

$$(m + 1)^{\text{th}} \text{ component of } \mathbf{g}(\mathbf{z}) = 1. \quad (2.3.54)$$

So doing produces a set of $m + 1$ autonomous (τ independent) equations of the form (3.46) where a dot now indicates $d/d\tau$. A solution of this autonomous set, after making the identification $t = \tau$, evidently produces a solution of the non-autonomous set (1.1).

Let us now apply the method (3.49) to (3.46) and examine the values of t at each stage. By the construction (3.53) and (3.54), the $(m + 1)^{\text{th}}$ component of \mathbf{g} is always 1. Next, using (3.48), show that the $(m + 1)^{\text{th}}$ component of every ℓ_i is also 1. Conclude that the $(m + 1)^{\text{th}}$ component of the argument of \mathbf{g} in (3.48) will be

$$(m + 1)^{\text{th}} \text{ component of } (\mathbf{z}^n + h \sum_{j=1}^s a_{ij} \ell_j) = [(m + 1)^{\text{th}} \text{ component of } \mathbf{z}^n] + h \sum_{j=1}^s a_{ij}. \quad (2.3.55)$$

Moreover, if the integration method is at least of order 1, it will have integrated the equation (3.50) exactly so that

$$(m + 1)^{\text{th}} \text{ component of } \mathbf{z}^n = t^n. \quad (2.3.56)$$

Thus, verify the result

$$(m + 1)^{\text{th}} \text{ component of } (\mathbf{z}^n + h \sum_{j=1}^s a_{ij} \ell_j) = t^n + h \sum_{j=1}^s a_{ij}. \quad (2.3.57)$$

Verify that the corresponding temporal argument on the right side of (3.7) is $t^n + c_i h$. Therefore, for consistency, verify that there must be the relation

$$t^n + c_i h = t^n + h \sum_{j=1}^s a_{ij}, \quad (2.3.58)$$

from which the consistency condition (3.16) follows.

2.3.11. As already mentioned in Exercise 3.10, Runge-Kutta methods, particularly those of high order, are difficult to discover. The purpose of this exercise is to explore some of the relations between Runge-Kutta formulas and quadrature formulas. For a review of quadrature formulas, see Section T.1.

Suppose the general Runge-Kutta method given by (3.6) and (3.7) is applied to the differential equation (1.1) in the special case that the right side is *independent* of \mathbf{y} . That is, consider differential equations of the special form

$$\dot{\mathbf{y}} = \mathbf{g}(t). \quad (2.3.59)$$

Show that in this case the relations (3.7) become

$$\mathbf{k}_i = \mathbf{g}(t^n + c_i h), \quad (2.3.60)$$

and the relation (3.6) becomes the stepping rule

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h \sum_{i=1}^s b_i \mathbf{g}(t^n + c_i h). \quad (2.3.61)$$

Suppose further that $t^0 = 0$ and $\mathbf{y}^0 = 0$, and set $n = 0$ so that (3.61) takes the form

$$\mathbf{y}^1 = h \sum_{i=1}^s b_i \mathbf{g}(c_i h). \quad (2.3.62)$$

Finally, suppose that $\mathbf{g}(t)$ has the special form

$$\mathbf{g}(t) = \boldsymbol{\alpha}_\ell t^\ell \quad (2.3.63)$$

where $\boldsymbol{\alpha}_\ell$ is some fixed vector. Then (3.62) becomes

$$\mathbf{y}^1 = h \sum_{i=1}^s b_i \boldsymbol{\alpha}_\ell (c_i h)^\ell = h^{\ell+1} \boldsymbol{\alpha}_\ell \sum_{i=1}^s b_i (c_i)^\ell. \quad (2.3.64)$$

Next verify that the *exact* solution to (3.59), with $t^0 = 0$ and $\mathbf{y}^0 = 0$ and \mathbf{g} given by (3.63), is

$$\mathbf{y}_e(t) = \boldsymbol{\alpha}_\ell t^{\ell+1}/(\ell+1), \quad (2.3.65)$$

and therefore

$$\mathbf{y}_e^1 = \mathbf{y}_e(h) = \boldsymbol{\alpha}_\ell h^{\ell+1}/(\ell+1). \quad (2.3.66)$$

Upon comparing (3.64) and (3.66), to the extent that \mathbf{y}^1 and \mathbf{y}_e^1 are to agree, we see that we should explore the possibilities

$$\sum_{i=1}^s b_i (c_i)^\ell \stackrel{?}{=} 1/(\ell+1) \quad (2.3.67)$$

for various choices of the b_i and c_i and various values of ℓ . Evidently the conditions (3.67) are the conditions for a quadrature formula with the b_i playing the role of the weights w_i

and the c_i playing the role of the sampling points x_i . See equation (T.1.2) in Appendix T. Note that the order conditions (3.42) through (3.44) are special cases of (3.67).

Verify that the b_i and c_i for the RK3 method (3.13) are those for the Newton-Cotes Simpson's rule $1 - 4 - 1$ formula, see (T.1.15) and (T.1.16), and therefore (3.67) holds for $\ell = 1, 2, 3$ but not $\ell > 3$. Verify that RK3, although only third-order accurate for differential equations of the general form (1.1), is fourth-order accurate for differential equations of the special form (3.59). See (T.1.66).

Verify that the b_i and c_i for the fourth-order method (3.15) are those for the Newton-Cotes Simpson's $3/8$ rule, see (T.1.20) and (T.1.21), and therefore (3.67) holds for $\ell = 1, 2, 3$ but not $\ell > 3$. Verify that this method, which is fourth-order accurate for differential equations of the general form (1.1), is also fourth-order (and not still higher-order) accurate for differential equations of the special form (3.59). See (T.1.69).

What about the b_i and c_i for the classic RK4 method (3.14)? Verify that the b_i and c_i for this case are *not* those for a Newton-Cotes quadrature. Verify that in this case (3.62) becomes

$$\begin{aligned}\mathbf{y}(h) &= h \sum_{i=1}^s b_i \mathbf{f}(c_i h) \\ &= (1/6)\mathbf{f}(0) + (2/6)\mathbf{f}(h/2) + (2/6)\mathbf{f}(h/2) + (1/6)\mathbf{f}(h) \\ &= (1/6)\mathbf{f}(0) + (4/6)\mathbf{f}(h/2) + (1/6)\mathbf{f}(h).\end{aligned}\tag{2.3.68}$$

Show that the right side of (3.68) *is* the Newton-Cotes quadrature rule corresponding to Simpson's $1 - 4 - 1$ formula, and therefore (3.68) is accurate through order 4, as we would at least expect since classic RK4 is supposed to be fourth order. See (T.1.66). Correspondingly, verify that (3.67) holds for $\ell = 1, 2, 3$ but not $\ell > 3$.

What about the b_i and c_i for the Gaussian Runge-Kutta methods (3.17), (3.19), and (3.20)? Verify that for these Butcher tableaux the b_i and c_i satisfy (3.67) through the advertised order. See Subsection T.1.3.

Finally, verify that the b_i and c_i for the Butcher tableaux (3.36) and (3.40) correspond to $k = 2$ closed Newton Cotes.

2.3.12. This exercise is a continuation of Exercise 3.11, which you should read. We have found and explored conditions to be satisfied by the b_i and the c_i . What can be said about the remaining matrix a_{ij} in the Butcher tableau (3.12)?

We will not consider the general case, but will describe a specific case. There is a class of Runge-Kutta methods that arises from a concept called *collocation*. For these methods, collocation is used to provide a stepping rule from \mathbf{y}^n to \mathbf{y}^{n+1} . Remarkably, for these methods, there is a formula that specifies the matrix a_{ij} in terms of the coefficients c_i . This formula makes possible the construction of a class of Runge-Kutta methods of arbitrary order.

We now describe the use of collocation to provide a stepping rule. Select s *distinct* quantities c_i with $i = 1, 2, \dots, s$. Let $\mathbf{P}_n(t)$ be a vector-valued polynomial in t of degree s specified by the $s + 1$ requirements that

$$\mathbf{P}_n(t^n) = \mathbf{y}^n,\tag{2.3.69}$$

$$\dot{\mathbf{P}}_n(t^n + c_i h) = \mathbf{f}[\mathbf{P}_n(t^n + c_i h), t^n + c_i h], \quad i = 1, 2, \dots, s. \quad (2.3.70)$$

The points $t^n + c_i h$ are called collocation points. According to dictionaries, *collocation* is defined as the result of “arranging” together. Here we have required that the time derivative of $\mathbf{P}_n(t)$ and the value of \mathbf{f} be equal at the collocation points. Since a polynomial of degree s requires $s+1$ conditions for its specification, we have indeed specified $\mathbf{P}_n(t)$.

Moreover since, according to (3.69) and (3.70), $\mathbf{P}_n(t)$ satisfies $s+1$ relations that are also satisfied by $\mathbf{y}(t)$, we expect that $\mathbf{P}_n(t)$ will be a good approximation to $\mathbf{y}(t)$. We therefore make the stepping rule

$$\mathbf{y}^{n+1} = \mathbf{P}_n(t^n + h). \quad (2.3.71)$$

It can be shown that if s quantities c_i and their associated b_i can be found such that (3.67) is satisfied for all $\ell < m$ (but not $\ell = m$), then use of (3.69) through (3.71) is equivalent to an s -stage Runge-Kutta method having order m . In other words,

$$m = \ell_{\max} + 1. \quad (2.3.72)$$

The Butcher tableau for this method contains the b_i and c_i . Moreover, as will be described shortly, with a knowledge of the c_i there is a recipe for constructing the matrix entries a_{ij} in the Butcher tableau. Thus, there is a procedure for constructing a class of Runge-Kutta formulas of arbitrary order.

Before describing the recipe for the matrix entries a_{ij} , we pause to elaborate briefly on the connection between collocation and Runge-Kutta. Begin with the truism that

$$\mathbf{y}^{n+1} - \mathbf{y}^n = \int_{t^n}^{t^{n+1}} dt \dot{\mathbf{y}}(t) = \int_{t^n}^{t^{n+1}} dt \mathbf{f}[\mathbf{y}(t), t]. \quad (2.3.73)$$

Next estimate the right side of (3.73) using a quadrature formula that employs the points $t^n + c_i h$ as sampling points and the b_i as weights. So doing gives the approximation

$$\mathbf{y}^{n+1} \simeq \mathbf{y}^n + h \sum_{i=1}^s b_i \mathbf{f}[\mathbf{y}(t^n + c_i h), t^n + c_i h]. \quad (2.3.74)$$

It can be shown that there is the correspondence

$$\mathbf{k}_i \simeq \mathbf{f}[\mathbf{y}(t^n + c_i h), t^n + c_i h], \quad (2.3.75)$$

and therefore

$$\mathbf{y}^{n+1} \simeq \mathbf{y}^n + h \sum_{i=1}^s b_i \mathbf{k}_i, \quad (2.3.76)$$

in agreement with (3.6).

We now describe the recipe for constructing a full Butcher tableau in terms of the c_i and based on the collocation Ansatz.¹¹ We already know how to construct the b_i in terms of the c_i . Given the c_i , we form the associated Lagrange polynomials $L_i(x)$ and then integrate

¹¹Ansatz is a German word, actually a noun, which in Mathematics and Physics literature means an initial hypothesis to be verified by further work. Since in German nouns are always capitalized no matter where they appear in a sentence, Ansatz should be written with a capital A. Its plural is Ansätze.

them over the interval $[0, 1]$ to find the b_i . See (T.1.4) through (T.1.9). It can be shown that the matrix entries a_{ij} associated with the collocation Ansatz are also given in terms of integrals of Lagrange polynomials by the rule

$$a_{ij} = \int_0^{c_i} dx L_j(x). \quad (2.3.77)$$

Further work, based on the result (3.77), shows that equivalently the matrix entries a_{ij} can be found from the c_i by a matrix algorithm: First, define an $s \times s$ matrix u by the rule

$$u_{jk} = c_j^{k-1} \text{ with } j, k = 1, \dots, s. \quad (2.3.78)$$

Next define an $s \times s$ matrix v by the rule

$$v_{ik} = c_i^k / k \text{ with } i, k = 1, \dots, s. \quad (2.3.79)$$

Then the matrix a is given by the rule

$$a = vu^{-1}. \quad (2.3.80)$$

For a proof of all these results, see the book *Geometric Numerical Integration* by Hairer et al. cited in the Bibliography at the end of this chapter.

Evidently there are two possible complications in executing the instructions (3.78) and (3.80). First it could happen that some $c_j = 0$, in which case use of (3.78) will involve the ambiguous quantity 0^0 . Indeed, this could well occur because $c_1 = 0$ for closed Newton Cotes. However, since x^0 is taken to represent the function $g(x) = 1$, we should make the choice

$$0^0 = g(0) = 1. \quad (2.3.81)$$

Second, one must verify that the matrix u has an inverse, which is equivalent to the condition

$$\det(u) \neq 0. \quad (2.3.82)$$

The c_j violate this condition if they are not all distinct. Note that the c_j for the RK4 Butcher tableau (3.14) have $c_2 = c_3$. Here we require that the c_j be distinct, and it can be shown that this condition is sufficient to guarantee the existence of u^{-1} .¹²

In summary, it can be shown that the quantities c_i , b_i , and a_{ij} , with the b_i constructed from the c_i using (T.1.5) and (T.1.9) and the matrix a given by (3.77) or (3.80), produce a Runge-Kutta method of order m provided (3.67) is satisfied for all $\ell < m$ (but not $\ell = m$). Do m values for this procedure, which according to (3.72) and (T.1.11) may be as large as $2s$, violate the claim of Table 3.1? The answer is *no* because the Runge-Kutta methods produced in this way are *implicit*.

We emphasize, of course, that not all Runge-Kutta methods are provided by this construction. In particular, the explicit Runge-Kutta methods fall outside this class.

Your task in this exercise is to use the matrix algorithm just described to construct the Butcher tableaux (3.17) for Gauss2, (3.40) for the trapezoidal rule, and (3.19) for Gauss4.

¹²Verify that $\det(u)$ is a *Vandermonde* determinant. See (17.2.23) and (17.2.29).

Consider first the $s = 1$ case of Gauss2 given by (3.17). In this case both u and v are 1×1 matrices. For b_1 and c_1 we use the values $b_1 = 1$ and $c_1 = 1/2$, which corresponds to the use of $k = 1$ Legendre Gauss. In this case (3.67) is satisfied for $\ell = 0$ and $\ell = 1$, but not $\ell = 2$. Thus we expect the method to have order $m = 2$. For $c_1 = 1/2$ show that

$$u_{11} = c_1^0 = (1/2)^0 = 1, \quad (2.3.83)$$

$$v_{11} = c_1^1 = (1/2)^1 = 1/2, \quad (2.3.84)$$

from which it follows that

$$a_{11} = v_{11}/u_{11} = 1/2, \quad (2.3.85)$$

in accord with the matrix entry in (3.17).

Consider next the $s = 2$ case of the trapezoidal rule given by (3.40). Since $s = 2$, we expect that u and v will be 2×2 . Suppose we use $k = 2$ Newton Cotes for which $b_1 = b_2 = 1/2$, $c_1 = 0$, and $c_2 = 1$. In this case we have $\ell_{\max} = 1$, see Table T.1.1, and we expect $m=2$. Verify the results

$$u_{11} = c_1^0 = 0^0 = 1, \quad (2.3.86)$$

$$u_{12} = c_1^1 = 0^1 = 0, \quad (2.3.87)$$

$$u_{21} = c_2^0 = 1^0 = 1, \quad (2.3.88)$$

$$u_{22} = c_2^1 = 1^1 = 1, \quad (2.3.89)$$

and therefore

$$u = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}. \quad (2.3.90)$$

Similarly, show that

$$v = \begin{pmatrix} 0 & 0 \\ 1 & 1/2 \end{pmatrix}. \quad (2.3.91)$$

Use these results to show that

$$a = vu^{-1} = \begin{pmatrix} 0 & 0 \\ 1/2 & 1/2 \end{pmatrix}, \quad (2.3.92)$$

in agreement with the matrix part of the Butcher tableau (3.40). Verify that the integrals (3.77) for the a_{ij} are easily evaluated in this case again yielding the result (3.92).

Your last challenge is to consider the $s = 2$ case of Gauss4 given by (3.19). In this case u and v are again 2×2 matrices. For the b_i and c_i choose values associated with $k = 2$ Legendre Gauss. That is, make the choice

$$(b_1, b_2) = (1/2, 1/2), \quad (2.3.93)$$

$$(c_1, c_2) = (1/2 - \sqrt{3}/6, 1/2 + \sqrt{3}/6). \quad (2.3.94)$$

For this choice (3.67) is satisfied for $\ell = 0, 1, 2$, and 3, but not $\ell = 4$; and we expect the order to be $m = 4$. Verify the results

$$u_{11} = c_1^0 = (1/2 - \sqrt{3}/6)^0 = 1, \quad (2.3.95)$$

$$u_{12} = c_1^1 = 1/2 - \sqrt{3}/6 \quad (2.3.96)$$

$$u_{21} = c_2^0 = (1/2 + \sqrt{3}/6)^0 = 1, \quad (2.3.97)$$

$$u_{22} = c_2^1 = 1/2 + \sqrt{3}/6, \quad (2.3.98)$$

and therefore

$$u = \begin{pmatrix} 1 & 1/2 - \sqrt{3}/6 \\ 1 & 1/2 + \sqrt{3}/6 \end{pmatrix}. \quad (2.3.99)$$

Also verify the results

$$v_{11} = c_1^1 = 1/2 - \sqrt{3}/6, \quad (2.3.100)$$

$$v_{12} = c_1^2/2 = (1/2 - \sqrt{3}/6)^2/2 = 1/6 - \sqrt{3}/12, \quad (2.3.101)$$

$$v_{21} = c_2^1 = 1/2 + \sqrt{3}/6, \quad (2.3.102)$$

$$v_{22} = c_2^2/2 = (1/2 + \sqrt{3}/6)^2/2 = 1/6 + \sqrt{3}/12, \quad (2.3.103)$$

and therefore

$$v = \begin{pmatrix} 1/2 - \sqrt{3}/6 & 1/6 - \sqrt{3}/12 \\ 1/2 + \sqrt{3}/6 & 1/6 + \sqrt{3}/12 \end{pmatrix}. \quad (2.3.104)$$

Next verify that

$$u^{-1} = \sqrt{3} \begin{pmatrix} 1/2 + \sqrt{3}/6 & -1/2 + \sqrt{3}/6 \\ -1 & 1 \end{pmatrix}. \quad (2.3.105)$$

Finally, show that

$$a = vu^{-1} = \begin{pmatrix} 1/4 & 1/4 - \sqrt{3}/6 \\ 1/4 + \sqrt{3}/6 & 1/4 \end{pmatrix}, \quad (2.3.106)$$

which agrees with the matrix part of (3.19).

2.4 Finite-Difference/Multistep/Multivalue Methods

2.4.1 Background

Motivation and Terminology

In Runge-Kutta methods, one essentially begins anew at each step, and (apart from the \mathbf{y} value that is already at hand) disregards any previously obtained information about the trajectory under study. Methods with this property are called *single-step* methods. This is fine, of course, when one is *beginning* a solution since all one has then is the initial conditions. However, once the integration is sufficiently underway, it clearly would be advantageous to make use of some of the “information” contained in previously obtained points. We now explore how this may be done.

Suppose we are willing to store results from $k = N + 1$ previous integration steps where N is an integer.¹³ That is, we are willing to store k previous successive values of \mathbf{y}^ℓ and k previous successive values of \mathbf{f}^ℓ . (Generally N ranges from 3 to 10. For purposes of

¹³Warning! The symbol N in this context has a different meaning than in Sections 2 and 3.

the present discussion, N is selected *once and for all*, and then held *fixed* throughout the integration run.) With these values at hand, we consider a relation of the form

$$\begin{aligned} \alpha_{N+1}\mathbf{y}^{n+1} + \alpha_N\mathbf{y}^n + \alpha_{N-1}\mathbf{y}^{n-1} + \cdots + \alpha_0\mathbf{y}^{n-N} &= \\ h(\beta_{N+1}\mathbf{f}^{n+1} + \beta_N\mathbf{f}^n + \beta_{N-1}\mathbf{f}^{n-1} + \cdots + \beta_0\mathbf{f}^{n-N}) \end{aligned} \quad (2.4.1)$$

which we rewrite in the (marching-order) form

$$\begin{aligned} \mathbf{y}^{n+1} &= -\alpha_N\mathbf{y}^n - \alpha_{N-1}\mathbf{y}^{n-1} - \cdots - \alpha_0\mathbf{y}^{n-N} \\ &\quad + h(\beta_{N+1}\mathbf{f}^{n+1} + \beta_N\mathbf{f}^n + \beta_{N-1}\mathbf{f}^{n-1} + \cdots + \beta_0\mathbf{f}^{n-N}). \end{aligned} \quad (2.4.2)$$

Here, without loss of generality, we have rescaled the α_ℓ and the β_ℓ so that $\alpha_{N+1} = 1$. The formula (4.2), with *fixed* h independent coefficients, is to be used to determine \mathbf{y}^{n+1} from the stored $\mathbf{y}^n \dots \mathbf{y}^{n-N}$ and the stored $\mathbf{f}^n \dots \mathbf{f}^{n-N}$. It is explicit if $\beta_{N+1} = 0$, and implicit otherwise. Methods of the form (4.2) are called *multistep* methods since they employ information from $k = N + 1$ previous steps. More precisely, methods of the form (4.2) are called *k-step* methods. They are also called *multivalue* methods since (4.2) involves k previous values of \mathbf{y}^ℓ and the k previous values of \mathbf{f}^ℓ .¹⁴ Sometimes they are also called *linear-multistep* or *linear-multivalue* methods since the relation (4.2) involves a *linear* combination of the \mathbf{y}^ℓ and the \mathbf{f}^ℓ . Finally, they are also called *finite-difference* methods because they can often be conveniently formulated in terms of finite differences.

Maximum Order

Suppose the coefficients in (4.2) are selected to obtain the highest possible local accuracy. What local accuracy can we hope to achieve? Imagine that \mathbf{y} is expanded in a Taylor series about $t = t^n$ and this Taylor series is used to determine $\mathbf{y}^{n+1} = \mathbf{y}(t^n + h)$. If this series is to be accurate through terms of order h^m , it must contain $m + 1$ terms since it begins with the constant term \mathbf{y}^n . On the other hand, we have $2k$ pieces of information available in the explicit case, and $2k + 1$ pieces of information in the implicit case. We therefore might hope, in the explicit case, to achieve a maximal local accuracy m_{\max} given by

$$m_{\max} = 2k - 1, \text{ explicit case}; \quad (2.4.3)$$

and, in the implicit case, a maximum local accuracy of

$$m_{\max} = 2k, \text{ implicit case}. \quad (2.4.4)$$

For example there is the $N = 1$, and therefore $k = 2$, two-step explicit formula

$$\mathbf{y}^{n+1} = -4\mathbf{y}^n + 5\mathbf{y}^{n-1} + 4h\mathbf{f}^n + 2h\mathbf{f}^{n-1}, \quad (2.4.5)$$

and the two-step implicit formula

$$\mathbf{y}^{n+1} = \mathbf{y}^{n-1} + (h/3)\mathbf{f}^{n+1} + (4h/3)\mathbf{f}^n + (h/3)\mathbf{f}^{n-1}. \quad (2.4.6)$$

¹⁴Some authors use the term *multivalue* to refer to the jet formulation described in Subsection 5.3.

Suppose we stipulate that the monomial

$$\mathbf{y}(t) = \mathbf{a}t^j \quad (2.4.7)$$

(where \mathbf{a} is a constant vector) be the exact solution to (1.1), from which it follows that

$$\mathbf{f}(\mathbf{y}, t) = j\mathbf{a}t^{j-1}. \quad (2.4.8)$$

Upon inserting (4.7) and (4.8) into (4.5) it is easily verified that (4.5) holds exactly for $j = 0, 1, 2, 3$ and fails to be exact for $j \geq 4$. The formula (4.5) is therefore locally accurate through terms of order h^3 , which according to (4.3) is the highest order that might be expected in the explicit case. Indeed, it is easy to verify that (4.5) is the unique explicit two-step formula having third-order accuracy. Similarly, it can be verified that (4.6) is exact for $j = 0, 1, 2, 3, 4$ and fails for $j \geq 5$. The formula (4.6) is therefore locally accurate through terms of order h^4 , which according to (4.4) is the highest order that might be expected in the implicit case.

Stability

At this point we can make a simple observation. Consider the polynomial $\rho(\zeta)$, which (for reasons that will become clear later) we call the *stability* polynomial, defined by the rule

$$\rho(\zeta) = \sum_{j=0}^{N+1} \alpha_j \zeta^j = \zeta^{N+1} + \sum_{j=0}^N \alpha_j \zeta^j. \quad (2.4.9)$$

Suppose that the marching rule (4.2) is exact for the monomial (4.7) with $j = 0$, in which case $\mathbf{f} = 0$. That is, we impose the requirement that (4.2) at least integrate the constant function $\mathbf{y} = \mathbf{a}$ exactly so that the Ansatz $\mathbf{y}^\ell = \mathbf{a}$ and $\mathbf{f}^\ell = 0$ satisfies (4.2) exactly. (This requirement is called *consistency* of order zero.) Doing so evidently yields the result

$$\mathbf{a} = -\mathbf{a}(\alpha_N + \alpha_{N-1} + \cdots + \alpha_0) \quad (2.4.10)$$

from which it follows that

$$1 + \sum_{j=0}^N \alpha_j = \rho(1) = 0. \quad (2.4.11)$$

Thus, the stability polynomial must have $\zeta = 1$ as a root for the method (4.2) to even be of minimal interest. In particular, if the method (4.2) has $m_{\max} \geq 1$, which are the cases of actual interest, then (4.11) must be satisfied. At this point, for convenient subsequent use and in analogy to (4.9), we also define a polynomial $\sigma(\zeta)$ by the rule

$$\sigma(\zeta) = \sum_{j=0}^{N+1} \beta_j \zeta^j. \quad (2.4.12)$$

To gain further insight into possible properties of multistep methods, let us now examine the use of the specific procedure (4.5) in more detail. Suppose it is used to integrate the scalar differential equation

$$\dot{y} = f(y, t) = \lambda y \quad (2.4.13)$$

with the initial condition

$$y(0) = 1. \quad (2.4.14)$$

(We suppose $t^0 = 0$.) The exact solution in this case is evidently

$$y(t) = \exp(\lambda t). \quad (2.4.15)$$

Let us study how the solution to the marching orders (4.5) approximates this exact solution.

For the case (4.13) we have $f^\ell = \lambda y^\ell$, and therefore the marching orders (4.5) become

$$y^{n+1} = -4y^n + 5y^{n-1} + 4h\lambda y^n + 2h\lambda y^{n-1} = (-4 + 4h\lambda)y^n + (5 + 2h\lambda)y^{n-1}. \quad (2.4.16)$$

Observe that (4.16) is a linear recursion relation. To solve it, try the Ansatz

$$y^n \propto (\zeta)^n \quad (2.4.17)$$

where the quantity ζ is to be determined and the notation is meant to be suggestive. The Ansatz (4.17), when inserted into (4.16), yields the *characteristic* equation

$$\zeta^2 + (4 - 4h\lambda)\zeta - (5 + 2h\lambda) = 0. \quad (2.4.18)$$

It follows that (4.16) has a general solution of the form

$$y^n = A[\zeta_1(h)]^n + B[\zeta_2(h)]^n \quad (2.4.19)$$

where ζ_1 and ζ_2 are the roots of (4.18), and the solution is made specific by selecting the coefficients A, B so that the conditions

$$y^0 = 1 \quad (2.4.20)$$

and

$$y^{-1} = \exp(-h\lambda) \quad (2.4.21)$$

are satisfied.

The roots of (4.18) are

$$\zeta = -2 + 2h\lambda \pm \sqrt{[9 - 6h\lambda + 4(h\lambda)^2]}, \quad (2.4.22)$$

and they have the expansions

$$\begin{aligned} \zeta_1(h) &= -2 + 2h\lambda + \sqrt{[9 - 6h\lambda + 4(h\lambda)^2]} \\ &= 1 + (h\lambda) + (h\lambda)^2/2! + (h\lambda)^3/3! + (h\lambda)^4/72 + \dots \\ &= \exp(h\lambda) + O(h^4), \end{aligned} \quad (2.4.23)$$

$$\begin{aligned} \zeta_2(h) &= -2 + 2h\lambda - \sqrt{[9 - 6h\lambda + 4(h\lambda)^2]} \\ &= -5 + 3(h\lambda) + O(h^2). \end{aligned} \quad (2.4.24)$$

Note from (4.23) that, as $h \rightarrow 0$, the root ζ_1 becomes $\zeta_1 = 1$. That $\zeta = 1$ is a root in this limit is to be expected: From (4.2) we see that the characteristic equation (4.18) can be written in the form

$$\rho(\zeta) - h\lambda\sigma(\zeta) = 0. \quad (2.4.25)$$

In the limit $h = 0$ the characteristic equation written as (4.25) becomes the relation

$$\rho(\zeta) = 0, \quad (2.4.26)$$

and we know from our previous discussion that $\zeta = 1$ is root of (4.26) since the method (4.5) has $m_{\max} = 3$ and therefore $m_{\max} \geq 1$.

Suppose we set $A = 1$ and $B = 0$ in (4.19). Then (4.20) is satisfied exactly, and from (4.23) we see that (4.21) is satisfied through terms of order h^3 . Moreover, in this case (4.19) can be rewritten in the form

$$y^n = (\zeta_1)^n = \exp[n \log(\zeta_1)] = \exp\{n[h\lambda + O(h^4)]\} = \exp(\lambda t^n) \exp[nO(h^4)]. \quad (2.4.27)$$

And, if we follow the marching orders to the time $t^n = T$ so that $n = T/h$, we obtain the result

$$y(T) = \exp(\lambda T) \exp[(T/h)O(h^4)] = \exp(\lambda T) \exp[TO(h^3)]. \quad (2.4.28)$$

Evidently, as comparison with (4.15) reveals, (4.28) becomes exact in the limit $h \rightarrow 0$.

Suppose instead we require (4.20) as before, but now require that (4.21) hold exactly. This would seem to be desirable because (4.20) and (4.21) are properties of the exact solution (4.15). Then we find the relations

$$A + B = 1, \quad (2.4.29)$$

$$A[\exp(-h\lambda) + O(h^4)] + B[-5 + O(h)]^{-1} = \exp(-h\lambda), \quad (2.4.30)$$

from which it follows that

$$A = 1 + O(h^4), \quad (2.4.31)$$

$$B = O(h^4) \neq 0. \quad (2.4.32)$$

Correspondingly, (4.19) becomes

$$y^n = A(\zeta_1)^n + B[-5 + O(h)]^n. \quad (2.4.33)$$

And, if we now if we follow the marching orders to the time T , we find the result

$$y(T) = [1 + O(h^3)] \exp(\lambda T) + O(h^4)(-5)^{T/h}. \quad (2.4.34)$$

We see that the first term in (4.34) becomes the exact solution in the limit $h \rightarrow 0$, but the second oscillates wildly with ever growing amplitude as $h \rightarrow 0$. For this reason, the method (4.5) is called *unstable*. Although the factor B in the second term of (4.33) and (4.34) vanishes as h^4 when $h \rightarrow 0$, the second factor grows (in amplitude) very rapidly because $|\zeta_2| > 1$. And this rapid growth dominates the vanishing of B so that their product also grows rapidly. On the other hand if it had happened that $|\zeta_2| < 1$, which might be the case for some other integration procedure, then both factors would vanish as $h \rightarrow 0$ so that only the first term would remain thereby producing the exact result for $y(T)$.

What have we learned from this example? First, the characteristic equation must have a root that is near $+1$, and this root produces a “desired” solution of the marching orders that approximates the exact solution of the associated differential equation. We will call this root the *good* root. In addition there are other roots, $k - 1$ in number because the characteristic equation is a polynomial of degree k , that produce other solutions. These solutions are called *parasitic* solutions. If their associated roots, which we will call parasitic roots, lie outside the unit circle in the complex plane, these solutions grow without bound and can eventually swamp the true solution. Finally, the nature of the roots can be found for small h by examining the roots of the stability polynomial $\rho(\zeta)$.

We conclude that a multistep method is generally of little interest if any roots of the stability polynomial $\rho(\zeta)$ lie outside the unit circle. A multistep method is defined to be *strongly stable* if its $\rho(\zeta)$ has $+1$ as a root and all other roots lie *inside* the unit circle. In general, unless a multistep procedure is initiated “just right”, some or all of the parasitic solutions will also be present in the result. Also, even when the procedure is initiated “just right”, the parasitic solutions will continually be “excited” during the march due to round-off errors. But if a method is strongly stable and h is small enough, then the parasitic-solution roots of the characteristic equation will lie within the unit circle and the parasitic solutions will decay to zero thereby leaving behind only the desired solution as $h \rightarrow 0$.

The First Dahlquist Barrier

What is the maximum local order m_{\max} that can be achieved with a strongly stable k -step method? It can be shown that if strong stability is required, then there is the result

$$m_{\max} = k, \text{ explicit case; } \quad (2.4.35)$$

$$m_{\max} = k + 2, \text{ implicit case and } k \text{ even, } \quad (2.4.36)$$

$$m_{\max} = k + 1, \text{ implicit case and } k \text{ odd. } \quad (2.4.37)$$

This limit is called the *first Dahlquist barrier*.¹⁵ A common practice is to employ order $m = k$ methods for the explicit case and order $m = k$ or $m = k + 1$ methods for the implicit case.

Strictly speaking, the order given by (4.36) cannot be reached unless all the roots of $\rho(\zeta)$ are on the unit circle, in which case it can be arranged that they are all distinct. By our definition, methods with this property are not strongly stable, but rather are a borderline case. However, they may be useful in some circumstances. In general, the first Dahlquist barrier for the implicit case, for both k even and k odd, is $m_{\max} = k + 1$.

Convergence

Again speaking strictly, our discussion of convergence so far holds for differential equations of the form (4.13). However, it can be proved that if a multistep method has local accuracy through terms of order h^m with $m \geq 1$ and is strongly stable, then the result of following the marching orders from $t = t^0$ to $t = t^0 + T$ converges to the exact result for $\mathbf{y}(t^0 + T)$

¹⁵There is also a *second* Dahlquist barrier that arises in the integration of so-called *stiff* equations by implicit methods. Their treatment is beyond the scope of this text.

as $h \rightarrow 0$ for any differential equation provided $\mathbf{f}(\mathbf{y}, t)$ has sufficiently many continuous derivatives and the stored starting values are exact.

We also note, still strictly speaking, that the concept of a characteristic equation applies only to cases of linear differential equations of the general form (4.13). However, if a method cannot integrate (4.13) well, then it is unlikely to be able to integrate more complicated nonlinear equations well.

2.4.2 Adams' Method

Suppose in (4.2) we set

$$\alpha_N = -1 \quad (2.4.38)$$

and

$$\alpha_\ell = 0 \text{ for } \ell = 0, 1, \dots, N-1. \quad (2.4.39)$$

In this case the stability polynomial becomes

$$\rho(\zeta) = \zeta^k - \zeta^{k-1} = (\zeta - 1)\zeta^{k-1}, \quad (2.4.40)$$

which evidently has the single root $\zeta_1 = 1$ and the multiple roots $\zeta_\ell = 0$ for $\ell = 2, 3, \dots, k$. This would seem to be a highly desirable state of affairs because with this choice for the α_ℓ all the parasitic roots of ρ vanish, and one might hope correspondingly that all the parasitic roots of the characteristic equation would be well within the unit circle providing h is not too large.

Upon taking into account the Ansatz specified by (4.38) and (4.39), the marching orders (4.2) take the form

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h(\beta_{N+1}\mathbf{f}^{n+1} + \beta_N\mathbf{f}^n + \beta_{N-1}\mathbf{f}^{n-1} + \dots + \beta_0\mathbf{f}^{n-N}).$$

The remaining task is to chose the β_ℓ in such a way that the order is maximized to bring it as close to the first Dahlquist barrier as is conveniently possible. The stepping methods thereby obtained are variously associated with the names *modified Adams*, *Adams-Basforth*, or *Adams-Moulton*. We shall simply call them *Adams*.

Because the derivation of Adams' method is fairly involved, we shall begin our discussion by describing the procedure for its use. Then, with the procedure well understood, we will give the derivations that justify the method. For convenience of description, we will assume the required stored starting values are obtain using Runge Kutta executed with a sufficiently small step size.

As in the cases of Crude Euler and Runge-Kutta, we begin with an initial vector \mathbf{y}^0 , and our task is to compute the successive vectors $\mathbf{y}^1, \mathbf{y}^2$ etc. The procedure for Adams' method is as follows:

1. Adams' method requires the storage of information about previously obtained points on a trajectory. In particular, since (4.38) and (4.39) hold but in general $\beta_\ell \neq 0$, it requires storage of the values $\mathbf{f}(\mathbf{y}, t)$ at these points and the most recent value of \mathbf{y} . As described earlier, let $N + 1$, where N is an integer, be the number of points whose “ \mathbf{f} ” values we are willing to store. For purposes of our present discussion, it is selected

once and for all, and then held fixed throughout the integration run. Thus, there is actually a whole family of Adams' methods with each member of the family having a different N . As we expect and will see later, the choice of N governs the accuracy of the method.

2. Using a Runge-Kutta method, compute the vectors $\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^N$ starting with \mathbf{y}^0 . At each point \mathbf{y}^n compute the vector $\mathbf{f}^n = \mathbf{f}(\mathbf{y}^n, t^n)$, and store the $N + 1$ vectors $\mathbf{f}^0, \mathbf{f}^1, \dots, \mathbf{f}^N$ as well as \mathbf{y}^N . Since the accuracy of these “ f ” values greatly affects the accuracy of the solution to be obtained later on, it is worth spending considerable effort on their accurate computation. One simple method is to run Runge-Kutta with a fractional step size h/m , where m is an integer, and then use every m th Runge-Kutta step for computing the desired \mathbf{y} 's and \mathbf{f} 's.
3. We are ready to switch to Adams' method. It consists of two stepping formulas called the *predictor* and the *corrector*. The predictor formula for marching from \mathbf{y}^n to \mathbf{y}^{n+1} reads

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h \sum_{k=0}^N \tilde{b}_k^N \mathbf{f}^{n-k}. \quad (\text{predictor})$$

It is an *explicit* formula. Here the \tilde{b}_k^N are a set of coefficients whose values will be derived and tabulated later on. Now, using the predictor formula and the stored \mathbf{f} 's, compute \mathbf{y}^{N+1} by putting $n = N$. This step is called *Predicting*, or P for short, and its result is called the predicted value of \mathbf{y}^{N+1} . An Adams' predictor formula is sometimes called an Adams-Basforth formula.

4. Using \mathbf{y}^{N+1} , compute $\mathbf{f}^{N+1} = \mathbf{f}(\mathbf{y}^{N+1}, t^{N+1})$. This step is called *Evaluating*, or E for short, since it requires an evaluation of the function \mathbf{f} .
5. The corrector formula reads

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h \sum_{k=0}^N \tilde{a}_k^N \mathbf{f}^{n+1-k}, \quad (\text{corrector})$$

where the \tilde{a}_k^N are another set of coefficients. It is an *implicit* formula and will be solved by iteration. Using the corrector formula, the stored \mathbf{f} 's, and \mathbf{f}^{N+1} from step 4, recompute \mathbf{y}^{N+1} by putting $n = N$ in the corrector formula. This step is called *Correcting*, or C for short, since, as we will later see, the corrector formula is more accurate. Its result is called the corrected value of \mathbf{y}^{N+1} . An Adams' corrector formula is sometimes called an Adams-Moulton formula.

6. Return to step 4, this time using the corrected value of \mathbf{y}^{N+1} . Repeat steps 4 and 5 until successive values of \mathbf{y}^{N+1} differ by less than some preassigned amount (usually the round-off accuracy of the computer). It can be shown that this iteration procedure converges if the step size h is small enough. Indeed, the operation EC can be shown to be a *contraction map*, and the operation P provides a first guess for the fixed point of

this contraction map.¹⁶ In actual practice the sequence *PECEC* is usually sufficient. A need for more iterations generally indicates a too large step size.

7. The procedure is finished. We have found \mathbf{y}^{N+1} . Now update the table of \mathbf{f} 's by adding to it the value of \mathbf{f}^{N+1} obtained from the last evaluation step, and discarding \mathbf{f}^0 .
8. To compute \mathbf{y}^{N+2} , relabel the \mathbf{y} 's and \mathbf{f} 's, and return to step 3. In this manner, proceed to compute $\mathbf{y}^{N+2}, \mathbf{y}^{N+3}$, etc. until the integration run is completed. Note again that, in each case, only the previous value of \mathbf{y} and the last $N + 1$ values of the \mathbf{f} 's are used.

2.4.3 Numerical Example

We show below in Exhibit 4.1 a program that illustrates the use of Adams' method with $N = 4$ for the differential equation set (2.7) through (2.9). Subsequently we will learn that $N = 4$ Adams is of order 5. That is, it is locally exact through terms of order h^5 and makes local errors of order h^6 . See (4.37) and (4.38). Therefore, it might appropriately be called Adams5.

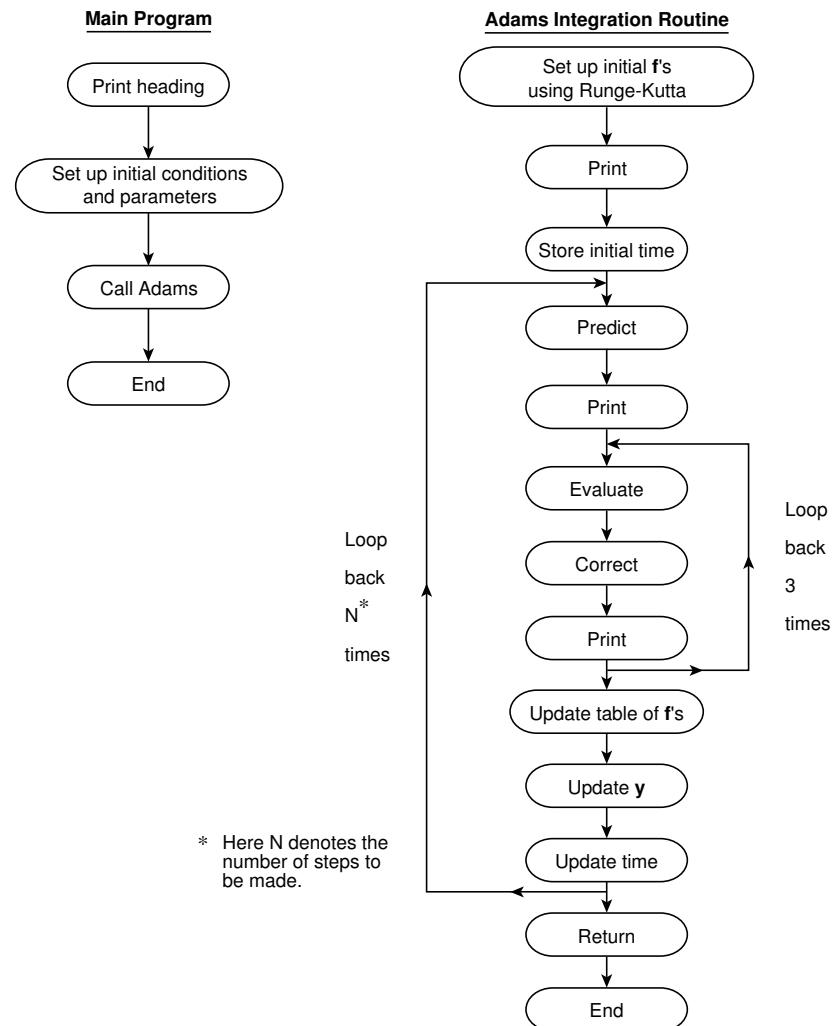
The time step is again $h = 1/10$. The program is written in double precision so that round-off errors are unimportant for this step size. That is, for pedagogical simplicity, we want to avoid the need to worry about round-off errors for this example. Runge-Kutta integration, with a step size of $h/10 = 1/100$, is used as a starting procedure.

The first values in the columns labeled *y1comp* and *y2comp* printed for each time in the Adams' routine are those found by the predictor. The next three lines are the result of successive corrector iterations. That is, we have used the sequence *PECECEC*. The convergence is good, and the sequence *PECEC* would have been sufficient. *Note that the solution is now accurate to almost eight significant figures.* A more efficient version of this program using vector arrays is given in Appendix B.

In passing, let us compare the accuracy of RK3 and Adams5, and the effort involved in each, for this simple example. From Exhibit 3.1 we saw that RK3 had an accuracy (with a step size $h = 1/10$) of five significant figures. And, according to (3.2) and (3.3), three function evaluations were required per step. By contrast, with the same step size, Adams5 has an accuracy of almost eight significant figures. And, when *PECEC* is used, only two function evaluations are required per step. Thus Adams5 is considerably more accurate and involves less effort than RK3.

¹⁶For a discussion of contraction maps, see the first paragraph of Section 29.4.3 and the references at the end of Chapter 29.

Exhibit 2.4.1: Fifth-Order Adams Integration



Computer Programs

```

c This is the main program for illustrating an Adams method
c for numerical integration.
c
      implicit double precision (a-h,o-z)
c
c Print heading.
c
      write(6,100)
100 format
  & (1h , 'time',4x,'y1comp',10x,'y2comp',10x,'y1true',
  & 10x,'y2true',/)
c
c Set up initial conditions and parameters. n is the number of integration
c steps we wish to make.
c
      t=0.d0
      h=.1d0
      n=15
      y1=0.d0
      y2=1.d0
c
      call adams(t,h,n,y1,y2)
c
      end
c
c This is a fifth-order Adams integration subroutine.
c
      subroutine adams(t,h,n,y1,y2)
      implicit double precision (a-h,o-z)
      dimension f1(5),f2(5)
c
      write(6,*) 'Starting with Runge-Kutta integration'
c
c Set up initial f values.
c
      call eval(y1,y2,t,f1(1),f2(1))
      call prints(t,y1,y2,y1true(t),y2true(t),0)
      do 10 i=2,5
      call rk3(t,h/10.d0,10,y1,y2)
      call eval(y1,y2,t,f1(i),f2(i))
      call prints(t,y1,y2,y1true(t),y2true(t),0)
10  continue
      write (6,*) 'Continuing with Adams integration'
      hdiv=h/720.d0
      n=n-4
      t=t+h
      tint=t
c
c Printing and integration loop.
c
      do 100 i=1,n
c

```

```

c Predictor step.
c
    p1=y1+hdiv*(1901.d0*f1(5)-2774.d0*f1(4)+2616.d0*f1(3)
& -1274.d0*f1(2)+251.d0*f1(1))
    p2=y2+hdiv*(1901.d0*f2(5)-2774.d0*f2(4)+2616.d0*f2(3)
& -1274.d0*f2(2)+251.d0*f2(1))
c
    call prints(t,p1,p2,y1true(t),y2true(t),0)
c
c Corrector steps.
c
    do 50 j=1,3
        call eval(p1,p2,t,g1,g2)
        c1=y1+hdiv*(251.d0*g1+646.d0*f1(5)-264.d0*f1(4)
& +106.d0*f1(3)-19.d0*f1(2))
        c2=y2+hdiv*(251.d0*g2+646.d0*f2(5)-264.d0*f2(4)
& +106.d0*f2(3)-19.d0*f2(2))
        p1=c1
        p2=c2
        call prints(t,c1,c2,0.,0.,1)
50 continue
c
c Update
c
    do 75 j=1,4
        f1(j)=f1(j+1)
        f2(j)=f2(j+1)
75 continue
    f1(5)=g1
    f2(5)=g2
    y1=c1
    y2=c2
    t=tint+float(i)*h
c
    100 continue
c
    return
end

```

Numerical Results

time	y1comp	y2comp	y1true	y2true
Starting with Runge-Kutta integration				
0.0000	0.00000000E+00	0.10000000E+01	0.00000000E+00	0.10000000E+01
0.1000	0.10016658E+00	0.10049958E+01	0.10016658E+00	0.10049958E+01
0.2000	0.20133067E+00	0.10199334E+01	0.20133067E+00	0.10199334E+01
0.3000	0.30447980E+00	0.10446635E+01	0.30447979E+00	0.10446635E+01
0.4000	0.41058166E+00	0.10789390E+01	0.41058166E+00	0.10789390E+01
Continuing with Adams integration				
0.5000	0.52057439E+00	0.11224171E+01	0.52057446E+00	0.11224174E+01
	0.52057446E+00	0.11224175E+01		
	0.52057448E+00	0.11224175E+01		
	0.52057448E+00	0.11224175E+01		

0.6000	0.63535744E+00	0.11746641E+01	0.63535753E+00	0.11746644E+01
	0.63535754E+00	0.11746644E+01		
	0.63535755E+00	0.11746644E+01		
	0.63535755E+00	0.11746644E+01		
0.7000	0.75578220E+00	0.12351576E+01	0.75578231E+00	0.12351578E+01
	0.75578234E+00	0.12351579E+01		
	0.75578235E+00	0.12351579E+01		
	0.75578235E+00	0.12351579E+01		
0.8000	0.88264379E+00	0.13032931E+01	0.88264391E+00	0.13032933E+01
	0.88264396E+00	0.13032934E+01		
	0.88264397E+00	0.13032934E+01		
	0.88264397E+00	0.13032934E+01		
0.9000	0.10166730E+01	0.13783898E+01	0.10166731E+01	0.13783900E+01
	0.10166732E+01	0.13783901E+01		
	0.10166732E+01	0.13783901E+01		
	0.10166732E+01	0.13783901E+01		
1.0000	0.11585289E+01	0.14596975E+01	0.11585290E+01	0.14596977E+01
	0.11585291E+01	0.14596978E+01		
	0.11585291E+01	0.14596978E+01		
	0.11585291E+01	0.14596978E+01		
1.1000	0.13087925E+01	0.15464037E+01	0.13087926E+01	0.15464039E+01
	0.13087928E+01	0.15464040E+01		
	0.13087928E+01	0.15464040E+01		
	0.13087928E+01	0.15464040E+01		
1.2000	0.14679608E+01	0.16376421E+01	0.14679609E+01	0.16376422E+01
	0.14679611E+01	0.16376423E+01		
	0.14679611E+01	0.16376423E+01		
	0.14679611E+01	0.16376423E+01		
1.3000	0.16364417E+01	0.17325011E+01	0.16364418E+01	0.17325012E+01
	0.16364420E+01	0.17325013E+01		
	0.16364420E+01	0.17325012E+01		
	0.16364420E+01	0.17325012E+01		
1.4000	0.18145501E+01	0.18300328E+01	0.18145503E+01	0.18300329E+01
	0.18145505E+01	0.18300329E+01		
	0.18145505E+01	0.18300329E+01		
	0.18145505E+01	0.18300329E+01		
1.5000	0.20025049E+01	0.19292627E+01	0.20025050E+01	0.19292628E+01
	0.20025052E+01	0.19292629E+01		
	0.20025052E+01	0.19292628E+01		
	0.20025052E+01	0.19292628E+01		

2.4.4 Derivation and Error Analysis

Calculus of Finite Differences

To reiterate, our remaining task is to choose the β_ℓ in such a way that the order is maximized to bring it as close to the first Dahlquist barrier as is conveniently possible. For this purpose it is useful to employ a *constructive* method based on the calculus of finite differences.

Let $\mathbf{y}(t)$ be any vector-valued function of t . We define a *backwards difference* operator ∇ by the rule

$$\nabla \mathbf{y}(t) = \mathbf{y}(t) - \mathbf{y}(t-h), \quad (2.4.41)$$

and in particular

$$\nabla \mathbf{y}^n = \mathbf{y}^n - \mathbf{y}^{n-1}. \quad (2.4.42)$$

Repeated applications of ∇ will be indicated by an exponent with the convention $\nabla^0 = 1$. Thus,

$$\nabla^2 \mathbf{y}^n = \nabla(\nabla \mathbf{y}^n) = \nabla \mathbf{y}^n - \nabla \mathbf{y}^{n-1} = \mathbf{y}^n - 2\mathbf{y}^{n-1} + \mathbf{y}^{n-2}, \quad (2.4.43)$$

and in general

$$\nabla^\ell \mathbf{y}^n = \sum_{k=0}^{\ell} (-1)^k \binom{\ell}{k} \mathbf{y}^{n-k}, \quad (2.4.44)$$

where the $\binom{\ell}{k}$ are the standard binomial coefficients.

Suppose $\mathbf{y}(t)$ is a polynomial in t with vector coefficients. Then it is easily checked that $\nabla \mathbf{y}$ is a polynomial of one order lower. We also have the relations

$$\nabla 1 = 0, \quad (2.4.45)$$

$$\nabla^k t^\ell = 0 \text{ if } k > \ell, \quad (2.4.46)$$

$$\nabla^\ell t^\ell = h^\ell \ell!, \quad (2.4.47)$$

where in this particular case t^ℓ denotes a power of t rather than the notation adopted in (1.2). Finally, we note that for \mathbf{y} polynomial in t , not only powers of ∇ are well defined; infinite series of the form $\sum_0^\infty a_k \nabla^k$ are also defined since by (4.46) the series must always terminate when applied to a polynomial.

From Taylor's theorem we know that

$$\mathbf{y}^{n-1} = \mathbf{y}(t^n - h) = \sum_{k=0}^{\infty} [(-h)^k / k!] (d^k \mathbf{y}^n / dt^k). \quad (2.4.48)$$

This relation can be written more compactly as

$$\mathbf{y}^{n-1} = e^{-hD} \mathbf{y}^n \quad (2.4.49)$$

where D denotes the differential operator

$$D = d/dt. \quad (2.4.50)$$

That is, if we expand e^{-hD} in a formal power series, we get (4.48). Combining (4.41) and (4.49), we find the result

$$\nabla \mathbf{y}^n = (1 - e^{-hD}) \mathbf{y}^n. \quad (2.4.51)$$

Watch closely! Since (4.51) is true for any \mathbf{y} whose functional dependence on t is polynomial, and since any continuous function can be approximated arbitrarily closely by polynomials, we can write the symbolic formula

$$\nabla = (1 - e^{-hD}) = hD - \frac{1}{2}h^2 D^2 + \dots. \quad (2.4.52)$$

Equation (4.52) should be viewed as a formal relation between two power series: one in ∇ and one in D . It becomes a true equation when applied to any polynomial. In this spirit, we may solve (4.52) for D to get the result

$$D = -h^{-1} \log(1 - \nabla). \quad (2.4.53)$$

Here $\log(1 - \nabla)$ denotes the formal series

$$\log(1 - \nabla) = - \sum_{k=1}^{\infty} \nabla^k / k. \quad (2.4.54)$$

Again, (4.53) becomes a true equation when applied to any polynomial.

Application of Finite Difference Calculus to Integration of Differential Equations

We now apply the calculus of difference operators we have just developed to the integration of differential equations. Observe that the differential equation under study, (1.1), can be written as

$$D \mathbf{y}^{n+1} = \mathbf{f}^{n+1}. \quad (2.4.55)$$

Suppose we knew how to *invert* the operator D . Then we might try writing

$$\mathbf{y}^{n+1} \stackrel{?}{=} D^{-1} \mathbf{f}^{n+1}. \quad (2.4.56)$$

However, we do not expect D^{-1} to be well defined since the inverse of differentiation is integration, and integration always involves the introduction of an arbitrary constant. This defect can be overcome by observing that the operator ∇D^{-1} is well defined since by (4.45) the operator ∇ annihilates any integration constant produced by D^{-1} . Thus we may convert (4.56) into the *integration formula*

$$\nabla \mathbf{y}^{n+1} = \nabla D^{-1} \mathbf{f}^{n+1}. \quad (2.4.57)$$

Now make another daring step. Since (4.53) and (4.54) express D as a formal series in ∇ , we might hope to get ∇D^{-1} as another series in ∇ by the operation of division. Assuming this is possible, use of (4.53) in (4.57) gives the result

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h[-\nabla / \log(1 - \nabla)] \mathbf{f}^{n+1}. \quad (2.4.58)$$

We shall verify shortly that the expression $[-\nabla / \log(1 - \nabla)]$ has a well-defined series expansion in ∇ so that (4.58) is formally correct, and actually true for polynomials. Before doing so, we continue on to derive another strange-looking expression. From the definition of ∇ we have the relation

$$\nabla \mathbf{f}^{n+1} = \mathbf{f}^{n+1} - \mathbf{f}^n. \quad (2.4.59)$$

Rearranging terms we find

$$\mathbf{f}^n = (1 - \nabla) \mathbf{f}^{n+1}. \quad (2.4.60)$$

Let us solve for \mathbf{f}^{n+1} . We have symbolically

$$\mathbf{f}^{n+1} = (1 - \nabla)^{-1} \mathbf{f}^n. \quad (2.4.61)$$

Now insert (4.61) into (4.58) to get the result

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h\{-\nabla / [(1 - \nabla) \log(1 - \nabla)]\} \mathbf{f}^n. \quad (2.4.62)$$

How are expressions such as (4.58) and (4.62) to be understood? Consider the functions $F(z)$ and $G(z)$ defined by

$$F(z) = -z / \log(1 - z), \quad (2.4.63)$$

$$G(z) = -z / [(1 - z) \log(1 - z)]. \quad (2.4.64)$$

Near $z = 0$ we know that $\log(1 - z) = -z + O(z^2)$ so that $F(0)$ and $G(0)$ are well defined. Furthermore, $\log(1 - z)$ and $(1 - z)^{-1}$ are singular only when $z = 1$. We conclude that F and G are analytic in the complex z plane within the unit disk $|z| < 1$. Consequently, we may write the series expansions

$$F(z) = \sum_0^\infty a_k z^k, \quad (2.4.65)$$

$$G(z) = \sum_0^\infty b_k z^k. \quad (2.4.66)$$

The first few coefficients are listed in Table 4.1 below. The ratio $|b_k/a_k|$ is also roughly tabulated for later use. The answer to our question is now clear. We use the series (4.65) and (4.66) to define the expressions in question, and in so doing obtain relations that are true for arbitrary polynomials.

Table 2.4.1: Expansion Coefficients for F and G .

k	0	1	2	3	4	5	6	7	8	9
a_k	1	$-\frac{1}{2}$	$-\frac{1}{12}$	$-\frac{1}{24}$	$-\frac{19}{720}$	$-\frac{3}{160}$	$-\frac{863}{60480}$	$-\frac{275}{24192}$	$-\frac{33953}{3628800}$	$-\frac{8183}{1036800}$
b_k	1	$\frac{1}{2}$	$\frac{5}{12}$	$\frac{3}{8}$	$\frac{251}{720}$	$\frac{95}{288}$	$\frac{19087}{60480}$	$\frac{5257}{17280}$	$\frac{1070017}{3628800}$	$\frac{25713}{89600}$
$ b_k/a_k $	1	1	5	9	~ 13	~ 17	~ 22	~ 27	~ 32	~ 36

Predictor-Corrector Formulas

With this brief explanation, we return to the problem of numerical integration. Suppose we wish to proceed from \mathbf{y}^n to \mathbf{y}^{n+1} on the basis of a polynomial fit in t of order $N + 1$. That is, $\mathbf{y}(t)$ is approximated by a polynomial of order $N + 1$, and we are willing to tolerate local errors of order h^{N+2} . Since $\mathbf{f} = \dot{\mathbf{y}}$, \mathbf{f} will be a polynomial of order N , and according to (4.46) we need to retain only N^{th} and lower differences. Thus, we may replace (4.58) and (4.62) by the two *truncated* formulas

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h \sum_0^N a_k \nabla^k \mathbf{f}^{n+1}, \quad (\text{corrector}) \quad (2.4.67)$$

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h \sum_0^N b_k \nabla^k \mathbf{f}^n. \quad (\text{predictor}) \quad (2.4.68)$$

As the reader may have guessed, we have given the formulas the names *corrector* and *predictor* in anticipation of their use. We may also write (4.67) and (4.68) in the expanded form

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h \sum_0^N \tilde{a}_k^N \mathbf{f}^{n+1-k}, \quad (\text{corrector}) \quad (2.4.69)$$

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h \sum_0^N \tilde{b}_k^N \mathbf{f}^{n-k} \quad (\text{predictor}) \quad (2.4.70)$$

where the coefficients $\tilde{a}_k^N, \tilde{b}_k^N$ are related to the earlier set a_k, b_k using (4.44). The coefficients \tilde{a}_k^N and \tilde{b}_k^N are listed in Tables 4.2 and 4.3 at the end of this section. Note that these coefficients depend on both k and N .

Error Analysis

Both formulas (4.67) and (4.68) are correct through terms of order h^{N+1} . However, in general the *truncation errors* involved in the *corrector* (4.67) are numerically smaller than those in the *predictor* (4.68). To see this, suppose that $\mathbf{y}(t)$ is approximated *exactly* by a polynomial of order $N + 2$,

$$\mathbf{y}(t) = \sum_0^{N+2} \mathbf{c}_j (t - t^n)^j. \quad (2.4.71)$$

Then, using a corrector series with upper summation index $(N+1)$, we would have the exact result

$$\mathbf{y}_{\text{true}}^{n+1} = \mathbf{y}^n + h \sum_0^{N+1} a_k \nabla^k \mathbf{f}^{n+1}. \quad (2.4.72)$$

[Note that in general the summation index in (4.72) should extend to infinity. However, because of the assumption (4.71), it may be cut off as indicated.] By contrast, using (4.67),

the actual corrector gives the approximate result

$$\mathbf{y}_{\text{corr}}^{n+1} = \mathbf{y}^n + h \sum_0^N a_k \nabla^k \mathbf{f}^{n+1}. \quad (2.4.73)$$

Upon subtracting the two results, we find the relation

$$\mathbf{y}_{\text{true}}^{n+1} - \mathbf{y}_{\text{corr}}^{n+1} = h a_{N+1} \nabla^{N+1} \mathbf{f}^{n+1}. \quad (2.4.74)$$

The right side of (4.74) is easily evaluated using $\mathbf{f} = \dot{\mathbf{y}}$, (4.71), (4.46), and (4.47). The result is the relation

$$\mathbf{y}_{\text{true}}^{n+1} - \mathbf{y}_{\text{corr}}^{n+1} = h^{N+2} a_{N+1} (N+2)! \mathbf{c}_{N+2}. \quad (2.4.75)$$

Finally, we observe that

$$(N+2)! \mathbf{c}_{N+2} = (d^{N+2} \mathbf{y} / dt^{N+2}), \quad (2.4.76)$$

so that the local error involved in using the corrector formula is given by

$$\mathbf{y}_{\text{true}}^{n+1} - \mathbf{y}_{\text{corr}}^{n+1} \approx h^{N+2} a_{N+1} (d^{N+2} \mathbf{y} / dt^{N+2})|_{t=t^n}. \quad (2.4.77)$$

Similarly, the predictor formula local error is given by

$$\mathbf{y}_{\text{true}}^{n+1} - \mathbf{y}_{\text{pred}}^{n+1} \approx h^{N+2} b_{N+1} (d^{N+2} \mathbf{y} / dt^{N+2})|_{t=t^n}. \quad (2.4.78)$$

Equations (4.77) and (4.78) are exact for polynomials of order $N+2$, and approximate otherwise. Now look at Table 1. We see that, for $N > 2$, a_N is considerably smaller than b_N and therefore the corrector formula has higher accuracy.

Since the corrector formula is more accurate, why did we bother to develop a predictor formula? The answer is that (4.67), as is evident from its expanded form (4.69), is an *implicit* or *closed* formula. That is, to employ it to compute \mathbf{y}^{n+1} , we need to know \mathbf{f}^{n+1} which itself depends on \mathbf{y}^{n+1} ! By contrast the predictor formula, although less accurate, is an *explicit* or *open* formula since we already presume to know the vectors \mathbf{f}^n back through \mathbf{f}^{n-N} from previous integration steps.

Finally, we note that the local orders described by (4.77) and (4.78) are close to the maximum order consistent with the first Dahlquist barrier. See Exercise 4.13.

Recapitulation of Adams' Method

At this point the reader should return to the first part of this section to review once again the procedure for Adams' method. He or she will see that it exploits the explicit nature of the predictor and the higher accuracy of the corrector by the following ingenious strategy:

- (a) (Step 3.) Suppose the vectors \mathbf{y}^n and $\mathbf{f}^n, \mathbf{f}^{n-1}, \dots, \mathbf{f}^{n-N}$ are known. Use formula (4.70) to *predict* a preliminary value for \mathbf{y}^{n+1} .
- (b) (Step 4.) Insert this \mathbf{y}^{n+1} and t^{n+1} into $\mathbf{f}(\mathbf{y}, t)$ to *evaluate* \mathbf{f}^{n+1} .
- (c) (Step 5.) With the \mathbf{f}^{n+1} thus obtained, recompute or *correct* \mathbf{y}^{n+1} using formula (4.69).

- (d) (Step 6.) Return to (b) and repeat (b) and (c) until convergence is achieved. Generally (see the discussion at the beginning of this section), the sequence *PECEC* should be sufficient. The net result is a value for \mathbf{y}^{n+1} that is correct within a local error given roughly by (4.77).
- (e) (Steps 7 and 8.) Update the table of \mathbf{f} 's, and go back to part (a) to compute \mathbf{y}^{n+2} , etc.

This strategy is often called the *predictor-corrector* method. Let us summarize what has been accomplished. Using (4.70) as a predictor and (4.69) as a corrector, we are able to compute \mathbf{y}^{n+1} through order h^{N+1} by generally making two and at most three computations (evaluations) of \mathbf{f} plus some simple additions. (That is, *PECEC* or at worst *PECECEC* should be sufficient. In practice it is common to use just *PECE*, and there are theoretical reasons to believe that it is best to end with an *E* operation.) All that is required is the storage of the previous $N + 1$ values $\mathbf{f}^n \cdots \mathbf{f}^{n-N}$ and the value \mathbf{y}^n . By contrast, the Runge-Kutta method (3.2) involves three evaluations of \mathbf{f} , and is correct only through order h^3 . Higher order Runge-Kutta schemes involve correspondingly more computations of \mathbf{f} . Since \mathbf{f} is usually a complicated function of \mathbf{y} and t , multiple computations of \mathbf{f} are generally made at the expense of considerable machine time and round-off error. We conclude that finite-difference methods give much higher accuracy for much less work, and are generally to be preferred once a solution is underway. There is, however, a caveat that makes the matter not quite so simple. One might be tempted, with a high order finite-difference method, to increase the step size in order to gain speed. That is, one might hope to trade accuracy for speed. However, as described in Subsection 7.3, finite-difference methods can become unstable if the step size is too large.¹⁷ See also Exercise 4.14. Consequently, Runge-Kutta may be preferable if only low accuracy is required, while finite-difference methods win if high accuracy is required.

¹⁷That is, if h is too large, some parasitic roots of the characteristic equation may lie outside the unit circle even when Adams is used to integrate equations of the form (4.13). And if these equations cannot be integrated well, there is doubt that more complicated nonlinear equations can be integrated well.

Table 2.4.2: The Adams' Corrector Coefficients \tilde{a}_l^N .

$k \ N$	2	3	4	5	6	7	8	9
0	$\frac{5}{12}$	$\frac{9}{24}$	$\frac{251}{720}$	$\frac{475}{1440}$	$\frac{19087}{60480}$	$\frac{36799}{120960}$	$\frac{1070017}{3628800}$	$\frac{2082753}{7257600}$
1	8	19	646	1427	65112	139849	4467094	9449717
2	-1	-5	-264	-798	-46461	-121797	-4604594	-11271304
3		1	106	482	37504	123133	5595358	16002320
4			-19	-173	-20211	-88547	-5033120	-17283646
5				27	6312	41499	3146338	13510082
6					-863	-11351	-1291214	-7394032
7						1375	312874	2687864
8							-33953	-583435
9								57281

The denominator of each of the coefficients of the first line is to be repeated for all the coefficients of the corresponding column.

Table 2.4.3: The Adams' Predictor Coefficients \tilde{b}_k^N .

$k \ N$	2	3	4	5	6	7	8	9
0	$\frac{23}{12}$	$\frac{55}{24}$	$\frac{1901}{720}$	$\frac{4277}{1440}$	$\frac{198721}{60480}$	$\frac{434241}{120960}$	$\frac{14097247}{3628800}$	$\frac{30277247}{7257600}$
1	-16	-59	-2774	-7923	-447288	-1152169	-43125206	-104995189
2	5	37	2616	9982	705549	2183877	95476786	265932680
3		-9	-1274	-7298	-688256	-2664477	-139855262	-454661776
4				251	2877	407139	2102243	137968480
5					-475	-134472	-1041723	-91172642
6						19087	295767	38833486
7							-36799	252618224
8								-9664106
9								1070017
								-2082753

Again the denominator of each of the coefficients of the first line is to be repeated for all the coefficients of the corresponding column.

Exercises

2.4.1. Verify that (4.5) holds exactly for $j = 0, 1, 2, 3$ and fails to be exact for $j \geq 4$. Verify that (4.5) is the unique explicit two-step formula having third-order accuracy. Find the error when $j = 4$. Verify that (4.6) is exact for $j = 0, 1, 2, 3, 4$ and fails for $j \geq 5$.

2.4.2. Review Exercise 4.1. Verify that the accuracy of (4.5) and (4.6) exceeds the limit specified by the first Dahlquist barrier. But, consistent with this barrier, (4.34) illustrates

that the method (4.5) is unstable. Verify that, in accord with the first Dahlquist barrier, the method (4.6) is also unstable.

2.4.3. Suppose that a multistep method has order $m_{\max} \geq 1$ so that, in particular, it is able to treat the case (4.7) and (4.8) exactly when $j = 1$. Show that then there is the relation

$$\rho'(1) = \sigma(1). \quad (2.4.79)$$

2.4.4. Verify (4.16), (4.18), (4.22), and the expansions (4.23) and (4.24).

2.4.5. Verify the relations (4.44) through (4.47).

2.4.6. Verify (4.67) and (4.68) by direct calculation using Table 4.1 in the case that $y(t) = t^3$. How large must N be in order to get exact results?

2.4.7. Compute the first few coefficients a_k in equation (4.65). [Hint: First try differentiating F to convince yourself that this is not a good method. Then try synthetic division using equation (4.54). Can you find any other good method?]

Show that the coefficients b_k satisfy the recursion relation

$$b_k - b_{k-1} = a_k, \quad (2.4.80)$$

and use this relation to compute the first few b 's.

2.4.8. Show that the coefficients \tilde{a}_k^N obey the relations

$$\tilde{a}_N^N = (-1)^N a_N, \quad (2.4.81)$$

$$\tilde{a}_k^N = 0 \text{ if } N < k, \quad (2.4.82)$$

$$\tilde{a}_k^{N+1} = \tilde{a}_k^N + (-1)^k \binom{N+1}{k} a_{N+1}. \quad (2.4.83)$$

Compute the first few \tilde{a}_k^N . Make a similar study of the \tilde{b} 's.

2.4.9. Use equation (4.77) to estimate the expected local truncation error for Example (4.1) and compare with the actual error. [Use the solution (2.10) and (2.11) to compute $(d^{N+2}\mathbf{y}/dt^{N+2})$.] Use both (4.37) and (4.38) to derive a formula for the corrector error that does not require a knowledge of $(d^{N+2}\mathbf{y}/dt^{N+2})$. Apply it to Example (4.1). [Ans: $\mathbf{y}_{\text{true}} - \mathbf{y}_{\text{corr}} \simeq a_{N+1}(\mathbf{y}_{\text{pred}} - \mathbf{y}_{\text{corr}})/(a_{N+1} - b_{N+1})$. This strategy is called the *Milne device*.]

2.4.10. Consider the differential equation set (2.7) through (2.9). You will see below a table of entries obtained from a very accurate Runge-Kutta starting routine. Using Adams, complete the table for $n = 4$. The step size is $h = 1/3$. Compare your answer with the exact result. How big do you expect your error to be?

n	t^n	y_1^n	$y_2^n = f_1^n$	f_2^n
0	0	0	1	0
1	1/3	.33947	1.05505	.32719
2	2/3	.71496	1.21412	.61837
3	1	1.15853	1.45970	.84147
4	4/3	?	?	?

2.4.11. Show that if one is integrating *linear* differential equations, then the corrector formula (4.69) can be made explicit so that it is in principle possible to integrate without a predictor. Whether or not one should actually do this is a matter of convenience and economy.

2.4.12. The use of the predictor-corrector method requires at least two evaluations of \mathbf{f} at each step. Explore the merits of integrating with a step size of $h/2$ and using just the predictor without correcting. That is, use just *PE* at each step. Both methods require the same number of \mathbf{f} evaluations to integrate over a given time interval. Which is more accurate? Answer the question first ignoring round-off error, and then taking it into account. Do not worry about stability.

2.4.13. Show that with the stored data $\mathbf{f}^n, \dots, \mathbf{f}^{n-N}$ one can set up the corrector formula

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h \sum_0^{N+1} a_k \nabla^k \mathbf{f}^{n+1}. \quad (2.4.84)$$

Verify that (4.84) has the expanded form

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h \sum_0^{N+1} \tilde{a}_k^{N+1} \mathbf{f}^{n+1-k}. \quad (2.4.85)$$

Show that the error associated with these formulas is given by the relation

$$\mathbf{y}_{\text{true}}^{n+1} - \mathbf{y}_{\text{corr}}^{n+1} \approx h^{N+3} a_{N+2} (d^{N+3} \mathbf{y} / dt^{N+3})|_{t=t^n}. \quad (2.4.86)$$

Comparison of (4.86) and (4.77) shows that use of (4.84), or equivalently (4.85), yields results of one order higher accuracy; and therefore we will refer to this corrector as a *higher-order corrector*.

What accuracy can be achieved if we use the corrector (4.85) in conjunction with the predictor (4.70)? Both make optimal use of the the stored data $\mathbf{f}^n, \dots, \mathbf{f}^{n-N}$. Whether or not the smaller error associated with the higher-order corrector is achieved in practice depends on the number of corrector iterations. It can be shown that *PEC* is not enough, but *PECEC* may suffice. If ending on an *E* step is deemed desirable, then one should use at least *PECECE*.

Your next task is to compare the accuracies specified by (4.77), (4.78), and (4.86) with that specified by the first Dahlquist barrier (4.35) through (4.37). Verify that, according to (4.78), the Adams predictor (4.70) makes local errors of order h^{N+2} and therefore is exact through order h^{N+1} . We also recall that $k = N + 1$ so that the Adams predictor (4.70) is exact through order h^k . According to (4.35) the highest local error m_{\max} that can be achieved by a strongly stable explicit k -step method is k . Therefore, the Adams predictor (4.70) achieves the first Dahlquist barrier limit. With regard to implicit formulas, verify that (4.77) shows that the corrector formula (4.69) is exact through order h^k . But, according to (4.36), (4.37), and the ensuing discussion, it should be possible, in the implicit case, to achieve results that are accurate through order h^{k+1} . Verify that, according to (4.86), the higher-order corrector (4.85) is exact through order h^{k+1} . Therefore in this case the first

Dahlquist barrier limit is also achieved assuming that the higher-order Adams corrector is employed.

Your last task is to consider two low-order cases. Show that use of (4.67), (4.68), and (4.84) for $N = 0$ gives the formulas

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h\mathbf{f}^n, \quad (\text{predictor}) \quad (2.4.87)$$

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h\mathbf{f}^{n+1}, \quad (\text{corrector}) \quad (2.4.88)$$

$$\mathbf{y}^{n+1} = \mathbf{y}^n + (h/2)(\mathbf{f}^{n+1} + \mathbf{f}^n), \quad (\text{higher-order corrector}). \quad (2.4.89)$$

Note that (4.87) is just the Euler method (2.3). The procedure (4.88) is sometimes called *backward* Euler, and in this context (4.87) is called *forward* Euler.

Show that use of (4.67), (4.68), and (4.84) for $N = 1$ gives the formulas

$$\mathbf{y}^{n+1} = \mathbf{y}^n + (h/2)(3\mathbf{f}^n - \mathbf{f}^{n-1}), \quad (\text{predictor}) \quad (2.4.90)$$

$$\mathbf{y}^{n+1} = \mathbf{y}^n + (h/2)(\mathbf{f}^{n+1} + \mathbf{f}^n), \quad (\text{corrector}) \quad (2.4.91)$$

$$\mathbf{y}^{n+1} = \mathbf{y}^n + (h/12)(5\mathbf{f}^{n+1} + 8\mathbf{f}^n - \mathbf{f}^{n-1}), \quad (\text{higher-order corrector}). \quad (2.4.92)$$

2.4.14. Suppose one were to continue making corrector steps, that is carry out the sequence of steps *PECECECECE* ..., until convergence is achieved to machine precision. Would there be any virtue in doing so? The answer is *no* because all that would be achieved is a result that differs from the exact result by the truncation error associated with the corrector formula. Verify that this happens for the example illustrated in Exhibit 4.1. Thus there is no point in making further corrector steps once convergence has been achieved to within the expected corrector formula truncation error.

2.4.15. The purpose of this exercise is to study the stability properties of the $N = 1$ Adams routine given by (4.90) through (4.92).

Let us begin with the predictor (4.90). Show that applying it to the differential equation (4.13) produces the characteristic equation

$$\zeta^2 - [1 + (3/2)(h\lambda)]\zeta + (1/2)(h\lambda) = 0, \quad (2.4.93)$$

and verify that the characteristic equation has the roots

$$\zeta = \{[1 + (3/2)(h\lambda)] \pm \sqrt{[1 + (h\lambda) + (9/4)(h\lambda)^2]}\}/2. \quad (2.4.94)$$

Show that (4.93) has the good root

$$\begin{aligned} \zeta_1 &= \{[1 + (3/2)(h\lambda)] + \sqrt{[1 + (h\lambda) + (9/4)(h\lambda)^2]}\}/2 \\ &= 1 + h\lambda + (h\lambda)^2/2! - (1/4)(h\lambda)^3 + \dots \\ &= \exp(h\lambda) + O(h^3) \end{aligned} \quad (2.4.95)$$

and the parasitic root

$$\begin{aligned} \zeta_2 &= \{[1 + (3/2)(h\lambda)] - \sqrt{[1 + (h\lambda) + (9/4)(h\lambda)^2]}\}/2 \\ &= (h\lambda)/2 - (h\lambda)^2/2 + O(h^3). \end{aligned} \quad (2.4.96)$$

Note that the good root goes to 1 as h goes to 0, as required; and the parasitic root goes to 0 as h goes to 0, as expected for Adams' method. Verify that the argument of the square root appearing in (4.95) and (4.96) is always positive when the quantity $h\lambda$ is real, and therefore there is no ambiguity involved in the definition of the square root. Show that $\zeta_2 < 1/2$ when $h\lambda$ is real. Verify that ζ_2 leaves the unit disk through $\zeta_2 = -1$ when $h\lambda = -1$, and becomes ever more negative than -1 as $h\lambda$ becomes ever more negative than -1 . Verify that $\zeta_1 = 1/2$ when $h\lambda = -1$.

Next consider the corrector (4.91). Show that applying it to the differential equation (4.13) produces the characteristic equation

$$\zeta^2 - \{[1 + (h\lambda)/2]/[1 - (h\lambda)/2]\}\zeta = 0. \quad (2.4.97)$$

Show that (4.97) has the good root

$$\begin{aligned} \zeta_1 &= [1 + (h\lambda)/2]/[1 - (h\lambda)/2] \\ &= 1 + h\lambda + (h\lambda)^2/2! + (1/4)(h\lambda)^3 + \dots \\ &= \exp(h\lambda) + O(h^3) \end{aligned} \quad (2.4.98)$$

and the parasitic root

$$\zeta_2 = 0. \quad (2.4.99)$$

Evidently, in this case, the parasitic root for the corrector has the optimum value of zero for all values of h ! To examine this matter further, show that applying (4.91) to the differential equation (4.13) produces the recursion relation

$$y^{n+1} = \zeta_1 y^n, \quad (2.4.100)$$

and show that this recursion relation has the unique solution

$$y^n = (\zeta_1)^n y^0. \quad (2.4.101)$$

Finally, consider the higher-order corrector (4.92). Show that applying it to the differential equation (4.13) produces the characteristic equation

$$[1 - (5/12)(h\lambda)]\zeta^2 - [1 + (2/3)(h\lambda)]\zeta + (h\lambda)/12 = 0. \quad (2.4.102)$$

Verify that (4.102) has the roots

$$\zeta = \{[1 + (2/3)(h\lambda)] \pm \sqrt{[1 + h\lambda + (7/12)(h\lambda)^2]}\}/\{2[1 - (5/12)(h\lambda)]\}. \quad (2.4.103)$$

Show that (4.102) has the good root

$$\begin{aligned} \zeta_1 &= \{[1 + (2/3)(h\lambda)] + \sqrt{[1 + h\lambda + (7/12)(h\lambda)^2]}\}/\{2[1 - (5/12)(h\lambda)]\}. \\ &= 1 + h\lambda + (h\lambda)^2/2! + (h\lambda)^3/3! + (1/12)(h\lambda)^4 + \dots \\ &= \exp(h\lambda) + O(h^4) \end{aligned} \quad (2.4.104)$$

and the parasitic root

$$\begin{aligned} \zeta_2 &= \{[1 + (2/3)(h\lambda)] - \sqrt{[1 + h\lambda + (7/12)(h\lambda)^2]}\}/\{2[1 - (5/12)(h\lambda)]\}. \\ &= (h\lambda)/12 - (7/144)(h\lambda)^2 + O(h^3). \end{aligned} \quad (2.4.105)$$

Verify that the argument of the square root appearing in (4.104) and (4.105) is always positive when the quantity $h\lambda$ is real, and therefore there is no ambiguity involved in the definition of the square root. Verify that ζ_2 remains within the unit disk when $h\lambda > -6$, has the value $\zeta_2 = -1$ when $h\lambda = -6$, and becomes ever more negative than -1 (leaves the unit disk through -1) as $h\lambda$ becomes ever more negative than -6 . Verify that $\zeta_1 = 1/7$ when $h\lambda = -6$.

2.4.16. This exercise extends that work on Finite Difference Calculus presented at the beginning of this subsection to derive what is called the *Euler-Maclaurin* formula. This formula is of use both in evaluating integrals and in summing series.

Suppose $f(t)$ is some function and we wish to calculate the integral I given by

$$I = \int_{t_0}^{t_N} f(t) dt. \quad (2.4.106)$$

If we subdivide the interval $[t_0, t_N]$ into N equal pieces as in Figure 1.1, then we may approximate the integral (4.106) by the areas of N trapezoids, each of base h , to obtain the *trapezoidal rule* result

$$\begin{aligned} I &\approx h[(1/2)(f_0 + f_1) + (1/2)(f_1 + f_2) + \cdots + (1/2)(f_{N-1} + f_N)] \\ &= h(1/2)(f_0 + f_N) + h \sum_{i=1}^{N-1} f_i = -h(1/2)(f_0 + f_N) + h \sum_{i=0}^N f_i. \end{aligned} \quad (2.4.107)$$

(Here we have employed subscript indices rather than the superscript indices used earlier in Subsection 4.4 because we will soon need superscript indices for another purpose.) In summary, the trapezoidal rule yields the approximation

$$\int_{t_0}^{t_N} f(t) dt \approx -h(1/2)(f_0 + f_N) + h \sum_{i=0}^N f_i. \quad (2.4.108)$$

We will now further develop the calculus of finite differences and employ it to improve on the accuracy of this approximation.

To begin define, in analogy to (4.4.1), a *forward difference* operator Δ by the rule

$$\Delta f(t) = f(t + h) - f(t). \quad (2.4.109)$$

Next define an operator J by the rule

$$Jf(t) = \int_t^{t+h} f(t') dt'. \quad (2.4.110)$$

[Here the symbol J is not to be confused with the J introduced in (1.7.11). At the expense of duplication of symbol use, we are trying to follow convention in both the subjects of Hamiltonian theory and finite difference calculus.] Show that there are the relations

$$DJf(t) = JDf(t) = f(t + h) - f(t) = \Delta f(t). \quad (2.4.111)$$

Consequently, there are the operator relations

$$DJ = JD = \Delta. \quad (2.4.112)$$

Also show, using reasoning similar to that which led to the result (4.52), that there is the operator relation

$$\Delta = \exp(hD) - 1. \quad (2.4.113)$$

Verify that (4.112) and (4.113) can be combined to produce the operator relation

$$h = \Delta^{-1}hDJ = \{hD/[\exp(hD) - 1]\}J. \quad (2.4.114)$$

How can we make sense of the right side of (4.114)? It can be shown that there is the analytic function result

$$\tau/[\exp(\tau) - 1] = \sum_{j=0}^{\infty} (B_j/j!) \tau^j \quad (2.4.115)$$

where the B_k are the *Bernoulli numbers*.¹⁸ These numbers have the property

$$B_3 = B_5 = B_7 = \dots = 0, \quad (2.4.116)$$

and the first few nonzero of them have the values

$$B_0 = 1, \quad B_1 = -1/2, \quad B_2 = 1/6, \quad B_4 = -1/36, \quad B_6 = 1/42, \quad B_8 = -1/30. \quad (2.4.117)$$

Consequently verify that, when acting on a function, (4.114) takes the forms

$$\begin{aligned} hf(t) &= \{hD/[\exp(hD) - 1]\}Jf(t) = \sum_{j=0}^{\infty} (B_j/j!) (hD)^j Jf(t) \\ &= \int_t^{t+h} f(t') dt' + \sum_{j=1}^{\infty} (B_j/j!) (hD)^j Jf(t) \\ &= \int_t^{t+h} f(t') dt' + \sum_{j=1}^{\infty} (B_j/j!) h^j D^{j-1} DJf(t) \\ &= \int_t^{t+h} f(t') dt' + \sum_{j=1}^{\infty} (B_j/j!) h^j D^{j-1} \Delta f(t) \\ &= \int_t^{t+h} f(t') dt' + \sum_{j=1}^{\infty} (B_j/j!) h^j D^{j-1} [f(t+h) - f(t)] \\ &= \int_t^{t+h} f(t') dt' + \sum_{j=1}^{\infty} (B_j/j!) h^j [f^{(j-1)}(t+h) - f^{(j-1)}(t)]. \end{aligned} \quad (2.4.118)$$

(Here a superscript index in parenthesis denotes a derivative of that order.) Note that if $f(t)$ is a polynomial in t , then the sum on the far right side of (4.118) terminates. Therefore

¹⁸Indeed, (4.115) defines the Bernoulli numbers.

(4.118) is well defined and exact for any polynomial f because no convergence questions arise.

Let us further manipulate (4.118). Suppose t in (4.118) is replaced by $t + h$. Verify that so doing yields the result

$$hf(t+h) = \int_{t+h}^{t+2h} f(t')dt' + \sum_{j=1}^{\infty} (B_j/j!)h^j[f^{(j-1)}(t+2h) - f^{(j-1)}(t+h)]. \quad (2.4.119)$$

Next add (4.119) to (4.118). Verify that so doing yields the result

$$h[f(t) + f(t+h)] = \int_t^{t+2h} f(t')dt' + \sum_{j=1}^{\infty} (B_j/j!)h^j[f^{(j-1)}(t+2h) - f^{(j-1)}(t)]. \quad (2.4.120)$$

Verify that this process can be generalized to yield the result

$$\begin{aligned} h[f(t) + f(t+h) + \cdots + f(t+Nh-h)] &= h \sum_{i=0}^{N-1} f(t+ih) \\ &= \int_t^{t+Nh} f(t')dt' + \sum_{j=1}^{\infty} (B_j/j!)h^j[f^{(j-1)}(t+Nh) - f^{(j-1)}(t)]. \end{aligned} \quad (2.4.121)$$

To manipulate still further, verify that

$$\begin{aligned} (B_1/1!)h[f^{(0)}(t+Nh) - f^{(0)}(t)] &= -(1/2)h[f^{(0)}(t+Nh) - f^{(0)}(t)] \\ &= -(1/2)h[f(t+Nh) - f(t)] = -hf(t+Nh) + h(1/2)[f(t) + f(t+Nh)]. \end{aligned} \quad (2.4.122)$$

Verify that employing (4.122) in (4.121) yields the result

$$\begin{aligned} &\int_t^{t+Nh} f(t')dt' \\ &= -h(1/2)[f(t) + f(t+Nh)] + h \sum_{i=0}^N f(t+ih) \\ &\quad - \sum_{j=2}^{\infty} (B_j/j!)h^j[f^{(j-1)}(t+Nh) - f^{(j-1)}(t)]. \end{aligned} \quad (2.4.123)$$

Finally set $t = t_0$ and make use of (4.116) to convert (4.123) to the relation

$$\begin{aligned} &\int_{t_0}^{t_N} f(t)dt = \\ &-h(1/2)(f_0 + f_N) + h \sum_{i=0}^N f_i - \sum_{k=1}^{\infty} [B_{2k}/(2k)!]h^{2k}[f_N^{(2k-1)} - f_0^{(2k-1)}]. \end{aligned} \quad (2.4.124)$$

It is also useful to express this result using the *initial* and *final* notation

$$t_{\text{in}} = t_0, \quad t_{\text{fin}} = t_N, \quad f_{\text{in}} = f_0, \quad f_{\text{fin}} = f_N, \quad (2.4.125)$$

so that (4.124) becomes

$$\begin{aligned} & \int_{t_{\text{in}}}^{t_{\text{fin}}} f(t) dt = \\ & -h(1/2)(f_{\text{in}} + f_{\text{fin}}) + h \sum_{i=0}^N f_i - \sum_{k=1}^{\infty} [B_{2k}/(2k)!] h^{2k} [f_{\text{fin}}^{(2k-1)} - f_{\text{in}}^{(2k-1)}]. \end{aligned} \quad (2.4.126)$$

We see that the accuracy of the trapezoidal rule (4.108) can be improved if one knows, in addition to the $(N+1)$ sampling-point values f_i , the end-point derivative values $f_{\text{fin}}^{(2k-1)}$ and $f_{\text{in}}^{(2k-1)}$.

Let us also introduce the notation

$$b_{2k} = [B_{2k}/(2k)!] [f_{\text{fin}}^{(2k-1)} - f_{\text{in}}^{(2k-1)}] = [B_{2k}/(2k)!] [f^{(2k-1)}(t)|_{t_{\text{in}}}^{t_{\text{fin}}}] \quad (2.4.127)$$

so that

$$\sum_{k=1}^{\infty} [B_{2k}/(2k)!] h^{2k} [f_{\text{fin}}^{(2k-1)} - f_{\text{in}}^{(2k-1)}] = \sum_{k=1}^{\infty} b_{2k} h^{2k} \quad (2.4.128)$$

and (4.126) can be written in the form

$$\int_{t_{\text{in}}}^{t_{\text{fin}}} f(t) dt = -h(1/2)(f_{\text{in}} + f_{\text{fin}}) + h \sum_{i=0}^N f_i - \sum_{k=1}^{\infty} b_{2k} h^{2k}. \quad (2.4.129)$$

In this form it is manifest that the correction to the trapezoidal rule is a Taylor series in h . This series will converge within some disc about $h = 0$ if the coefficients b_{2k} are sufficiently well behaved. But the convergence radius could also be zero if the coefficients b_{2k} are not sufficiently well behaved.

We can also infer from the previous discussion that (4.129) is exact when f is a polynomial in t because the sum over k then terminates. More precisely, as is evident from (4.127), in this case all the $b_{2k} = 0$ once k is sufficiently large. If f is not polynomial and the sum over k is terminated when $k = m$, then may write

$$\int_{t_{\text{in}}}^{t_{\text{fin}}} f(t) dt = -h(1/2)(f_{\text{in}} + f_{\text{fin}}) + h \sum_{i=0}^N f_i - \sum_{k=1}^m b_{2k} h^{2k} - E_m \quad (2.4.130)$$

where E_m is a *remainder/error* term. Note that E_m is well defined because all the other terms in (4.130) are well defined (assuming h and N are finite). Define a function $\hat{E}_m(\tau)$ by the rule

$$\begin{aligned} \hat{E}_m(\tau) &= (Nh)^{2m+2} \{B_{2m+2}/[(2m+2)!]\} f^{(2m+2)}(\tau) \\ &= (t_{\text{fin}} - t_{\text{in}})^{2m+2} \{B_{2m+2}/[(2m+2)!]\} f^{(2m+2)}(\tau). \end{aligned} \quad (2.4.131)$$

It can be shown that

$$E_m = \hat{E}_m(\tau) \quad (2.4.132)$$

for some $\tau \in (t_{\text{in}}, t_{\text{fin}})$. Verify that if

$$\max_{\tau \in [t_{\text{in}}, t_{\text{fin}}]} |\hat{E}_m(\tau)| \rightarrow 0 \text{ as } m \rightarrow \infty, \quad (2.4.133)$$

then the series over k will converge and (4.129) is well defined and exact. By contrast, verify that if

$$\min_{\tau \in [t_{\text{in}}, t_{\text{fin}}]} |\hat{E}_m(\tau)|$$

does not approach zero as $m \rightarrow \infty$, then the series over k is divergent.

Frequently, when f is not polynomial, the relation (4.129) has an *asymptotic* character: The E_m do not tend to zero as $m \rightarrow \infty$, but rather there is an optimum value of m for which $|E_m|$ takes on a minimum (but generally nonzero) value. For m values smaller or larger than this optimum value the remainder/error is larger.

In closing this part of the discussion we note that if f is *periodic* and analytic, and integration is to be done over a full period, then, for any k ,

$$f_{\text{fin}}^{(2k-1)} - f_{\text{in}}^{(2k-1)} = 0 \quad (2.4.134)$$

from which it follows that $b_{2k} = 0$ for all k . In this case the only correction to the trapezoidal rule is the remainder/error term E_m . We then conclude that for this case the error associated with the trapezoidal rule vanishes as $h \rightarrow 0$ and $N \rightarrow \infty$ faster than any power of h . For further discussion of the application of the trapezoidal rule to the integration of analytic periodic functions see the paragraph on “Performing the Forward $\phi \rightarrow m$ Fourier Transform” right after (19.1.27) and Exercises 19.1.2 through 19.1.4. See also Section 19.2.4 and Exercises 19.2.2 through 19.2.4. Finally, see the references to “Angular Integrals” at the end of Chapter 19.

On some occasions it is useful to rewrite (4.124) in the form

$$\begin{aligned} \sum_{i=0}^N f_i &= \\ (1/h) \int_{t_0}^{t_N} f(t) dt + (1/2)(f_0 + f_N) + \sum_{k=1}^{\infty} [B_{2k}/(2k)!] h^{2k-1} [f_N^{(2k-1)} - f_0^{(2k-1)}]. \end{aligned} \quad (2.4.135)$$

If the integral and sum on the right side of (4.135) can be evaluated, then the sum on the left side has been computed. Even if this cannot be accomplished, verify that (4.124) can be rewritten in the form

$$\begin{aligned} \sum_{i=0}^N f_i &= \\ (1/h) \int_{t_0}^{t_N} f(t) dt + (1/2)(f_0 + f_N) + \sum_{k=1}^m [B_{2k}/(2k)!] h^{2k-1} [f_N^{(2k-1)} - f_0^{(2k-1)}] + E_m/h. \end{aligned} \quad (2.4.136)$$

This relation can be used to compute the sum on the left within an error that can be estimated using (4.132).

If we set $N = \infty$ and keep h finite, and also assume that

$$f_\infty = 0 \text{ and all } f_\infty^{(2k-1)} = 0, \quad (2.4.137)$$

show that then (4.136) becomes

$$\begin{aligned} \sum_{i=0}^{\infty} f_i &= \\ (1/h) \int_{t_0}^{\infty} f(t) dt + (1/2)f_0 - \sum_{k=1}^m [B_{2k}/(2k)!] h^{2k-1} f_0^{(2k-1)} + E_m/h. \end{aligned} \quad (2.4.138)$$

In this case (4.132) cannot be used to estimate E_m . However, there are other more complicated estimates that can be used, and their use provides a value for the infinite sum on the left of (4.138) within a computable error estimate.

2.5 (Automatic) Choice and Change of Step Size and Order

In our initial discussion concerning the choice of step size h , we were a bit cavalier. We merely stated that h should be small compared to the characteristic time scale of the physical system under study. This statement is somewhat vague since the time scale may be different for different parts of the trajectory. Consider, for example, the orbit of a comet about the sun. When it is far away from the sun, it nearly moves in a straight line. This part of the trajectory could be integrated with a large h . By contrast, the trajectory changes rapidly near the sun and a small time step is required for this part of the orbit.

Ideally, two things are needed: a method for automatically estimating the local truncation error at each integration step and a procedure for adjusting the step size or the order of the integration routine (or both) to keep the error within acceptable bounds. These ideals require some effort to realize in practice. Methods that accomplish at least one of these ideals are called *adaptive*.

2.5.1 Adaptive Change of Step Size in Runge-Kutta

In the case of Runge-Kutta, one can carry out a step using a step size h , and also carry out two steps using a step size $h/2$. By comparing the results of these two procedures, it is possible to estimate the local error, and then adjust the step size accordingly. See Exercise 5.1. Alternatively, as first discovered by *Fehlberg*, there are some pairs of Runge-Kutta procedures whose orders differ (usually by one) and that, in making one integration step, share many or all intermediate evaluation points. For these so-called *embedded* pairs, one can carry out both procedures simultaneously with little added expense. Then, by subtracting

the higher-order result from the lower-order result, one can estimate the *local* error in the lower-order result, and adjust the step size accordingly. See Appendix B for further detail.

We have seen that it is possible to adjust the Runge-Kutta step size automatically during the course of an integration run. In principle, with more complicated procedures, it is also possible to change the order as well. This is not now done in common practice, but is a subject of current research.

In summary, there are Runge-Kutta routines for which one specifies the initial and final times (t^0 and $t^0 + T$), the initial conditions $\mathbf{y}(t^0)$, and the acceptable local error. The routine then automatically selects and dynamically adjusts the step size to compute $\mathbf{y}(t^0 + T)$ with a minimal number of integration steps and with a global error that can be estimated from the allowed local error and the number of integration steps.

2.5.2 Adaptive Finite-Difference Methods

In the case of finite-difference methods it is possible, with some effort, to adjust both the step size and the order. We will now describe how this can be done.

Change of Order

In the Adams' method we have been discussing, it is easy to raise or lower the order. Suppose we are at $t = t^n$, and wish to step to t^{n+1} . We have at our disposal \mathbf{y}^n and the $N + 1$ \mathbf{f} values $\mathbf{f}^n \dots \mathbf{f}^{n-N}$. To lower the order by one, throw away the stored \mathbf{f}^{n-N} , and continue the integration using the N values $\mathbf{f}^n \dots \mathbf{f}^{n-N+1}$ with one order *lower* predictor and corrector formulas. Suppose we are at $t = t^{n+1}$ and have just completed a converged corrector step. Then the $N + 2$ \mathbf{f} values $\mathbf{f}^{n+1}, \mathbf{f}^n, \dots, \mathbf{f}^{n-N}$ are momentarily available. To raise the order by one, keep \mathbf{f}^{n-N} rather than discarding it, as would normally be done. Then, after relabeling the \mathbf{f} 's, we have available the $N + 2$ \mathbf{f} values $\mathbf{f}^n \dots \mathbf{f}^{n-N-1}$, and can make all future integration steps using one order *higher* formulas.

Change of Time Step

Changing the time step is more difficult. The simplest procedure is to stop the finite-difference routine. Then a Runge-Kutta routine with a different step size is begun using the previously obtained point as an initial condition. After a few starting values have been computed, one again returns to a finite-difference method. This finite-difference method would have the modified step size, and could also have a different order. Thus, a typical integration run could consist of several finite-difference segments of various step sizes and orders joined together by short pieces of Runge-Kutta.

Is there a more sophisticated way to change the time step? There is, but it is complicated. Given the $\mathbf{f}^n \dots \mathbf{f}^{n-N}$ at times $t^n \dots t^{n-N}$ separated by h , it is in principle possible by interpolation to find an equivalent set of \mathbf{f}' values $\mathbf{f}'^n \dots \mathbf{f}'^{n-N}$ at times $t'^n \dots t'^{n-N}$ separated by h' in such a way that the current times t^n and t'^n agree. The interpolated \mathbf{f}' values can then be used to make Adams' steps with a step size h' .

2.5.3 Jet Formulation

Is there a reformulation of the Adams' method that would facilitate changes in the time step and, at the same time, still make it easy to change orders? There is, but its description requires some explanation and discussion. In so doing, we will also learn about *jets* and classify all finite-difference/multistep methods.

As described earlier, Adams' method is a special case of multistep/multivalue methods where some combination of both previous \mathbf{f} values and previous \mathbf{y} values are stored. How much information about a trajectory is contained in these stored values? Take \mathbf{y}^n as given. Suppose there are M previously stored values (counting both \mathbf{y} and \mathbf{f} values). Then from this information, by suitable Taylor expansions, we might hope to compute $\dot{\mathbf{y}}^n, \ddot{\mathbf{y}}^n, \dots, \mathbf{y}^{(M)n}$ where $\mathbf{y}^{(m)n}$ denotes an approximation to the m 'th derivative of \mathbf{y} evaluated at t^n . Arrange these quantities in an $M + 1$ dimensional vector $\vec{\mathbf{j}}^n$ in the form

$$\vec{\mathbf{j}}^n = \begin{pmatrix} \mathbf{y}^n \\ h\dot{\mathbf{y}}^n \\ (h^2/2)\ddot{\mathbf{y}}^n \\ \vdots \\ (h^M/M!)\mathbf{y}^{(M)n} \end{pmatrix}. \quad (2.5.1)$$

If we wish, we can ensure that the $\dot{\mathbf{y}}^n$ entry in $\vec{\mathbf{j}}^n$ is exact by using (1.1) to compute $\dot{\mathbf{y}}(t^n)$. The remaining derivatives will be approximate. In keeping with terminology to be employed in subsequent chapters, we will refer to $\vec{\mathbf{j}}$ as a *jet*. More precisely we will refer to $\vec{\mathbf{j}}^n$ as given by (5.1) as an M -jet.

Conversion of Adams' Data into Jet Data

As an example of the procedure just described, let us convert stored Adams' data into jet data.¹⁹ Consider the case $N = 2$. Then at t^n we have the stored values \mathbf{f}^{n-1} and \mathbf{f}^{n-2} . If we imagine that these values are exact, we may make the Taylor expansions

$$\begin{aligned} h\mathbf{f}^{n-1} &= h\mathbf{f}(t^n - h) = h\dot{\mathbf{y}}(t^n - h) \\ &= h\dot{\mathbf{y}}(t^n) - h^2\ddot{\mathbf{y}}(t^n) + (h^3/2)\dddot{\mathbf{y}}(t^n) + \dots \\ &= h\dot{\mathbf{y}}^n - 2(h^2/2)\ddot{\mathbf{y}}^n + 3(h^3/6)\dddot{\mathbf{y}}^n + \dots, \end{aligned} \quad (2.5.2)$$

$$\begin{aligned} h\mathbf{f}^{n-2} &= h\mathbf{f}(t^n - 2h) = h\dot{\mathbf{y}}(t^n - 2h) \\ &= h\dot{\mathbf{y}}(t^n) - 2h^2\ddot{\mathbf{y}}(t^n) + [(2h)^3/2]\dddot{\mathbf{y}}(t^n) + \dots \\ &= h\dot{\mathbf{y}}^n - 4(h^2/2)\ddot{\mathbf{y}}^n + 12(h^3/6)\dddot{\mathbf{y}}^n + \dots \end{aligned} \quad (2.5.3)$$

Define a vector $\bar{\mathbf{s}}^n$ by writing

$$\bar{\mathbf{s}}^n = \begin{pmatrix} \mathbf{y}^n \\ h\mathbf{f}^n \\ h\mathbf{f}^{n-1} \\ h\mathbf{f}^{n-2} \end{pmatrix}. \quad (2.5.4)$$

¹⁹There is an alternate approach due to *Nordsieck* that essentially amounts to the same thing. Instead of storing Adams' data, one stores their finite differences.

We will refer to \vec{s} as *spread* data (Adams' in this case) since it refers to data at different times. Corresponding to (5.4) we expect to have a jet \vec{J}^n of the form

$$\vec{J}^n = \begin{pmatrix} \mathbf{y}^n \\ h\dot{\mathbf{y}}^n \\ (h^2/2)\ddot{\mathbf{y}}^n \\ (h^3/6)\dddot{\mathbf{y}}^n \end{pmatrix}. \quad (2.5.5)$$

Indeed, upon neglecting higher order terms, the relations (5.2) and (5.3) along with (1.1) can be written in the form

$$\vec{s}^n = R\vec{J}^n, \quad (2.5.6)$$

where R is the matrix

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & -2 & 3 \\ 0 & 1 & -4 & 12 \end{pmatrix}. \quad (2.5.7)$$

The matrix R has the inverse

$$R^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 3/4 & -1 & 1/4 \\ 0 & 1/6 & -1/3 & 1/6 \end{pmatrix}, \quad (2.5.8)$$

and therefore we may also write

$$\vec{J}^n = R^{-1}\vec{s}^n. \quad (2.5.9)$$

Jet Version of Adams' Predictor Formula

The Adams' predictor formula for $N = 2$ is

$$\mathbf{y}^{n+1} = \mathbf{y}^n + (h/12)(23\mathbf{f}^n - 16\mathbf{f}^{n-1} + 5\mathbf{f}^{n-2}). \quad (2.5.10)$$

See (4.70) and Table 2.4.3. Let us also find a formula for \mathbf{f}^{n+1} based only on Taylor expansions. We have the relation

$$\begin{aligned} h\mathbf{f}^{n+1} &= h\mathbf{f}(t^n + h) = h\dot{\mathbf{y}}(t^n + h) \\ &= h\dot{\mathbf{y}}(t^n) + h^2\ddot{\mathbf{y}}(t^n) + (h^3/2)\ddot{\mathbf{y}}(t^n) + \dots \\ &= h\dot{\mathbf{y}}^n + 2(h^2/2) + \ddot{\mathbf{y}}^n + 3(h^3/6)\ddot{\mathbf{y}}^n + \dots \end{aligned} \quad (2.5.11)$$

The quantities on the right side of (5.11) are components of \vec{J}^n . Use (5.9) to re-express them in terms of components of \vec{s}^n . Doing so gives the result

$$h\mathbf{f}^{n+1} = 3h\mathbf{f}^n - 3h\mathbf{f}^{n-1} + h\mathbf{f}^{n-2}. \quad (2.5.12)$$

According to (5.4), \vec{s}^{n+1} has the components

$$\vec{s}^{n+1} = \begin{pmatrix} \mathbf{y}^{n+1} \\ h\mathbf{f}^{n+1} \\ h\mathbf{f}^n \\ h\mathbf{f}^{n-1} \end{pmatrix}. \quad (2.5.13)$$

We see from (5.10), (5.12), and (5.13) that the relation between \vec{s}^n and \vec{s}^{n+1} can be written in the form

$$\vec{s}^{n+1} = A^{(2)} \vec{s}^n \quad (2.5.14)$$

where $A^{(2)}$, the $N = 2$ Adams' matrix, is defined by the relation

$$A^{(2)} = \begin{pmatrix} 1 & \frac{23}{12} & -\frac{16}{12} & \frac{5}{12} \\ 0 & 3 & -3 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (2.5.15)$$

What does the Adams' predictor step (5.14) correspond to in terms of jets? Using (5.6) and (5.9) we may write (5.14) in the equivalent form

$$R^{-1} \vec{s}^{n+1} = R^{-1} A^{(2)} \vec{s}^n = R^{-1} A^{(2)} R R^{-1} \vec{s}^n, \quad (2.5.16)$$

or

$$\vec{j}^{n+1} = T \vec{j}^n, \quad (2.5.17)$$

where T is the matrix

$$T = R^{-1} A^{(2)} R. \quad (2.5.18)$$

From (5.7), (5.8), and (5.15) we find for T the explicit result

$$T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (2.5.19)$$

Jet Version of Adams' Predictor Formula Is Simply Taylor's Theorem

Suppose we simply compute \vec{j}^{n+1} from a Taylor series. For \vec{y}^{n+1} we have the result

$$\vec{y}^{n+1} = \vec{y}(t^n + h) = \vec{y}(t^n) + h \dot{\vec{y}}(t^n) + (h^2/2) \ddot{\vec{y}}(t^n) + (h^3/6) \dddot{\vec{y}}(t^n) + \dots. \quad (2.5.20)$$

Also, from (5.11) we have the expansions

$$h \dot{\vec{y}}^{n+1} = h \dot{\vec{y}}^n + 2(h^2/2) \ddot{\vec{y}}^n + 3(h^3/6) \dddot{\vec{y}}^n + \dots. \quad (2.5.21)$$

Similarly, we have the expansions

$$\begin{aligned} (h^2/2) \ddot{\vec{y}}^{n+1} &= (h^2/2) \ddot{\vec{y}}(t^n + h) = (h^2/2) \ddot{\vec{y}}(t^n) + (h^3/2) \dddot{\vec{y}}(t^n) + \dots \\ &= (h^2/2) \ddot{\vec{y}}^n + 3(h^3/6) \dddot{\vec{y}}^n + \dots, \end{aligned} \quad (2.5.22)$$

$$(h^3/6) \dddot{\vec{y}}^{n+1} = (h^3/6) \dddot{\vec{y}}^n + \dots. \quad (2.5.23)$$

Upon comparing the coefficients in (5.17), and (5.19) through (5.23), we see that the jet relation (5.17) is simply Taylor's theorem. For this reason we will refer to T as the *Taylor*

matrix. We note that the entries in T are simply related to the binomial coefficients by the formula

$$T_{k\ell} = \binom{\ell}{k}, \quad (2.5.24)$$

with the understanding that

$$\binom{\ell}{k} = 0 \text{ when } k > \ell. \quad (2.5.25)$$

[Here, for convenience, the matrix elements in T are labeled starting from 0. That is, the elements (from left to right) in the first row of T are T_{00} , T_{01} , T_{02} , T_{03} , etc.] See Exercise 5.8. Indeed, the upper triangular portion of T is just *Pascal's triangle* turned on its side.²⁰

Effect of Evaluation on a Jet

So far we have seen how a jet changes under the operation of simple *prediction* P , and have found that the result (5.17) is just Taylor's theorem in disguise. Suppose we now add the *evaluation* operation E as well since it is the operation PE that is required for integration using only the predictor. See Exercise 4.12. What effect does PE have on a jet?

As before, let us first see what the E operation does to the spread vector \vec{s}^{n+1} . The E operation requires that we replace the $h\mathbf{f}^{n+1}$ entry in the spread vector (5.13) with $h\mathbf{f}(\mathbf{y}^{n+1}, t^{n+1})$. All other entries are unchanged. For simplicity of notation let us introduce the definition

$$h\tilde{\mathbf{f}}^{n+1} = 3h\mathbf{f}^n - 3h\mathbf{f}^{n-1} + h\mathbf{f}^{n-2}. \quad (2.5.26)$$

See (5.12). Also, define a quantity Δ by the rule

$$\Delta(\vec{s}^{n+1}, t^{n+1}) = h\mathbf{f}(\mathbf{y}^{n+1}, t^{n+1}) - h\tilde{\mathbf{f}}^{n+1}. \quad (2.5.27)$$

[Note that the vector Δ defined by (5.27) is not to be confused with the forward-difference operator Δ employed in (4.109).] Here it is understood that the \mathbf{y}^{n+1} in (5.27) is given by the predictor formula (5.10), and consequently also by the first component of \vec{s}^{n+1} in the relation (5.14). Note also that $h\tilde{\mathbf{f}}^{n+1}$ is the second component of \vec{s}^{n+1} . With these definitions, we see that under the full PE operation the vector \vec{s}^n is sent to the vector \vec{s}^{n+1} according to the rule

$$\vec{s}^{n+1} = A^{(2)}\vec{s}^n + \vec{e}, \quad (2.5.28)$$

where \vec{e} is the vector

$$\vec{e} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \Delta. \quad (2.5.29)$$

We observe that the *evaluation* vector \vec{e} , as desired, changes the second entry in (5.13) and leaves all the other entries unchanged.

²⁰Google *Pascal's triangle*.

Now that we know from (5.28) the effect of PE on a spread vector, we are ready to find the equivalent effect of the operation PE on a jet vector. Using (5.6) and (5.9) as before, we find that (5.28) takes the form

$$R^{-1}\vec{s}^{n+1} = R^{-1}A^{(2)}RR^{-1}\vec{s}^n + R^{-1}\vec{e}, \quad (2.5.30)$$

and consequently we have the relation

$$\vec{J}^{n+1} = T\vec{J}^n + \vec{r}, \quad (2.5.31)$$

where \vec{r} is given by

$$\vec{r} = R^{-1}\vec{e}. \quad (2.5.32)$$

If we use (5.7) and (5.29), we find that \vec{r} has the explicit form

$$\vec{r} = \begin{pmatrix} 0 \\ 1 \\ 3/4 \\ 1/6 \end{pmatrix} \Delta(\vec{J}^{n+1}, t^{n+1}). \quad (2.5.33)$$

Note that in terms of the jet \vec{J}^{n+1} , Δ as given by (5.27) takes the form

$$\Delta(\vec{J}^{n+1}, t^{n+1}) = h\mathbf{f}(\mathbf{y}^{n+1}, t^{n+1}) - h\tilde{\mathbf{f}}^{n+1}, \quad (2.5.34)$$

where \mathbf{y}^{n+1} and $h\mathbf{f}^{n+1}$ are given by the relations

$$\mathbf{y}^{n+1} = \mathbf{y}^n + \dot{\mathbf{y}}^n + (h^3/2)\ddot{\mathbf{y}}^n + (h^2/6)\ddot{\mathbf{y}}^n, \quad (2.5.35)$$

$$h\tilde{\mathbf{f}}^{n+1} = h\dot{\mathbf{y}}^n + 2(h^2/2)\ddot{\mathbf{y}}^n + 3(h^3/6)\ddot{\mathbf{y}}^n. \quad (2.5.36)$$

These relations follow from (5.6) and (5.10), and from (5.6) and (5.26), respectively. Note also that (5.35) and (5.36) are just the first two components of the predicted \vec{J}^{n+1} given by (5.17) and (5.19).

Effect of Corrector on a Jet

We have found the effect of Adams' prediction P and evaluation E on both spread vectors \vec{s} and jets \vec{J} . What about the corrector operation C ? The $N = 2$ corrector formula is

$$\mathbf{y}^{n+1} = \mathbf{y}^n + (h/12)(5\mathbf{f}^{n+1} + 8\mathbf{f}^n - \mathbf{f}^{n-1}). \quad (2.5.37)$$

See Table 4.2. As before, use (5.12) for \mathbf{f}^{n+1} . Doing so gives the result that the " \mathbf{f} " factor on the right side of (5.37) can be rewritten in the form

$$5h\mathbf{f}^{n+1} + 8h\mathbf{f}^n - h\mathbf{f}^{n-1} = 23h\mathbf{f}^n - 16h\mathbf{f}^{n-1} + 5h\mathbf{f}^{n-2}. \quad (2.5.38)$$

In view of (5.10), (5.37), and (5.38), we see that the spread vector relation (5.14) still holds with the same matrix $A^{(2)}$ given by (5.15). But now we have to take into account successive E and C operations. Their effect is to replace the $h\mathbf{f}^{n+1}$ in (5.37) and in the second component of the spread vector (5.13) by $\mathbf{f}(\mathbf{y}^{n+1}, t^{n+1})$, with \mathbf{y}^{n+1} defined by (5.37). Recall that we had

used (5.12) for \mathbf{f}^{n+1} . Define Δ as before in (5.27) but with the understanding that \mathbf{y}^{n+1} is now defined by (5.37). Then we see from (5.37) and (5.13) that the first component of \mathbf{s}^{n+1} is altered by $(5/12)\Delta$ and the second component, as before, is altered by Δ . Thus, when the converged correction operation is taken into account, (5.14) is modified to take the form

$$\bar{\mathbf{s}}^{n+1} = A^{(2)}\mathbf{s}^n + \vec{\mathbf{c}}, \quad (2.5.39)$$

where the *correction* vector $\vec{\mathbf{c}}$ is given by

$$\vec{\mathbf{c}} = \begin{pmatrix} 5/12 \\ 1 \\ 0 \\ 0 \end{pmatrix} \Delta. \quad (2.5.40)$$

We are now ready to determine the effect of correction on jets. As before, we multiply (5.39) by R^{-1} to get the results

$$R^{-1}\bar{\mathbf{s}}^{n+1} = R^{-1}A^{(2)}RR^{-1}\bar{\mathbf{s}}^n + R^{-1}\vec{\mathbf{c}}, \quad (2.5.41)$$

or

$$\bar{\mathbf{j}}^{n+1} = T\bar{\mathbf{j}}^n + \vec{\mathbf{r}} \quad (2.5.42)$$

where the vector $\vec{\mathbf{r}}$ is now defined by the relation

$$\vec{\mathbf{r}} = R^{-1}\vec{\mathbf{c}}. \quad (2.5.43)$$

By use of (5.8) and (5.40) we find the explicit result

$$\vec{\mathbf{r}} = \begin{pmatrix} 5/12 \\ 1 \\ 3/4 \\ 1/6 \end{pmatrix} \Delta. \quad (2.5.44)$$

So that there is no possible source of confusion, let us try to be perfectly clear about what is meant by Δ in (5.44). With the use of (5.6) we have the relation

$$(h/12)(8\mathbf{f}^n - \mathbf{f}^{n-1}) = (1/12)[7h\dot{\mathbf{y}}^n + 2(h^2/2)\ddot{\mathbf{y}}^n - 3(h^3/6)\dddot{\mathbf{y}}^n]. \quad (2.5.45)$$

Also, we use (5.36). Then, from (5.34) we have the result

$$\Delta = h\mathbf{f}(\mathbf{y}^{n+1}, t^{n+1}) - h\dot{\mathbf{y}}^n - 2(h^2/2)\ddot{\mathbf{y}}^n - 3(h^3/6)\dddot{\mathbf{y}}^n, \quad (2.5.46)$$

where, according to (5.37) and (5.45), \mathbf{y}^{n+1} satisfies the equation

$$\mathbf{y}^{n+1} = \mathbf{y}^n + (h/12)\mathbf{f}(\mathbf{y}^{n+1}, t^{n+1}) + (1/12)[7h\dot{\mathbf{y}}^n + 2(h^2/2)\ddot{\mathbf{y}}^n - 3(h^3/6)\dddot{\mathbf{y}}^n]. \quad (2.5.47)$$

We note that (5.47) can be solved by iteration just as was done before in Section 2.4. We begin by putting the predicted value (5.35) into the right side of (5.47), and then iterate. If the operations *PECE* were deemed adequate in the original spread vector variables, then the same holds true for the jet variables since (5.47) and (5.37) are actually the same equations. When the iterations have converged and the smoke has cleared, the first two components of $\bar{\mathbf{j}}^{n+1}$ are given by \mathbf{y}^{n+1} and $h\mathbf{f}(\mathbf{y}^{n+1}, t^{n+1})$, respectively. Now that \mathbf{y}^{n+1} and $h\mathbf{f}(\mathbf{y}^{n+1}, t^{n+1})$ are known, Δ and $\vec{\mathbf{r}}$ can be evaluated using (5.46) and (5.44). Finally, (5.42) can now also be used. By construction, it gives the same first two components of $\bar{\mathbf{j}}^{n+1}$ as found before, namely \mathbf{y}^{n+1} and $h\mathbf{f}(\mathbf{y}^{n+1}, t^{n+1})$. It also determines the remaining components of $\bar{\mathbf{j}}^{n+1}$.

2.5.4 Virtues of Jet Formulation

Overview

What are the virtues of using a jet formulation? First, there is a conceptual or theoretical advantage. It can be shown that *all* multistep/multivalue methods can be brought to jet variable form, and when this is done one always obtains results of the form (5.42). As might be guessed, the matrix T is universal. The various multistep/multivalue methods differ only in the choice of \vec{r} . Consequently, multistep/multivalue methods can be classified by their \vec{r} vectors. For example, the $N = 2$ Adams' predictor-evaluator method has the \vec{r} given by (5.33), and the $N = 2$ Adams' corrector method has the \vec{r} given by (5.44).

From a programming perspective, it is only necessary to build into a code the required \vec{r} vectors if it is to be able to run for various orders. (Note that if we program in the \vec{r} vectors for a variety of *methods*, the code can then also run for a variety of methods, and can even switch between methods!) Because of its simple form (5.24), the Taylor matrix can easily be computed and stored by the code itself as needed.

With regard to speed, the predictor part in the jet formulation consists in computing the first entry in \vec{J}^{n+1} as given by (5.17), which is just (5.20). This is no more difficult to compute than its spread-vector counterpart (4.70). As seen earlier, the evaluation and corrector operations required to compute the first two entries in \vec{J}^{n+1} are essentially the same in both the spread-vector and jet-vector formulations, and it is these operations that are the most time consuming. Finally, we need to compare the time required to compute the remaining entries in the spread vector \vec{s}^{n+1} with that required for the jet vector \vec{J}^{n+1} . As is evident from (5.15), (5.28), and (5.29), all that is required in the spread-vector case is a simple relabelling (what we have called updating) of stored \mathbf{f} values, which is very fast. In the jet-vector case inspection of (5.19), (5.42), and (5.44) shows that we must carry out some matrix-vector multiplies and some vector addition, which is a bit slower.

Change of Order

Changing the order for any method in the jet-vector formulation is also as easy as it was for the Adams' method in the spread-vector formulation. Suppose we wish to lower the order by one in the jet formulation. Simply delete the last component of the jet vector, and continue the integration by using the lower order version of (5.42) — which amounts to deleting the right-most column and bottom row from T and selecting the \vec{r} appropriate to the lower order. Raising the order is not much more difficult. Suppose, before the order is to be raised, that \vec{J} is an M -jet: the last entry in \vec{J} is $(h^M/M!)\mathbf{y}^{(M)}$. See (5.1). Store this entry for two or more successive steps and form the difference. Observe that we have the relation

$$\begin{aligned} [1/(M+1)](h^M/M!)[\mathbf{y}^{(M)}(t^n) - \mathbf{y}^{(M)}(t^n - h)] &= \\ [1/(M+1)](h^M/M!)[h\mathbf{y}^{(M+1)}(t^n) - (h^2/2)\mathbf{y}^{(M+2)}(t^n) + \dots] &\simeq \\ [h^{M+1}/(M+1)!]\mathbf{y}^{(M+1)n}. \end{aligned} \tag{2.5.48}$$

If desired, one can use more accurate formulas involving higher order differences and based on the relations (4.53) and (4.54). For an example of the use of these relations, see Exercise

5.3.] Equation (5.48) gives a value for $[h^{M+1}/(M+1)!]\mathbf{y}^{(M+1)n}$ which can be appended to the end of $\vec{\mathbf{j}}^n$ to convert it into an $(M+1)$ -jet. We can now continue the integration by using the one order higher version of (5.42) — which amounts to enlarging the T matrix using (5.24) and selecting the $\vec{\mathbf{r}}$ appropriate to the higher order.

Change of Step Size

The *main* virtue of the jet formulation is that it *easy* to change the step size. Observe that the step size appears nowhere in the marching orders (5.42) except for a simple dependency in Δ as given by (5.46) or (5.34). All the major h dependence occurs in the definition of $\vec{\mathbf{j}}$ as given by (5.1). Suppose we wish to change the step size from h to h' . Form the diagonal *scaling* matrix S defined by

$$S = \begin{pmatrix} 1 & & & \\ & (h'/h) & & \\ & & (h'/h)^2 & \\ & & & \ddots \end{pmatrix}. \quad (2.5.49)$$

Given the jet vector $\vec{\mathbf{j}}^n$ corresponding to step size h , form the corresponding jet vector $\vec{\mathbf{j}}'^n$ corresponding to step size h' by the relation

$$\vec{\mathbf{j}}'^n = S\vec{\mathbf{j}}^n. \quad (2.5.50)$$

We are now ready to continue the integration using (5.42) with h' and the $\vec{\mathbf{j}}'$ jet vectors.

Interpolation/Dense Output

There is yet another virtue to the jet formulation. As it runs, a numerical integration scheme only produces \mathbf{y} values at discrete points t^n . It may happen (particularly if the step size is being controlled dynamically by the integration program) that we need to know \mathbf{y} at some time τ that lies between two points, say t^m and t^{m+1} . This is easily done using the jet vector. Define a small quantity ϵ by the relation

$$\tau = t^m + \epsilon. \quad (2.5.51)$$

Also define an $(M+1)$ component vector $\vec{\delta}$ by the rule

$$\vec{\delta} = \begin{pmatrix} 1 \\ (\epsilon/h) \\ (\epsilon/h)^2 \\ \vdots \\ (\epsilon/h)^M \end{pmatrix}. \quad (2.5.52)$$

Then, by Taylor's theorem, we have the result

$$\begin{aligned} \mathbf{y}(\tau) &= \mathbf{y}(t^m + \epsilon) = \sum_k (\epsilon/h)^k (h^k/k!) \mathbf{y}^{(k)}(t^m) \\ &\simeq \vec{\delta} \cdot \vec{\mathbf{j}}^m. \end{aligned} \quad (2.5.53)$$

Adaptive Error Control

We have seen how the use of jets makes it possible to change the order and the time step at will with a fairly modest overhead. The size of the truncation error can also be estimated. If the jet formulation is based on the Adams' method, as we have been describing in our examples, then the error estimates (4.77) and (4.78) still hold. Consequently, if the order of the predictor and the corrector are the same, the error can be estimated by comparing predictor and corrector results. See Exercise 4.9. If the corrector is one order higher than the predictor, see Exercise 4.13, then the predictor error can be estimated directly simply by subtracting the corrector result from the predictor result. Finally, the error can also be estimated directly from (4.77) or (4.78) by using finite-difference relations such as (5.48) to compute the required derivatives.

With an error estimate in hand, it is possible to construct a jet-based code that will automatically select and dynamically adjust both step size and order to achieve a solution within the allowed error and with a minimal number of integration steps. Like the Runge-Kutta codes described at the beginning of this section, all it requires in principle is a specification of initial and final times (t^0 and $t^0 + T$), the initial condition $\mathbf{y}(t^0)$, and the acceptable error. A typical strategy is to have the program estimate from time to time the error currently being made at each step. If the error is too large, or if the error is too small (which means that too much effort is being spent in achieving unnecessary accuracy), the program computes what the step size should be for the error to be within the allowed bounds. This calculation is done both for the current order and for orders one higher and one lower. The program then shifts to the order that allows the largest step size, adjusts the step size to the largest value allowed, and continues to run for some time with this order and step size.

Self Starting

We observe that an integration routine having the features just described can be *self starting*. That is, unlike the finite-difference methods described in Section 2.4, such a program does not need a Runge-Kutta or other starting routine. Rather, it can begin with the $N = 0$ Adams' procedure (but in jet form) given by (4.85) and (4.86) or (4.87) since all this routine needs to start is the initial condition $\mathbf{y}(t^0)$. See Exercise 4.13. It can also automatically choose the step size to make sure that the accuracy of the first few steps is sufficiently high. Once the program is underway, it will then automatically adjust the order and step size to optimal values.

2.5.5 Advice to the Novice

As might be imagined, it is not a simple matter to write a variable order and variable step size program that will actually run in an optimal fashion for a wide variety of differential equations. Much time has been spent by professional mathematicians and numerical analysts in writing such programs. We have presented enough of the theory behind these programs to make them intelligible to readers and possible users; but they are advised not to try writing such programs on their own without exploring existing programs and without being prepared to expend considerable time and effort.

Exercises

2.5.1. The result of numerically integrating a differential equation from t^0 to $t^0 + T$ depends in general on the step size h . We express this fact by writing the result as $\mathbf{y}(t^0 + T; h)$. Neglecting round-off error, we expect $\mathbf{y}(t^0 + T; h)$ to approach the exact result as $h \rightarrow 0$. Consider an integration method that has a *cumulative* truncation error of order h^m . To be more precise, *assume* (what really requires proof and need not always be true) that we have

$$\mathbf{y}(t^0 + T; h) = \mathbf{y}_e(t^0 + T) + \mathbf{c}h^m + O(h^{m+1}), \quad (2.5.54)$$

where the subscript “e” stands for “exact”, and \mathbf{c} is independent of h , but otherwise unknown. Show that \mathbf{y}_e can be approximated by the formula

$$\mathbf{y}_e(t^0 + T) = \mathbf{y}(t^0 + T; h) + (1 - 2^{-m})^{-1}[\mathbf{y}(t^0 + T; h/2) - \mathbf{y}(t^0 + T; h)]. \quad (2.5.55)$$

Show that \mathbf{c} can be approximated by the formula

$$\mathbf{c}h^m = -(1 - 2^{-m})^{-1}[\mathbf{y}(t^0 + T; h/2) - \mathbf{y}(t^0 + T; h)]. \quad (2.5.56)$$

You see below a line of output for Example 3.1 run with a step size of $h = 1/20$.

time	$y1comp$	$y2comp$
1.5000	.20025125+01	.19292636+01

What should m be for RK3? Estimate $\mathbf{y}_e(1.5)$ and compare with the exact result. Devise a procedure that could be used if one had results for three different step sizes. You are studying *Richardson* extrapolation.

2.5.2. Verify (5.6) through (5.9).

2.5.3. Equation (5.7) for R is a direct consequence of Taylor’s theorem as used in (5.2) in (5.3). Equation (5.8) for R^{-1} was then found by inverting R . The entries in R^{-1} can also be found directly by requiring (5.9). For example, from (1.1) and (4.50) we have the result

$$\ddot{\mathbf{y}}^n = D\mathbf{f}^n. \quad (2.5.57)$$

Next use (4.53) and (4.54) to get the result

$$h\ddot{\mathbf{y}}^n = \sum_{k=1}^{\infty} (1/k) \nabla^k \mathbf{f}^n. \quad (2.5.58)$$

Discard terms in this series beyond $k = 2$, and verify that doing so reproduces the third row in (5.8). Similarly, we may write $\ddot{\mathbf{y}}^n = D^2 \mathbf{f}^n$. Use this result to reproduce the fourth row in (5.8).

2.5.4. Verify (5.12) using (5.9). See also Exercise 5.11.

2.5.5. Verify (5.14) and (5.15).

2.5.6. Verify (5.16) through (5.19).

2.5.7. Verify (5.20) through (5.23).

2.5.8. Verify (5.24) for the case (5.19). Let $g(t)$ be any (analytic) function. With D defined by (4.50), verify the formula

$$e^{hD}g(t) = g(t + h). \quad (2.5.59)$$

Verify the formal power series [in (hD)] identity

$$\exp(hD)(h^i/i!)D^i = \sum_{j=0}^{\infty} \binom{j}{i} (h^j/j!) D^j = \sum_{j=0}^{\infty} T_{ij}(h^j/j!) D^j. \quad (2.5.60)$$

Apply both sides of (5.60) to $\mathbf{y}(t)$ to derive (5.17) with the definition (5.24).

2.5.9. Verify (5.28) and (5.29).

2.5.10. Verify (5.30) through (5.36).

2.5.11. Study Exercise 4.13. Show that the higher corrector corresponding to the predictor (5.10) and the corrector (5.37) is given by the formula

$$\mathbf{y}^{n+1} = \mathbf{y}^n + (h/24)(9\mathbf{f}^{n+1} + 19\mathbf{f}^n - 5\mathbf{f}^{n-1} + \mathbf{f}^{n-2}). \quad (2.5.61)$$

See Table 2.2. Show that (5.12) can be derived by subtracting (5.10) from either (5.37) or (5.61). Show that (5.12) can also be derived by subtracting (5.37) from (5.61). Let α , β , γ be any three constants satisfying $\alpha + \beta + \gamma = 0$ and $10\beta + 9\gamma \neq 0$. Form the linear combination of *equations* given by the suggestive expression

$$\alpha(5.10) + \beta(5.37) + \gamma(5.61),$$

and use the result to verify (5.12).

2.5.12. Verify (5.38).

2.5.13. Verify (5.39) through (5.47).

2.5.14. Suppose that (5.10) and (5.61) are used as a predictor-corrector pair. What will the local truncation error be in this case? Show that (5.39) and (5.40) hold in this case providing that \vec{c} is the vector

$$\vec{c} = \begin{pmatrix} 3/8 \\ 1 \\ 0 \\ 0 \end{pmatrix} \Delta. \quad (2.5.62)$$

Suppose this Adams' method is reformulated in terms of jets. Show that the associated vector \vec{r} for (5.42) in this case is

$$\vec{r} = \begin{pmatrix} 3/8 \\ 1 \\ 3/4 \\ 1/6 \end{pmatrix} \Delta. \quad (2.5.63)$$

2.5.15. Verify (5.53). Suppose we wish to estimate the entire jet $\vec{J}(\tau)$. Let $S(h', h)$ denote the matrix (5.49). Verify the result

$$\vec{J}(\tau) = S(h, \epsilon) TS(\epsilon, h) \vec{J}^n. \quad (2.5.64)$$

2.6 Extrapolation Methods

2.6.1 Overview

In the previous section we have learned how it is possible to construct multistep methods that adjust both the order and the step size dynamically. We also learned how some Runge-Kutta methods (which are single step) can be modified to include dynamic step size control. In this section we will describe a single-step method that adjusts both order and step size dynamically.

The problem we desire to solve is the same: given the differential equation (1.1) with the initial condition $\mathbf{y}(t^0)$ at time $t = t^0$ and some acceptable error, we wish to find the final condition $\mathbf{y}(t^0 + T)$ within that error. In the methods described so far we have sought to achieve this goal by a march composed of many small steps, which we will call *micro* steps, of typical size h . In the method to be described now we will try to achieve the same goal by making fewer but larger steps, which we will call *meso* steps, whose typical size will be denoted by the symbol H . The procedure for making each meso step will have the feature of being self starting in that no information will be needed about the previous step (save for its ultimate result!); the meso step procedure is therefore a single-step method. Since the meso step size H will be relatively large, we can anticipate expending considerable effort in making each such step. The first meso step will take us from $\mathbf{y}(t^0)$ at time t^0 to $\mathbf{y}(t^0 + H)$ at time $(t^0 + H)$. Subsequent meso steps, perhaps with different sizes, will take us to subsequent times.

2.6.2 Making a Meso Step

How is such a step to made? We will describe the first meso step. Subsequent meso steps are made in the same way.

Simple Micro Step Formula

As before, divide up the time axis over the interval $[t^0, t^0 + H]$ into M equal *micro* steps of duration h . Then we have the relations

$$t^m = t^0 + mh, \quad h = H/M. \quad (2.6.1)$$

Refer back to Figure (1.1) with H now playing the role of T and intermediate times labelled as t^m . Next, in this interval, we will construct an apparently simple but actually quite subtle approximation to the values \mathbf{y}^m , the values of $\mathbf{y}(t)$ at the times t^m , that we will call $\boldsymbol{\eta}^m$. For $m = 0$ and $m = 1$ we will use the prescription

$$\boldsymbol{\eta}^0 = \mathbf{y}^0, \quad (2.6.2)$$

$$\boldsymbol{\eta}^1 = \boldsymbol{\eta}^0 + h\mathbf{f}(\boldsymbol{\eta}^0, t^0). \quad (2.6.3)$$

Comparison with (2.2) shows that (6.3) is simply an Euler step, and involves a local error of order h^2 . With $\boldsymbol{\eta}^0$ and $\boldsymbol{\eta}^1$ in hand, we define successive $\boldsymbol{\eta}^m$ by the *midpoint rule*

$$\boldsymbol{\eta}^{m+1} = \boldsymbol{\eta}^{m-1} + 2h\mathbf{f}(\boldsymbol{\eta}^m, t^m). \quad (2.6.4)$$

It is easily verified that the procedure (6.4) makes local errors of order h^3 . See Exercise 3.2. Continue the march (6.4) until $\boldsymbol{\eta}^{M-1}$ and $\boldsymbol{\eta}^M$ have been found. Finally, approximate $\mathbf{y}(t^0 + H)$ using $\boldsymbol{\eta}^{M-1}$ and $\boldsymbol{\eta}^M$ by the formula

$$\mathbf{y}(t^0 + H; M) = (1/2)[\boldsymbol{\eta}^M + \boldsymbol{\eta}^{M-1} + h\mathbf{f}(\boldsymbol{\eta}^M, t^M)]. \quad (2.6.5)$$

A Taylor series expansion of this last step again reveals a possible error of order h^2 , just as for the first step. Here we have used the notation $\mathbf{y}(t^0 + H; M)$ to indicate that (6.5) is an approximation to $\mathbf{y}(t^0 + H)$ that naturally depends on the size h of the micro steps and therefore, assuming that H is held fixed, on the number of micro steps M .

What is the virtue of this process? Let us estimate the *global* error for the formula (6.5). As already described, the first and last steps involve errors of order h^2 . Note that we may write

$$h^2 = (H/M)^2 = H^2(1/M)^2.$$

Here we have used (6.1). The intervening $(M - 1)$ midpoint rule steps (6.4) each involve local errors of order h^3 , and hence their cumulative effect should behave as

$$(M - 1)h^3 \approx Mh^3 \approx Hh^2 \approx H^3(1/M)^2.$$

Thus the total global meso step error should behave as

$$\text{meso step error} \approx (1/M)^2. \quad (2.6.6)$$

Consequently, if $\mathbf{y}_e(t^0 + H)$ denotes the “exact” solution, we might expect a relation, perhaps only asymptotic, of the form

$$\mathbf{y}(t^0 + H; M) - \mathbf{y}_e(t^0 + H) = \mathbf{c}_2(1/M)^2 + \mathbf{c}_3(1/M)^3 + \mathbf{c}_4(1/M)^4 + \dots \quad (2.6.7)$$

where the coefficients $\mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4 \dots$ are hoped to be independent of M . Here we reiterate that it is to be understood that H is held fixed, but h varies by changing M in (6.1).

Remarkably, it can be shown that the procedure given by (6.2) through (6.5) has the extraordinary property that the coefficients of the odd powers of $(1/M)$ in (6.7) all vanish!²¹ Thus, (6.7) actually has the form

$$\mathbf{y}(t^0 + H; M) - \mathbf{y}_e(t^0 + H) = \sum_{k=1}^{\infty} \mathbf{c}_{2k}(1/M)^{2k}. \quad (2.6.8)$$

To be honest, our discussion has been oversimplified. What is actually true is that there are asymptotic expansions of the form

$$\mathbf{y}(t^0 + H; M) - \mathbf{y}_e(t^0 + H) = \sum_{k=1}^{\infty} \mathbf{d}_{2k}(1/M)^{2k}, \quad M \text{ odd}; \quad (2.6.9)$$

²¹The choice of the integration procedure (6.2) through (6.5), called a *modified* midpoint rule because of the starting and ending steps (6.3) and (6.5), and the realization that this procedure would lead to the vanishing of all odd powers in (6.7), are due to *Gragg*.

$$\mathbf{y}(t^0 + H; M) - \mathbf{y}_e(t^0 + H) = \sum_{k=1}^{\infty} \mathbf{e}_{2k} (1/M)^{2k}, \quad M \text{ even.} \quad (2.6.10)$$

That is, the nature of the expansion depends on whether M is odd or even. (In view of this discovery, the assumption made in Exercise 5.1 requires proof!) The proof of this result is beyond the scope of this text, as also appears to be the case for many books on numerical analysis. However, it is proved in the Extrapolation Methods references listed at the end of the chapter. See also Exercise 6.1, which treats the special case for which $\mathbf{f}(\mathbf{y}, t)$ is, in fact, not dependent on \mathbf{y} .

Extrapolation

The background has now been provided to present remarkable ideas associated variously with the names *Richardson*, *Gragg*, *Bulirsch*, and *Stoer*. According to either (6.9) or (6.10) we have the result

$$\lim_{M \rightarrow \infty} \mathbf{y}(t^0 + H; M) = \mathbf{y}_e(t^0 + H), \quad (2.6.11)$$

as is desired for any integration scheme. But now suppose we evaluate $\mathbf{y}(t^0 + H; M)$ for a finite number of M values (all odd or all even), and from these results try to *extrapolate* to a limiting result for $M = \infty$. This process is an example of what is called *Richardson extrapolation*. Bulirsch and Stoer originally proposed that the extrapolation be based on the sequence of (even) M values given by the list

$$M = 2, 4, 6, 8, 12, 16, 24, \dots, (M_{j+2} = 2M_j \text{ when } j > 1). \quad (2.6.12)$$

Subsequent work by *Deuflhard* recommends using simply the even integers

$$M = 2, 4, 6, 8, \dots, (M_j = 2j). \quad (2.6.13)$$

In some realizations of the procedure the first few integers near the beginning of either list are discarded at some stage, and the extrapolation is based on the remaining larger integers.

One possible extrapolation method is to assume a polynomial fit of the form

$$\mathbf{y}(t^0 + H; M) = \mathbf{e}_0 + \sum_{k=1}^K \mathbf{e}_{2k} (1/M)^{2k} \quad (2.6.14)$$

in the $(K + 1)$ unknowns $\mathbf{e}_0, \mathbf{e}_2, \mathbf{e}_4, \dots, \mathbf{e}_{2K}$, which amounts to truncating the sum (6.10) at $k = K$. We then evaluate (6.14) for $(K + 1)$ different values of M (and hence h) selected from (6.12) or (6.13), and use the results to solve \mathbf{e}_0 . Finally, we make the extrapolation

$$\mathbf{y}_e(t^0 + H) \simeq \mathbf{e}_0. \quad (2.6.15)$$

According to (6.10) the error involved in this extrapolation should be on the order of $\mathbf{e}_{(2K+2)} (1/M_{\min})^{(2K+2)}$ where M_{\min} is the smallest M value used in the lists (6.12) or (6.13). Indeed, during the course of the extrapolation we have available (approximate) values for the coefficients $\mathbf{e}_2, \mathbf{e}_4 \dots, \mathbf{e}_{2K}$, and from these we can form the quantities $\mathbf{e}_{2k} (1/M_{\min})^{2k}$ for $k = 1, 2, \dots, K$. These quantities should approach zero as k increases, and we can use the

last few of them to estimate the error in (6.15). Alternatively (and preferably) we can solve (6.14) for \mathbf{e}_0 using successive values of K , beginning with $K = 1$, and observe how these values of \mathbf{e}_0 converge as K is allowed to increase.

The polynomial extrapolation method presupposes analyticity in h or, equivalently, analyticity in $(1/M)$. For differential equations whose right sides are analytic we expect, by Poincaré's theorem, that there will be analyticity along the real $(1/M)$ axis. However, there might be singularities somewhere off the real axis in the complex $(1/M)$ plane, and such singularities could affect the extrapolation process. Another extrapolation method, originally proposed by Bulirsch and Stoer, consists of using *rational function* or *Padé* approximation fits to $\mathbf{y}(t^0 + H; M)$ as a function of M rather than the fits of the form (6.14). Such a procedure should be more effective than polynomial extrapolation if there are pole singularities in the complex $(1/M)$ plane.²² Describing the use of rational function approximation will require some additional notation. As in (1.4.4), let y_j denote the j^{th} component of \mathbf{y} . For the case of even M , as in either (6.12) or (6.13), we make fits of the form

$$\begin{aligned} y_j(t^0 + H; M) &= \frac{p_j^{(0)} + p_j^{(2)}(1/M)^2 + p_j^{(4)}(1/M)^4 + \dots}{1 + q_j^{(2)}(1/M)^2 + q_j^{(4)}(1/M)^4 + \dots} \\ &= \left[\sum_{k=0}^L p_j^{(2k)}(1/M)^{2k} \right] / \left[1 + \sum_{k=1}^L q_j^{(2k)}(1/M)^{2k} \right]. \end{aligned} \quad (2.6.16)$$

These fits are called *diagonal* rational function approximations because the numerator and denominator have equal degree. For fixed L the relation (6.16) may be viewed as a fit in the $(2L+1)$ unknowns $p_j^{(0)}, p_j^{(2)}, p_j^{(4)}, \dots, p_j^{(2K)}, q_j^{(2)}, q_j^{(4)}, \dots, q_j^{(2K)}$. We next evaluate (6.16) for $(2L+1)$ different values of M , selected from (6.12) or (6.13), and solve for $p_j^{(0)}$. Finally, letting $\mathbf{p}^{(0)}$ denote a vector with components $p_j^{(0)}$, we make the extrapolation

$$\mathbf{y}_e(t^0 + H) \simeq \mathbf{p}^{(0)}. \quad (2.6.17)$$

If the rational function (6.16) were to be expanded as a Taylor series in $(1/M)^2$, we would get an expression whose initial coefficients might be expected to agree with those of (6.14) through $K = 2L$. Thus, we may expect the error in (6.17) to be of order $\mathbf{e}_{(4L+2)}(1/M_{\min})^{(4L+2)}$. As before we can estimate the error in $\mathbf{p}^{(0)}$ directly by solving (6.16) for successive values of L , beginning with $L = 1$, and observing how these values of $\mathbf{p}^{(0)}$ converge as L is allowed to increase.

The calculation of $\mathbf{p}^{(0)}$ using (6.16) is obviously more work than the calculation of \mathbf{e}_0 using (6.14). However, the rational function (6.16) might be expected to be a somewhat better fit to $\mathbf{y}(t^0 + H; M)$ than the polynomial (6.14) for small values of M , and therefore the convergence of the $\mathbf{p}^{(0)}$ for successive L might be expected to be somewhat better than that of the \mathbf{e} for corresponding K values. This has indeed been observed to be the case for a variety of differential equations. But is not yet clear whether the extra effort involved in rational function approximation is generally worth the improved convergence. Several authors find, for example applications they have examined, that it is not.²³

²²Padé was a student of Hermite.

²³It is interesting to note that in the Kepler problem there are singularities in the complex t plane, but they are branch points.

Finally, we observe that in either case the convergence is remarkably *fast*. Thanks to the occurrence of only even powers of $(1/M)$ in (6.9) or (6.14), we gain an extra power of $(1/M)^2$, which is equivalent to *two* powers of h , for each unit increase in K . When (6.16) is used, we gain an extra power of $(1/M)^4$, which is equivalent to *four* powers of h , for each unit increase in L . Of course, for a given K there are $(K + 1)$ values of M that must be used in (6.14) while for a given L value there are $(2L + 1)$ values of M that must be used in (6.16). Thus, apart from more refined considerations concerning behavior at small M values, the convergence rates of both extrapolation methods are roughly the same.

2.6.3 Summary

Looking back over what has been described so far, we have seen that the *order* of truncation in the single meso step that takes us from t^0 to $(t^0 + H)$ can be adjusted by the choice of K or L . Also, we clearly have the choice of H at our disposal, and therefore we may also adjust the macro step size at will. Finally, we have built-in error estimates based on the observed convergence of the extrapolation procedure. Thus, we have all the ingredients for a method that can adjust both order and step size dynamically. Typically, one chooses a macro step size H , and then begins an extrapolation process. The K or L values involved are successively increased until convergence is achieved within the specified error bounds. The program has specified maximum values of K or L , call them K_{\max} or L_{\max} , that are not allowed to be exceeded in this process in order to keep the process under control and in order to avoid excessive round-off error. If satisfactory convergence is not achieved within the allowed K or L values, the chosen meso step size is rejected, and the extrapolation process is tried again with a smaller step size. This process is repeated, if necessary, until convergence is finally achieved. When convergence is achieved, the results of this step are accepted and stored. Note is also made of the satisfactory meso step size H and the ease of convergence (the K or L values required to achieve the desired accuracy) of the extrapolation process. The size of the next meso step to be attempted is then selected based on this information, and the extrapolation process is begun anew.

2.6.4 Again, Advice to the Novice

As might be imagined (and just as for the case of the jet or multivalue methods described in the previous section), the procedure for implementing in detail the ideas of the previous paragraphs are quite involved. For this reason, potential users of extrapolation methods are advised to begin with existing programs written for this purpose; and then they should make modifications on these programs, if necessary, only when their algorithms and performance are well understood.

Exercises

2.6.1. The aim of this exercise is to examine some special cases for which the asymptotic expansion (6.10) can be verified explicitly. For this purpose we will need the following

identities:

$$S(M, 0) = \sum_{n=0}^M n^0 = \sum_{n=0}^M 1 = M + 1, \quad (2.6.18)$$

$$S(M, 1) = \sum_{n=0}^M n^1 = M^2/2 + M/2, \quad (2.6.19)$$

$$S(M, 2) = \sum_{n=0}^M n^2 = M^3/3 + M^2/2 + M/6, \quad (2.6.20)$$

$$S(M, 3) = \sum_{n=0}^M n^3 = M^4/4 + M^3/2 + M^2/4, \quad (2.6.21)$$

$$S(M, 4) = \sum_{n=0}^M n^4 = M^5/5 + M^4/2 + M^3/3 - M/30. \quad (2.6.22)$$

These identities can easily be found by the method of undetermined coefficients. Show that there is the recursion relation

$$S(M + 1, \ell) = S(M, \ell) + (M + 1)^\ell \quad (2.6.23)$$

with the starting condition

$$S(0, \ell) = \delta_{0,\ell}. \quad (2.6.24)$$

Make, for example, the Ansatz

$$S(M, 4) = AM^5 + BM^4 + CM^3 + DM^2 + EM \quad (2.6.25)$$

where the coefficients A through E are to be determined. Show that insertion of this Ansatz into the recursion relation (6.23) determines the coefficients A through E to yield the result (6.22).

In terms of the notation (1.4), the Gragg micro-step procedure (6.2) through (6.5) reads

$$h = H/M, \quad (2.6.26)$$

$$\boldsymbol{\eta}^0 = \mathbf{y}^0, \quad (2.6.27)$$

$$\boldsymbol{\eta}^1 = \boldsymbol{\eta}^0 + h\mathbf{f}^0, \quad (2.6.28)$$

$$\boldsymbol{\eta}^{m+1} = \boldsymbol{\eta}^{m-1} + 2h\mathbf{f}^m, \quad (2.6.29)$$

$$\mathbf{y}(t^0 + H; M) = (1/2)[\boldsymbol{\eta}^M + \boldsymbol{\eta}^{M-1} + h\mathbf{f}^M]. \quad (2.6.30)$$

For M even, show that the net result of this procedure is the relation

$$\mathbf{y}(t^0 + H; M) = \mathbf{y}^0 - (h/2)(\mathbf{f}^0 + \mathbf{f}^M) + h \sum_{n=0}^M \mathbf{f}^n. \quad (2.6.31)$$

Consider, for simplicity, the case where \mathbf{f} is one dimensional and of the simple form

$$f(y, t) = Nt^{N-1}. \quad (2.6.32)$$

That is, f is independent of y and has only a monomial dependence on t . When f is of the form (6.32), show that the exact solution to $\dot{y} = f$ is

$$\begin{aligned} y_e(t^0 + H) &= y^0 \text{ when } N = 0, \\ &= y^0 + H^N \text{ when } N > 0. \end{aligned} \quad (2.6.33)$$

Let us examine the results of using (6.31) in the cases $N = 0$ and $N = 1$. When $N = 0$, use of (6.32) gives $f^n = 0$ so that (6.31) yields the numerical result

$$y(t^0 + H; M) = y^0, \quad (2.6.34)$$

which agrees with the exact result. When $N = 1$, verify that use of (6.32) gives $f^n = 1$ and that use of (6.31) yields the numerical result

$$y(t^0 + H; M) = y^0 + Mh = y^0 + H, \quad (2.6.35)$$

which again agrees with the exact result.

To examine the cases $N > 1$, Suppose further that

$$t^0 = 0. \quad (2.6.36)$$

Then we have the general result

$$f^n = N(nh)^{N-1} = N(H/M)^{N-1}n^{N-1} \quad (2.6.37)$$

with the particular results

$$hf^0 = 0 \quad (2.6.38)$$

and

$$hf^M = (H/M)N(H/M)^{N-1}M^{N-1} = NH^N/M. \quad (2.6.39)$$

Correspondingly, show that (6.31) then takes the form

$$\begin{aligned} y(t^0 + H; M) &= y^0 - (1/2)N(H^N/M) + (H/M)N(H/M)^{N-1} \sum_{n=0}^M n^{N-1} \\ &= y^0 - (1/2)N(H^N/M) + N(H/M)^N \sum_{n=0}^M n^{N-1} \\ &= y^0 + H^N[-(1/2)(N/M) + N/M^N \sum_{n=0}^M n^{N-1}]. \end{aligned} \quad (2.6.40)$$

Evaluate (6.40) for the case $N = 2$ to show that there is again the result

$$y(t^0 + H; M) = y_e(t^0 + H). \quad (2.6.41)$$

We have learned that the numerical solution is exact for all the cases $N = 0, 1, 2$.

Now consider the case $N = 3$. Verify that in this case

$$\begin{aligned} N/M^N \sum_{n=0}^M n^{N-1} &= (3/M^3)[M^3/3 + M^2/2 + M/6] \\ &= 1 + 3/(2M) + 1/(2M^2). \end{aligned} \quad (2.6.42)$$

Correspondingly, show that

$$-(1/2)(N/M) + N/M^N \sum_{n=0}^M n^{N-1} = -3/(2M) + 1 + 3/(2M) + 1/(2M^2) = 1 + 1/(2M^2). \quad (2.6.43)$$

Consequently, show that in this case (6.40) takes the form

$$y(t^0 + H; M) = y^0 + H^3 + H^3/(2M^2) = y_e(t^0 + H) + H^3/(2M^2). \quad (2.6.44)$$

Next, for the cases $N = 4$ and $N = 5$, show that (6.40) gives the results

$$y(t^0 + H; M) = y^0 + H^4 + H^4/M^2 = y_e(t^0 + H) + H^4/M^2, \quad (2.6.45)$$

and

$$\begin{aligned} y(t^0 + H; M) &= y^0 + H^5 + H^5[5/(3M^2) - 1/(6M^4)] \\ &= y_e(t^0 + H) + 5H^5/(3M^2) - H^5/(6M^4), \end{aligned} \quad (2.6.46)$$

respectively. Observe that (6.44) through (6.46) are of the claimed form (6.10).

Finally, show that the assumption (6.36) can be dropped and that f can be any polynomial of degree 4 in t without changing the conclusions of this exercise: the result (6.10) still holds.

We close this exercise with an important remark. Observe that (6.31) can be rewritten in the form

$$y(t^0 + H; M) = \mathbf{y}^0 + (h/2)(\mathbf{f}^0 + \mathbf{f}^1) + (h/2)(\mathbf{f}^1 + \mathbf{f}^2) + \cdots + (h/2)(\mathbf{f}^{M-1} + \mathbf{f}^M), \quad (2.6.47)$$

and that each term in parentheses on the right side (6.47) is the result of applying the *trapezoidal rule* over an interval of duration h . It is known, say from the *Euler-Maclaurin sum formula* (see Exercise 4.15), that the error associated with this *extended trapezoidal rule* has the properties (6.10) for any polynomial and, by extension, any analytic function. Thus, you have explicitly verified specific cases of a general result.

2.7 Things Not Covered

We have given the rudiments of numerical integration, and their mastery should provide sufficient knowledge to handle most problems. However, there are several additional topics whose study we commend to the reader who wishes to become truly expert. We list them below along with brief explanatory paragraphs. Further detail may be found in the books listed at the end of the chapter.

2.7.1 Størmer-Cowell and Nyström Methods

The differential equations of classical mechanics often contain only second derivatives with *no* first derivatives present. In this case it is possible to work directly with the second-order equations instead of converting them into a first-order set of twice the dimensionality. The result can be a saving in computer time and an increase in accuracy. Appendix A describes a predictor-corrector method due to *Størmer* and *Cowell* and modified Runge-Kutta methods due to *Nyström* that have this feature.

2.7.2 Other Starting Procedures

In this chapter we have always started Adams' solutions using Runge-Kutta. Other techniques, such as the use of Taylor series or various iterative processes, are also sometimes used. Of course, starting procedures are not required for the methods of Sections 2.5 and 2.6.

2.7.3 Stability

An introductory discussion of order, stability, and convergence was given at the beginning of Section 2.4. Much more can be found on the subject in some of the references listed at the end of this chapter. The finite-difference equations of Adams and Størmer-Cowell are special examples of the whole class of multistep/multivalue equations. To recapitulate, in general a multistep/multivalue equation (when applied to a linear differential equation) has several solutions, and only one of these solutions approximates the solution of the differential equation being integrated. It is important to be sure that the other so called *parasitic* solutions do not enter the calculation and eventually swamp the main solution. The reader should be warned that many of the numerical methods described in older books, such as *Milne's* and *Nyström's* multistep/multivalue methods, have parasitic solutions that grow exponentially. Thus, if a small amount of a parasitic solution happens to be introduced due to round-off errors or improper initial conditions, it will soon grow to the point where it completely dominates the main solution, and the accuracy of the numerical solution is completely destroyed. By contrast, the parasitic solutions in Adams and Størmer-Cowell are exponentially damped (if the step size is small enough) so that even if they happen to enter a calculation, their effect rapidly dies away. But there is a complication: the higher the order the smaller this step size must be to guarantee stability. For example, when integrating the simple harmonic oscillator with unit frequency ($x'' + x = 0$) using the `adams10` method given in Appendix B, at least 50 steps per oscillation are required before stability is safely achieved and the error analysis of Section 2.4.2 becomes relevant. Finally, if a multistep/multivalue method cannot integrate a linear differential equation well, it is unlikely to be able to integrate more complicated nonlinear differential equations well.

2.7.4 Regularization, Etc.

We have already discussed in Sections 2.5 and 2.6 something about the choice of step size h and how it may be varied during the course of integration. An alternative procedure to

making frequent changes in h is to analytically regularize the equations of motion before integration by the introduction of a new independent variable in place of the time. This can often be done while remaining within a Hamiltonian framework.²⁴ See Exercise 1.6.5 and the regularization references at the end of this chapter. It is known, for example, how to regularize the Kepler problem and the Størmer problem (the problem of finding the motion of a charged particle in the external field of a point magnetic dipole, of interest for Van Allen radiation).²⁵ We also mention that in some cases it is worthwhile to change rather radically the form of the differential equation by introducing new dependent variables. For example, if a solution $y(t)$ is known to be highly oscillatory, one should try making an *eikonal* or *Madelung* transformation by writing $y(t) = a(t) \sin b(t)$ and then integrating the differential equations for a and b .

Differential equations whose solutions contain both rapidly and slowly varying terms are colorfully referred to by numerical analysts as being *stiff*. The presence of a rapidly varying part forces the integration time step to be small when the usual integration methods are used. But the features of physical interest may reside in the slowly varying part, and thus to explore these features one may be forced to integrate for many very small steps. In the case that a stiff equation cannot be regularized easily, it may be possible to use directly certain integration methods devised especially for stiff equations. These methods are beyond the scope of this text, but are described in some of the references.

2.7.5 Solutions with Few Derivatives

Our discussion has always assumed that $\mathbf{y}(t)$ has a large number of continuous derivatives. Although this is true for many problems, there are important examples where this is not the case. Consider a space ship outside the Earth's atmosphere. As long as it is subject only to gravitational forces, it can be shown that its trajectory vector $\mathbf{r}(t)$ has arbitrarily many continuous derivatives. However, suppose the space ship's rocket engine is fired at a time t_f . Then, according to Newton, $\ddot{\mathbf{r}}(t)$ is discontinuous at t_f . To handle this situation numerically, one possible procedure is to terminate any finite difference scheme slightly before t_f and integrate through t_f using Runge-Kutta. The Runge-Kutta routine should be used in such a way that an integration step is initiated at t_f and at any other time at which the rocket thrust either changes discontinuously or has discontinuous changes in its first few time derivatives.

2.7.6 Symplectic and Geometric/Structure-Preserving Integrators

We will see in Chapter 6 that Hamiltonian systems have special properties. Their transfer maps are symplectic. Symplectic integrators are integrators specifically constructed to

²⁴It is also possible in some cases to arrange, by a suitable change of variables, that the final Hamiltonian will be of the form $T(p) + V(q)$. Sometimes one can also arrange that $T(p) = p \cdot p/2$. Hamiltonians of this form are desirable because there are special integration methods for them that are particularly efficient. See Chapter 12 and Appendix A.

²⁵Surprisingly, there is a regularization transformation that converts the Kepler problem to that of a two-dimensional harmonic oscillator.

preserve these properties. They produce maps that, while still approximations to the exact transfer map, are at least exactly symplectic. Symplectic integrators are an example of so-called *geometric* integrators. A second example is integration on *manifolds*. Many differential equations have the property that their solutions lie on manifolds, often manifolds associated with groups. In this case one seeks numerical integration methods that, despite truncation errors, still guarantee that the numerical solutions they generate also lie on the these manifolds. Extensive work has been done on both these aspects of geometric integration. See Chapters 11 and 12.

2.7.7 Error Analysis

In discussing Runge-Kutta errors, we have given only their expected order in h without any mention of the coefficient multiplying h^m . The analysis of the local truncation error committed in Runge-Kutta is considerably more complicated than in the case of predictor-corrector methods. Estimates, however, are available. Of course, one may also use the methods for error estimation described at the beginning of Section 2.5. A more complicated question with regard to both Runge-Kutta and finite difference methods is how an error propagates through successive time steps after its initial introduction. This question is particularly difficult with regard to round-off error, and is still a topic of study.

One way to reduce round-off error with only a small increase in machine time is to use *partial double precision*. In this method \mathbf{f} is evaluated with the usual number of significant figures. However, in the Adams' routine for example, the addition (4.69) is carried out with additional significant figures and the \mathbf{y}^n are stored with additional significant figures. (See Appendix B for an analogous treatment of the Runge-Kutta routine RK3.) Of course even a further reduction in round-off error is realized if all calculations are carried out in *double precision*, i.e. using twice the usual number of significant figures. But the use of full double precision may require considerably more computer time.

It is often difficult to ascertain rigorously the total error at the end of a long integration run. However, there are several informal procedures. In the case of fixed step size methods, one procedure is to make several runs with different values of h , and then study how the \mathbf{y} 's at the end of the trajectory depend upon h . The magnitude of the error should at first decrease with decreasing h , and then again increase due to round-off error. For variable step size methods, one can change the specified error to see what effect it has on the solution. Another procedure is to first integrate a trajectory forward in time, and then reintegrate it backward to see how close one comes to the original initial conditions.²⁶ In the case that the differential equations have known constants of motion such as energy or angular momentum, one can and always should check to see to what extent they are actually preserved by the numerical solution. Finally, the accuracy of an integration routine always should be checked on equations whose solutions are known exactly. These equations should include both those leading to oscillatory functions such as sines and cosines and those leading to growing and damped exponentials.

²⁶To integrate backwards, simply replace h by $-h$. Truncation errors associated with forward integration followed by backward integration are not expected to cancel unless the integration method is *symmetric*. See Section 12.1. In any case, round-off errors are not expected to cancel.

2.7.8 Backward Error Analysis

Suppose x is some input, g is some function, and we wish to compute $g(x)$. If g is a complicated function, as is often the case, the best that we are able or willing to do is to compute some approximating function \hat{g} . What is then called the *forward* error associated with such a computation is the difference $[\hat{g}(x) - g(x)]$. Turn the situation around. We may ask if there is a modified input \bar{x} near x such that $g(\bar{x})$ gives the result $\hat{g}(x)$. That is, there is the requirement that the exact calculation applied to the modified input should agree with the approximate calculation applied to the original input: $g(\bar{x}) = \hat{g}(x)$. We would then call the difference $[\bar{x} - x]$ the *backward* error.

Somewhat the same philosophy may be applied to the numerical integration of ordinary differential equations. Suppose, as in Section 2.1, we are given as input the vector $\mathbf{f}(\mathbf{y}, t)$ to be used as the right side of an ordinary differential equation, the vector \mathbf{y}^0 to be used as an initial condition, and the quantity h to be used as a step size. We wish to compute the vectors \mathbf{y}^n . What we actually accomplish, because of truncation error, is the computation of a set of approximate vectors $\hat{\mathbf{y}}^n$. (Here we assume that round-off error is negligible.) Instead of examining the forward error vectors $[\hat{\mathbf{y}}^n - \mathbf{y}^n]$, we might ask if there is a modified differential equation with right side $\tilde{\mathbf{f}}(\mathbf{y}, t; h)$ (which will in general depend on h) and exact solution $\bar{\mathbf{y}}^n$ such that $\bar{\mathbf{y}}^n = \hat{\mathbf{y}}^n$. That is, the exact solution of the modified differential equation should agree with the approximate solution of the original differential equation at the times t^n . For some integration methods it can be shown that it is indeed possible to find such a modified differential equation. Moreover, in some cases there is the further possibility of modifying the original differential equation [its right side becomes $\tilde{\mathbf{f}}(\mathbf{y}, t; h)$] so that its approximate solution agrees with the exact solution of the original differential equation (well, not perfectly, but in principle to any desired finite order in h). That is, the original differential equation can be modified in such a way as to compensate (at least to any desired finite order in h) for the errors produced by the integration method.

These considerations are of particular interest for symplectic integrators applied to Hamiltonian differential equations. In that case it can be shown that the use of a symplectic integrator produces the exact solution of some modified Hamiltonian differential equation. Let $\mathcal{S}_{\text{exact}}$ be an integrator that solves any differential equation exactly. It could be viewed as the result of using any integrator in the limit that $h \rightarrow 0$, correspondingly the number of integration steps becomes indefinitely large, and all results are carried out with unlimited precision. Also, let $\mathcal{S}_{\text{approx}}$ be some integrator that is correct only through some order in h , but is exactly symplectic. Then, according to the discussion above, we have the result

$$\mathcal{S}_{\text{approx}}(H) \equiv \mathcal{S}_{\text{exact}}(H_{\text{mod}}). \quad (2.7.1)$$

Here the notation $\mathcal{S}_{\text{approx}}(H)$ denotes the trajectory that results from integrating the equations of motion associated with H using the integrator $\mathcal{S}_{\text{approx}}$, the notation $\mathcal{S}_{\text{exact}}(H_{\text{mod}})$ denotes the trajectory that results from integrating the equations of motion associated with H_{mod} using the integrator $\mathcal{S}_{\text{exact}}$, and the symbol \equiv means *equivalent to*. If the modified Hamiltonian H_{mod} is deemed to be sufficiently close to the original Hamiltonian H in some sense (small backward error), then the results of symplectic integration might also be deemed to have some special merit.

Moreover, if $H(q, p, t)$ is the Hamiltonian of interest and some particular symplectic

integrator is being used, one might try to find a modified Hamiltonian $\tilde{H}(q, p, t; h)$ such that symplectically integrating its equations of motion produces results that are closer to the exact results for the original Hamiltonian H . That is, if we can master relationships of the form (7.1), then we might be able to arrange the relation

$$\mathcal{S}_{\text{approx}}(\tilde{H}) \equiv \mathcal{S}_{\text{exact}}(H), \quad (2.7.2)$$

at least through terms of some high order in h .

2.7.9 Comparison of Methods

There is an extensive literature comparing the virtues of various integration methods and computer codes. The matter is complicated. The criteria for which method is “best” vary from problem to problem, and may also be machine dependent. Moreover, the manner in which a particular method is implemented also affects the over-all performance of a computer code. A typical discussion of such matters is given in the review article of *Shampine et al.* For relatively simple problems, those for which the characteristic time scale varies relatively little over a trajectory or those which can be regularized, the fixed step Adams’ method started with Runge-Kutta is satisfactory, easy to program, and easy to use. More difficult problems may well benefit from the use of jet or extrapolation methods as described in Sections 2.5 and 2.6. In this case one is well advised to begin with professionally written programs. These programs should, however, be used with care and understanding. Some may produce unpleasant surprises. (For example, it is not uncommon that programs which automatically adjust step size have difficulties with some kinds of problems.) At present there is some indication and fairly widespread opinion that extrapolation methods (at least when high accuracy is required) may well be the method of choice for a wide variety of problems. Seek advice from a local Computer Center if it has resident experts. Much work has gone into writing good integration programs.

Bibliography

Books on General Numerical Analysis

- [1] F.B. Hildebrand, *Introduction to Numerical Analysis*. (McGraw-Hill 2nd Edition, 1974) QA 297.H54. A standard reference book on numerical analysis, which has recently been reprinted by Dover.
- [2] J. Todd (editor), *Survey of Numerical Analysis*. (McGraw-Hill 1962) QA 297.T6. Contains a chapter on differential equations with numerous references.
- [3] S.D. Conte, *Elementary Numerical Analysis*. (McGraw-Hill 1965) QA 297.C62. Describes use of partial double precision.
- [4] B. Carnahan et al., *Applied Numerical Methods*. (John Wiley 1969) QA 297.C34. Gives Fortran programs.
- [5] L. Collatz, *Functional Analysis and Numerical Mathematics*. (Academic Press 1966) QA 297.C58. Discusses the theoretical aspects of numerical analysis.
- [6] F.S. Acton, *Numerical Methods That (Usually) Work*. (Mathematical Association of America 1990) QA 297.A33. Discusses Madelung and other transformations.
- [7] L.B. Rall, ed., *Error in Digital Computation*. Vol. 1, QA 3.U45 No. 14 (Wiley 1965). Contains a chapter by P. Henrici on error in the integration of differential equations, and gives an extensive bibliography.
- [8] R.W. Hamming, *Numerical Methods for Scientists and Engineers*. (McGraw-Hill 1962) QA 297.H28.
- [9] J.M. Ortega, *Numerical Analysis; a Second Course*. (Academic Press 1972) QA 297.078.
- [10] G.N. Lance, *Numerical Methods for High Speed Computers*. (Iliffe and Sons 1960) QA 76.L27.
- [11] A. Ralston and H.F. Wilf, ed., *Mathematical Methods for Digital Computers*, Vol. 1 and 2. (Wiley 1960) QA 76.5.R3, Vol. 1 and 2.
- [12] A. Ralston, *A First Course in Numerical Analysis*. (McGraw-Hill 1965) QA 297.R3.

- [13] I.S. Berezin and N.P. Zhidkov, *Computing Methods* (2 Vols), QA 297.B4213, 1965 (Pergamon Press and Addison-Wesley 1965). A thorough, readable, and scholarly presentation. Some of its predictor and corrector coefficient listings contain typographical errors.
- [14] E. Isaacson and H.B. Keller, *Analysis of Numerical Methods*, John Wiley (1966) and Dover (1994).
- [15] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes, the Art of Scientific Computing*, Cambridge University Press (2003). These authors are enthusiastic about Richardson Extrapolation and the Bulirsch-Stoer Method.
- [16] D. Kahaner, C. Moler, and S. Nash, *Numerical Methods and Software*, Prentice Hall (1989).
- [17] G. Forsythe, C. Moler, and M. Malcom, *Computer Methods for Mathematical Computations*, Prentice Hall (1977). See also the Web site <http://www.pdas.com/fmm.html>.
- [18] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, third edition, Springer-Verlag (2002).

Books and Articles on the Numerical Solution of Differential Equations

- [19] L. Fox, *Numerical Solution of Ordinary and Partial Differential Equations*. (Addison-Wesley 1962) QA 371.L758, 1962. Also discusses integral equations.
- [20] L. Collatz, *The Numerical Treatment of Differential Equations*, Springer-Verlag (1966). A standard reference work.
- [21] W.E. Milner, *Numerical Solution of Differential Equations*. (John Wiley 1953 and Dover 1970) QA 371.M57, 1970. A standard older work and a bargain in paperback.
- [22] P. Henrici, *Discrete Variable Methods in Ordinary Differential Equations*, Wiley (1962). P. Henrici, *Error Propagation for Difference Methods*, Wiley (1963). These two books are the standard works on round-off and truncation error and their propagation. The first book gives Nyström's versions of Runge-Kutta. They are the Runge-Kutta analog of Störmer-Cowell in that they work directly with second-order equations when first derivatives are absent.
- [23] I. Babuska et al., *Numerical Processes in Differential Equations*, John Wiley (1966) QA 371.B2313. Gives numerical examples of the effect of round-off error.
- [24] F. Ceschino and J. Kuntzmann, *Numerical Solution of Initial Value Problems*, Prentice-Hall (1966). Contains useful tabulations of coefficients for various integration schemes including very high-order Runge-Kutta.
- [25] G.A. Chebotarev, *Analytical and Numerical Methods of Celestial Mechanics*, American Elsevier (1967). Describes the use of perturbation series in practical celestial mechanics.

- [26] C.W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice-Hall (1971). Describes Richardson extrapolation and general “extrapolation methods”. Also describes methods of automatic change of order and step size.
- [27] J.C. Butcher, *The numerical analysis of ordinary differential equations: Runge-Kutta and general linear methods*, John Wiley (1987).
- [28] J.C. Butcher, *Numerical Methods for Ordinary Differential Equations*, First Edition, John Wiley (2003).
- [29] J.C. Butcher, *Numerical Methods for Ordinary Differential Equations*, Second Edition, John Wiley (2008). <http://www.math.auckland.ac.nz/~butcher/ODE-book-2008/>.
- [30] J.C. Butcher, “Runge-Kutta Methods”, *Scholarpedia* (2011). http://www.scholarpedia.org/article/Runge-Kutta_methods.
- [31] G. Hall and J.M. Watt (eds.), *Modern numerical methods for ordinary differential equations*, Oxford University Press, Oxford (1976).
- [32] L.F. Shampine and M.K. Gordon, *Computer Solution of Ordinary Differential Equations: The Initial Value Problem*, W.H. Freeman, San Francisco (1975).
- [33] L.F. Shampine, *Numerical Solution of Ordinary Differential Equations*, Chapman and Hall (1994).
- [34] J.D. Lambert, *Computational methods in ordinary differential equations*, Wiley, New York (1973).
- [35] J.D. Lambert, *Numerical Methods for Ordinary Differential Systems: The Initial Value Problem*, Wiley (1991).
- [36] H. Stetter, *Analysis of Discretization Methods for Ordinary Differential Equations*, Springer Verlag (1973).
- [37] S.O. Fatunla, *Numerical methods for initial value problems in ordinary differential equations*, Academic Press, London (1989).
- [38] E. Hairer, S. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations I. Non-stiff Problems*, Springer (1993).
- [39] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II. Stiff and Differential Algebraic Problems*, Springer (2002).
- [40] A. Iserles, *A First Course in the Numerical Analysis of Differential Equations*, Second Edition, Cambridge University Press (2009).
- [41] D. F. Griffiths and D. J. Higham, *Numerical Methods for Ordinary Differential Equations: Initial Value Problems*, Springer (2010).

- [42] J.B. Rosser, “A Runge-Kutta for All Seasons”, *SIAM Review* **9**, 417-452 (1967). Describes an improved Runge-Kutta method.
- [43] P. Deuflhard and F. Bornemann, *Scientific Computing with Ordinary Differential Equations*, Springer (2002).
- [44] L.F. Shampine, H.A. Watts, and S.M. Davenport, “Solving Nonstiff Ordinary Differential Equations - The State of the Art”, *SIAM Review* **18**, 376-411 (1976). Compares merits of various methods and computer codes.
- [45] A.M. Stuart and A.R. Humphries, *Dynamical Systems and Numerical Analysis*, Cambridge University Press (1996).
- [46] A.C. Hindmarsh, “ODEPACK: A Systematized Collection of ODE Solvers”, in *Scientific Computing*, R.S. Stepleman et al. eds., North-Holland, Amsterdam (1983).
- [47] L.R. Petzold, “Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations”, *SIAM Journal of Scientific and Statistical Computing* **4**, pp. 136-148 (1983).
- [48] S. Herrick, *Astrodynamicics*, vols. 1 and 2, Van Nostrand Reinhold (1971).
- [49] J. Dormand and P. Prince, “A family of embedded Runge-Kutta formulae”, *J. Comp. Appl. Math.* **6**, 19-26 (1980).
- [50] M. Sofroniou and G. Spaletta, “Construction of explicit runge-kutta pairs with stiffness detection”, *Mathematical and Computer Modelling* **40**, 1157-1169 (2004).
- [51] G. Wanner, “Germund Dahlquist’s classical papers on Stability Theory”, <http://www.unige.ch/~wanner/DQsem.pdf>.
- [52] The program *Mathematica* provides a command `NDSolve` that implements many different integration methods including Runge-Kutta, Adams, and extrapolation. For a discussion of Runge-Kutta in *Mathematica*, Google “`NDSolve` explicit Runge-Kutta” and follow related links.

Extrapolation Methods

- [53] R. Bulirsch and J. Stoer, “Numerical Treatment of Ordinary Differential Equations by Extrapolation Methods”, *Numerische Mathematik* **8**, 1-13, 93-104 (1966).
- [54] P. Deuflhard, “Order and Step-size Control in Extrapolation Methods”, *Numerische Mathematik* **41**, 399-422 (1983).
- [55] P. Deuflhard, “Recent Progress in Extrapolation Methods for Ordinary Differential Equations”, *SIAM Review* **27**, 505-535 (1985).
- [56] E. Hairer, S. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations I. Non-stiff Problems*, Springer (1993).

- [57] E. Hairer and C. Lubich, "Asymptotic expansions of the global error of fixed stepsize methods", *Numer. Math.* **45**, 345-360 (1984).
- [58] H. Stetter, *Analysis of Discretization Methods for Ordinary Differential Equations*, Springer Verlag (1973).
- [59] H. Stetter, "Symmetric Two-Step Algorithms for Ordinary Differential Equations", *Computing* **5**, 267-280 (1970).
- [60] P. Deuflhard and F. Bornemann, *Scientific Computing with Ordinary Differential Equations*, Springer (2002).

References to Regularization of Kepler and Størmer Problems

- [61] E.L. Stiefel and G. Scheifele, *Linear and Regular Celestial Mechanics*, Springer Verlag (1971).
- [62] D. Boccaletti and G. Pucacco, *Theory of Orbits*, 2 vols., Springer-Verlag (1996). This excellent 2-volume set is full of interesting material including a discussion of Lie methods.
- [63] A.J. Dragt, Trapped Orbits in a Magnetic Dipole Field, *Rev. Geophys.* **3**, p. 255-298 (1965).
- [64] A.J. Dragt and J.M. Finn, Insolubility of Trapped Particle Motion in a Magnetic Dipole Field, *J. Geophys. Res.* **81**, p. 2327-2340 (1976).
- [65] A.J. Dragt and J.M. Finn, Normal Form for Mirror Machine Hamiltonians, *J. Math. Physics* **20**, p. 2649-2660 (1979).

References to Symplectic Integration and Backward Error Analysis

See also the references at the end of Chapters 11 and 12.

- [66] J.M. Sanz-Serna and M.P. Calvo, *Numerical Hamiltonian Problems*, Chapman and Hall (1994) or Dover (1994).
- [67] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, Second Edition, Springer (2010).
- [68] A. Iserles, H. Munthe-Kaas, S. Nørsett, and A. Zanna, "Lie-group methods", *Acta Numerica* **14** (2005), pp. 1-148, Cambridge University Press, 1999.
- [69] P. Chartier, E. Hairer, and G. Vilmart, "Numerical integrators based on modified differential equations", *Mathematics of Computation* S 0025-5718(07)01967-9.
- [70] Sebastian Reich, "Backward error analysis for numerical integrators", *SIAM J. Numer. Anal.* **36**, 1549-1570 (1999).

- [71] E. Hairer and C. Lubich, “The Life-Span of Backward Error Analysis for Numerical Integrators”, *Numer. Math.* **76**, 441-462 (1996).

Manuals on Computer Programming

- [72] D.D. McCracken, *A Guide to FORTRAN IV Programming*. QA 76.5M1872 (Wiley 1965).

- [73] E.I. Organick, *A FORTRAN IV Primer*. QA 76.5.072 (Addison-Wesley 1966).

- [74] G.B. Davis and T.R. Hoffman, *FORTRAN 77, A Structured, Disciplined Style* (McGraw-Hill 1988).

- [75] American National Standard Programming Language FORTRAN (77), American National Standards Institute (ANSI), New York, NY (1978).

Original Sources

- [76] C. Runge, “Über die numerische Auflösung von Differentialgleichungen”, *Math. Ann.* **46**, 167-178 (1895).

- [77] G. Coriolis, “Mémoire sur le degré d’approximation qu’on obtient pour les valeurs numériques d’une variable qui satisfait à une équation différentielle, en employant pour calculer ces valeurs diverses équations aux différences plus ou moins approchées”, *Journal de Mathématiques Pures et Appliquées* **2**, 229-244 (1837). This paper reveals that some of what we now call Runge-Kutta methods were, in fact, known much earlier to Coriolis.

Chapter 3

Symplectic Matrices and Lie Algebras/Groups

Lie theory is in the process of becoming the most important part of modern mathematics. Little by little it became obvious that the most unexpected theories, from arithmetic to quantum physics, came to encircle this Lie field like a gigantic axis.

Jean Dieudonne

We will learn in subsequent chapters that symplectic matrices play an important role in the advanced treatment of Hamiltonian systems. Briefly put, Hamiltonian motion produces symplectic maps. Also, symplectic maps preserve the Hamiltonian form of the equations of motion. Finally, symplectic maps are characterized by symplectic matrices. The purpose of this chapter is to define symplectic matrices and to explore some of their properties in preparation for future use. This exploration also provides a context for the discussion of Lie algebras and Lie groups.

In his youth, and for publishing his first mathematical paper (1869), Sophus Lie received a travel grant from the Norwegian University of Christiania to visit the mathematical capitals of Europe. One such capital was Paris where he visited and worked with Klein, who himself was visiting there from Prussia.

While Lie and Klein thought deeply about mathematics in Paris, the political situation between France and Prussia deteriorated. The popularity of the French emperor Napoleon III was declining and he thought war with Prussia, which his advisors said the French army was sure to win, might change his political fortunes. Bismarck, the Prussian chancellor, saw a war with France as an opportunity to unite the South German states. With both sides feeling that a war was to their advantage, the Franco-Prussian war became inevitable. On July 14, 1870, Bismarck sent a telegram which infuriated the French government. On July 19, France declared war on Prussia. For Klein there was then only one possibility: he had to return quickly to Berlin.

However, Lie was a Norwegian and he was finding mathematical discussions in Paris very stimulating. He decided to remain there, but became anxious as the German offensive met with only ineffectual French response. In August, when the German army trapped part of the French army in Metz, Lie decided it was also time for him to leave Paris, and he planned

to hike (on foot!) to Italy. He made as far as Fontainebleau, just south of Paris, when the French police spotted him as a suspicious-looking young man wandering in lonely places in the forest, stopping now and then to make notes and drawings in his notebook. “*He was of tall stature and had the classic Nordic appearance. A full blond beard framed his face and his grey-blue eyes sparkled behind his glasses. He gave the impression of unusual physical strength*” (Élie Cartan). The police searched him and found a map, letters in German, and papers full of mysterious formulas, complexes, diagrams, and names. He was suspected of being a German spy and imprisoned.

Lie had to stay in prison in Fontainebleau for 4 weeks before his French colleague Gaston Darboux learned about the incident and arrived on behalf of the French Academy of Sciences with a release order signed by the Minister of Home Affairs. Lie himself had taken things truly philosophically and made good use of his time in prison. For, as he recounted later, in these forced leisure days he had plenty of peace and quiet to concentrate on his problems and advance them essentially. In a letter to his Norwegian friend Ernst Motzfeldt, written directly after his release, Lie remarked: “*I think that a mathematician is well suited to be in prison.*”

The French army surrendered on September 2 but, after a September 4 coup d'état against Napoleon III, France resumed the war on September 6. On September 19 the German army began to blockade Paris. This time Lie successfully fled to Italy, then from there he made his way back to Christiania via Germany so that he could again meet and discuss mathematics with Klein. Thus began the work of Sophus Lie.¹

3.1 Definitions

To define symplectic matrices, it is first necessary to introduce a certain *fundamental* $2n \times 2n$ antisymmetric matrix J . It is defined by the equation

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}. \quad (3.1.1)$$

Here each entry in J is an $n \times n$ matrix, I denotes the $n \times n$ identity matrix, and all other entries are zero. The observant reader will recognize this J as the Poisson matrix already defined in Section 1.7 in connection with Poisson brackets.

With this background, a $2n \times 2n$ matrix M is said to be *symplectic* if

$$M^T JM = J. \quad (3.1.2)$$

Here M^T denotes the transpose of M . Observe that symplectic matrices must be of even dimension by definition. Usually we will be interested in real symplectic matrices. However, in some cases we will also be interested in symplectic matrices with complex entries.

Finally, we remark that the use of the adjective *symplectic* in this general context is due to Hermann Weyl (1885–1955). *Symplectic*, the Greek equivalent of the Latin-based word

¹See the Web site <http://www-history.mcs.st-andrews.ac.uk/Biographies/Lie.html>. See also the “Overview and History of the Theory of Lie Algebras and Lie Groups” references given at the end of Chapter 27.

complex, comes from $\sigma\nu\mu\pi\lambda\epsilon\kappa\tau\iota\kappa\circ\varsigma$, which means *intertwined* or *braided*. Weyl had in mind the *symplectic 2-form* associated with J when introducing this adjective.² We may view it as intertwining the components of two vectors, call them w and z , with the components of J . See (2.3). We may also view (1.2) as an intertwining of J with M^T and M .

Exercises

3.1.1. Show that the matrix J has the following properties:

$$J^T = -J, \quad (3.1.3)$$

$$J^2 = -I \text{ or } J^{-1} = -J, \quad (3.1.4)$$

$$\det(J) = 1, \quad (3.1.5)$$

$$J^T J = J J^T = I. \quad (3.1.6)$$

3.1.2. Suppose that $n = 1$ (in which case J is 2×2) and suppose A is any 2×2 matrix. Write A in the form

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \quad (3.1.7)$$

Then A has determinant

$$\det(A) = ad - bc. \quad (3.1.8)$$

Verify that

$$A^{-1} = [1/\det(A)] \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \quad (3.1.9)$$

and

$$-JA^TJ = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \quad (3.1.10)$$

Verify that

$$A^TJA = [\det(A)]J. \quad (3.1.11)$$

3.1.3. By taking the determinant of both sides of (1.2), show that any symplectic matrix M has the property

$$\det(M) = \pm 1. \quad (3.1.12)$$

It follows that symplectic matrices are always invertible.

Comment: It will be shown in Subsection 3.3 that $\det(M)$ actually always equals $+1$ for a symplectic matrix. Also, as is easily seen from (1.11), in the 2×2 case the necessary and sufficient condition for a matrix to be symplectic is that it have determinant $+1$.

²Weyl would have preferred to use the Latin-based word *complex* because the vanishing of the 2-form defines what is called a line complex, and even did so for a time. However he abandoned this usage because of the confusion it created with complex numbers. It is also of historical interest to note that the term “Lie algebra” itself was first introduced in 1934 by Weyl. Prior to that time terms like “infinitesimal group” had been employed.

3.1.4. Show that any symplectic matrix M has the following properties:

$$M^{-1} = -JM^TJ = J^{-1}M^TJ = JM^TJ^{-1}, \quad (3.1.13)$$

$$MJM^T = J, \quad (3.1.14)$$

$$(M^{-1})^T = -JMJ. \quad (3.1.15)$$

Note that for a symplectic matrix (1.13) makes it possible to compute M^{-1} using only matrix operations without computing the determinant and minors of M .

3.1.5. Show that the matrices I , J , and $(1/\sqrt{2})(I \pm J)$ are symplectic.

3.1.6. Suppose M is a symplectic matrix. Show that $\pm M$, $\pm M^{-1}$, and $\pm M^T$ are then also symplectic matrices.

3.1.7. Suppose M and N are symplectic matrices. Show that the product MN is then also a symplectic matrix. Taken together, Exercises 1.5, 1.6, and this exercise show, among other things, that the set of all $2n \times 2n$ symplectic matrices forms a *group*. See Section 3.6.

3.1.8. Show that a symplectic matrix cannot have $\lambda = 0$ as an eigenvalue.

3.1.9. Let M be any $2n \times 2n$ matrix. Define its *symplectic transpose* M^S by the rule

$$M^S = JM^TJ^{-1}. \quad (3.1.16)$$

Show that, similar to the case for the ordinary transpose, there are the relations

$$I^S = I, \quad J^S = -J, \quad (M^S)^S = M, \quad (MN)^S = N^S M^S. \quad (3.1.17)$$

Show that the symplectic condition (1.2) can be written in the form

$$M^S M = M M^S = I. \quad (3.1.18)$$

3.1.10. Here are some things it is assumed you know about matrices: Let A and B be any two $n \times n$ matrices. The determinant function has the properties $\det(A^T) = \det(A)$ and $\det(AB) = \det(A)\det(B)$. The trace function has the properties $\text{tr}(A^T) = \text{tr}(A)$ and $\text{tr}(AB) = \text{tr}(BA)$. The matrix A has an inverse, which is unique and is both a left and right inverse, iff $\det(A) \neq 0$. There is the relation $\det(A^{-1}) = [\det(A)]^{-1}$. The transposition, Hermitian conjugation, and inversion operations have the properties $(AB)^T = B^T A^T$, $(AB)^\dagger = B^\dagger A^\dagger$, $(AB)^{-1} = B^{-1} A^{-1}$. The operations of inversion and transposition, and inversion and Hermitian conjugation, commute: $(A^T)^{-1} = (A^{-1})^T$ and $(A^\dagger)^{-1} = (A^{-1})^\dagger$. If these results are unfamiliar to you, consult the *Matrix Theory* references provided at the end of this chapter.

3.1.11. Suppose K is a real 2×2 matrix. Find all real solutions to the equation

$$K^2 = -I. \quad (3.1.19)$$

3.2 Variants

There are other possible choices for the form of the matrix J . One important variant is described in this section. All possible variants are discussed in Section 3.13.

Let x and y be two n -dimensional vectors with real entries. Define a $2n$ -component real vector z by the rule

$$z = (z_1 \cdots z_n, z_{n+1} \cdots z_{2n}) = (x_1 \cdots x_n, y_1 \cdots y_n). \quad (3.2.1)$$

Similarly, let u and v be another pair of real n -dimensional vectors, and define the $2n$ -component real vector w by the rule

$$w = (w_1 \cdots w_n, w_{n+1} \cdots w_{2n}) = (u_1 \cdots u_n, v_1 \cdots v_n). \quad (3.2.2)$$

Then one has the relation

$$(w, Jz) = (u, y) - (v, x). \quad (3.2.3)$$

This quadratic form is called the *fundamental symplectic 2-form*.³ Note that the inner product on the left of (2.3) is that for $2n$ -dimensional vectors, and those on the right of (2.3) are for n -dimensional vectors.

Define a $2n$ -component vector z' in terms of the vector z by requiring that z' have the entries

$$z' = (x_1, y_1, x_2, y_2, \dots, x_n, y_n). \quad (3.2.4)$$

Evidently, z' is related to z by a linear transformation. Indeed, the entries in z' are a *permutation* of those in z . Consequently, there is a matrix P , with entries 0 and 1, such that

$$z' = Pz. \quad (3.2.5)$$

See, for example, Exercise 2.5. Let w' be defined in terms of w by an analogous relation,

$$w' = Pw. \quad (3.2.6)$$

It follows from (2.4) and its counterpart for w' that one has the relation

$$(w', z') = (u, x) + (v, y) = (w, z). \quad (3.2.7)$$

But, by (2.5) and (2.6), there is also the relation

$$(w', z') = (Pw, Pz) = (w, P^T Pz). \quad (3.2.8)$$

Comparison of (2.7) and (2.8) shows that P is *orthogonal*,

$$P^T P = I \quad \text{or} \quad P^T = P^{-1}. \quad (3.2.9)$$

³Other authors take (Jw, z) to be the fundamental symplectic 2-form. In view of (1.3), (w, Jz) and (Jw, z) differ only by a sign. A 2-form is a function that takes as inputs two vectors, is linear in both inputs, and delivers a number. It is also usually required to be odd under the interchange of the two vector inputs. See (1.3).

Let J' be the $2n \times 2n$ matrix defined by the equation

$$J' = \begin{pmatrix} J_2 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_2 \end{pmatrix}. \quad (3.2.10)$$

That is, all the entries of J' are zero save for n 2×2 blocks on the diagonal. These blocks are identical, and are specified by the equation

$$J_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad (3.2.11)$$

The matrix J' has been defined in such a way as to satisfy the relation

$$(w', J'z') = (u, y) - (v, x) = (w, Jz). \quad (3.2.12)$$

By (2.5) and (2.6) there is also the relation

$$(w', J'z') = (Pw, J'Pz) = (w, P^T J' P z). \quad (3.2.13)$$

It follows from (2.12) and (2.13) that J and J' are related by the orthogonal similarity transformation

$$P^T J' P = J \quad \text{or} \quad J' = PJP^T. \quad (3.2.14)$$

To complete the story, suppose that M is any symplectic matrix with respect to J . See (1.2). Consider the matrix M' defined by the orthogonal similarity transformation

$$M' = PMP^T. \quad (3.2.15)$$

Then it is easily checked using (1.2) and (2.15) that M' is symplectic with respect to J' ,

$$(M')^T J' M' = J'. \quad (3.2.16)$$

Indeed, if M and N are any two matrices (not even necessarily symplectic), and M' and N' are their counterparts defined by relations of the form (2.15), then it follows from the orthogonality condition (2.9) that

$$(MN)' = PMNP^T = PMP^TPNP^T = M'N'. \quad (3.2.17)$$

The results of this section and the exercises below show that for the most part, *mutatis mutandis*, one may use either J or J' when defining or working with symplectic matrices. Generally we shall drop the prime notation, and use the symbol J to denote either J or J' . Sometimes, however, a particular choice of J may give simpler or more interesting results. When this is the case, we shall be more specific.

Exercises

3.2.1. Consider the properties of J given in Exercise (1.1). Show that J' has the same properties.

3.2.2. Show that I and J' are symplectic with respect to J' .

3.2.3. Exercises 1.3, 1.5, 1.6, and 1.7 describe properties of matrices symplectic with respect to J . Show that matrices symplectic with respect to J' have directly analogous properties.

3.2.4. Let M be any matrix. Define the operation of “priming” a matrix by (2.15). Show that the operations of priming and transposing commute, $(M^T)' = (M')^T$. Show that the operations of priming and inverting also commute, $(M^{-1})' = (M')^{-1}$. Finally, show that inverting and transposing also commute, $(M^{-1})^T = (M^T)^{-1}$.

3.2.5. Compute P explicitly in the 4×4 and 6×6 cases. For each case verify that P is orthogonal, and find its eigenvalues and determinant. You should find $\det(P) = -1$ in both cases. In the 4×4 case you should find the result

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (3.2.18)$$

and in the 6×6 case you should find the result

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (3.2.19)$$

3.2.6. Recall the definition of the collection of variables z given by (1.9.9). Compare (2.1) and (2.4) and identify x_j with q_j and y_j with p_j to show that z' is the collection of variables

$$z' = (q_1, p_1, q_2, p_2, \dots, q_n, p_n). \quad (3.2.20)$$

Verify that the variables z' obey the Poisson bracket relations

$$[z'_a, z'_b] = J'_{ab}. \quad (3.2.21)$$

3.2.7. Let x and y be a pair of real n -component vectors. Define a real $2n$ -component vector z by the rule

$$z = (z_1 \cdots z_n, z_{n+1} \cdots z_{2n}) = (x_1 \cdots x_n, y_1 \cdots y_n). \quad (3.2.22)$$

Also define a complex n -component vector w by the rule

$$w = (w_1 \cdots w_n) = (x_1 + iy_1 \cdots x_n + iy_n) = x + iy. \quad (3.2.23)$$

Let $w' = x' + iy'$ be another such vector. Form the *complex* inner product (w, w') . Obtain the result

$$\begin{aligned}(w, w') &= (x + iy, x' + iy') \\ &= (x, x') + (y, y') + i[(x, y') - (y, x')] \\ &= (z, z') + i(z, Jz').\end{aligned}\tag{3.2.24}$$

You have shown that the symplectic 2-form (z, Jz') may be obtained as the *imaginary* part of a complex inner product. Suppose we make the correspondence

$$w \leftrightarrow z\tag{3.2.25}$$

as described by (2.22) and (2.23). This correspondence is a bijective mapping between the complex vector space C^n and the real vector space R^{2n} . Show that there is then the correspondence

$$-iw \leftrightarrow Jz.\tag{3.2.26}$$

Thus, in some ways, J acts like $-i$. See (1.4) and Exercise 8.1. For this reason, J is said to provide phase space with an *almost complex structure*. Suppose one instead defines w by the rule

$$w = (w_1 \cdots w_n) = (y_1 + ix_1 \cdots y_n + in_n) = y + ix,\tag{3.2.27}$$

and then again makes the correspondence (2.25). Show that now there is the correspondence

$$iw \leftrightarrow Jz,\tag{3.2.28}$$

so that now J acts like $+i$.

3.2.8. Suppose two phase-space points (vectors) w and z are sent under the action of a linear symplectic map, described by the symplectic matrix M , to the points w' , z' :

$$w' = Mw,\tag{3.2.29}$$

$$z' = Mz.\tag{3.2.30}$$

Show that the fundamental symplectic 2-form (2.3) is preserved under a symplectic transformation. That is, the relation

$$(w', Jz') = (Mw, JMz) = (w, Jz)\tag{3.2.31}$$

holds for any real symplectic matrix M and any pair of points w, z . It follows that a real matrix is symplectic if and only if it preserves the fundamental symplectic 2-form.

3.3 Simple Symplectic Restrictions and Symplectic Factorization

3.3.1 Large-Block Formulation

Suppose M is a $2n \times 2n$ matrix. Then it can be written in the form

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}\tag{3.3.1}$$

where the matrices A through D are $n \times n$ blocks. Correspondingly, M^T can be written as

$$M^T = \begin{pmatrix} A^T & C^T \\ B^T & D^T \end{pmatrix}. \quad (3.3.2)$$

Now require that M be symplectic with respect to the J of (1.1). It then follows from (1.2) that the matrices A through D must satisfy the conditions

$$A^T C = C^T A, \quad (3.3.3)$$

$$B^T D = D^T B, \quad (3.3.4)$$

$$A^T D - C^T B = I. \quad (3.3.5)$$

If M is symplectic with respect to J , so is M^T . See (1.14). From (1.14) it follows that A through D must also satisfy the conditions

$$AB^T = BA^T, \quad (3.3.6)$$

$$CD^T = DC^T, \quad (3.3.7)$$

$$AD^T - BC^T = I. \quad (3.3.8)$$

3.3.2 Symplectic Block Factorization

Consider matrices having the block forms

$$M = \begin{pmatrix} I & B \\ 0 & I \end{pmatrix}, \quad (3.3.9)$$

$$M = \begin{pmatrix} I & 0 \\ C & I \end{pmatrix}, \quad (3.3.10)$$

$$M = \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix}, \quad (3.3.11)$$

Then it is readily verified from (3.3) through (3.8) that the matrices M in (3.9) and (3.10) are symplectic if

$$B^T = B \text{ and } C^T = C. \quad (3.3.12)$$

Also, M in (3.11) is symplectic if

$$A^T D = I \text{ or } D = (A^T)^{-1}. \quad (3.3.13)$$

Observe that all matrices of the form (3.9) and (3.10) have determinant +1. Moreover the matrix M given by (3.11) also has determinant +1 if it is symplectic: Simple calculation and use of (3.13) gives the result

$$\begin{aligned} \det(M) &= \det(A) \det(D) = \det(A) \det[(A^T)^{-1}] \\ &= \det(A) \det(A^{-1}) = \det(AA^{-1}) = \det(I) = 1. \end{aligned} \quad (3.3.14)$$

See Exercise 3.4.

Let M be any matrix written in the form (3.1). Suppose A and/or D are invertible [$\det(A) \neq 0$ and/or $\det(D) \neq 0$]. Then, as can be easily checked by direct matrix multiplication, M has the block factorizations

$$M = \begin{pmatrix} I & 0 \\ CA^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{pmatrix} \begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix}, \quad (3.3.15)$$

$$M = \begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} I & 0 \\ D^{-1}C & I \end{pmatrix}. \quad (3.3.16)$$

Next, suppose that M is symplectic. Then remarkably each of the factors appearing in (3.15) and (3.16) is separately symplectic. We prove this assertion for the factorization (3.15). The proof for the factorization (3.16) is similar. Before so doing, we note that the three factors appearing in (3.15) and (3.16) are of the forms (3.9) through (3.11) and therefore, if symplectic, have determinant +1. Therefore M in this case has determinant +1.

To prove that the factors in (3.15) are symplectic, begin by observing that (3.3) can be rewritten in the form

$$CA^{-1} = (CA^{-1})^T. \quad (3.3.17)$$

That is, the matrix CA^{-1} is symmetric. It follows that the first factor in (3.15) is symplectic. Similarly, observe that (3.6) can be rewritten in the form

$$A^{-1}B = (A^{-1}B)^T, \quad (3.3.18)$$

and consequently the matrix $A^{-1}B$ is also symmetric. It follows that the third factor in (3.15) is symplectic. Finally, with the aid of (3.3), the relation (3.5) can be rewritten in the form

$$A^T(D - CA^{-1}B) = I. \quad (3.3.19)$$

It follows that the second factor in (3.15) is also symplectic.

Even if a symplectic M cannot be written as a product of three symplectic factors as in (3.15) and (3.16), it can always be written as a product of a finite number of symplectic factors of the form (3.9) through (3.11). That is, symplectic matrices of the form (3.9) through (3.11) *generate* all symplectic matrices.⁴

To verify this assertion, suppose M is written in the form (3.1). We distinguish two cases: either the block A vanishes identically or it does not. Suppose A does vanish. Then we have the relations

$$\begin{pmatrix} I & I \\ 0 & I \end{pmatrix} \begin{pmatrix} 0 & B \\ C & D \end{pmatrix} = \begin{pmatrix} C & B + D \\ C & D \end{pmatrix}, \quad (3.3.20)$$

⁴The word *generate* has many meanings depending on context. Here it means that any symplectic matrix can be expressed as a product of a finite number of symplectic matrices having a specific form. In the Lie algebraic context it happens that some Lie group elements G can be written in the form $G = \exp(g)$ where g is in the associated Lie algebra. See Section 3.7. In that case, but with a different meaning, we also say that g generates G .

$$M = \begin{pmatrix} 0 & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & -I \\ 0 & I \end{pmatrix} \begin{pmatrix} C & B+D \\ C & D \end{pmatrix}. \quad (3.3.21)$$

Moreover, the matrix C must satisfy $\det(C) \neq 0$. For if $\det(C) = 0$, then the n columns of C must be linearly dependent, which implies that the first n columns of M must be linearly dependent, which implies $\det(M) = 0$ contrary to the result of Exercise (1.3). [The same conclusion, $\det(C) \neq 0$, also follows directly from (3.5).] Also, according to Exercise (1.6), the matrix on the right side of (3.20) is symplectic. It follows that this matrix has a factorization of the form (3.15), and correspondingly according to (3.21) M can be written as a product of four factors of the form (3.9) through (3.11).

Next suppose that A does not vanish. For any nonsingular matrix W let V_W denote the matrix

$$V_W = \begin{pmatrix} W & 0 \\ 0 & W^* \end{pmatrix}, \quad (3.3.22)$$

where

$$W^* = (W^T)^{-1}. \quad (3.3.23)$$

According to (3.23) and (3.13) V_W is symplectic. Pre and post multiply M by V_X and V_Y , where X and Y are nonsingular matrices to be determined, to get the result

$$M' = V_X M V_Y = \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix}, \quad (3.3.24)$$

with A' given by the relation

$$A' = XAY. \quad (3.3.25)$$

According to a standard result in matrix theory, nonsingular matrices X and Y can be selected in such a way that A' takes the block form

$$A' = \begin{pmatrix} I_\ell & 0 \\ 0 & 0_{n-\ell} \end{pmatrix}. \quad (3.3.26)$$

Here I_ℓ is the $\ell \times \ell$ identity matrix (with $\ell \geq 1$), and $0_{n-\ell}$ is a complementary zero matrix. (The integer ℓ is the rank of A .) Correspondingly, C' can be written in the block form

$$C' = \begin{pmatrix} C'_{11} & C'_{12} \\ C'_{21} & C'_{22} \end{pmatrix}. \quad (3.3.27)$$

Then use of the symplectic condition (3.3) when applied to A' and C' gives the result

$$C'_{12} = 0. \quad (3.3.28)$$

It follows that $\det(C'_{22}) \neq 0$. For if $\det(C'_{22}) = 0$, then the $(n - \ell)$ columns of C'_{22} must be linearly dependent, which implies that $(n - \ell)$ columns of M' must be linearly dependent, which implies $\det(M') = 0$ contrary to Exercises (1.3) and (1.6). Let T_λ , where λ is an arbitrary real parameter, denote the symplectic matrix

$$T_\lambda = \begin{pmatrix} I & \lambda I \\ 0 & I \end{pmatrix}. \quad (3.3.29)$$

Multiply M' by T_λ on the left to get the result

$$M'' = T_\lambda M' = \begin{pmatrix} I & \lambda I \\ 0 & I \end{pmatrix} \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix} = \begin{pmatrix} A'' & B'' \\ C'' & D'' \end{pmatrix} \quad (3.3.30)$$

with A'' given by the relation

$$A'' = A' + \lambda C' = \begin{pmatrix} I_\ell + \lambda C'_{11} & 0 \\ \lambda C'_{21} & \lambda C'_{22} \end{pmatrix}. \quad (3.3.31)$$

A little thought shows that λ can be selected in such a way that $\det(A'') \neq 0$. By inverting the relations (3.24) and (3.30), we see that M can be written in the form

$$M = V_X^{-1} T_\lambda^{-1} M'' V_Y^{-1}. \quad (3.3.32)$$

And, according to the previous discussion, M'' has a factorization of the form (3.15). Thus, M can again be written as a product of factors (this time six in number) of the form (3.9) through (3.11).⁵

3.3.3 Symplectic Matrices Have Determinant +1

Moreover, as a bonus, we observe that since each factor has determinant +1, the matrix M itself must have determinant +1. We conclude that every symplectic matrix M (real or complex) must satisfy the relation

$$\det(M) = +1. \quad (3.3.33)$$

Here is a topological perspective on the relation (3.33): Suppose it can be established that symplectic matrices written in the form (3.31) and having $\det(A) \neq 0$ are *dense* in the set of all symplectic matrices. That is for any symplectic matrix M' , written in the form (3.31) and having $\det(A) = 0$, there is a symplectic matrix M arbitrarily nearby with $\det(A) \neq 0$. For this matrix we know from the factorization (3.15) that $\det(M) = 1$. But from (1.8) we know that $\det(M') = \pm 1$. Since the determinant of a matrix is a continuous function of its entries, it follows from the density hypothesis that $\det(M') = +1$.

3.3.4 Small-Block Formulation

Equally interesting are the results of requiring M to be symplectic with respect to the J' of (2.10). For simplicity in this case, and for later use, we restrict our discussion to 6×6 matrices. Then M can be written in the form

$$M = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \quad (3.3.34)$$

⁵Subsequently, using *polar decomposition*, we will learn that any real symplectic matrix M can be written as a product of two real symplectic matrices with each having a special form. See Subsection 8.2 and Section 4.2. See also Exercise 5.10.16 for analogous results for the complex case.

where the matrices a through i are all 2×2 . Correspondingly, M^T can be written as

$$M^T = \begin{pmatrix} a^T & d^T & g^T \\ b^T & e^T & h^T \\ c^T & f^T & i^T \end{pmatrix}. \quad (3.3.35)$$

Now require that M satisfy the condition (1.2) with J replaced by J' . We find that the matrices a through i must satisfy the conditions

$$a^T J_2 a + d^T J_2 d + g^T J_2 g = J_2, \quad (3.3.36)$$

$$b^T J_2 b + e^T J_2 e + h^T J_2 h = J_2,$$

$$c^T J_2 c + f^T J_2 f + i^T J_2 i = J_2,$$

$$b^T J_2 a + e^T J_2 d + h^T J_2 g = 0, \quad (3.3.37)$$

$$c^T J_2 a + f^T J_2 d + i^T J_2 g = 0,$$

$$c^T J_2 b + f^T J_2 e + i^T J_2 h = 0.$$

Note that, because of (1.11), the relations (3.36) can also be written in the form

$$\det a + \det d + \det g = 1, \quad (3.3.38)$$

$$\det b + \det e + \det h = 1,$$

$$\det c + \det f + \det i = 1.$$

As before, M must also satisfy (1.14) with J replaced by J' . As a consequence, the matrices a through i must also satisfy the conditions

$$\det a + \det b + \det c = 1, \quad (3.3.39)$$

$$\det d + \det e + \det f = 1,$$

$$\det g + \det h + \det i = 1,$$

$$d J_2 a^T + e J_2 b^T + f J_2 c^T = 0, \quad (3.3.40)$$

$$g J_2 a^T + h J_2 b^T + i J_2 c^T = 0,$$

$$g J_2 d^T + h J_2 e^T + i J_2 f^T = 0.$$

Exercises

3.3.1. Subsections 3.1 through 3.4 and elsewhere presumed the reader is familiar with how to add and multiply matrices written in block form. Addition is obvious, but multiplication requires more thought. The purpose of this exercise and the next is to explore/recall rules for the multiplication of matrices that are *partitioned* into “submatrices”. We wish to express the multiplication rules for such matrices in terms of addition/multiplication rules for the submatrices. To see what is involved, here we consider the simplest case of 4×4 matrices partitioned into four 2×2 blocks.

Suppose \mathcal{L} and \mathcal{M} are two linear operators whose associated matrices L and M are of the form

$$L = \begin{pmatrix} {}^1L & {}^2L \\ {}^3L & {}^4L \end{pmatrix} \quad (3.3.41)$$

and

$$M = \begin{pmatrix} {}^1M & {}^2M \\ {}^3M & {}^4M \end{pmatrix} \quad (3.3.42)$$

where the matrices jL and kM are 2×2 . Also, suppose \mathcal{N} is a linear operator with associated matrix N and that

$$\mathcal{N} = \mathcal{L}\mathcal{M}. \quad (3.3.43)$$

Consequently there will be the relation

$$N = LM. \quad (3.3.44)$$

Also write N in the 2×2 block form

$$N = \begin{pmatrix} {}^1N & {}^2N \\ {}^3N & {}^4N \end{pmatrix}. \quad (3.3.45)$$

Then we may write (3.44) in the form

$$\begin{pmatrix} {}^1N & {}^2N \\ {}^3N & {}^4N \end{pmatrix} = \begin{pmatrix} {}^1L & {}^2L \\ {}^3L & {}^4L \end{pmatrix} \begin{pmatrix} {}^1M & {}^2M \\ {}^3M & {}^4M \end{pmatrix}. \quad (3.3.46)$$

Our problem for this exercise is to find the matrices kN in terms of the contents of the matrices iL and jM .

To proceed, introduce 4×4 matrices ${}^i\hat{L}$ by the rules

$${}^1\hat{L} = \begin{pmatrix} {}^1L & O \\ O & O \end{pmatrix}, \quad (3.3.47)$$

$${}^2\hat{L} = \begin{pmatrix} O & {}^2L \\ O & O \end{pmatrix}, \quad (3.3.48)$$

$${}^3\hat{L} = \begin{pmatrix} O & O \\ {}^3L & O \end{pmatrix}, \quad (3.3.49)$$

$${}^4\hat{L} = \begin{pmatrix} O & O \\ O & {}^4L \end{pmatrix}, \quad (3.3.50)$$

where O denotes the 2×2 matrix with all entries zero. Verify that there is the relation

$$L = \sum_j {}^j \hat{L}. \quad (3.3.51)$$

Write analogous relations for M and N . Verify that

$$N = \sum_\ell {}^\ell \hat{N} = \sum_{jk} {}^j \hat{L} {}^k \hat{M}. \quad (3.3.52)$$

Evidently, to find the ${}^\ell \hat{N}$, we must find the matrix products ${}^j \hat{L} {}^k \hat{M}$ and then compose the results into blocks.

There are 16 matrices of the form ${}^j \hat{L} {}^k \hat{M}$. Eventually you will be asked to prove the relations

$${}^1 \hat{L} {}^1 \hat{M} = \begin{pmatrix} {}^1 L & O \\ O & O \end{pmatrix} \begin{pmatrix} {}^1 M & O \\ O & O \end{pmatrix} = \begin{pmatrix} {}^1 L {}^1 M & O \\ O & O \end{pmatrix}, \quad (3.3.53)$$

$${}^1 \hat{L} {}^2 \hat{M} = \begin{pmatrix} {}^1 L & O \\ O & O \end{pmatrix} \begin{pmatrix} O & {}^2 M \\ O & O \end{pmatrix} = \begin{pmatrix} O & {}^1 L {}^2 M \\ O & O \end{pmatrix}, \quad (3.3.54)$$

$${}^1 \hat{L} {}^3 \hat{M} = \begin{pmatrix} {}^1 L & O \\ O & O \end{pmatrix} \begin{pmatrix} O & O \\ {}^3 M & O \end{pmatrix} = \begin{pmatrix} O & O \\ O & O \end{pmatrix}, \quad (3.3.55)$$

$${}^1 \hat{L} {}^4 \hat{M} = \begin{pmatrix} {}^1 L & O \\ O & O \end{pmatrix} \begin{pmatrix} O & O \\ O & {}^4 M \end{pmatrix} = \begin{pmatrix} O & O \\ O & O \end{pmatrix}, \quad (3.3.56)$$

$${}^2 \hat{L} {}^1 \hat{M} = \begin{pmatrix} O & {}^2 L \\ O & O \end{pmatrix} \begin{pmatrix} {}^1 M & O \\ O & O \end{pmatrix} = \begin{pmatrix} O & O \\ O & O \end{pmatrix}, \quad (3.3.57)$$

$${}^2 \hat{L} {}^2 \hat{M} = \begin{pmatrix} O & {}^2 L \\ O & O \end{pmatrix} \begin{pmatrix} O & {}^2 M \\ O & O \end{pmatrix} = \begin{pmatrix} O & O \\ O & O \end{pmatrix}, \quad (3.3.58)$$

$${}^2 \hat{L} {}^3 \hat{M} = \begin{pmatrix} O & {}^2 L \\ O & O \end{pmatrix} \begin{pmatrix} O & O \\ {}^3 M & O \end{pmatrix} = \begin{pmatrix} {}^2 L {}^3 M & O \\ O & O \end{pmatrix}, \quad (3.3.59)$$

$${}^2 \hat{L} {}^4 \hat{M} = \begin{pmatrix} O & {}^2 L \\ O & O \end{pmatrix} \begin{pmatrix} O & O \\ O & {}^4 M \end{pmatrix} = \begin{pmatrix} O & O \\ O & {}^2 L {}^4 M \end{pmatrix}, \quad (3.3.60)$$

$${}^3 \hat{L} {}^1 \hat{M} = \begin{pmatrix} O & O \\ {}^3 L & O \end{pmatrix} \begin{pmatrix} {}^1 M & O \\ O & O \end{pmatrix} = \begin{pmatrix} O & O \\ {}^3 L {}^1 M & O \end{pmatrix}, \quad (3.3.61)$$

$${}^3 \hat{L} {}^2 \hat{M} = \begin{pmatrix} O & O \\ {}^3 L & O \end{pmatrix} \begin{pmatrix} O & {}^2 M \\ O & O \end{pmatrix} = \begin{pmatrix} O & O \\ O & {}^3 L {}^2 M \end{pmatrix}, \quad (3.3.62)$$

$${}^3 \hat{L} {}^3 \hat{M} = \begin{pmatrix} O & O \\ {}^3 L & O \end{pmatrix} \begin{pmatrix} O & O \\ {}^3 M & O \end{pmatrix} = \begin{pmatrix} O & O \\ O & O \end{pmatrix}, \quad (3.3.63)$$

$${}^3 \hat{L} {}^4 \hat{M} = \begin{pmatrix} O & O \\ {}^3 L & O \end{pmatrix} \begin{pmatrix} O & O \\ O & {}^4 M \end{pmatrix} = \begin{pmatrix} O & O \\ O & O \end{pmatrix}, \quad (3.3.64)$$

$${}^4 \hat{L} {}^1 \hat{M} = \begin{pmatrix} O & O \\ O & {}^4 L \end{pmatrix} \begin{pmatrix} {}^1 M & O \\ O & O \end{pmatrix} = \begin{pmatrix} O & O \\ O & O \end{pmatrix}, \quad (3.3.65)$$

$${}^4\hat{L} {}^2\hat{M} = \begin{pmatrix} O & O \\ O & {}^4L \end{pmatrix} \begin{pmatrix} O & {}^2M \\ O & O \end{pmatrix} = \begin{pmatrix} O & O \\ O & O \end{pmatrix}, \quad (3.3.66)$$

$${}^4\hat{L} {}^3\hat{M} = \begin{pmatrix} O & O \\ O & {}^4L \end{pmatrix} \begin{pmatrix} O & O \\ {}^3M & O \end{pmatrix} = \begin{pmatrix} O & O \\ {}^4L {}^3M & O \end{pmatrix}, \quad (3.3.67)$$

$${}^4\hat{L} {}^4\hat{M} = \begin{pmatrix} O & O \\ O & {}^4L \end{pmatrix} \begin{pmatrix} O & O \\ O & {}^4M \end{pmatrix} = \begin{pmatrix} O & O \\ O & {}^4L {}^4M \end{pmatrix}. \quad (3.3.68)$$

Assuming for the moment that (3.53) through (3.68) are correct, use these relations, (3.45), and (3.52) to show that

$${}^1N = {}^1L {}^1M + {}^2L {}^3M, \quad (3.3.69)$$

$${}^2N = {}^1L {}^2M + {}^2L {}^4M, \quad (3.3.70)$$

$${}^3N = {}^3L {}^1M + {}^4L {}^3M, \quad (3.3.71)$$

$${}^4N = {}^3L {}^2M + {}^4L {}^4M. \quad (3.3.72)$$

It follows that the relation (3.46) becomes the relation

$$\begin{aligned} \begin{pmatrix} {}^1N & {}^2N \\ {}^3N & {}^4N \end{pmatrix} &= \begin{pmatrix} {}^1L & {}^2L \\ {}^3L & {}^4L \end{pmatrix} \begin{pmatrix} {}^1M & {}^2M \\ {}^3M & {}^4M \end{pmatrix} \\ &= \begin{pmatrix} {}^1L {}^1M + {}^2L {}^3M & {}^1L {}^2M + {}^2L {}^4M \\ {}^3L {}^1M + {}^4L {}^3M & {}^3L {}^2M + {}^4L {}^4M \end{pmatrix}. \end{aligned} \quad (3.3.73)$$

Conversely verify that if (3.73) is correct, then (3.53) through (3.68) are correct because they are special cases of (3.73).

It is time to verify that (3.53) through (3.68) are correct. How might one do so? One procedure is to write out the matrices explicitly in 4×4 form, multiply them, and then deduce the block structures of the results. We will work out a few sample cases. You can work out the rest.

Begin with the case (3.53). For this case we find

$$\begin{aligned} {}^1\hat{L} {}^1\hat{M} &= \begin{pmatrix} {}^1L & O \\ O & O \end{pmatrix} \begin{pmatrix} {}^1M & O \\ O & O \end{pmatrix} \\ &= \begin{pmatrix} {}^1L_{11} & {}^1L_{12} & 0 & 0 \\ {}^1L_{21} & {}^1L_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} {}^1M_{11} & {}^1M_{12} & 0 & 0 \\ {}^1M_{21} & {}^1M_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} {}^1L_{11} {}^1M_{11} + {}^1L_{12} {}^1M_{21} & {}^1L_{11} {}^1M_{12} + {}^1L_{12} {}^1M_{22} & 0 & 0 \\ {}^1L_{21} {}^1M_{11} + {}^1L_{22} {}^1M_{21} & {}^1L_{21} {}^1M_{12} + {}^1L_{22} {}^1M_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} ({}^1L {}^1M)_{11} & ({}^1L {}^1M)_{12} & 0 & 0 \\ ({}^1L {}^1M)_{21} & ({}^1L {}^1M)_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} {}^1L {}^1M & O \\ O & O \end{pmatrix}, \end{aligned} \quad (3.3.74)$$

as claimed.

Continue with the case (3.54). For this case we find

$$\begin{aligned}
 {}^1\hat{L} {}^2\hat{M} &= \begin{pmatrix} {}^1L & O \\ O & O \end{pmatrix} \begin{pmatrix} O & {}^2M \\ O & O \end{pmatrix} \\
 &= \begin{pmatrix} {}^1L_{11} & {}^1L_{12} & 0 & 0 \\ {}^1L_{21} & {}^1L_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & {}^2M_{11} & {}^2M_{12} \\ 0 & 0 & {}^2M_{21} & {}^2M_{22} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
 &= \begin{pmatrix} 0 & 0 & {}^1L_{11} {}^2M_{11} + {}^1L_{12} {}^2M_{21} & {}^1L_{11} {}^2M_{12} + {}^1L_{12} {}^2M_{22} \\ 0 & 0 & {}^1L_{21} {}^2M_{11} + {}^1L_{22} {}^2M_{21} & {}^1L_{21} {}^2M_{12} + {}^1L_{22} {}^2M_{22} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
 &= \begin{pmatrix} 0 & 0 & ({}^1L {}^2M)_{11} & ({}^1L {}^2M)_{12} \\ 0 & 0 & ({}^1L {}^2M)_{21} & ({}^1L {}^2M)_{22} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
 &= \begin{pmatrix} O & {}^1L {}^2M \\ O & O \end{pmatrix}, \tag{3.3.75}
 \end{aligned}$$

as claimed.

As a third example, consider the case (3.55). For this case we find

$$\begin{aligned}
 {}^1\hat{L} {}^3\hat{M} &= \begin{pmatrix} {}^1L & O \\ O & O \end{pmatrix} \begin{pmatrix} O & O \\ {}^3M & \end{pmatrix} \\
 &= \begin{pmatrix} {}^1L_{11} & {}^1L_{12} & 0 & 0 \\ {}^1L_{21} & {}^1L_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ {}^3M_{11} & {}^3M_{12} & 0 & 0 \\ {}^3M_{21} & {}^3M_{22} & 0 & 0 \end{pmatrix} \\
 &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
 &= \begin{pmatrix} O & O \\ O & O \end{pmatrix}, \tag{3.3.76}
 \end{aligned}$$

as claimed.

There is one final observation that is worth noting. Suppose ℓ , m , and n are all 2×2 matrices with (*numerical*) entries:

$$\ell = \begin{pmatrix} {}^1\ell & {}^2\ell \\ {}^3\ell & {}^4\ell \end{pmatrix}, \text{ etc.} \tag{3.3.77}$$

Suppose also there is the matrix relation

$$n = \ell m. \tag{3.3.78}$$

Verify that, in terms of components, this relation becomes

$$\begin{aligned} \begin{pmatrix} 1n & 2n \\ 3n & 4n \end{pmatrix} &= \begin{pmatrix} 1\ell & 2\ell \\ 3\ell & 4\ell \end{pmatrix} \begin{pmatrix} 1m & 2m \\ 3m & 4m \end{pmatrix} \\ &= \begin{pmatrix} 1\ell 1m + 2\ell 3m & 1\ell 2m + 2\ell 4m \\ 3\ell 1m + 4\ell 3m & 3\ell 2m + 4\ell 4m \end{pmatrix}. \end{aligned} \quad (3.3.79)$$

Compare (3.73) and (3.79) to observe they are *identical* in form! Their only difference is that the additions and multiplications appearing on the right side of (3.79) are “ordinary” operations on numbers while those appearing in the right side of (3.73) are *matrix* operations on matrices.

3.3.2. Review Exercise 3.2. It considered the case of 4×4 matrices partitioned into four 2×2 blocks. This exercise considers 6×6 matrices partitioned into nine 2×2 blocks. This case is of particular interest for dealing with linear transformations of six-dimensional phase space.

Suppose L , M , and N are all 6×6 matrices partitioned into 2×2 blocks: Express this supposition by writing

$$L = \begin{pmatrix} 1L & 2L & 3L \\ 4L & 5L & 6L \\ 7L & 8L & 9L \end{pmatrix}, \quad (3.3.80)$$

and write M and N in the same way. Also suppose there is the relation

$$N = LM. \quad (3.3.81)$$

Show that, in terms of blocks, the relation (3.81) takes the form

$$\begin{aligned} N &= \begin{pmatrix} 1N & 2N & 3N \\ 4N & 5N & 6N \\ 7N & 8N & 9N \end{pmatrix} = \begin{pmatrix} 1L & 2L & 3L \\ 4L & 5L & 6L \\ 7L & 8L & 9L \end{pmatrix} \begin{pmatrix} 1M & 2M & 3M \\ 4M & 5M & 6M \\ 7M & 8M & 9M \end{pmatrix} \\ &= \begin{pmatrix} 1L 1M + 2L 4M + 3L 7M & 1L 2M + 2L 5M + 3L 8M & 1L 3M + 2L 6M + 3L 9M \\ 4L 1M + 5L 4M + 6L 7M & 4L 2M + 5L 5M + 6L 8M & 4L 3M + 5L 6M + 6L 9M \\ 7L 1M + 8L 4M + 9L 7M & 7L 2M + 8L 5M + 9L 8M & 7L 3M + 8L 6M + 9L 9M \end{pmatrix}. \end{aligned} \quad (3.3.82)$$

3.3.3. Verify the relations (3.3) through (3.8).

3.3.4. Verify that M as given by (3.11) can be written in the form

$$M = \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & D \end{pmatrix}. \quad (3.3.83)$$

and therefore

$$\det(M) = \det(A) \det(D). \quad (3.3.84)$$

3.3.5. Verify the block factorizations (3.15) and (3.16). Work out in detail the proof that each factor in (3.15) and (3.16) is separately symplectic if M is symplectic.

3.3.6. Verify in detail all the steps required to show that matrices of the form (3.9) through (3.11) generate all symplectic matrices.

3.3.7. Verify the relations (3.36) through (3.40).

3.3.8. Consider the 4×4 matrix M given by

$$M = \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix} \quad (3.3.85)$$

where I is the 2×2 identity matrix. Observe that M has determinant +1. But is M symplectic? Verify that

$$M^T JM = -J, \quad (3.3.86)$$

so that M is *not* symplectic. Any $2n \times 2n$ matrix with the property (3.86) is called *anti-symplectic*. See Exercise 13.8.

3.4 Eigenvalue Spectrum

Suppose some map \mathcal{M} acts on some space with coordinates z and suppose \mathcal{M} has a fixed point z_f ,

$$\mathcal{M}z_f = z_f.$$

What can be said about the behavior of points near this fixed point under repeated application of \mathcal{M} ? In lowest approximation, this behavior is controlled by the matrix M that specifies the linear part of \mathcal{M} when it is expanded about z_f . It can be shown, in the linear approximation, that points near z_f remain near z_f under repeated application of \mathcal{M} if all the eigenvalues of M are within the unit circle in the complex plane or are on the unit circle and distinct. In this case z_f is said to be *stable*. On the other hand, if any eigenvalue lies outside the unit circle, there are points near z_f that, again in the linear approximation, are mapped away from z_f exponentially fast under repeated application of \mathcal{M} . In this case, z_f is said to be *unstable*. The case where all eigenvalues are on the unit circle but not all distinct can be stable or unstable depending on additional conditions. See Subsection 5.8.

By definition, the linear part of a symplectic map is specified by a symplectic matrix. See Section 6.1.2. We are therefore particularly interested in the eigenvalues of M when M is symplectic.

Finally, in the context of accelerator physics, suppose \mathcal{M} is the one-turn map for a circular machine (ring). In this case, a fixed point of \mathcal{M} corresponds to a closed orbit. In order to accelerate/store a large number of particles, it is essential that this closed orbit (fixed point) be stable. That is, if one fails to inject onto the closed orbit, as will be the case for most of any injected beam, one desires that particles near the closed orbit will remain so for very large times (in some cases equivalent in terms of the number of oscillations about the closed orbit to the number of trips of the earth around the sun since the Big Bang). Therefore, for successful accelerator design and operation, it is essential to know and control the eigenvalues of M .

3.4.1 Background

The *characteristic polynomial* $P(\lambda)$ of any matrix M is defined by the equation

$$P(\lambda) = \det(M - \lambda I). \quad (3.4.1)$$

Evidently $P(\lambda)$ is a polynomial with real coefficients if the matrix M is real. Also, the eigenvalues of M are the roots of the equation

$$P(\lambda) = 0. \quad (3.4.2)$$

It follows that if M is a *real* matrix, then its eigenvalues must also be real or must occur in complex conjugate pairs $\lambda, \bar{\lambda}$.

Suppose M is a symplectic matrix. Then it follows from (1.13) that

$$J^{-1}(M^T - \lambda I)J = M^{-1} - \lambda I = -\lambda M^{-1}(M - \lambda^{-1}I). \quad (3.4.3)$$

Since M is symplectic, we also have the relation

$$\det(M) = +1. \quad (3.4.4)$$

See Section 3.3. Now take the determinant of both sides of (4.3). The result is the relation

$$P(\lambda) = \lambda^{2n} P(1/\lambda). \quad (3.4.5)$$

It follows that if λ is an eigenvalue of a symplectic matrix, so is the reciprocal $1/\lambda$. [Note that according to Exercise 1.7, $\lambda = 0$ is not an eigenvalue, so we need not be concerned about multiplying or dividing by zero.] Consequently, the eigenvalues of a symplectic matrix must form reciprocal pairs. This property is called *reflexivity*.

The symmetry between λ and $1/\lambda$ exhibited by (4.5) can be further displayed by rewriting (4.5) in the form

$$\lambda^{-n} P(\lambda) = \lambda^n P(1/\lambda). \quad (3.4.6)$$

Now define another function $Q(\lambda)$ by writing

$$Q(\lambda) = \lambda^{-n} P(\lambda). \quad (3.4.7)$$

The functions P and Q evidently have the same zeroes. Moreover, the condition (4.6) requires that Q have the symmetry property

$$Q(\lambda) = Q(1/\lambda). \quad (3.4.8)$$

Equation (4.8) shows not only that the eigenvalues of a symplectic matrix must occur in reciprocal pairs; it shows that they must also occur with the same multiplicity. That is, if the root λ_0 has multiplicity k , so must the root $1/\lambda_0$. Indeed, the eigenvalues λ_0 and λ_0^{-1} must have the same Jordan block structure. See Exercise 4.6.

Also, if either $+1$ or -1 is a root, then this root must have *even* multiplicity. To see this, suppose for example that $\lambda = 1$ is a root. Introduce the variable μ by writing $\lambda = \exp \mu$.

Then (4.8) shows that Q is an *even* function of the variable μ and hence near $\lambda = 1$ (near $\mu = 0$) Q must have an expansion of the form

$$Q = \sum_{m=0}^{\infty} c_m \mu^{2m}. \quad (3.4.9)$$

Moreover, when λ is near 1, λ and μ are related by the expansion

$$\mu = \log \lambda = \log[1 + (\lambda - 1)] = (\lambda - 1)[1 - (\lambda - 1)/2 + \dots]. \quad (3.4.10)$$

Comparison of (4.9) and (4.10) shows that $\lambda = 1$ is not a root unless $c_0 = 0$. If $c_0 = 0$, then $\lambda = 1$ is a root of multiplicity 2. If $c_1 = 0$ as well, then $\lambda = 1$ is a root of multiplicity 4, etc. A similar argument holds near $\lambda = -1$ upon making the substitution $\lambda = -\exp \mu$.

In summary, it has been shown that the eigenvalues of a real symplectic matrix must satisfy the following properties:

1. They must be real or occur in complex conjugate pairs.
2. They must occur in reciprocal pairs, and each member of the pair must have the same multiplicity.
3. If either ± 1 is an eigenvalue, it must have even multiplicity.

When combined, the conditions just enumerated place strong restrictions on the possible eigenvalues of a real symplectic matrix. Among them is the fact that the eigenvalues cannot all lie inside or all lie outside the unit circle. We will learn in later chapters that, by definition, the linear part of a symplectic map at a fixed point is a symplectic matrix. Therefore, fixed points of a symplectic map cannot be attractors or repellers. We will also learn that Hamiltonian systems produce symplectic maps. It follows that Hamiltonian systems have neither attractors nor repellers.

3.4.2 The 2×2 Case

Consider first the simplest case of a 2×2 symplectic matrix ($n = 1$). Call the eigenvalues λ_1 and λ_2 . Then, by the reciprocal property, it follows that

$$\lambda_1 \lambda_2 = 1. \quad (3.4.11)$$

Suppose, now, that λ_1 is real, positive, and greater than 1. Then λ_2 is real, positive, and less than 1. Similarly, if λ_1 is real, negative, and less than -1 , then λ_2 is real, negative, and greater than -1 . On the other hand, if λ_1 is complex, then $\lambda_2 = \bar{\lambda}_1$. This condition, when combined with (4.11), shows that in this case λ_1 and λ_2 must lie on the unit circle in the complex plane. Finally, there are the two special cases $\lambda_1 = \lambda_2 = 1$ and $\lambda_1 = \lambda_2 = -1$.

Altogether, there are five possible cases. They are listed below along with names and designations whose significance will become clear later on. See also Figure 4.1.

1. Hyperbolic case (unstable): $\lambda_1 > 1$ and $0 < \lambda_2 < 1$.

2. Inversion hyperbolic case (unstable): $\lambda_1 < -1$ and $-1 < \lambda_2 < 0$.
3. Elliptic case (stable): $\lambda_1 = e^{i\phi}, \lambda_2 = e^{-i\phi}$. (Eigenvalues are complex conjugates and lie on the unit circle).
4. Parabolic case (generally linearly unstable): $\lambda_1 = \lambda_2 = +1$.
5. Inversion parabolic case (generally linearly unstable): $\lambda_1 = \lambda_2 = -1$.⁶

Note that in all cases both eigenvalues cannot lie inside the unit circle nor can both eigenvalues lie outside the unit circle.

3.4.3 The 4×4 and Remaining $2n \times 2n$ Cases

The next simplest case is that of a 4×4 symplectic matrix ($n = 2$). In this case, one has to deal with four possible eigenvalues and then apply reasoning analogous to the 2×2 case. Figures 4.2 illustrate the various possibilities that can occur. Analysis of the possible spectrum of the $2n$ eigenvalues for the general $2n \times 2n$ real symplectic matrix proceeds in a similar fashion. Again, in particular, it can never happen that all the eigenvalues lie inside the unit circle, nor can it happen that they all lie outside the unit circle. See Exercise 4.1.

⁶It is easily checked that inversion hyperbolic or inversion parabolic symplectic matrices can be written as the products of hyperbolic or parabolic symplectic matrices with the negative identity matrix, respectively. The negative identity matrix (which is also symplectic) acts on phase space to produce *inversion* through the origin. Some authors use the terminology *reflection* hyperbolic and *reflection* parabolic rather than *inversion* hyperbolic and *inversion* parabolic. This terminology is less precise: Reflection can mean reflection in/about the origin, in which case it is the same as inversion through the origin. In other contexts, reflections refer to transformations, like mirror reflections, that change the sign of some components of a vector, but not all components.

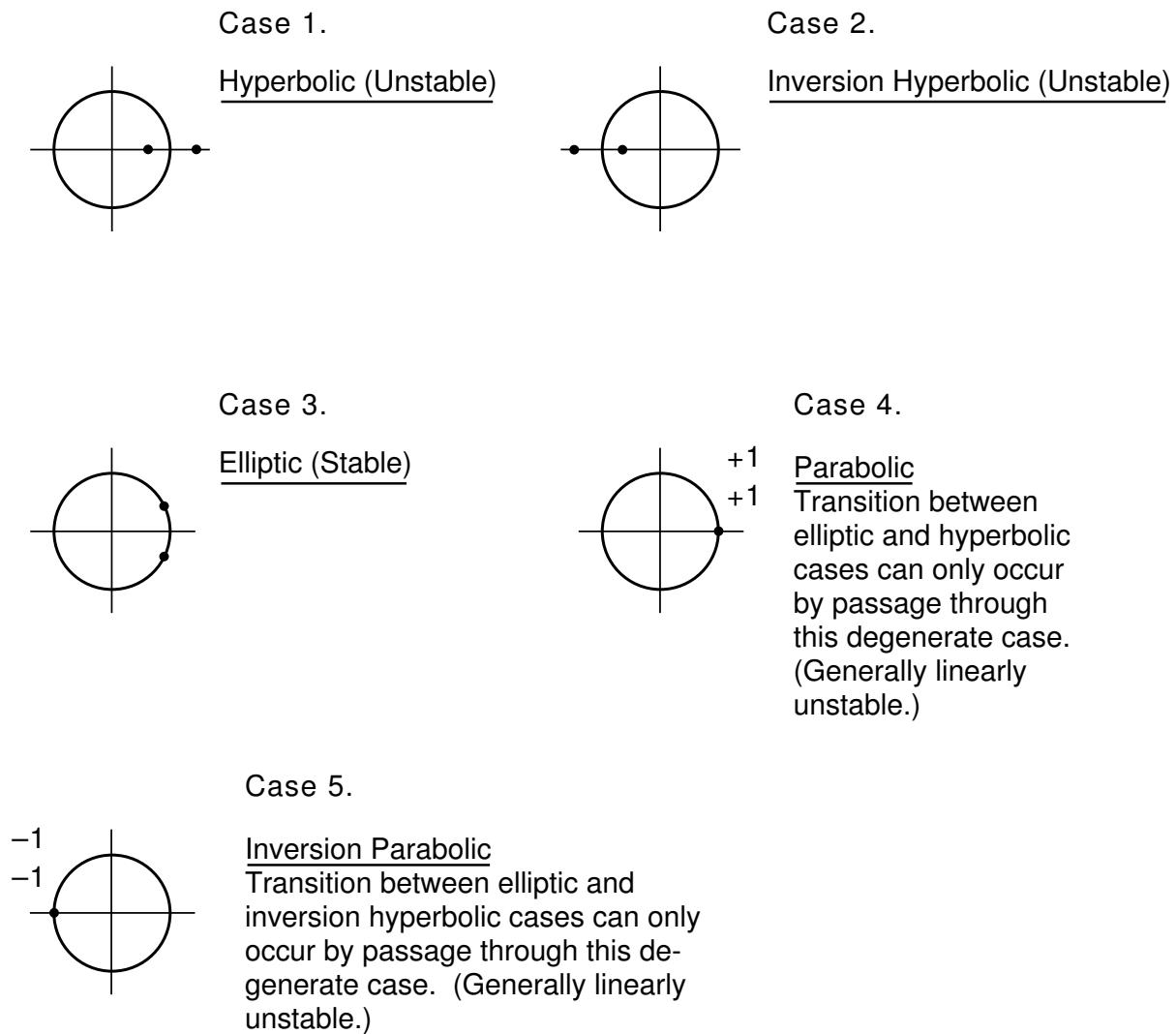


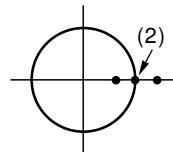
Figure 3.4.1: Possible cases for the eigenvalues of a 2×2 real symplectic matrix.

Figure 3.4.2: Possible eigenvalue configurations for a 4×4 real symplectic matrix. The mirror image of each configuration is also a possible configuration, and therefore is not shown in order to save space. Various authors have given these configurations various names. Notably, Case 1 is commonly called a *Krein quartet*.

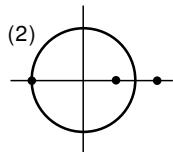
A. Generic Configurations

- | | |
|--|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | Case 1. All eigenvalues complex and off the unit circle.
All eigenvalues can be obtained from a single one by the operations of complex conjugation and taking reciprocals. <u>Unstable</u> . |
| | Case 2. All eigenvalues real, off the unit circle, and of same sign. Eigenvalues form reciprocal pairs.
<u>Unstable</u> . |
| | Case 3. All eigenvalues real, off the unit circle, and of differing sign. Eigenvalues form reciprocal pairs.
<u>Unstable</u> . |
| | Case 4. Two eigenvalues complex and confined to unit circle. Two eigenvalues real. Eigenvalues form reciprocal pairs. Complex eigenvalues are also complex conjugate. <u>Unstable</u> . |
| | Case 5. All eigenvalues complex and confined to the unit circle. Eigenvalues form reciprocal pairs that are also complex conjugate. <u>Stable</u> . |

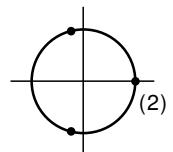
B. Degenerate Configurations. Transitions between generic configurations can only occur by passage through a degenerate configuration. Mirror image configurations are again possible, but not shown.



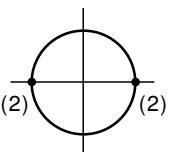
- Case 1. Two eigenvalues equal, and two eigenvalues real. All of same sign. Occurs in transition between generic cases 2 and 4. Unstable.



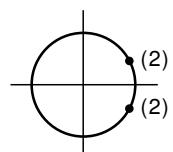
- Case 2. Two eigenvalues equal, and two eigenvalues real. Signs differ. Occurs in transition between generic cases 3 and 4. Unstable.



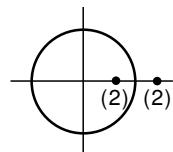
- Case 3. Two eigenvalues equal, and two eigenvalues confined to unit circle. Occurs in transition between generic cases 4 and 5. Generally linearly unstable.



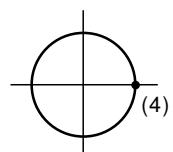
- Case 4. Two eigenvalues equal +1 and two equal -1. Occurs in transition between generic cases 3 and 5, or 3 and 4, or 4 and 5. Also occurs in transition between degenerate cases 2 and 3. Generally linearly unstable.



- Case 5. Two pairs of eigenvalues equal, and confined to unit circle. Occurs in transition between generic cases 1 and 5. Not, however, a sufficient condition to guarantee that such a transition is possible. Stability also undetermined in absence of further conditions.



- Case 6. Two pairs of eigenvalues equal and real. Occurs in transition between generic cases 1 and 2. Unstable.



- Case 7. All eigenvalues equal and have value ± 1 . Occurs in transitions between generic cases 1, 2, 4, and 5 and degenerate cases 1, 3, 5, and 6. Generally linearly unstable.

3.4.4 Further Symplectic Restrictions

Background

We will next see that the symplectic condition not only simplifies the computation of eigenvalues, but also influences how the eigenvalues depend on various parameters. For the general case of a $2n \times 2n$ matrix M , the characteristic polynomial (4.1) is of degree $2n$. Correspondingly, one might think that the determination of the eigenvalues from (4.2) would require finding the roots of a $2n$ degree polynomial. However, if M is symplectic, we can use the fact that P and Q have the same zeroes and the symmetry property (4.8) to reduce the problem to that of finding the roots of an n degree polynomial. Introduce a variable w by the relation

$$w = \lambda + 1/\lambda. \quad (3.4.12)$$

Equation (4.12) can be inverted to give the result

$$\lambda = [w \pm (w^2 - 4)^{1/2}]/2. \quad (3.4.13)$$

Since $P(\lambda)$ is a polynomial of degree $2n$, it follows from (4.4), (4.7), and (4.8) that Q must be of the form

$$Q(\lambda) = Q_r(w) = \sum_{m=0}^n b_m w^m \quad (3.4.14)$$

with

$$b_n = 1. \quad (3.4.15)$$

Note that Q_r , which we will call the *reduced* characteristic polynomial of M , has degree n . The eigenvalues of M can now be determined by finding the n roots of the equation

$$Q_r(w) = 0, \quad (3.4.16)$$

and substituting these roots into (4.13).

Let us see how the results just described work out in the cases $n = 1$ and $n = 2$.

The 2×2 Case

Suppose $n = 1$. Then, if M is symplectic, we have the result

$$P(\lambda) = \lambda^2 - A\lambda + 1, \quad (3.4.17)$$

with the coefficient (parameter) A given by the relation

$$A = \text{tr } (M). \quad (3.4.18)$$

For $Q_r(w)$ we find the result

$$Q_r(w) = w - A. \quad (3.4.19)$$

Evidently the solution to (4.16) in this case is simply $w = A$, and we find from (4.13) the eigenvalues

$$\lambda = [A \pm (A^2 - 4)^{1/2}]/2. \quad (3.4.20)$$

Note that (4.20) gives eigenvalues on the unit circle (stability) when

$$-2 < A < 2, \quad (3.4.21)$$

and real eigenvalues (instability) otherwise. Figure 4.3, which is to be compared with Figure 4.1, illustrates the nature of the eigenvalues λ as a function of A .

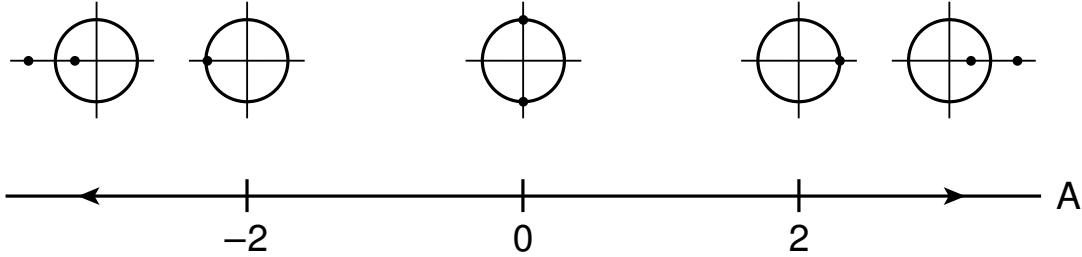


Figure 3.4.3: Eigenvalues of a 2×2 real symplectic matrix M as a function of $A = \text{tr}(M)$.

The 4×4 Case

Suppose $n = 2$. Then, if M is symplectic, the characteristic polynomial has the form

$$P(\lambda) = \lambda^4 - A\lambda^3 + B\lambda^2 - A\lambda + 1. \quad (3.4.22)$$

The coefficients (parameters) A and B can be found from the relations

$$A = [P(-1) - P(1)]/4, \quad (3.4.23)$$

$$B = [P(-1) + P(1)]/2 - 2. \quad (3.4.24)$$

They can also be found directly in terms of M using the relations

$$A = \text{tr}(M), \quad (3.4.25)$$

$$B = \{\text{tr}(M)^2 - \text{tr}(M^2)\}/2. \quad (3.4.26)$$

See Exercise 3.7.17. For $Q_r(w)$ we find the result

$$Q_r(w) = w^2 - Aw + B - 2. \quad (3.4.27)$$

The solutions to (4.16) in this case are

$$w = [A \pm (A^2 - 4B + 8)^{1/2}]/2. \quad (3.4.28)$$

These solutions are to be substituted into (4.13). Observe that in general there are four choices of signs to be made corresponding to the four possible eigenvalues expected for M . Figure 4.4, which is to be compared with Figure 4.2, illustrates the nature of the eigenvalues as a function of A and B . We note that the region of stability is the arrow-head shaped domain in which the following conditions are satisfied simultaneously:

$$B \geq 2A - 2, \quad B \geq -2A - 2,$$

$$B \leq A^2/4 + 2, \quad B \leq 6. \quad (3.4.29)$$

Transitions from stability to instability through the points $\lambda = \pm 1$ occur across the line segments

$$\lambda = +1 : B = 2A - 2 \text{ and } B \in [-2, 6], \quad (3.4.30)$$

$$\lambda = -1 : B = -2A - 2 \text{ and } B \in [-2, 6]. \quad (3.4.31)$$

Transitions to instability through Krein collisions, see Case 5 of Figure 4.2B and Section 3.5, occur across the parabolic segment

$$B = A^2/4 + 2, \quad (3.4.32)$$

with

$$A \in [-4, 4]. \quad (3.4.33)$$

The 6×6 Case

The 6×6 case can be treated in a manner analogous to the 2×2 and 4×4 cases. In the 6×6 case one needs to solve a cubic equation to find the eigenvalues. See Exercise 4.14.

Dimension Counting

We close this subsection with a remark on dimension counting. We see from (4.14) and (4.15) that the spectrum of a $2n \times 2n$ symplectic matrix is determined by the n parameters b_0, b_1, \dots, b_{n-1} . We will learn in Section 3.7 that symplectic matrices form a Lie group, and that the dimensionality of this group is $n(2n + 1)$. Since $n(2n + 1)$ is much larger than n , it follows that many different symplectic matrices have the same spectrum. This fact is relevant to accelerator design. As outlined at the beginning of this section, the linear stability of closed orbits in an accelerator is governed by the spectrum of the linear part of its one-turn transfer map. It is therefore important to be able to control the spectrum, and there are typically many knobs in an accelerator control room for this purpose. Despite these many knobs, accelerator operators often discover to their dismay that they are unable to adjust the spectrum at will. The dimension counting comparison tells us why. Much of the possible knob turning simply leads to different symplectic matrices having the same or nearly the same spectrum.

3.4.5 In Praise of and Gratitude for the Symplectic Condition

We have learned that the symplectic condition guarantees that there are symplectic maps \mathcal{M} whose linear parts M have all their eigenvalues on the unit circle and distinct. Thus, it is in principle possible to build circular (ring) accelerators and storage rings with a stable closed orbit.

Moreover, the set of $2n \times 2n$ real symplectic matrices whose eigenvalues lie on the unit circle and are distinct is *open* in the set of all $2n \times 2n$ symplectic matrices. By this we mean that if M is a real symplectic matrix all of whose eigenvalues lie on the unit circle and are distinct, then the same is true of all real symplectic matrices sufficiently near M . To

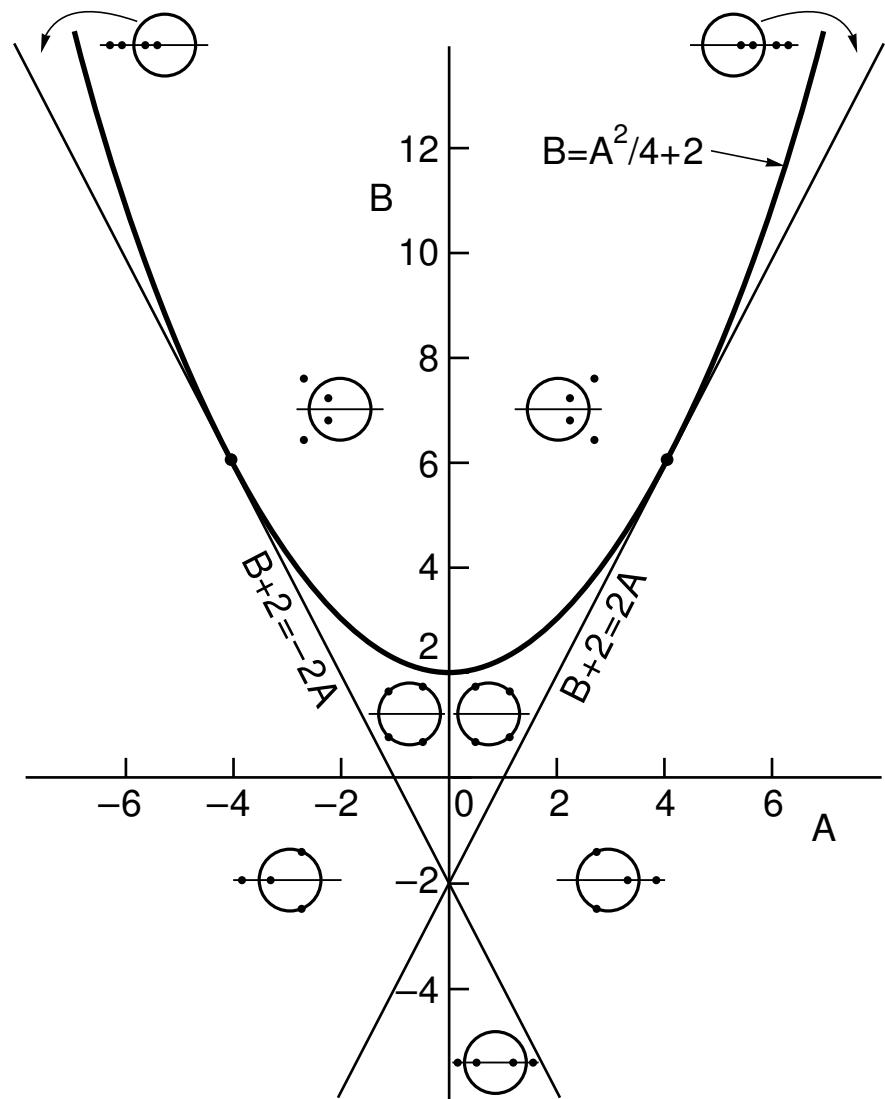


Figure 3.4.4: Eigenvalues of a 4×4 real symplectic matrix M as a function of the coefficients A and B in its characteristic polynomial.

see this, suppose that M is changed slightly, but in such a way that it remains symplectic. The eigenvalues of a matrix are the roots of a polynomial whose coefficients are continuous functions of the entries in the matrix. See (4.1) and (4.2). Also, the roots of a polynomial are continuous functions of the coefficients in the polynomial. It follows that the eigenvalues of a matrix are continuous functions of the entries in the matrix. That is, if the matrix is slightly changed, its eigenvalues are also only slightly changed. But if the eigenvalues are initially on the unit circle and distinct, there are no nearby eigenvalue configurations for a symplectic matrix where the eigenvalues are not all distinct or at least one eigenvalue is outside the unit circle. Thus, if the change in M is finite but small enough, then the eigenvalues must remain on the unit circle and must still be distinct.

The fact that this stability cannot be destroyed by small and symplectic perturbations should be of comfort to accelerator designers and builders because it means that, at least in the linear approximation, the stability of orbits will not be damaged by small errors in machine construction and operation. That is, thanks to the symplectic condition, accelerator performance is robust under small fabrication and control parameter errors.

Even more can be said. In our discussion we have implicitly assumed the existence of a closed orbit. That is, we have assumed that \mathcal{M} has a fixed point z_f . It can be shown that if a symplectic map has a stable fixed point (a fixed point for which M has all its eigenvalues on the unit circle and distinct), then all nearby symplectic maps will also have a stable fixed point. See Subsection 29.4.5. Thus, if a circular accelerator or storage ring is *designed* to have a stable closed orbit, both the existence and the stability of a closed orbit for the actual machine are guaranteed even in the presence of small fabrication and control parameter errors providing these errors are not too large.

Exercises

3.4.1. Verify (4.5) starting with (4.3).

3.4.2. Show, using (3.33), that the eigenvalues of a symplectic matrix cannot all have absolute value less than 1, nor can they all have absolute value greater than 1.

3.4.3. Show that, for a real 4×4 symplectic matrix M , that all the generic eigenvalue configurations of Figure 4.2 are unchanged by small perturbations of M providing the perturbed M are also symplectic. That is why these configurations are called *generic*.

3.4.4. Show that the eigenvalues of J are all $\pm i$.

3.4.5. Suppose M' is defined in terms of M by (2.15). Show that M and M' have the same spectrum.

3.4.6. Suppose λ_0 is a complex eigenvalue of a real symplectic matrix. Show that the Jordan block structures for the eigenvalues λ_0 and $\bar{\lambda}_0$ are the same. Suppose λ_0 is an eigenvalue of a (possibly complex) symplectic matrix. Use (1.9) to show that the Jordan block structures for the eigenvalues λ_0 and λ_0^{-1} are the same.

3.4.7. Given (4.12), verify (4.13).

3.4.8. Using (4.4), (4.7), and (4.8), verify (4.14) and (4.15).

3.4.9. Verify (4.17) through (4.21). Let O and F be the matrices

$$O = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \text{ and } F = \begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix}. \quad (3.4.34)$$

The matrix O is the 2×2 version of the transfer matrix for a drift of length a , and the matrix F is the 2×2 version of the transfer matrix for a focusing element with focal length

$$f = -1/b. \quad (3.4.35)$$

Therefore we expect that $a > 0$ and (assuming F is focusing) $b < 0$. See Chapter 13. Suppose M is the 2×2 matrix

$$M = OFO. \quad (3.4.36)$$

It is the 2×2 version of the transfer matrix for an OFO cell. Verify that M and its factors in (4.37) are symplectic matrices. Verify that

$$A = \text{tr}(M) = 2(1 + ab). \quad (3.4.37)$$

By referring to Figures 4.1 and 4.3 verify that for M there are the following cases:

$$\begin{aligned} &\text{hyperbolic when } ab > 0, \\ &\text{elliptic when } -2 < ab < 0, \\ &\text{inversion hyperbolic when } ab < -2. \end{aligned} \quad (3.4.38)$$

Suppose M' is the matrix

$$M' = FOF. \quad (3.4.39)$$

It is the 2×2 version of the transfer matrix for a FOF cell. Verify that M' is symplectic. Verify that

$$A' = \text{tr}(M') = 2(1 + ab). \quad (3.4.40)$$

Verify that for M' there are also the cases (4.38).

3.4.10. Verify (4.22) through (4.24). Verify (4.27) and (4.28).

3.4.11. Study Figure 4.4. Verify the statements made in connection with (4.29) through (4.33).

3.4.12. Where do the eigenvalues λ lie when A and B are on the portion of the parabola (4.32) having $A > 4$ or $A < -4$? Where do the eigenvalues lie when A and B are on the portions of the lines $B = \pm 2A - 2$ and $B \notin [-2, 6]$? Find where the eigenvalues lie for the cases $A = 4, B = 6; A = -4, B = 6; A = 0, B = -2$.

3.4.13. Consider a 4-dimensional phase space. Suppose the phase-space variables are arranged according to (2.4) rather than (2.1). Verify that doing so makes no difference for the discussion of the present section. Suppose, with the arrangement (2.4), that a 4×4 symplectic matrix M is written in the 2×2 block form (3.1), and the blocks B and C are

identically zero. In this case the x_1, y_1 space is mapped into itself, and the x_2, y_2 space is mapped into itself. See (2.4). With this assumption, show that the characteristic polynomial for M takes the form

$$P(\lambda) = (\lambda^2 - \alpha\lambda + 1)(\lambda^2 - \delta\lambda + 1). \quad (3.4.41)$$

Here α is the trace of the upper left 2×2 block in M , and δ is the trace of the lower right 2×2 block in M . Show that, according to (4.21), the quantities α, δ must lie in the square

$$-2 < \alpha < 2, -2 < \delta < 2 \quad (3.4.42)$$

in order that all eigenvalues of M lie on the unit circle (stability). Show that, when multiplied out, (4.41) takes the form

$$P(\lambda) = \lambda^4 - (\alpha + \delta)\lambda^3 + (2 + \alpha\delta)\lambda^2 - (\alpha + \delta)\lambda + 1. \quad (3.4.43)$$

Now compare (4.22) and (4.43) to get the results

$$A = \alpha + \delta, \quad (3.4.44)$$

$$B = 2 + \alpha\delta. \quad (3.4.45)$$

Show that the interior of the square (4.42) maps into the arrow-head shaped domain (4.29) of Figure 4.4 under the transformation given by (4.44) and (4.45). Also show that the exterior of the square maps to points outside the arrow-head shaped domain. Does one get all points outside the arrow-head shaped domain?

3.4.14. Consider a 6-dimensional phase space. Show that in this case $P(\lambda)$ for a symplectic matrix can be written in the form

$$P(\lambda) = \lambda^6 - A\lambda^5 + B\lambda^4 - C\lambda^3 + B\lambda^2 - A\lambda + 1, \quad (3.4.46)$$

and $Q_r(w)$ takes the form

$$Q_r(w) = w^3 - Aw^2 + (B - 3)w + (2A - C). \quad (3.4.47)$$

What region of the A, B, C parameter space gives stability (all eigenvalues on the unit circle)? That is, what is the 3-dimensional analog of the arrow-head shaped domain of Figure 4.4?

3.4.15. Look at the coefficients appearing in (4.46). Listing them from left to right, we see that they have the values $1, -A, B, -C, B, -A, 1$. This sequence is a *palindrome*. That is, when read backwards, the result is the same as reading forwards. Observe that this feature also appears in (4.17) and (4.22). Also observe that the first and last coefficients always have the value $+1$. Prove that these results hold for any phase-space dimension.

3.5 Eigenvector Structure, Normal Forms, and Stability

3.5.1 Eigenvector Basis

Let M be a $2n \times 2n$ real symplectic matrix. Suppose its eigenvalues are all distinct. Call them $\lambda_1, \lambda_2, \dots, \lambda_{2n}$, and call the associated eigenvectors $\psi_1, \psi_2, \dots, \psi_{2n}$. Then we have $2n$ relations of the form

$$M\psi_j = \lambda_j \psi_j. \quad (3.5.1)$$

Note that if any λ_j is complex, the corresponding ψ_j must also have complex entries. Finally, since the λ_j are assumed to be distinct, the $2n$ vectors ψ_j must be linearly independent, and must consequently form a basis.

3.5.2 J -Based Angular Inner Product

Let $(,)$ denote the usual *complex* scalar product.⁷ Introduce an *angular inner product* \langle , \rangle by the rule

$$\langle \chi, \theta \rangle = (\chi, K\theta) \quad (3.5.2)$$

with K defined by the relation

$$K = -iJ. \quad (3.5.3)$$

Here χ and θ are any two vectors. We note that K is Hermitian,

$$K^\dagger = K, \quad (3.5.4)$$

with respect to the standard complex scalar product $(,)$. Consequently, we have the relation

$$\langle \theta, \chi \rangle = \overline{\langle \chi, \theta \rangle}. \quad (3.5.5)$$

Finally we observe that for real vectors the angular inner product, apart from a factor of $-i$, is just the fundamental symplectic 2-form (2.3).⁸

3.5.3 Use of Angular Inner Product

What is the angular inner product good for? Let ψ_j and ψ_k be two eigenvectors. We have the result

$$\begin{aligned} \langle \psi_j, M\psi_k \rangle &= (\psi_j, KM\psi_k) = \lambda_k(\psi_j, K\psi_k) \\ &= \lambda_k \langle \psi_j, \psi_k \rangle. \end{aligned} \quad (3.5.6)$$

From the symplectic condition (1.2) we conclude that

$$KM = (M^T)^{-1}K. \quad (3.5.7)$$

⁷We adopt the usual physicists' convention that $(\alpha\phi, \beta\psi) = \bar{\alpha}\beta(\phi, \psi)$. Mathematicians frequently follow the convention that $(\alpha\phi, \beta\psi) = \alpha\bar{\beta}(\phi, \psi)$.

⁸Apart from a factor of $-i$, the angular inner product (5.2) is sometimes called the Lagrange bracket of $\bar{\chi}$ and θ .

Consequently, the quantity $\langle \psi_j, M\psi_k \rangle$ can also be written in the form

$$\begin{aligned}\langle \psi_j, M\psi_k \rangle &= (\psi_j, KM\psi_k) = (\psi_j, (M^T)^{-1}K\psi_k) \\ &= (M^{-1}\psi_j, K\psi_k) = \bar{\lambda}_j^{-1}(\psi_j, K\psi_k) \\ &= \bar{\lambda}_j^{-1}\langle \psi_j, \psi_k \rangle.\end{aligned}\tag{3.5.8}$$

Here we have used the relation

$$M^{-1}\psi_j = \lambda_j^{-1}\psi_j,\tag{3.5.9}$$

which follows from (5.1). Comparison of the relations (5.6) and (5.8) gives the result

$$(\bar{\lambda}_j^{-1} - \lambda_k)\langle \psi_j, \psi_k \rangle = 0.\tag{3.5.10}$$

Consequently, we have the *orthogonality* relation

$$\langle \psi_j, \psi_k \rangle = 0 \text{ if } \bar{\lambda}_j^{-1} \neq \lambda_k.\tag{3.5.11}$$

The exact consequences of the orthogonality relation depend on the nature of the spectrum. Suppose, for the purposes of this section, that all the eigenvalues of M are complex, distinct, and lie on the unit circle. Then the λ_j can be written in the form

$$\lambda_j = e^{i\phi_j}\tag{3.5.12}$$

where the phases ϕ_j are real. In this case we have the relation

$$\bar{\lambda}_j^{-1} = \lambda_j.\tag{3.5.13}$$

Correspondingly, the orthogonality relation (5.11) becomes the relation

$$\langle \psi_j, \psi_k \rangle = 0 \text{ if } \lambda_j \neq \lambda_k.\tag{3.5.14}$$

Further, we claim that

$$\langle \psi_k, \psi_k \rangle \neq 0 \text{ for all } k.\tag{3.5.15}$$

For suppose $\langle \psi_k, \psi_k \rangle$ did vanish for some k . Then, by (5.14), $\langle \psi_j, \psi_k \rangle$ would vanish for all j . Correspondingly, as a result of the definitions (5.2) and (5.3), we would conclude that

$$(\psi_j, J\psi_k) = 0 \text{ for all } j,\tag{3.5.16}$$

and consequently

$$J\psi_k = 0\tag{3.5.17}$$

since the ψ_j form a basis. But the matrix J is invertible. Thus (5.17) implies that ψ_k itself must vanish, which is impossible because the vectors ψ_j form a basis.

3.5.4 Definition and Use of Signature

Observe that, according to (5.5), the quantities $\langle \psi_j, \psi_j \rangle$ must be real. Since they cannot vanish, they must be positive or negative. Suppose we rephase and renormalize the vectors ψ_j to produce new vectors ψ'_j defined by the relations

$$\psi'_j = r_j e^{i\chi_j} \psi_j. \quad (3.5.18)$$

Here the r_j are real, positive quantities, and the phases χ_j are arbitrary. We then find the relations

$$\langle \psi'_j, \psi'_j \rangle = r_j^2 \langle \psi_j, \psi_j \rangle. \quad (3.5.19)$$

We see that the r_j can be selected in such a way that

$$\langle \psi'_j, \psi'_j \rangle = \sigma_j \quad (3.5.20)$$

where σ_j has the (possible) values

$$\sigma_j = \pm 1. \quad (3.5.21)$$

Note that the sign of σ_j is independent of the phase χ_j . Thus, the sign is an *intrinsic* property of the vector ψ_j . It is called the *signature* of ψ_j . From now on, we drop the prime notation. With this understanding, and recalling that the λ_j are assumed to be distinct, we may require that the ψ_j be normalized in such a way that they obey the orthogonality relation

$$\langle \psi_j, \psi_k \rangle = \sigma_j \delta_{jk}. \quad (3.5.22)$$

Consider some eigenvalue λ_k . Since $\bar{\lambda}_k$ is also an eigenvalue, it must be one of the λ_j . Let $\lambda_{k'}$ denote this particular λ_j . Then, we have the relations

$$M\psi_k = \lambda_k \psi_k, \quad (3.5.23)$$

$$M\psi_{k'} = \lambda_{k'} \psi_{k'} = \bar{\lambda}_k \psi_{k'}. \quad (3.5.24)$$

Complex conjugate (5.23) to get the relation

$$M\bar{\psi}_k = \bar{\lambda}_k \bar{\psi}_k = \lambda_{k'} \bar{\psi}_{k'}. \quad (3.5.25)$$

We observe from (5.24) and (5.25) that $\psi_{k'}$ and $\bar{\psi}_k$ are both eigenvectors of M with the same eigenvalue $\lambda_{k'}$. Consider the vector $\bar{\psi}_k$. By the same argument that led to (5.14), it must be orthogonal to all ψ_j with $j \neq k'$. Since the ψ_j form a basis, it follows that $\bar{\psi}_k$ must be proportional to $\psi_{k'}$. Thus, there is a relation of the form

$$\bar{\psi}_k = \alpha_k \psi_{k'} \quad (3.5.26)$$

where α_k is some proportionality constant yet to be determined. Consider the quantity $\sigma_k = \langle \psi_k, \psi_k \rangle$. Working it out in component form gives the result

$$\sigma_k = \langle \psi_k, \psi_k \rangle = \sum_{\alpha} \bar{\psi}_{k,\alpha} (K\psi_k)_{\alpha} = \sum_{\alpha, \beta} \bar{\psi}_{k,\alpha} K_{\alpha\beta} \psi_{k,\beta}. \quad (3.5.27)$$

Here the quantities $\psi_{k,\beta}$ are the components of ψ_k . Next consider the quantity $\langle \bar{\psi}_k, \bar{\psi}_k \rangle$. It evidently has the value

$$\begin{aligned}\langle \bar{\psi}_k, \bar{\psi}_k \rangle &= \sum_{\alpha,\beta} \psi_{k,\alpha} K_{\alpha\beta} \bar{\psi}_{k,\beta} = - \sum_{\alpha,\beta} \bar{\psi}_{k,\beta} K_{\beta\alpha} \psi_{k,\alpha} \\ &= -\langle \psi_k, \psi_k \rangle = -\sigma_k.\end{aligned}\quad (3.5.28)$$

Here use has been made of the antisymmetry of K . But from (5.26) we have the relation

$$\langle \bar{\psi}_k, \bar{\psi}_k \rangle = |\alpha_k|^2 \langle \psi_{k'}, \psi_{k'} \rangle = |\alpha_k|^2 \sigma_{k'}. \quad (3.5.29)$$

Comparison of (5.28) and (5.29) gives the result

$$|\alpha_k|^2 \sigma_{k'} = -\sigma_k. \quad (3.5.30)$$

Two relations follow from (5.30) and (5.21). First, we have the relation

$$\sigma_{k'} = -\sigma_k. \quad (3.5.31)$$

We have learned that if $\lambda_{k'} = \bar{\lambda}_k$, then ψ_k and $\psi_{k'}$ have opposite signature. It follows that half the ψ_j have signature +1, and half have signature -1. Second, we also have the relation

$$|\alpha_k|^2 = 1. \quad (3.5.32)$$

Thus, α_k must be just a phase factor. Since the vectors ψ_k and $\psi_{k'}$ are only defined up to overall phase factors, we may set $\alpha_k = 1$ without loss of generality to get the relation

$$\bar{\psi}_k = \psi_{k'}. \quad (3.5.33)$$

The preceding discussion makes it possible to improve our notation. Suppose the ψ_j are relabeled in such a way that the vectors ψ_ℓ with $\ell = 1, 2, \dots, n$ have positive signature. Let the corresponding ψ_j with negative signature be labeled as $\psi_{-\ell}$. That is, arrange the labeling scheme so that the following relations hold with $\ell, m = 1, 2, \dots, n$:

$$\langle \psi_\ell, \psi_m \rangle = \delta_{\ell,m}, \quad (3.5.34)$$

$$\langle \psi_{-\ell}, \psi_{-m} \rangle = -\delta_{\ell,m}, \quad (3.5.35)$$

$$\langle \psi_\ell, \psi_{-m} \rangle = \langle \psi_{-\ell}, \psi_m \rangle = 0, \quad (3.5.36)$$

$$\bar{\lambda}_\ell = \lambda_{-\ell}, \quad (3.5.37)$$

$$\bar{\psi}_\ell = \psi_{-\ell}. \quad (3.5.38)$$

By their association with the ψ_ℓ , the eigenvalues λ_ℓ are also said to have positive signature. Correspondingly, the eigenvalues $\lambda_{-\ell}$ are said to have negative signature.

3.5.5 Definition of Phase Advances and Tunes

Consider the eigenvalues λ_ℓ corresponding to the vectors ψ_ℓ having *positive* signature. That is, consider the eigenvalues with positive signature. Define phases ϕ_ℓ by the relation

$$\lambda_\ell = e^{i\phi_\ell}. \quad (3.5.39)$$

Evidently these phases are defined modulo 2π . For the discussion that follows, it is convenient to take them to lie in the range $(-\pi, \pi)$. The quantities ϕ_ℓ with $\ell = 1, 2, \dots, n$ are called the *phase advances* of M . Also, define corresponding quantities T_ℓ by the relations

$$T_\ell = \phi_\ell/(2\pi). \quad (3.5.40)$$

Evidently the T_ℓ are defined modulo 1, but our choice of the range $(-\pi, \pi)$ for phases places the T_ℓ in the range $(-1/2, 1/2)$. The quantities T_ℓ are called the *tunes* of M .⁹

Example 5.1: Let M be the 2×2 matrix

$$M = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix}. \quad (3.5.41)$$

Since M is 2×2 and has determinant +1, it must be symplectic. See Exercise 1.3. A simple calculation shows that M has the eigenvectors

$$\psi_{+1} = (1/\sqrt{2}) \begin{pmatrix} 1 \\ i \end{pmatrix}, \quad (3.5.42)$$

$$\psi_{-1} = (1/\sqrt{2}) \begin{pmatrix} 1 \\ -i \end{pmatrix}, \quad (3.5.43)$$

with eigenvalues $e^{+i\phi}$ and $e^{-i\phi}$, respectively. Also, it is easily checked that ψ_{+1} and ψ_{-1} have signatures +1 and -1, respectively. It follows that the phase advance of M is ϕ , and the tune is $\phi/(2\pi)$.

3.5.6 The Krein-Moser Theorem and Krein Collisions

The discussion so far has been restricted to the case for which the eigenvalues of M are distinct and lie on the unit circle. Suppose M is varied in such a way that two eigenvalues collide. In actuality (when $n \geq 2$), two pairs must collide. See Case 5 of the degenerate configurations of Figure 4.2. Then, as M is varied further, the eigenvalues can pass over each other to give Case 5 of the generic configurations, or they can leave the unit circle to give Case 1 of the generic configurations. See Figure 5.1. It can be shown that if the

⁹In Chapter 30 we will see that if M can be viewed as the product of many symplectic matrices, all of which are near the identity, then the phase advances and tunes of M can be defined in such a way that they may lie *outside* the ranges $(-\pi, \pi)$ and $(-1/2, 1/2)$, respectively. However, modulo 2π or 1, respectively, these phase advances and tunes still agree with those defined above.

colliding eigenvalues have the *same* signature, which is the case of nearly equal tunes, then they cannot leave the unit circle and must pass over each other. Also, when the eigenvalues do collide, M remains diagonalizable even though its eigenvalues are no longer distinct. This result is called the *Krein-Moser* theorem or condition.

The same signature case is the case of nearly equal tunes. By contrast, if the eigenvalues have opposite signatures, then there are small perturbations of M that will cause the eigenvalues to collide and then leave the unit circle thereby forming a Krein quartet. Such a collision is called a *Krein collision*. This is the case of nearly equal and opposite tunes. For a proof of these assertions, see Exercise 3.8.18.

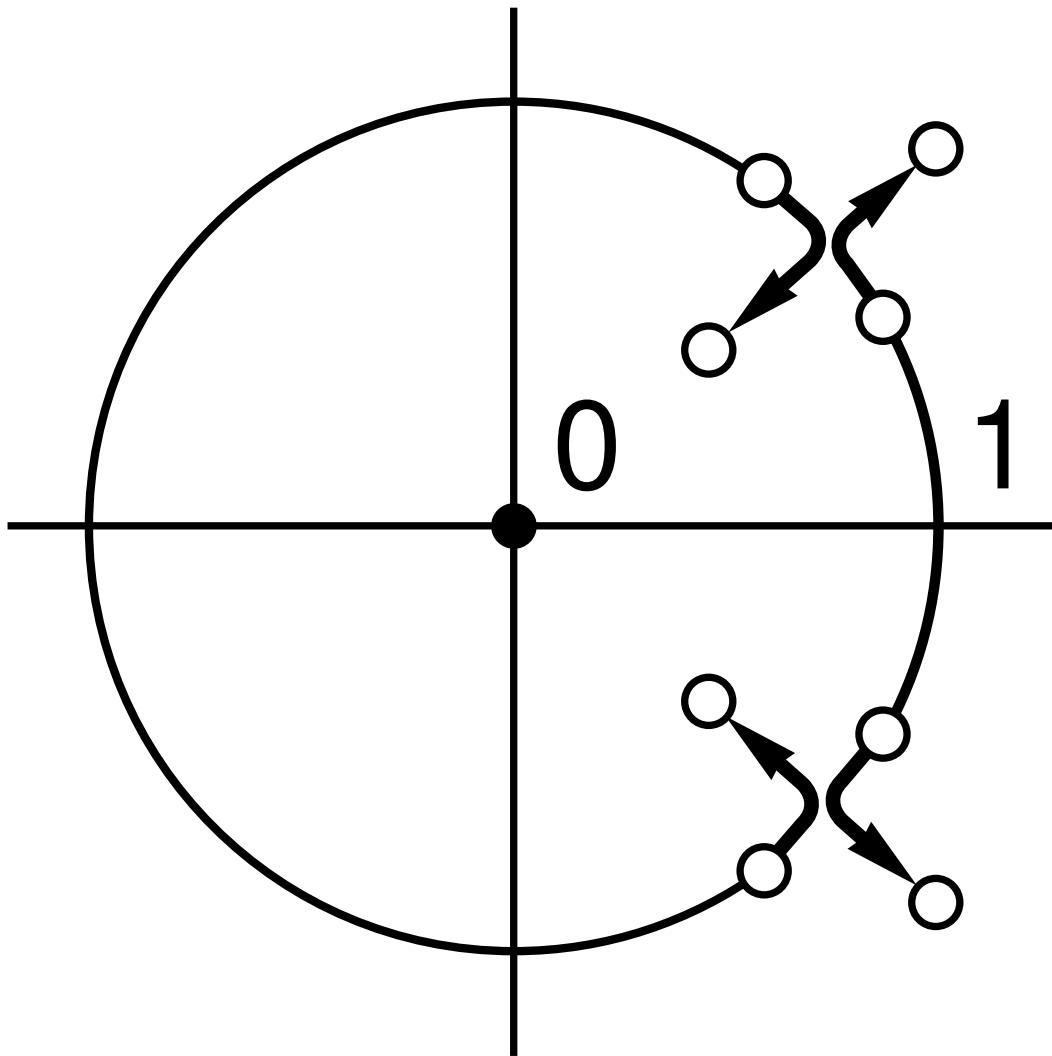


Figure 3.5.1: Illustration of eigenvalues colliding and then leaving the unit circle to form what is called a Krein quartet.

The Krein-Moser theorem is a remarkable result. Consider, for example, the case $n = 2$ corresponding to 4×4 symplectic matrices M . Suppose M is such that two eigenvalues of the *same* signature collide. In this case the values A and B given by (4.25) and (4.26) must lie on the parabolic segment specified by (4.32) and (4.33). Now consider all symplectic

matrices near M . They form a 10-dimensional space. See (7.35). Compute the values of A and B for these matrices. According to the Krein-Moser theorem, *all* these values must also lie on the parabolic segment or in the arrow-head shaped domain of Figure 4.4 *below* the parabolic segment, and *none* will be *above* the parabolic segment. No matter which way we move in M space (at least locally), we *cannot* get points in the A, B image space that lie above the parabolic segment. By contrast, suppose we go to another region of M space where two eigenvalues again collide, but now have *opposite* signature. In this case the A, B values will again lie on the parabolic segment. However, if we now consider all symplectic matrices near M , we will find that their image in A, B space lies on the parabolic segment, or in the arrow-head shaped domain, *or* in the region immediately *above* the parabolic segment. Thus in this case, by making a proper move in M space, we *can* get anywhere (locally) in A, B space.

Recall that the situation in which all the eigenvalues lie on the unit circle and are distinct corresponds to stability (in the linear approximation), and the case of any eigenvalue off the unit circle corresponds to instability. See the discussion in the beginning of Section 3.4. Consequently, to achieve qualitative *insensitivity* to small perturbations, the case of nearly equal and opposite tunes should be avoided.¹⁰

3.5.7 Normal Forms

The last topic to be discussed in this section is that of a *normal form* for M . As will be seen, a normal form for M is a particularly simple form for M achieved by a symplectic similarity transformation. Recall that we have assumed that all eigenvalues are complex, distinct, and lie on the unit circle. Thus, we restrict our attention here to this case. (Normal forms are also known for the other cases, but their discussion is more complicated.) We also assume, without loss of generality, that M is symplectic with respect to the J of (2.10).

Suppose the eigenvectors ψ_ℓ are decomposed into real and imaginary parts by writing the relations

$$\psi_\ell = \xi_\ell + i\eta_\ell, \quad (3.5.44)$$

where the vectors ξ_ℓ and η_ℓ are real. From (5.38) we conclude that the $\psi_{-\ell}$ have the decomposition

$$\psi_{-\ell} = \xi_\ell - i\eta_\ell. \quad (3.5.45)$$

Insert the representations (5.44) and (5.45) into (5.34) and (5.36), and equate real and imaginary parts. Doing so, and use of (5.3), gives the results

$$(\xi_\ell, J\xi_m) = 0 , \quad (\eta_\ell, J\eta_m) = 0, \quad (3.5.46)$$

$$2(\xi_\ell, J\eta_m) = \delta_{\ell m} , \quad 2(\eta_\ell, J\xi_m) = -\delta_{\ell m}. \quad (3.5.47)$$

Also insert the representation (5.44) into the relation

$$M\psi_\ell = \lambda_\ell\psi_\ell = e^{i\phi_\ell}\psi_\ell, \quad (3.5.48)$$

¹⁰Instability can also occur if the eigenvalues are on the unit circle but M cannot be diagonalized. This can occur when the eigenvalues are ± 1 and as well as at Krein collisions. Moreover, the eigenvalues can also leave the unit circle through these degenerate configurations. Therefore, integer and half-integer tunes should also be avoided.

and equate real and imaginary parts. Doing so gives the result

$$M\xi_\ell = (\cos \phi_\ell)\xi_\ell - (\sin \phi_\ell)\eta_\ell, \quad (3.5.49)$$

$$M\eta_\ell = (\sin \phi_\ell)\xi_\ell + (\cos \phi_\ell)\eta_\ell. \quad (3.5.50)$$

Consider the matrix A defined by the equation

$$A = \sqrt{2}(\xi_1, \eta_1, \xi_2, \eta_2, \dots, \xi_n, \eta_n). \quad (3.5.51)$$

Here each of the vectors ξ_ℓ and η_ℓ are to be viewed as column vectors so that the collection (5.51) forms a real $2n \times 2n$ matrix. Then it is easily verified that the relations (5.46) and (5.47) are equivalent to the matrix relation

$$A^TJA = J \quad (3.5.52)$$

providing the form (2.10) is employed for J . Thus, A is a symplectic matrix with respect to this J .

Finally, consider the matrix N defined by the equation

$$N = A^{-1}MA. \quad (3.5.53)$$

The matrix MA can be computed using (5.49), (5.50), and (5.51). One finds the result

$$\begin{aligned} MA &= \sqrt{2}(M\xi_1, M\eta_1, \dots, M\xi_n, M\eta_n) \\ &= \sqrt{2}(c_1\xi_1 - s_1\eta_1, s_1\xi_1 + c_1\eta_1, \dots, c_n\xi_n - s_n\eta_n, s_n\xi_n + c_n\eta_n). \end{aligned} \quad (3.5.54)$$

Here use has been made of the abbreviations

$$c_\ell = \cos \phi_\ell, \quad s_\ell = \sin \phi_\ell. \quad (3.5.55)$$

Since A is symplectic, the matrix A^{-1} may be formed using (1.9),

$$A^{-1} = -JA^TJ. \quad (3.5.56)$$

With this observation, we can continue the calculation. From (5.54) we find that JMA has the representation

$$JMA = \sqrt{2}(c_1J\xi_1 - s_1J\eta_1, s_1J\xi_1 + c_1J\eta_1, \dots, c_nJ\xi_n - s_nJ\eta_n, s_nJ\xi_n + c_nJ\eta_n). \quad (3.5.57)$$

The matrix A^TJMA can now be computed using (5.46), (5.47), (5.51), and (5.57). The result is

$$A^TJMA = \begin{pmatrix} B_1 & & & \\ & B_2 & & \\ & & \ddots & \\ & & & B_n \end{pmatrix}. \quad (3.5.58)$$

That is, all entries are zero save for n 2×2 blocks on the diagonal. The blocks themselves are given by the equations

$$B_\ell = \begin{pmatrix} -\sin \phi_\ell & \cos \phi_\ell \\ -\cos \phi_\ell & -\sin \phi_\ell \end{pmatrix}. \quad (3.5.59)$$

Finally, $N = A^{-1}MA = -JAT^TJMA$ can be computed by applying $-J$ to (5.58). The result is

$$N = \begin{pmatrix} R_1 & & & \\ & R_2 & & \\ & & \ddots & \\ & & & R_n \end{pmatrix}. \quad (3.5.60)$$

Again, all entries in N are zero save for n 2×2 blocks on the diagonal. The blocks themselves are given by the equations

$$R_\ell = \begin{pmatrix} \cos \phi_\ell & \sin \phi_\ell \\ -\sin \phi_\ell & \cos \phi_\ell \end{pmatrix}. \quad (3.5.61)$$

We conclude that, given any (real) symplectic matrix M whose eigenvalues are distinct and all lie on the unit circle, there is then a real symplectic similarity transformation (5.53) that brings M to the simple form (5.60). We call N the *normal form* of M , and say that M has been brought to normal form by the transforming matrix A . To reiterate, we have the key relations

$$N = A^{-1}MA \quad (3.5.62)$$

and

$$M = ANA^{-1}. \quad (3.5.63)$$

We also observe that the normal form is unique up to permutations of the ϕ_ℓ . There is somewhat more freedom available in the choice of the transforming matrix A . This freedom will be discussed later in Section 23.*. Finally, if we consider a two-dimensional phase space $z_\ell = (q_\ell; p_\ell)$ and define the action of R_ℓ as

$$z'_\ell = R_\ell z_\ell, \quad (3.5.64)$$

then we find the relations

$$q'_\ell = q_\ell \cos \phi_\ell + p_\ell \sin \phi_\ell, \quad (3.5.65)$$

$$p'_\ell = -q_\ell \sin \phi_\ell + p_\ell \cos \phi_\ell. \quad (3.5.66)$$

We see that the effect of R_ℓ is a clockwise rotation in the $(q_\ell; p_\ell)$ plane by the phase-advance angle ϕ_ℓ . Evidently, each R_ℓ , and therefore also N , is a real orthogonal matrix.

We close this subsection by remarking that there are also normal forms for symplectic matrices whose eigenvalues lie on the unit circle but are not distinct, or some or all of whose eigenvalues do not lie on the unit circle. The discussion of the general case is quite complicated, and falls outside the scope of this book. For a discussion of the 2×2 case, see Exercise 5.7. Further information may be found in the references listed at the end of this chapter. See also Subsection 27.2.2.

3.5.8 Stability

Suppose, as sketched at the beginning of Section 3.4, that a map \mathcal{M} acts on some space with coordinates z and suppose \mathcal{M} has a fixed point z_f ,

$$\mathcal{M}z_f = z_f. \quad (3.5.67)$$

In this subsection we will verify some of the claims made in Section 3.4 about the repeated action of \mathcal{M} on points near z_f .

A point near z_f can be written in the form $z_f + \delta$ where δ is a small vector. By the definition of *linear part* we assume the existence of an expansion of the form

$$\mathcal{M}(z_f + \delta) = z_f + M\delta + O(\delta^2) \quad (3.5.68)$$

where the matrix M describes the linear part of \mathcal{M} about z_f . It follows from repeated application of (5.68) that

$$\mathcal{M}^m(z_f + \delta) = z_f + M^m\delta + O(\delta^2). \quad (3.5.69)$$

Therefore, to analyze the stability of z_f in the linear approximation, we must examine the behavior of $M^m\delta$ for large m .

It can be shown that if all the eigenvectors of M lie within the unit circle in the complex plane, then

$$\lim_{m \rightarrow \infty} M^m = 0. \quad (3.5.70)$$

See Exercise 5.10. Therefore in this case, and neglecting terms of order δ^2 , we find that

$$\lim_{m \rightarrow \infty} \mathcal{M}^m(z_f + \delta) = z_f. \quad (3.5.71)$$

Thus, in this case and in linear approximation, z_f is an attractor.¹¹

For the case of symplectic maps, we have seen that not all eigenvalues of M can lie within the unit circle. For symplectic maps we are interested in the next best possibility, the case where all eigenvalues lie on the unit circle. Suppose all the eigenvalues of M lie on the unit circle and are distinct. Then, employing (5.63), we may write

$$M^m = (ANA^{-1})^m = AN^m A^{-1}. \quad (3.5.72)$$

Next, with the aid of vector and matrix norms, we find that

$$\|M^m\delta\| = \|AN^m A^{-1}\delta\| \leq \|A\| \|N^m\| \|A^{-1}\| \|\delta\|. \quad (3.5.73)$$

If we use the Euclidean norm, see Exercise 7.1, and observe that N^m is orthogonal, we find the result

$$\|N^m\| = (2n)^{1/2}. \quad (3.5.74)$$

Here we have used the group property that since N is $2n \times 2n$ and orthogonal, then so is N^m . See Subsection 6.1. Combining (5.73) and (5.74) gives the estimate

$$\|M^m\delta\| \leq (2n)^{1/2} \|A\| \|A^{-1}\| \|\delta\|. \quad (3.5.75)$$

We conclude that $M^m\delta$ remains bounded for all m , and therefore z_f is stable in the linear approximation.¹²

¹¹According to a theorem of Hartman, in this case z_f is also an attractor even if all nonlinear terms are taken into account.

¹²For a discussion of what occurs when the effect of the neglected nonlinear terms is concluded, see Chapter 35.

One might wonder whether the requirement that the eigenvalues be distinct is essential. It is. There are unstable counter examples for which the eigenvalues lie on the unit circle but are not distinct. The simplest 2×2 case is the matrix

$$M = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \quad (3.5.76)$$

where α is any nonzero real number. Since this M is 2×2 and has determinant $+1$, it is symplectic. Also, it has the non-distinct eigenvalue $+1$, and no eigenvalues off the unit circle. However, it is easily verified that

$$M^m = \begin{pmatrix} 1 & m\alpha \\ 0 & 1 \end{pmatrix}. \quad (3.5.77)$$

Therefore, if

$$\delta = \begin{pmatrix} 0 \\ \epsilon \end{pmatrix} \quad (3.5.78)$$

where ϵ is any small number, we have the result

$$M^m \delta = \begin{pmatrix} m\alpha\epsilon \\ \epsilon \end{pmatrix}. \quad (3.5.79)$$

We see that in this case $M^m \delta$ has entries that grow *linearly* in m as m increases, and therefore we may say that the fixed point z_f is linearly unstable.

We close this subsection with a simple example for which one eigenvalue is outside the unit circle and for which z_f is manifestly unstable. Consider the symplectic 2×2 case

$$M = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} \quad (3.5.80)$$

where $\lambda > 1$. In this case, if

$$\delta = \begin{pmatrix} \epsilon \\ 0 \end{pmatrix}, \quad (3.5.81)$$

we find the result

$$M^m \delta = \begin{pmatrix} \lambda^m \epsilon \\ 0 \end{pmatrix}. \quad (3.5.82)$$

Note that

$$\lambda^m = \exp(m \log \lambda) \quad (3.5.83)$$

and $\log \lambda > 0$ when $\lambda > 1$. Thus, now $M^m \delta$ has entries that grow *exponentially* in m as m increases, and therefore we may say that the fixed point z_f is exponentially unstable.

By the above examples we have demonstrated that there are cases where instability occurs when the eigenvalues are on the unit circle but not distinct, or some eigenvalue lies outside the unit circle. These result holds in general, and can be proved with the aid of normal forms for these cases.

Exercises

3.5.1. Show that if M is any matrix with distinct eigenvalues, then the corresponding eigenvectors must form a basis.

3.5.2. Carry out the calculations required for Example (5.1).

3.5.3. Show that if two tunes of a symplectic matrix M , call them T_1 and T_2 , are nearly equal (modulo the integers), then there are two associated eigenvalues that are nearly equal and have the same signature, and vice versa. In this case we have a relation of the form

$$T_1 - T_2 \simeq n \quad (3.5.84)$$

where n is an integer, and say that we are dealing with a potential *difference* resonance. By the Krein-Moser theorem, we know that under perturbation M remains diagonalizable and its eigenvalues remain on the unit circle. Therefore a difference resonance is harmless as far as stability is concerned.

Show that if two tunes are nearly equal and opposite (again modulo the integers), then there are two related eigenvalues that are nearly equal and have opposite signatures, and vice versa. In this case we have a relation of the form

$$T_1 + T_2 \simeq n \quad (3.5.85)$$

where n is an integer, and say that we are dealing with a potential *sum* resonance. By the Krein-Moser theorem, we know that in this case the eigenvalues can leave the unit circle under perturbation of M , and therefore a sum resonance is likely harmful.

3.5.4. Verify (5.46), (5.47), (5.49), and (5.50).

3.5.5. Verify (5.52)

3.5.6. Verify (5.54) and (5.57) through (5.61).

3.5.7. Suppose M and N are two matrices that are related by an equation of the form (5.53) where A is yet another matrix. If such a relation exists for some (invertible) matrix A , the matrices M and N are said to be *conjugate*, and we write $M \sim N$. It can be shown that conjugacy is an *equivalence* relation. This equivalence relation can be used to partition the set of all matrices into disjoint *equivalence classes*, which in this case are called *conjugacy classes*. See Exercise 5.12.7. Suppose M and N are symplectic, and a symplectic A can be found such that (5.53) holds. Then we will say that M and N are *symplectically conjugate*. Consider the case of all 2×2 symplectic matrices. Suppose that two such matrices, call them M and N , have the same Jordan normal form. Show that they are then symplectically conjugate. Hint: See the comment following Exercises 1.2 and 1.3. Show that the matrices

$$M = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad (3.5.86)$$

$$N = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad (3.5.87)$$

are symplectically conjugate, and find the conjugating matrix A .

3.5.8. Let χ and θ be any two (possibly complex) vectors, and let M be any real symplectic matrix. Define transformed vectors χ' and θ' by the rule

$$\chi' = M\chi, \quad (3.5.88)$$

$$\theta' = M\theta. \quad (3.5.89)$$

Show that the inner product $\langle \cdot, \cdot \rangle$ defined by (5.2) has the invariance property

$$\langle \chi', \theta' \rangle = \langle \chi, \theta \rangle. \quad (3.5.90)$$

3.5.9. Take matrix elements of (1.2) using the eigenvectors (5.1) to obtain the relation

$$(\psi_j, M^T JM \psi_k) = (\psi_j, J \psi_k). \quad (3.5.91)$$

Verify the manipulations

$$\begin{aligned} (\psi_j, M^T JM \psi_k) &= (M\psi_j, JM\psi_k) = (\lambda_j \psi_j, J\lambda_k \psi_k) = \bar{\lambda}_j \lambda_k (\psi_j, J\psi_k) \\ &= i\bar{\lambda}_j \lambda_k (\psi_j, K\psi_k) = i\bar{\lambda}_j \lambda_k \langle \psi_j, \psi_k \rangle, \end{aligned} \quad (3.5.92)$$

$$(\psi_j, J\psi_k) = i(\psi_j, K\psi_k) = i\langle \psi_j, \psi_k \rangle. \quad (3.5.93)$$

Show that (5.91) through (5.93) yield the result

$$(\bar{\lambda}_j \lambda_k - 1)\langle \psi_j, \psi_k \rangle = 0. \quad (3.5.94)$$

Verify that (5.94) is equivalent to (5.10) and (5.11).

3.5.10. Suppose that M is any matrix all of whose eigenvalues lie inside the unit circle. The aim of this exercise is to prove (5.70). Begin by assuming the eigenvalues of M are distinct. In this case there is an invertible matrix A such that

$$M = ADA^{-1} \quad (3.5.95)$$

where D is a diagonal matrix with the eigenvalues of M on its diagonal. Show from (5.95) that

$$M^m = AD^m A^{-1}. \quad (3.5.96)$$

Verify that

$$\lim_{m \rightarrow \infty} \lambda^m = 0 \text{ if } |\lambda| < 1, \quad (3.5.97)$$

and therefore

$$\lim_{m \rightarrow \infty} D^m = 0, \quad (3.5.98)$$

and thus, from (5.96), (5.70) holds.

If the eigenvalues of M are not distinct, it may not be diagonalizable. If M is not diagonalizable, it may still be brought to *Jordan* normal form,

$$M = ANA^{-1}, \quad (3.5.99)$$

so that we may write

$$M^m = AN^m A^{-1}. \quad (3.5.100)$$

Here N is a matrix having all zeroes except for possessing the eigenvalues of M on the diagonal and possibly ones just above the diagonal. For example, if M is 4×4 and not diagonalizable and all eigenvalues are the same, the most degenerate case would be that for which N has the form

$$N = \begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{pmatrix}. \quad (3.5.101)$$

In this case write

$$N = D + K \quad (3.5.102)$$

where D is diagonal and K is the matrix

$$K = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (3.5.103)$$

Verify that K is *nilpotent*, and in particular satisfies the relation

$$K^4 = 0. \quad (3.5.104)$$

By the binomial theorem for commuting entities, show that

$$\begin{aligned} N^m &= (D + K)^m \\ &= D^m + mD^{m-1}K + [m(m-1)/2!]D^{m-2}K^2 + [m(m-1)(m-2)/3!]D^{m-3}K^3. \end{aligned} \quad (3.5.105)$$

Verify that each term in (5.105) vanishes in the limit $m \rightarrow \infty$ if $|\lambda| < 1$, and thus, from (5.100), (5.70) holds. Proof of (5.70) for the general nondiagonalizable case follows similarly.

3.5.11. Scan Subsection 3.7.1 and Exercise 7.1. Verify (5.74) for the case of the Euclidean norm. Verify that for the spectral norm

$$\|N^m\| = 1. \quad (3.5.106)$$

3.6 Group Properties, Dyadic and Gram Matrices, and Bases

In this section we will describe what a group is, and will show that symplectic and orthogonal matrices form groups. Closely related to the symplectic and orthogonal groups are special bases called symplectic and orthonormal bases. The treatment of bases is facilitated by the introduction of dyadic and Gram matrices. Finally, given some basis, we will explore ways of specifying associated orthonormal and symplectic bases.

3.6.1 Group Properties

Abstract Groups

Arnold once asked and answered

What is a group? Algebraists teach that this is supposedly a set with two operations that satisfy a load of easily-forgettable axioms

We will begin with the abstract definition of a group. Then we will define matrix groups.

Abstractly, a group G is a set of elements subject to some rule of combination, usually called multiplication. For the moment, let us denote multiplication by the symbol \circ . Then we require the following properties:

1. If M and N are in G , so is the product $M \circ N$.
2. Multiplication is associative, $L \circ (M \circ N) = (L \circ M) \circ N$.
3. G contains a unique *identity* element I such that $I \circ M = M \circ I = M$ for all M in G .
4. If M is in G , there is a unique *inverse* element M^{-1} that is also in G such that $M \circ M^{-1} = M^{-1} \circ M = I$.

We remark that requirements 3 and 4 above can be weakened. For example, requirement 3 can be weakened to just require that there are left and right identity elements. These elements can then be proven to be unique and the same. Also, requirement 4 can be weakened to just require that there are left and right inverses. These elements can then be proven to be unique and the same.

A subgroup H of G is a subset of G whose elements also satisfies the above group properties. Any group G always has the identity element I as a subgroup. Whether it has any other nontrivial subgroups depends on the nature of G .

Matrix Groups

For matrices (assumed to be $n \times n$) we may take for the combination (multiplication) rule the ordinary operation of matrix multiplication. This automatically makes multiplication associative. Also, we may take for the identity element the identity matrix I , and for the inverse element the inverse matrix. With these provisos, a set of $n \times n$ matrices G forms a group if it satisfies the following properties:

1. If M and N are in G , so is the product MN .
2. The identity matrix I is in G .
3. If M is in G , M^{-1} exists and is also in G .

Note, in the matrix case, that iff a matrix M satisfies $\det(M) \neq 0$, then there is a unique matrix denoted by M^{-1} such that $MM^{-1} = M^{-1}M = I$.

Evidently, according to Exercise 1.4, Equation (1.9), and Exercises 1.5 and 1.6, the set of all $2n \times 2n$ symplectic matrices (for any particular value of n) forms a group. This group is

often denoted by the symbol $Sp(2n)$. More precisely, if we are working with *real* symplectic matrices, they form a group denoted by $Sp(2n, \mathbb{R})$; and if we are working with *complex* symplectic matrices, they form a group denoted by $Sp(2n, \mathbb{C})$. Where there is no possibility of confusion, we will use the notation $Sp(2n)$ to mean $Sp(2n, \mathbb{R})$.¹³

We remark that the symplectic condition (1.2) is a set of *algebraic* (polynomial) relations among the entries in M . For this reason, the symplectic group is an *algebraic group*.

An $n \times n$ matrix O that satisfies the condition

$$O^T O = I \quad (3.6.1)$$

is called *orthogonal*. It is easy to check that the set of all such matrices also forms a group called the orthogonal group, and denoted by the symbols $O(n, \mathbb{R})$ or $O(n, \mathbb{C})$ depending on the choice of field. Evidently, the orthogonal group is also an algebraic group. From (6.1) it follows that orthogonal matrices have the property

$$\det(O) = \pm 1. \quad (3.6.2)$$

Since the determinant of a matrix is a continuous function of the entries in the matrix, we conclude that the set of orthogonal matrices consists of two disjoint (and disconnected) subsets: those orthogonal matrices having determinant +1, and those having determinant -1. The subset of all orthogonal matrices with determinant +1 (called *proper* orthogonal matrices) forms a connected subgroup of the orthogonal group. This subgroup is called the *special* orthogonal group, and is referred to by the symbols $SO(n, \mathbb{R})$ or $SO(n, \mathbb{C})$. Note that the condition (6.1) can be written in the expanded form

$$O^T I O = I, \quad (3.6.3)$$

and this form is analogous to (1.2) with J replaced by I . Also, compare (6.1) and (3.1.14). This analogy results in some similarities in the ways that $O(n)$ and $Sp(2n)$ can be analyzed. However, in another sense, the two groups are polar opposites because I is symmetric and J is antisymmetric.

We remark for future use that the matrix J is both symplectic and special orthogonal. That is, J belongs both to $Sp(2n, \mathbb{R})$ and $SO(2n, \mathbb{R})$. See Exercises 1.1 and 1.5.

At this point one might wonder about generality. According to Exercise 2.7, the symplectic group consists of all linear transformations that preserve the fundamental symplectic 2-form (2.3). The matrix J in this 2-form has the property that it is antisymmetric and nonsingular. What happens if one replaces J by any (but real) antisymmetric nonsingular matrix? Does one still get a group, and is this group something new, or merely the symplectic group in disguise? Section 3.12 shows that one simply gets a variant of the symplectic group. It follows that the group $Sp(2n, \mathbb{R})$ is as general as might be desired.

Transformation Groups

We close this subsection with the comment that many groups arise naturally as *transformation groups*. Let \mathcal{Z} be some set/space and consider mappings/transformations \mathcal{M} of \mathcal{Z} into

¹³Warning! Some authors, particularly Mathematicians, use the notation $Sp(2n)$ to denote $U\text{Sp}(2n)$, the *unitary symplectic* group. See Section 5.10. Some other authors use $Sp(n)$ to stand for $Sp(2n, \mathbb{R})$ or $U\text{Sp}(2n)$.

itself. Two such mappings may be combined by letting them act on \mathcal{Z} successively, and this composition operation may be taken to be a rule for multiplying mappings. By the nature of composition, this rule automatically satisfies the associative property, and thus a set of mappings of \mathcal{Z} into itself has the potential of forming a group. Naturally, we will require that the product of any two mappings in the set will also be in the set. Furthermore, we may take the identity mapping \mathcal{I} , which leaves each element of \mathcal{Z} unchanged, to be the identity element in the potential group. Finally, if require that every mapping in the set have an inverse, we may regard the set as forming a group.

By this definition we see that all matrix groups are transformation groups because each group element is also a transformation of some vector space into itself. That is, in this case, \mathcal{Z} is some vector space. Moreover, in Section 5.12, we will learn that the symplectic group can also be viewed as providing a set of transformations of a generalized upper half plane, called a *Siegel space*, into itself. In this case, \mathcal{Z} is a generalized upper half plane. And, as described in Chapter 6, the group of all symplectic maps is a transformation group with \mathcal{Z} being phase space.

Finally, an abstract group G can always be thought of acting on itself by left or right multiplication:

$$h \rightarrow gh \text{ or } h \rightarrow hg^{-1}; \quad g \in G, \quad h \in \mathcal{Z} = G.$$

Here g is any element in G ; and h is any element in \mathcal{Z} , where, in fact, \mathcal{Z} is also G . Thus, by either of these constructions, every group can also be viewed as being a transformation group. Moreover, each action is *transitive*. That is, any element h in G can be sent to any other element \bar{h} in G . For example, consider left multiplication. Form the element $g = \bar{h}h^{-1}$. This element will be in G because \bar{h} and h^{-1} are in G . Then we find

$$h \rightarrow gh = (\bar{h}h^{-1})h = \bar{h}(h^{-1}h) = \bar{h}.$$

Moreover, for right multiplication, form the element $g = \bar{h}^{-1}h$ so that $g^{-1} = h^{-1}\bar{h}$. Then we find

$$h \rightarrow hg^{-1} = h(h^{-1}\bar{h}) = (hh^{-1})\bar{h} = \bar{h}.$$

3.6.2 Dyadic and Gram Matrices, Bases and Reciprocal Bases

The remaining concern of this section is a study of bases. To do so, we will first develop the tools of dyadic and Gram matrices, and then apply them in the study of various bases.

We begin with dyadic notation, which is an application of Dirac notation for the case of a finite dimensional real vector space. For a review of the needed Dirac notation, see Exercises 6.1 and 6.2 at the end of this subsection.

Suppose we are given a set of *real* linearly independent vectors w^1, w^2, \dots, w^N . By definition, such a set constitutes a basis for an N -dimensional vector space. Let e^1, e^2, \dots

e^N denote the standard column unit vectors

$$e^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, \quad e^2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, \quad e^3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}, \quad \text{etc.} \quad (3.6.4)$$

Define a linear operator W by the rule

$$We^j = w^j. \quad (3.6.5)$$

Its matrix elements are given by the relation

$$W_{ij} = (e^i, We^j) = (e^i, w^j) \quad (3.6.6)$$

where $(,)$ denotes the usual scalar product but *without* complex conjugation. (Indeed, in the following, all vectors and matrices will be assumed to be *real*. And, in any symplectic context, all vector spaces will be assumed to be *even* dimensional.) In view of (6.4) and (6.5), W can be written in terms of the w^j in the form

$$W = (w^1, w^2, w^3, \dots, w^N) \quad (3.6.7)$$

where each w^j is regarded as a column vector so that the collection (6.7) forms an $N \times N$ matrix. Since the w^j are assumed to be linearly independent, we must have the relation

$$\det W \neq 0, \quad (3.6.8)$$

and we conclude that W^{-1} exists. Thus, for every *real* invertible $N \times N$ matrix W there is a basis of *real* vectors w^1, w^2, \dots, w^N , and vice versa.

Let us also use the notation $|w^j\rangle$ to denote the column vector w^j , and let $\langle w^j|$ denote its dual row vector. With this notation we define the matrix $D(W)$ associated with the w^j by the rule

$$D(W) \leftrightarrow \sum_k |w^k\rangle \langle w^k|. \quad (3.6.9)$$

Note that the right side of (6.9) is a sum of symmetric dyads. See Exercise 6.1. Simple matrix manipulation shows that D can also be written in the form

$$D(W) = WW^T. \quad (3.6.10)$$

See Exercise 6.3. Next define the *Gram* matrix $G(W)$ by the rule

$$G_{ij} = (w^i, w^j). \quad (3.6.11)$$

Matrix manipulation shows that G can also be written in the form

$$G(W) = W^TW. \quad (3.6.12)$$

See Exercise 6.4.

We have seen that for every real basis there are associated real dyadic and Gram matrices D and G . It follows from (6.10) and (6.12) that D and G are conjugate under the action of W ,

$$W^{-1}DW = G \text{ and } WGW^{-1} = D. \quad (3.6.13)$$

We also note that D and G are symmetric,

$$D^T = D, \quad G^T = G. \quad (3.6.14)$$

D and G are also invertible, and have positive determinant if W is real,

$$\det D = \det G = \det(W) \det(W^T) = (\det W)^2 > 0. \quad (3.6.15)$$

[We remark that $\det W$ is the (oriented) volume V of the parallelepiped with edges w^j , and consequently (6.15) is equivalent to the statement $\det D = \det G = V^2$.] Finally, we see from their forms (6.10) and (6.12) that both D and G are positive definite if W is real. That is, we have the relations

$$(v, Dv) > 0, \quad (v, Gv) > 0 \quad (3.6.16)$$

for any real nonzero vector v .

We have also seen that a basis w^j specifies an associated nonsingular matrix W . Using this matrix, define the related vectors ${}^r w^j$ by the rule

$${}^r w^j = (W^{-1})^T e^j. \quad (3.6.17)$$

In view of (6.5) we may also write

$${}^r w^j = (W^{-1})^T (W^{-1}) w^j. \quad (3.6.18)$$

The ${}^r w^j$ have the pleasing property

$$({}^r w^i, w^j) = ([W^{-1}]^T e^i, We^j) = (e^i, W^{-1}We^j) = (e^i, e^j) = \delta_{ij}, \quad (3.6.19)$$

and are called the *reciprocal* basis to the original basis. Note that in analogy to (6.5) we may write

$${}^r We^j = {}^r w^j \quad (3.6.20)$$

with

$${}^r W = (W^{-1})^T. \quad (3.6.21)$$

That is, the columns of ${}^r W$ are the ${}^r w^j$. Also, since there is the relation

$${}^r ({}^r W) = \{[(W^{-1})^T]^{-1}\}^T = W, \quad (3.6.22)$$

it follows that the reciprocal basis of the reciprocal basis is the original basis.

It is easily verified that D and G have the property

$$D({}^r W) = [D(W)]^{-1}, \quad G({}^r W) = [G(W)]^{-1}. \quad (3.6.23)$$

Also suppose U , V , and W are all invertible matrices. Define rU and rV in analogy to (6.21). Then the relation

$$W = UV \quad (3.6.24)$$

implies the relation

$${}^rW = {}^rU {}^rV, \quad (3.6.25)$$

and vice versa. Thus, group properties are preserved under the r operation.

Simple calculation shows that the identity operator I has two dyadic representations in terms of the original and reciprocal bases,

$$I = \sum_j |{}^r w^j)(w^j|, \quad (3.6.26)$$

$$I = \sum_j |w^j)({}^r w^j|. \quad (3.6.27)$$

See Exercise 6.5. The representations (6.26) and (6.27) can be used to expand an arbitrary vector v in terms of either the w^j basis or the ${}^r w^j$ basis. From (6.26) we have the result

$$v = Iv = \sum_j |{}^r w^j)(w^j, v), \quad (3.6.28)$$

which is an expansion of v in the reciprocal basis. From (6.27) we have the relation

$$v = Iv = \sum_j |w^j)({}^r w^j, v), \quad (3.6.29)$$

which is an expansion of v in the w^j basis.

Finally, it is interesting to have dyadic representations for W and W^{-1} . From (6.5) we find the representation

$$W = \sum_j |w^j)(e^j|. \quad (3.6.30)$$

Similarly, (6.17) gives the result

$$(W^{-1})^T = \sum_j |{}^r w^j)(e^j|, \quad (3.6.31)$$

from which it follows that

$$W^{-1} = \sum_j |e^j)({}^r w^j|. \quad (3.6.32)$$

3.6.3 Orthonormal and Symplectic Bases

So far we have been discussing general bases and their associated reciprocal bases. Now we want to consider two special kinds of bases: *orthonormal* bases and *symplectic* bases. A set of vectors v^j is called an orthonormal basis if it has the property

$$(v^i, v^j) = \delta_{ij}. \quad (3.6.33)$$

A set of vectors v^j (now necessarily *even* in number) is called a symplectic basis if it has the property

$$(v^i, Jv^j) = J_{ij}. \quad (3.6.34)$$

(Note that the basis set e^j , assuming the e^j are even in number, is both orthonormal and symplectic.) We shall discuss the properties of these two kinds of bases in turn.

Orthonormal Bases and Orthogonal Matrices

We begin with orthonormal bases. Suppose a set of *real* vectors v^j satisfies (6.33). Then it follows that they are linearly independent, and therefore entitled to be called a basis. For suppose there is a relation of the form

$$\sum_j \alpha_j v^j = 0. \quad (3.6.35)$$

Then, using (6.33), we find

$$\sum_j \alpha_j (v^i, v^j) = \alpha_i = 0 \text{ for all } i. \quad (3.6.36)$$

As before, define a linear operator V by writing

$$v^j = Ve^j. \quad (3.6.37)$$

Then we find that V has the property

$$(e^i, V^T Ve^j) = (Ve^i, Ve^j) = (v^i, v^j) = \delta_{ij}, \quad (3.6.38)$$

or, in matrix notation,

$$V^T V = I. \quad (3.6.39)$$

This is the condition for V to be an orthogonal matrix. Conversely, suppose V is orthogonal, and define vectors v^j using (6.37). Then we find

$$(v^i, v^j) = (Ve^i, Ve^j) = (e^i, V^T Ve^j) = (e^i, e^j) = \delta_{ij}, \quad (3.6.40)$$

and conclude that the v^j form an orthonormal basis. Put another way, the columns of an orthogonal matrix are orthonormal, and any matrix whose columns are orthonormal is orthogonal. Moreover, since the transpose of an orthogonal matrix is also orthogonal, the rows of an orthogonal matrix are orthonormal; and any matrix whose rows are orthonormal is orthogonal.

As an immediate consequence of the orthogonality condition (6.40) and the definitions of D and G we have the results

$$D(V) = G(V) = I. \quad (3.6.41)$$

Moreover, from (6.18) and (6.39) we see that an orthonormal basis is self reciprocal,

$${}^r v^j = v^j. \quad (3.6.42)$$

Next suppose R is an orthogonal matrix and that the v^j form an orthonormal basis. Then the vectors u^j defined by

$$u^j = Rv^j \quad (3.6.43)$$

also form an orthonormal basis. Indeed, we find

$$(u^i, u^j) = (Rv^i, Rv^j) = (v^i, R^T R v^j) = (v^i, v^j) = \delta_{ij}. \quad (3.6.44)$$

Conversely, any two orthonormal bases u^j and v^j are related by a *unique* orthogonal transformation R . Indeed, from (6.37) and the analogous relation

$$u^j = Ue^j, \quad (3.6.45)$$

we find the result

$$u^j = Ue^j = UV^{-1}v^j = Rv^j \quad (3.6.46)$$

with

$$R = UV^{-1}. \quad (3.6.47)$$

Since orthogonal matrices form a group, we conclude that R is also orthogonal. What we have shown is that the orthogonal group acts *transitively* on the set of orthonormal bases.

Symplectic Bases

Now consider symplectic bases (in which case the dimensionality must be *even*). The discussion in this case has many parallels to the orthonormal case. Suppose a set of *real* vectors v^j satisfies (6.34). Suppose there is also an alleged linear dependency (6.35). Then use of (6.34) gives the relation

$$0 = \sum_j \alpha_j(v^i, Jv^j) = \sum_j J_{ij}\alpha_j = (J\alpha)_i \text{ for all } i. \quad (3.6.48)$$

Since J is invertible, it must again be the case that all α_j vanish, and the v^j must be linearly independent.

In analogy with the orthogonal case, there is a close connection between symplectic bases and symplectic matrices. Given a symplectic basis v^j we define V using (6.37) and find the relation

$$(e^i, V^T JV e^j) = (V e^i, JV e^j) = (v^i, Jv^j) = J_{ij}, \quad (3.6.49)$$

and conclude that V is symplectic,

$$V^T JV = J. \quad (3.6.50)$$

Conversely, if V is symplectic and the v^j are defined by (6.37), then the v^j comprise a symplectic basis:

$$(v^i, Jv^j) = (V e^i, JV e^j) = (e^i, V^T JV e^j) = (e^i, Je^j) = J_{ij}. \quad (3.6.51)$$

Put another way, the columns of a symplectic matrix form a symplectic basis; and any matrix whose columns form a symplectic basis is symplectic. Moreover, since the transpose

of a symplectic matrix is also symplectic, the rows of a symplectic matrix form a symplectic basis; and any matrix whose rows form a symplectic basis is symplectic.

Next suppose R is a symplectic matrix and that the v^j form a symplectic basis. Then the vectors u^j defined by (6.43) also form a symplectic basis:

$$(u^i, Ju^j) = (Rv^i, JRv^j) = (v^i, R^T JRv^j) = (v^i, Jv^j) = J_{ij}. \quad (3.6.52)$$

Conversely, any two symplectic bases are related by a unique symplectic transformation. Consideration of this matter again leads to (6.45) through (6.47) with U and V now being symplectic matrices. Since symplectic matrices form a group, we conclude that the R given by (6.47) is also symplectic. What we have shown now is that the symplectic group acts *transitively* on the set of symplectic bases.

There are a few remaining observations to be made about the symplectic case. From the V analog of (6.17) we see that the ${}^r v^j$ form a symplectic basis if the v^j form a symplectic basis. Also, from their definitions and the symplectic group properties, we see that D and G are both symplectic; and (6.13) shows that they are symplectically conjugate. Finally, suppose that a basis v^j is *both* orthonormal and symplectic. In this case the V appearing in (6.37) is both orthogonal and symplectic. It will be shown in Section 3.9 that such V also form a group, which is in fact the unitary subgroup $U(n)$ of $Sp(2n, \mathbb{R})$.¹⁴ It follows that all bases that are both orthonormal and symplectic are in one-to-one correspondence with the elements of $U(n)$, and $U(n)$ acts transitively on the set of all such bases.

3.6.4 Construction of Orthonormal Bases

Thanks to the work of Subsection 8.1 yet to come, we know how to construct all orthogonal matrices in terms of exponentials of antisymmetric matrices. Call such a matrix R . We also know that the rows (and columns) of any orthogonal matrix form an orthonormal basis, and therefore we know how to construct all orthonormal bases. Here we consider a related problem: Given a set of *real* basis vectors w^j , specify a method to construct from them a set of basis vectors v^j that is orthonormal. Strictly speaking, as just posed, this question is meaningless. Having found such a method, we can specify an infinite number of other methods. We take the vectors v^j produced by the found method and use them as the columns of a matrix we will call V . This V will be an orthogonal matrix since the necessary and sufficient conditions for a matrix to be orthogonal is for its columns to form an orthonormal. Next form the product $\bar{V} = RV$. The matrix \bar{V} , by the group property, will also be orthogonal. Therefore its column vectors, call them \bar{v}^j , will form an orthonormal basis. In this way (assuming $R \neq I$) we have found a different method for constructing an orthonormal basis \bar{v}^j starting with the basis w^j . And the number of such methods is infinite since the number of matrices $R \neq I$ is infinite.

A better question is this: Given some *real* basis w^j , are there natural or useful ways of associating particular orthonormal bases with the given w^j basis? In what follows we will describe various known ways of doing so, and examine some of their features. Also, in preparation for the next subsection, Subsection 6.5, we will gain insights into the more difficult problem of constructing *symplectic* bases starting with the w^j .

¹⁴Unitary matrices are defined in Subsection 7.6. As illustrated in Section 3.9, there are *real* matrices whose group properties are those of $U(n)$.

Gram-Schmidt Orthogonalization

The most commonly known procedure for constructing an orthonormal basis, given a set of *real* basis vectors w^j , is *Gram-Schmidt* orthogonalization.¹⁵ Starting with w^1 , we construct intermediate vectors, call them u^j , and then final normalized vectors v^j by the rules

$$\begin{aligned} u^1 &= w^1, \quad v^1 = u^1 / \| u^1 \|; \\ u^2 &= w^2 - (v^1, w^2)v^1, \quad v^2 = u^2 / \| u^2 \|; \\ u^3 &= w^3 - (v^1, w^3)v^1 - (v^2, w^3)v^2, \quad v^3 = u^3 / \| u^3 \|; \\ &\vdots \\ u^N &= w^N - (v^1, w^N)v^1 - \cdots - (v^{N-1}, w^N)v^{N-1}, \quad v^N = u^N / \| u^N \| . \end{aligned} \quad (3.6.53)$$

Here we employ the usual notation

$$\| u^j \| = [(u^j, u^j)]^{1/2}, \quad (3.6.54)$$

and note that all the scalar products appearing in (3.6.53) and (33,6,54) are *real* scalar products. It is easy to verify that at each step there is the relation $\| u^j \| \neq 0$ as is required for the definition of each v^j . In fact, if $\| u^j \| = 0$ for some j , then the vectors w^1 to w^j would be linearly dependent contrary to assumption. Finally, it is easy to check that the v^j are orthonormal.

Since the w^j are given and the v^j have been determined, we have their associated matrices W and V , both of which are invertible. Because the w^j form a basis, the v^j can be expanded in terms of them to give a relation of the form

$$v^i = \sum_j \alpha_{ij} w^j \quad (3.6.55)$$

Indeed, (6.53) is just such a relation. Using the matrices W and V this relation can be written in the compact form

$$V = WA^T \quad (3.6.56)$$

where A is the matrix given by the relation

$$A_{ij} = \alpha_{ij}. \quad (3.6.57)$$

Similarly, by considering rows rather than columns, there is a matrix B such that

$$V = BW. \quad (3.6.58)$$

Both A and B are unique and invertible, and satisfy the relations

$$I = V^T V = AW^T WA^T = AG(W)A^T, \quad (3.6.59)$$

$$I = VV^T = BWW^T B^T = BD(W)B^T. \quad (3.6.60)$$

¹⁵The method is named after Jørgen Pedersen Gram and Erhard Schmidt, but Laplace had been familiar with it before Gram and Schmidt.

We say that G is *congruent* to I under the action of A , and say that A is the *intertwining* transformation or matrix.¹⁶ (Note that A is generally not orthogonal, and therefore generally $A^T \neq A^{-1}$.) Similarly, D is congruent to I under the action of B . Note that there are many pairs A, B satisfying (6.59) and (6.60) with one such pair for each orthogonalization process. Indeed, if we replace A and B by

$$A' = RA, \quad B' = RB \quad (3.6.61)$$

where R is any orthogonal matrix, we find the relations

$$A'G(A')^T = RAGA^TR^T = RR^T = I, \quad (3.6.62)$$

$$B'D(B')^T = RBDB^TR^T = RR^T = I. \quad (3.6.63)$$

Conversely, if A and A' satisfy (6.59) and (6.62) respectively, then R defined by (6.61) is orthogonal, etc.

QR Decomposition

Closely related to Gram-Schmidt orthogonalization is what is called *QR decomposition*. It can be shown that any square nonsingular matrix W can be written in the factored form $W = QR$ where Q is orthogonal and R is upper triangular.¹⁷ This factorization is unique if we require that the diagonal entries of R be positive. Evidently (6.56) can be rewritten in the form $W = V(A^T)^{-1}$. It can be verified that $(A^T)^{-1}$ is upper triangular with all diagonal entries positive, and we know that V is orthogonal. Thus we may make the identifications $Q = V$ and $R = (A^T)^{-1}$ to observe that the Gram-Schmidt process is one way to produce a *QR* decomposition. Given W there are other ways besides Gram-Schmidt to produce a *QR* decomposition (with the diagonal entries in R positive) including *Householder* transformations and *Givens* rotations.¹⁸ And, by uniqueness, they all produce the same matrices Q and R that would be produced by applying Gram-Schmidt to W . Hence, by setting $V = Q$, these other ways can be also be used to produce the orthogonal matrix V that would also have resulted from applying Gram-Schmidt to W .

Democratic Polar Decomposition of Real Matrices

Although Gram-Schmidt orthogonalization is straight forward and often natural, the result depends on the order in which the w^j are labeled, and does not treat all the w^j on an equal footing. For example, v^1 is always in the direction of w^1 , but in general none of the other v^j are in the direction of the w^j . Sometimes it is desirable to have a procedure that treats all the w^j democratically. There are many ways to do this. The first uses *polar decomposition*.¹⁹

¹⁶If two matrices U and V are related by an equation of the form $V = AUA^{-1}$, they are said to be *similar* or *conjugate*. If they are related by an equation of the form $V = AUA^T$, they are said to be *congruent*. Here A is assumed to be nonsingular.

¹⁷Here there is an unfortunate conflict of notation. We have been using the symbol R to denote an orthogonal matrix. But in this paragraph, since *QR* is already standard notation in the mathematics literature, R will denote an upper triangular matrix.

¹⁸There are a variety of numerical *QR* packages.

¹⁹Polar decomposition was discovered by Cauchy.

Let M be any *real* nonsingular matrix. It can be shown that any such M can be written uniquely in the form (called a polar decomposition)

$$M = PO. \quad (3.6.64)$$

Here P is a real positive-definite symmetric matrix, and O is a real orthogonal matrix. See Section 4.2. Intuitively, polar decomposition may be regarded as the matrix analog of expressing a complex number z in the polar form $z = r \exp(i\phi)$. Assuming the representation (6.64), we find that

$$D(M) = MM^T = POO^TP = P^2. \quad (3.6.65)$$

Since $D(M)$ is real, symmetric, and positive definite, it has a unique square root that is also positive definite and symmetric, and we may write

$$P = [D(M)]^{1/2}. \quad (3.6.66)$$

With this information we can solve (6.64) for O to find the result

$$O = [D(M)]^{-1/2}M. \quad (3.6.67)$$

Apply this result to the case $M = W$, where W is given by (6.5), to find the orthogonal matrix

$$O(W) = [D(W)]^{-1/2}W. \quad (3.6.68)$$

Now generate the v^j using (6.37) with

$$V = [D(W)]^{-1/2}W. \quad (3.6.69)$$

Note that (6.69) treats all the w^j on the same footing. It also has the feature (as does the Gram-Schmidt procedure) that if the w^j are already orthonormal, then

$$v^j = w^j, \quad (3.6.70)$$

for in this case $D = I$. See (6.41). In addition, the prescription (6.69) has the feature that the V it produces is the orthogonal matrix O that is *closest* to W in the sense of *minimizing* $\|W - O\|_E$ in the Euclidean matrix norm. See Exercise 7.1 and Section 4.4.2. Therefore V may be viewed as the solution to a *variational* problem.

The polar decomposition (6.64) can also be written in reverse order:

$$M = PO = OO^{-1}PO = OP', \quad (3.6.71)$$

where

$$P' = O^{-1}PO = O^TPO. \quad (3.6.72)$$

Evidently P' is also real positive-definite symmetric. Note that the orthogonal factor O is the same in both orders. Using the representation (6.71), we find

$$G(M) = M^T M = P' O^T O P' = (P')^2. \quad (3.6.73)$$

Thus, upon setting $M = W$, we find the equally valid relation

$$V = W[G(W)]^{-1/2}. \quad (3.6.74)$$

This relation also follows directly from (6.69) with the use of (6.13). Comparison of (6.50) and (6.58) with (6.69) and (6.74) shows that for this normalization process there are the relations

$$A = \{[G(W)]^{-1/2}\}^T = [G(W)]^{-1/2}, \quad (3.6.75)$$

$$B = [D(W)]^{-1/2}. \quad (3.6.76)$$

For the V defined by (6.69) we find the result

$$V^T D(W) V = W^T [D(W)]^{-1/2} D(W) [D(W)]^{-1/2} W = W^T W = G(W). \quad (3.6.77)$$

Since $V^T = V^{-1}$, this result can also be rewritten in the form

$$VG(W)V^T = D(W). \quad (3.6.78)$$

Thus, D and G are also conjugate under the action of the orthogonal matrix V . Compare (6.77) and (6.78) with (6.13).

Other Democratic Orthogonalizations

Having found one particular pleasing orthogonalization process, let us see if there are others. Based on the earlier discussion, without loss of generality we may consider all U of the form

$$U = VR \quad (3.6.79)$$

where R is any orthogonal matrix. If R is chosen at random, then all correlation of U with W is lost. However, if R is fixed ($R = I$ in the previous example) or is itself related to W , then U will also be related to W . Alternatively, U itself may be related to W in some direct way.

From (6.77) we find the result

$$U^T D(W) U = R^T V^T D(W) V R = R^T G(W) R. \quad (3.6.80)$$

Since G is real symmetric, we know there is an orthogonal transformation that diagonalizes it. Select R to be such a transformation,

$$R = R_G \quad (3.6.81)$$

where

$$R_G^T G(W) R_G = \Delta_G. \quad (3.6.82)$$

Here we have used the notation Δ_G to denote a diagonal form of G , and R_G to denote an orthogonal transformation that accomplishes this diagonalization. We remark that Δ_G is unique up to permutations of its diagonal entries, and R_G is unique up to (orthogonal)

permutation matrices providing the entries of Δ_G (the eigenvalues of G) are distinct. Upon setting $U_D = VR_G$, and using (6.80) and (6.82), we find the result

$$U_D^T D(W) U_D = \Delta_G. \quad (3.6.83)$$

We see that U_D is an orthogonal transformation that diagonalizes D ,

$$U_D^T D(W) U_D = \Delta_D, \quad (3.6.84)$$

and there is the relation

$$\Delta_D = \Delta_G. \quad (3.6.85)$$

It is interesting to recognize that an orthogonal U_D that accomplishes (6.84) may also be viewed as a solution to a variational problem.²⁰ Let \mathcal{F}_D be the functional

$$\mathcal{F}_D[U] = \sum_k [(U^T D U)_{kk}]^2. \quad (3.6.86)$$

(Note that \mathcal{F}_D is *quartic* in the u^j . See Exercise 6.10.) There is the familiar algebraic result

$$\text{tr} \{(U^T D U)^T (U^T D U)\} = \sum_{ij} [(U^T D U)_{ij}]^2. \quad (3.6.87)$$

See Exercise 6.8. But we also find by direct evaluation the result

$$\text{tr} \{(U^T D U)^T (U^T D U)\} = \text{tr} (U^T D^T U U^T D U) = \text{tr} (U^T D^T D U) = \text{tr} (D^T D) = \text{tr} (D^2). \quad (3.6.88)$$

Here we have used the facts that U is orthogonal and D is symmetric, and standard properties of the trace operation. See Exercise 6.7. By combining (6.86) through (6.88) we find the relation

$$\text{tr} (D^2) = \sum_{ij} [(U^T D U)_{ij}]^2 = \sum_k [(U^T D U)_{kk}]^2 + \sum_{i \neq j} [(U^T D U)_{ij}]^2, \quad (3.6.89)$$

and therefore

$$\mathcal{F}_D[U] = \text{tr} (D^2) - \sum_{i \neq j} [(U^T D U)_{ij}]^2. \quad (3.6.90)$$

Evidently the maximum possible value of $\mathcal{F}_D[U]$ is $\text{tr} (D^2)$, and this maximum can be reached if there is a $U \in SO(N)$ such that

$$(U^T D U)_{ij} = 0 \text{ for all } i, j \text{ satisfying } i \neq j. \quad (3.6.91)$$

According to (6.83) there is such a U , namely $U = U_D$. Thus we have the result

$$\max_{U \in SO(N)} \mathcal{F}_D[U] = \text{tr} (D^2), \quad (3.6.92)$$

and this maximum is achieved when

$$U = U_D = VR_G. \quad (3.6.93)$$

At this point it should be evident that there are several other possibilities for constructing orthogonal U matrices that are related to W . For example, after any construction, one could replace U by U^T . Or, one could require that U diagonalize $G(W)$ instead of $D(W)$. See Exercise 6.11.

²⁰See the reference to the paper of H.C. Schweinler and E.P. Wigner listed in the references at the end of this chapter.

The Complex Case and the Polar Decomposition of Complex Matrices

So far the discussion has been devoted to real vectors and real matrices. Some of it can be readily extended to the complex case. For example, Gram-Schmidt can be extended to the complex case simply by replacing the usual real scalar product with the usual complex scalar product. A second example of extension to the complex case is that there is an analogous polar decomposition (also simply called polar decomposition) for the case of complex matrices. A factorization of the form (6.64) still holds but now M is complex, P is Hermitian and positive definite, and O is unitary. See Exercise 4.2.5.

3.6.5 Construction of Symplectic Bases

There are some resemblances between the construction of symplectic bases and the construction of orthonormal bases. We begin with the resemblances. Thanks to the work of Subsection 8.2 again yet to come, we know how to construct all symplectic matrices, call them M , in terms of exponentials of certain well defined matrices:

$$M = \exp(JS^a) \exp(JS^c).$$

Here S^c is any symmetric matrix that *commutes* with J and S^a is any symmetric matrix that anticommutes with J . We also know that the rows (and columns) of any symplectic matrix form a symplectic basis, and therefore we know how to construct all symplectic bases. And again we consider a related problem: Given a set of *real* basis vectors w^j , specify a method to construct from them a set of vectors v^j that form a symplectic basis.

Again as just posed, this question is meaningless. Having found such a method, we can specify an infinite number of other methods. We take the vectors v^j produced by the found method and use them as the columns of a matrix and call it V . This V will be a symplectic matrix since the necessary and sufficient conditions for a matrix to be symplectic is for its columns to form an symplectic basis. Next form the product $\bar{V} = MV$. The matrix \bar{V} , by the group property, will also be symplectic. Therefore its column vectors, call them \bar{v}^j , will form a symplectic basis. In this way (assuming $M \neq I$) we have found a different method for constructing a symplectic basis \bar{v}^j starting with the basis w^j . And the number of such methods is infinite since the number of symplectic matrices $M \neq I$ is infinite.

Again a better question is this: Given some *real* basis w^j , are there natural or useful ways of associating particular symplectic bases with the given w^j basis? In what follows we will describe various known ways of doing so, and examine some of their properties. All of them have the feature that they can be applied to any *real* basis w^j .

Subsequently, in Chapter 4 under the rubric of Matrix Symplectification, we will consider additional methods that can be viewed as constructing symplectic bases. Some of them have the property that they do not succeed for all *real* bases w^j , but they do have certain redeeming virtues.

Darboux Construction

We will first describe an analog of the Gram-Schmidt procedure, which we will call *Darboux* construction. Suppose the w^j are a set of $2n$ linearly independent vectors and we wish to

construct from them a symplectic basis v^j . For this purpose it is convenient to use the form (2.10) for J . Below is an algorithm for constructing the v^j :

1. Define v^1 by the simple rule

$$v^1 = w^1. \quad (3.6.94)$$

2. Starting with w^2 , search through the w^j with $j \geq 2$ to find the first j , call it k , with the property

$$(v^1, Jw^j) \neq 0. \quad (3.6.95)$$

[Better yet, if one is working numerically and therefore only to finite precision, select j so that $|(v^1, Jw^j)|$ is maximized. The analogous choices should also be made in steps 6, 10, etc. below.] Renumber the vectors $w^2 \cdots w^{2n}$ so that w^k becomes w^2 .

3. Define v^2 by the rule

$$v^2 = w^2 / [(v^1, Jw^2)]. \quad (3.6.96)$$

We then have the result

$$(v^1, Jv^2) = 1 = J_{12}. \quad (3.6.97)$$

And, since J is antisymmetric, at this stage we have the result

$$(v^i, Jv^j) = J_{ij} \text{ for } i, j = 1 \text{ to } 2. \quad (3.6.98)$$

4. Using the remaining vectors $w^3 \cdots w^{2n}$, define new vectors ${}^1w^j$ with $j \geq 3$ by the rule

$${}^1w^j = w^j + (v^2, Jw^j)v^1 - (v^1, Jw^j)v^2. \quad (3.6.99)$$

As a result of this rule there are the relations

$$(v^i, J {}^1w^j) = 0 \text{ for } i = 1, 2 \text{ and } j = 3, 4, \dots, 2n. \quad (3.6.100)$$

5. Define v^3 by the rule

$$v^3 = {}^1w^3. \quad (3.6.101)$$

6. Starting with ${}^1w^4$, search through the ${}^1w^j$ with $j \geq 4$ to find the first j , call it k , with the property

$$(v^3, J {}^1w^j) \neq 0. \quad (3.6.102)$$

Renumber the vectors ${}^1w^4 \cdots {}^1w^{2n}$ so that ${}^1w^k$ becomes ${}^1w^4$.

7. Define v^4 by the rule

$$v^4 = {}^1w^4 / [(v^3, J {}^1w^4)]. \quad (3.6.103)$$

At this stage we have the results

$$(v^i, Jv^j) = J_{ij} \text{ for } i, j = 1 \text{ to } 4. \quad (3.6.104)$$

8. Using the remaining vectors ${}^1w^5 \dots {}^1w^{2n}$, define new vectors ${}^2w^j$ with $j \geq 5$ by the rule

$${}^2w^j = {}^1w^j + (v^4, J {}^1w^j)v^3 - (v^3, J {}^1w^j)v^4. \quad (3.6.105)$$

Now we have the relations

$$(v^i, J {}^2w^j) = 0 \text{ for } i = 1 \text{ to } 4 \text{ and } j = 5, 6, \dots, 2n. \quad (3.6.106)$$

9. Define v^5 by the rule

$$v^5 = {}^2w^5. \quad (3.6.107)$$

10. Starting with ${}^2w^6$, search through the ${}^2w^j$ with $j \geq 6$ to find the first j , call it k , with the property

$$(v^5, J {}^2w^j) \neq 0. \quad (3.6.108)$$

Renumber the vectors ${}^2w^6 \dots {}^2w^{2n}$ so that ${}^2w^k$ becomes ${}^2w^6$.

11. Define v^6 by the rule

$$v^6 = {}^2w^6 / [(v^5, J {}^2w^6)]. \quad (3.6.109)$$

At this stage we have the results

$$(v^i, J v^j) = J_{ij} \text{ for } i, j = 1 \text{ to } 6. \quad (3.6.110)$$

12. Proceed with the obvious extension of the above process to construct $v^7, v^8, \dots, v^{2n-2}$.

Then at the last stage we have

$$v^{2n-1} = {}^m w^{2n-1}, \quad (3.6.111)$$

$$v^{2n} = {}^m w^{2n} / [(v^{2n-1}, J {}^m w^{2n})], \quad (3.6.112)$$

with

$$m = n - 1. \quad (3.6.113)$$

At this point several comments are in order. First, if we are working only with two, four, or six-dimensional phase space, as is the case for accelerator physics, then we may terminate the algorithm at steps 3, 7, or 11. Second, how does one know that the required vectors ${}^m w^k$ described in steps 2, 6, 10, etc. exist? Third, how does one know that the vectors $v^3, v^5, \dots, v^{2n-1}$ given in steps 5, 9, etc. are nonzero? As was the case with the Gram-Schmidt orthogonalization process, we are saved from such embarrassment because the w^j are assumed to be linearly independent and J is invertible. See Exercise 6.14. Finally, we close this discussion with the comment that, like the Gram-Schmidt orthogonalization procedure described in Section 3.6.4, Darboux Symplectification does not treat the vectors m^j democratically. Perhaps this defect could be overcome by a more complicated procedure like those given in Section 3.6.4.

Modified Darboux Construction

Suppose one is given a matrix M whose determinant is nonzero. The columns of M may be regarded as vectors m^1, m^2, m^3, \dots , and the condition $\det(M) \neq 0$ is equivalent to the statement that the vectors m^j are linearly independent. Given a set of linearly independent vectors m^j , there is the Darboux process for constructing an associated set of symplectic vectors r^j . Finally, the vectors r^j may be viewed as the columns of a matrix R , and this matrix will be symplectic. Thus, given any nonsingular matrix M , there is a procedure for constructing a corresponding symplectic matrix R . Moreover, if M itself is nearly symplectic, then R will be near M . Indeed, if M happens to be symplectic, then R will coincide with M . See Sections 3.6.3 and 3.6.5. In what follows we will describe what we will call *modified* Darboux symplectification, and will examine how close R is to M if M is nearly symplectic.

Let M be a $2n \times 2n$ matrix. Rather than using (4.1), we will describe the *failure* of M to be symplectic in terms of an antisymmetric matrix F defined by the relation

$$F = M^T JM - J. \quad (3.6.114)$$

From (4.1) and (6.1) we have the result

$$\begin{aligned} \| F \| &= \| (M^T JM J^T - I)J \| \leq \| M^T JM J^T - I \| \| J \| \\ &\leq \| M^T JM J^T - I \| = \| M^T J(M^T)^T J^T - I \| = f(M^T). \end{aligned} \quad (3.6.115)$$

Here we have assumed that the matrix norm employed has the property

$$\| J \| = 1, \quad (3.6.116)$$

which is true for the maximum column sum norm (3.7.15) and the spectral norm (3.7.17). If the norm also has the property (3.7.97), which we shall also assume, then the matrix elements of F are bounded by the relation

$$|F_{jk}| \leq f(M^T). \quad (3.6.117)$$

Suppose we view M as a collection of column vectors m^1, m^2, \dots, m^{2n} . Let m_i^j denote the i th component of the j th such vector. Then, following the usual matrix element labelling scheme, we have the relation

$$m_i^j = M_{ij}. \quad (3.6.118)$$

In terms of the vectors m^j , the relation (6.1) can be rewritten in the form

$$(M^T JM)_{jk} = (m^j, Jm^k) = J_{jk} + F_{jk}. \quad (3.6.119)$$

Correspondingly if R is a symplectic matrix and we view it as a collection of column vectors r^j , then the symplectic condition (3.1.2) can be written in the form

$$(r^j, Jr^k) = J_{jk}. \quad (3.6.120)$$

Assume we are given an M for which $f(M^T)$ is sufficiently small. From M we extract the vectors m^j using (6.5). Our task is to use these m^j , which obey (6.6), to construct a set

of vectors r^j that obey (6.7). Moreover, this construction is to be made in such a way that the corresponding symplectic matrix R is near M in the sense that

$$\| M - R \| \sim f(M^T). \quad (3.6.121)$$

We will construct the vectors r^j two at a time, beginning with r^1 and r^2 . To simplify our presentation, we will use a J matrix of the form (3.2.10). For this choice we have the relation

$$J_{12} = 1, \quad (3.6.122)$$

and (6.6) gives the result

$$(m^1, Jm^2) = 1 + F_{12}. \quad (3.6.123)$$

According to (6.4) and (6.10), if $f(M^T)$ is sufficiently small, the quantity (m^1, Jm^2) will be positive and hence will have a positive square root γ_{12} ,

$$\gamma_{12} = +[(m^1, Jm^2)]^{1/2}. \quad (3.6.124)$$

We can therefore define “normalized” vectors r^1 and r^2 by the rules

$$r^1 = m^1 / \gamma_{12}, \quad (3.6.125)$$

$$r^2 = m^2 / \gamma_{12}. \quad (3.6.126)$$

Note that by (6.4) and (6.10), γ_{12} will be near 1 if $f(M^T)$ is sufficiently small. Correspondingly r^1 and r^2 will be near m^1 and m^2 , respectively. By construction, these vectors satisfy the relation

$$(r^1, Jr^2) = 1 = J_{12}, \quad (3.6.127)$$

as required by (6.7). Also, because J is antisymmetric, we automatically get from (6.14) the relations

$$(r^j, Jr^k) = J_{jk} \text{ when } j = 1, 2 \text{ and } k = 1, 2, \quad (3.6.128)$$

as is also required by (6.7).

Next we construct the vectors r^3 and r^4 . We begin by defining intermediate vectors s^3 and s^4 according to the rule

$$s^3 = m^3 + \alpha_{31}r^1 + \alpha_{32}r^2, \quad (3.6.129)$$

$$s^4 = m^4 + \alpha_{41}r^1 + \alpha_{42}r^2, \quad (3.6.130)$$

where the α 's are coefficients still to be determined. According to (6.7) we must have the relations

$$(r^j, Js^k) = 0 \text{ when } j = 1, 2 \text{ and } k = 3, 4. \quad (3.6.131)$$

Let us therefore require the relations

$$(r^j, Js^k) = 0 \text{ when } j = 1, 2 \text{ and } k = 3, 4. \quad (3.6.132)$$

Doing so determines the values of the coefficients α :

$$\alpha_{31} = (r^2, Jm^3), \quad (3.6.133)$$

$$\alpha_{32} = -(r^1, Jm^3), \quad (3.6.134)$$

$$\alpha_{41} = (r^2, Jm^4), \quad (3.6.135)$$

$$\alpha_{42} = -(r^1, Js^4). \quad (3.6.136)$$

If $f(M^T)$ is sufficiently small then, according to (6.4), (6.6), (6.11) through (6.13), and (6.20) through (6.23), all the α 's are of order $f(M^T)$. It also follows that the quantity (s^3, Js^4) will be positive and consequently will have the positive square root

$$\gamma_{34} = +[(s^3, Js^4)]^{1/2}. \quad (3.6.137)$$

Finally, we define the normalized vectors r^3 and r^4 by the rules

$$r^3 = s^3/\gamma_{34}, \quad (3.6.138)$$

$$r^4 = s^4/\gamma_{34}. \quad (3.6.139)$$

Upon reflection we see that we have now constructed four vectors r^1 through r^4 that are, respectively, near m^1 through m^4 if $f(M^T)$ is small; and these vectors satisfy the relations

$$(r^j, Jr^k) = J_{jk} \text{ when } j, k = 1, 2, 3, 4. \quad (3.6.140)$$

Moreover the general pattern is now clear. We see that the construction can be continued to include r^5 and r^6 (and still more r 's if we are dealing with more than a 6-dimensional phase space). We simply write the analogs of (6.16) and (6.17), for example

$$s^5 = m^5 + \alpha_{51}r^1 + \alpha_{52}r^2 + \alpha_{53}r^3 + \alpha_{54}r^4, \quad (3.6.141)$$

$$s^6 = m^6 + \alpha_{61}r^1 + \alpha_{62}r^2 + \alpha_{63}r^3 + \alpha_{64}r^4, \quad (3.6.142)$$

determine the α 's, and then normalize the results. Finally, we may view all the r^j we have constructed in this manner as the columns of a matrix R . This matrix will be symplectic and will be close to M in the sense of satisfying (6.8).

There is one last nuisance to be resolved. All our estimates have involved the quantity $f(M^T)$ whereas it would be more pleasant to work with $f(M)$. This defect can be overcome by using the modified Darboux procedure just described to symplectify the matrix M^T instead of M . Call the resulting symplectic matrix R' . Using (3.7.51) and (6.8) we will then have the result

$$|(M^T)_{jk} - R'_{jk}| \sim f(M). \quad (3.6.143)$$

Finally we define R , which is to be the symplectification of M , by writing

$$R = (R')^T. \quad (3.6.144)$$

Combining (6.30) and (6.31) then gives the desired result

$$|M_{jk} - R_{jk}| \sim f(M). \quad (3.6.145)$$

We close this discussion with the comment that, like the Darboux construction procedure described in Section 3.6.5, the modified Darboux construction procedure also does not treat the vectors m^j democratically.

Application: Transitive Action of $Sp(2n)$ on Phase Space

Review our earlier discussion of the Darboux construction procedure. It has a consequence that is worth observing. Suppose α and β are any two nonzero vectors. Let M^α be a symplectic matrix whose first column is the vector α . We know that such a matrix exists because we may set $w^1 = \alpha$ in (6.94). Then we have the relations

$$\alpha = M^\alpha e^1 \text{ and } e^1 = (M^\alpha)^{-1} \alpha. \quad (3.6.146)$$

Similarly, let M^β be a symplectic matrix whose first column is the vector β . Now define a matrix M by the rule

$$M = M^\beta (M^\alpha)^{-1}. \quad (3.6.147)$$

By the group property this matrix will be symplectic, and by construction it will have the property

$$M\alpha = M^\beta (M^\alpha)^{-1} \alpha = M^\beta e^1 = \beta. \quad (3.6.148)$$

We have found the remarkable result that, with the exception of the origin, any point in phase space can be sent into any other point by a symplectic matrix. (The origin is obviously sent into itself.) Following the terminology elaborated on in Section 5.12, we say that, with the exception of the origin, $Sp(2n)$ acts *transitively* on phase space.

Other Constructions

Let us now explore briefly additional methods for constructing symplectic bases. One of them makes use of “symplectic” polar decomposition, and is the subject of Sections 4.3 and 4.4. To consider others introduce, in imitation of the orthogonal case, analogous dyadic and Gram matrices by the definitions

$$D_J(W) = W J W^T. \quad (3.6.149)$$

$$G_J(W) = W^T J W. \quad (3.6.150)$$

Note that both D_J and G_J are antisymmetric. It is easy to see that there are again unique nonsingular matrices A and B such that the relations (6.56) and (6.57) hold. Then, from (6.56) and (6.58) and the requirement that V be symplectic, we find the results

$$J = V^T J V = A W^T J W A^T = A G_J(W) A^T, \quad (3.6.151)$$

$$J = V J V^T = B W J W^T B^T = B D_J(W) B^T. \quad (3.6.152)$$

We see that G_J is congruent to J under the action of A , and D_J is congruent to J under the action of B .

We already know that (6.119) and (6.120) have a full infinity of solutions for A and B . One simply solves (6.56) or (6.58) for A or B using any symplectic V . This matter is considered from a broader perspective in Section 3.13.

Finally we remark that, since both D and G as given by (6.10) and (6.12) are symmetric and positive definite, it can be shown there are symplectic matrices U and V such that

$$UD(W)U^T = \text{Williamson diagonal form}, \quad (3.6.153)$$

$$VG(W)V^T = \text{ Williamson diagonal form.} \quad (3.6.154)$$

Moreover, since J is orthogonal, the matrices $JD(W)J^T$ and $JG(W)J^T$ are also symmetric and positive definite. Therefore there are also symplectic matrices, again call them U and V , such that

$$UJD(W)J^TU^T = \text{ Williamson diagonal form,} \quad (3.6.155)$$

$$VJG(W)J^TV^T = \text{ Williamson diagonal form.} \quad (3.6.156)$$

See Section 33.6.3. The columns (or rows) of these symplectic matrices U and V may also be regarded as symplectic bases related in a specific way to W .

Exercises

3.6.1. This is an exercise which introduces and employs Dirac notation for vectors and linear operators. It is based on special dyads composed of real vectors denoted by $|k\rangle$ which are orthonormal and therefore satisfy the inner product relations

$$\langle j|k\rangle = \delta_{jk}. \quad (3.6.157)$$

We begin with the introduction: For simplicity, assume we are working with real vectors in a 4-dimensional space and with linear operators represented by 4×4 matrices. Since we are working with 4×4 matrices, we will need four *ket* vectors $|k\rangle$ and four *bra* vectors $\langle j|$. Ket vectors $|k\rangle$ are associated with column vectors having all entries 0 save for a 1 entry at location k counting from the top of the column to the bottom. For example, there are the correspondences

$$|1\rangle \leftrightarrow \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad (3.6.158)$$

$$|2\rangle \leftrightarrow \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \text{ etc.} \quad (3.6.159)$$

Bra vectors $\langle j|$ are associated with row vectors with all entries 0 save for a 1 entry at location j counting from the left end of the row to the right end. For example, there are the correspondences

$$\langle 1| \leftrightarrow (1000), \quad (3.6.160)$$

$$\langle 2| \leftrightarrow (0100), \text{ etc.} \quad (3.6.161)$$

Associated with these definitions, there is the inner product (*bracket*) relation (6.157).²¹

²¹Note that in this exercise we are using angular brackets \langle and \rangle rather than round brackets $($ and $)$ because we wish to reserve round brackets for use as parentheses.

We next define special dyads, which we denote in Dirac notation by the symbols $|i\rangle\langle j|$.²² These dyads are linear operators that map kets into kets (or the zero vector), and are associated with certain square matrices. A square matrix may be viewed as a square array made of column vectors labelled, from left to right, by an integer j . Alternatively, it may be viewed as a collection of row vectors labelled, from top to bottom, by an integer i .

Let $|k\rangle$ be some ket. The action of $|i\rangle\langle j|$ on this ket is defined by the rule

$$(|i\rangle\langle j|)|k\rangle = |i\rangle\langle j||k\rangle = |i\rangle\langle j|k\rangle$$

which, using (10.65), we rewrite as

$$(|i\rangle\langle j|)|k\rangle = \delta_{jk}|i\rangle. \quad (3.6.162)$$

Evidently, the matrix associated with $|i\rangle\langle j|$ has all entries 0 save for a 1 entry at the intersection of the i 'th row and the j 'th column,

$$\langle m|(|i\rangle\langle j|)|n\rangle = \langle m||i\rangle\langle j||n\rangle = \langle m|i\rangle\langle j|n\rangle = \delta_{mi}\delta_{jn}. \quad (3.6.163)$$

Put another way, corresponding to $|i\rangle\langle j|$ is a matrix with zeroes everywhere save for a 1 at the location ij . For example, there are the correspondences

$$|1\rangle\langle 1| \leftrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (3.6.164)$$

$$|1\rangle\langle 2| \leftrightarrow \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (3.6.165)$$

$$|2\rangle\langle 1| \leftrightarrow \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (3.6.166)$$

$$|3\rangle\langle 2| \leftrightarrow \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (3.6.167)$$

²²Recall Subsection 6.2 where use of dyads is described. Dirac notation is both ingenious and somewhat confusing. It is ingenious because it capitalizes on the already existing notation for the inner product. It is somewhat confusing because, when necessary, in use it deletes parentheses and replaces $||$ by $|$.

$$|3\rangle\langle 4| \leftrightarrow \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (3.6.168)$$

$$|4\rangle\langle 4| \leftrightarrow \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \text{ etc.} \quad (3.6.169)$$

We are now ready for some applications. The first involves vectors. Suppose u is a column vector with entries $u_1 \dots$,

$$|u\rangle \leftrightarrow \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix}. \quad (3.6.170)$$

Also suppose v is a row vector with entries $v_1 \dots$,

$$v \leftrightarrow (v_1, v_2, v_3, v_4). \quad (3.6.171)$$

Using Dirac notation, verify we may write

$$|u\rangle = \sum_i u_i |i\rangle \quad (3.6.172)$$

and

$$\langle v| = \sum_j v_j \langle j|. \quad (3.6.173)$$

Also, verify that there is the *inner product* result

$$\langle v|u\rangle = \sum_{ji} v_j u_i \langle j|i\rangle = \sum_{ji} v_j u_i \delta_{ji} = \sum_k v_k u_k, \quad (3.6.174)$$

as expected. Finally, verify that there is the dyadic/*outer product* result

$$|u\rangle\langle v| = \sum_{ij} u_i v_j |i\rangle\langle j|. \quad (3.6.175)$$

Show, in view of correspondences of the kind (6.164) through (6.169), that there is the correspondence

$$|u\rangle\langle v| \leftrightarrow \begin{pmatrix} u_1 v_1 & u_1 v_2 & u_1 v_3 & u_1 v_4 \\ u_2 v_1 & u_2 v_2 & u_2 v_3 & u_2 v_4 \\ u_3 v_1 & u_3 v_2 & u_3 v_3 & u_3 v_4 \\ u_4 v_1 & u_4 v_2 & u_4 v_3 & u_4 v_4 \end{pmatrix}. \quad (3.6.176)$$

In the case $v = u$ there is the dyad $|u\rangle\langle u|$, which we will call a *symmetric* dyad because its corresponding matrix is symmetric.

3.6.2. Evidently (6.176) is a matrix result. Let us find more matrix results. Suppose \mathcal{F} is a linear operator which acts on a vector space to send it into itself, and suppose it is described by a square matrix with entries F_{ij} . Using the operators $|i\rangle\langle j|$, make the definition

$$\mathcal{F} = \sum_{ij} F_{ij}|i\rangle\langle j|. \quad (3.6.177)$$

From this construction verify that

$$\langle m|\mathcal{F}|n\rangle = \sum_{ij} F_{ij}\langle m|i\rangle\langle j|n\rangle = \sum_{ij} F_{ij}\delta_{mi}\delta_{jn} = F_{mn}, \quad (3.6.178)$$

as desired.

As a second application suppose \mathcal{G} and \mathcal{H} are two linear operators, with associated matrices G and H , defined by

$$\mathcal{G} = \sum_{ab} G_{ab}|a\rangle\langle b| \quad (3.6.179)$$

and

$$\mathcal{H} = \sum_{cd} H_{cd}|c\rangle\langle d|. \quad (3.6.180)$$

Also suppose that \mathcal{H} is the product of \mathcal{F} and \mathcal{G} ,

$$\mathcal{H} = \mathcal{F}\mathcal{G}. \quad (3.6.181)$$

Verify that employing (10.78) and (10.80) in (10.82) yields the result that \mathcal{H} can be written in the form

$$\mathcal{H} = \sum_{abcd} F_{ab}G_{cd}(|a\rangle\langle b|)(|c\rangle\langle d|) = \sum_{abcd} F_{ab}G_{cd}|a\rangle\delta_{bc}\langle d| = \sum_{abd} F_{ab}G_{bd}|a\rangle\langle d|. \quad (3.6.182)$$

Note that in writing (6.183) we have implicitly made the step

$$(|a\rangle\langle b|)(|c\rangle\langle d|) = |a\rangle\langle b||c\rangle\langle d| = |a\rangle\langle b|c\rangle\langle d| = |a\rangle\delta_{bc}\langle d| = \delta_{bc}|a\rangle\langle d|, \quad (3.6.183)$$

which follows from the rules for dyad multiplication. Verify/conclude that \mathcal{H} can be written in the form

$$\mathcal{H} = \sum_{ad} H_{ad}|a\rangle\langle d| \quad (3.6.184)$$

with

$$H_{ad} = \sum_b F_{ab}G_{bd} = (FG)_{ad}, \quad (3.6.185)$$

the expected matrix multiplication rule.

Here is another derivation of this result: Define the operator \mathcal{I} by the rule

$$\mathcal{I} = \sum_{bc} \delta_{bc}|b\rangle\langle c| = \sum_b |b\rangle\langle b|. \quad (3.6.186)$$

Verify that for any ket $|k\rangle$

$$\mathcal{I}|k\rangle = |k\rangle, \quad (3.6.187)$$

and therefore, since the kets $|k\rangle$ form a basis, \mathcal{I} is the *identity* operator. Consequently there is the result

$$\mathcal{H} = \mathcal{F}\mathcal{G} = \mathcal{F}\mathcal{I}\mathcal{G}, \quad (3.6.188)$$

which is an instance of what we call a judicious insertion of the identity. Using (10.87) and (10.89), verify that

$$\begin{aligned} H_{ad} &= \langle a|\mathcal{H}|d\rangle = \langle a|\mathcal{F}\mathcal{I}\mathcal{G}|d\rangle = \langle a|\mathcal{F}\left(\sum_b|b\rangle\langle b|\right)\mathcal{G}|d\rangle \\ &= \sum_b \langle a|\mathcal{F}|b\rangle\langle b|\mathcal{G}|d\rangle = \sum_b F_{ab}G_{bd}, \end{aligned} \quad (3.6.189)$$

as before.

3.6.3. Show that orthogonal matrices, matrices satisfying (6.1), have the property (6.2).

3.6.4. Suppose that O is an orthogonal matrix. Show that $-O$, O^T , and O^{-1} are also orthogonal matrices. Show that orthogonal matrices form a group. Show that O and O^T commute, $O^TO = OO^T$.

3.6.5. Verify (6.10) by showing that D has the matrix elements

$$\begin{aligned} D_{ij} &= \sum_k (e^i, w^k)(w^k, e^j) = \sum_k (e^i, w^k)(e^j, w^k) \\ &= \sum_k W_{ik}W_{jk} = \sum_k W_{ik}(W^T)_{kj} = (WW^T)_{ij}. \end{aligned} \quad (3.6.190)$$

3.6.6. Verify (6.12) by showing that G can be written in the form

$$G(W) = \sum_{k\ell} |e^k)(w^k, w^\ell)(e^\ell|, \quad (3.6.191)$$

and has the matrix elements

$$G_{ij} = \sum_{k\ell} (e^i, e^k)(w^k, w^\ell)(e^\ell, e^j) = (w^i, w^j) = (We^i, We^j) = (e^i, W^TWe^j). \quad (3.6.192)$$

3.6.7. Verify (6.26) and (6.27). Note that (6.5) implies the relation

$$(w^j| = (e^j|W^T. \quad (3.6.193)$$

3.6.8. Verify that the v^j given by (6.53) are orthonormal.

3.6.9. Show that the trace operation has the properties

$$\text{tr}(A) = \text{tr}(A^T), [\text{tr}(A)]^* = \text{tr}(A^\dagger), \quad (3.6.194)$$

$$\text{tr}(AB) = \text{tr}(BA). \quad (3.6.195)$$

Here a $*$ denotes complex conjugation.

3.6.10. Verify the relations

$$\mathrm{tr}(A^T A) = \sum_{ij} (A_{ij})^2, \quad \mathrm{tr}(A^\dagger A) = \sum_{ij} |A_{ij}|^2. \quad (3.6.196)$$

3.6.11. Verify (6.88) through (6.90).

3.6.12. Verify that \mathcal{F}_D has the explicit form

$$\mathcal{F}_D[U] = \sum_k \left\{ \sum_\ell [(u^k, w^\ell)]^2 \right\}^2. \quad (3.6.197)$$

3.6.13. As an alternative to (6.81), consider the option of writing

$$U = RV, \quad (3.6.198)$$

and then working with (6.80). Show that one can require that U^T diagonalize $G(W)$, and find an associated variational problem.

3.6.14. This exercise studies the Darboux (Gram-Schmidt like) method of constructing a symplectic basis. Assume that the w^j are linearly independent and recall that J is invertible. Show that the vectors ${}^m w^k$ exist and the vectors $v^3, v^5, \dots, v^{2n-1}$ are nonzero. For example, at step 2 of the algorithm, show that the possibility

$$(v^1, Jw^j) = 0 \text{ for all } j \quad (3.6.199)$$

would imply that the w^j are linearly dependent. Similarly, at steps 5 and 6, show that the vectors $v^1, v^2, {}^1 w^3, {}^1 w^4, \dots, {}^1 w^{2n}$, are linearly independent and the vector v^3 is nonzero. Continue on to show that steps 9, 13, \dots and steps 10, 14, \dots succeed. Alternatively, verify by induction on n that the Darboux construction is always possible:

- a) Verify the case of dimension 2.
- b) Assume the result holds in dimension $(2n - 2)$. Consider the case of dimension $2n$ as in Section 3.6.5. Show that a w^j can be found that satisfies (6.95) because the w^i are assumed to be linearly independent.
- c) Verify that the $(2n - 2)$ vectors ${}^1 w^j$ defined by (6.99) satisfy (6.100) and are linearly independent. Therefore, by the induction hypothesis, a symplectic basis can be found for this set of vectors. Show that these $(2n - n)$ symplectic basis vectors, along with v^1 and v^2 , then form a set of $2n$ symplectic basis vectors.

3.6.15. Suppose M is a $2n \times 2n$ matrix. Regard M as a collection of column vectors m^a so that it can be written in the form

$$M = \begin{pmatrix} M_{1,1} & M_{1,2} & M_{1,3} & \cdots & M_{1,2n} \\ M_{2,1} & M_{2,2} & M_{2,3} & \cdots & M_{2,2n} \\ M_{3,1} & M_{3,2} & M_{3,3} & \cdots & M_{3,2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ M_{2n,1} & M_{2n,2} & M_{2n,3} & \cdots & M_{2n,2n} \end{pmatrix} = (m^1, m^2, m^3, \dots, m^{2n}). \quad (3.6.200)$$

Verify that, with this convention, the column vectors m^a will have entries m_c^a given by the relations

$$m_c^a = M_{ca}. \quad (3.6.201)$$

Correspondingly, verify that there is the relation

$$m^a = Me^a. \quad (3.6.202)$$

Next, suppose M is a real symplectic matrix that satisfies (1.1) with J specified by (1.2). Verify that, in terms of matrix elements, (1.1) takes the form

$$(e^a, M^T J M e^b) = (e^a, J e^b) = J_{ab}, \quad (3.6.203)$$

from which it follows that

$$J_{ab} = (e^a, M^T J M e^b) = (Me^a, JM e^b) = (m^a, Jm^b). \quad (3.6.204)$$

Thus, as expected from the discussion of Section 3.6.3, the vectors m^a form a symplectic basis. Verify that if

$$a = i \text{ with } i \in [1, n] \quad (3.6.205)$$

and

$$b = i + n, \quad (3.6.206)$$

then

$$(m^a, Jm^b) = J_{ab} = J_{i,i+n} = 1. \quad (3.6.207)$$

Verify that

$$(m^a, m^a) = \sum_c (M_{ca})^2. \quad (3.6.208)$$

Suppose that instead M is symplectic with respect to the J' defined by (2.10). Show that the general discussion of this exercise goes through as before except that we should now set

$$a = i \text{ with } i = 1, 3, 5, \dots \quad (3.6.209)$$

and

$$b = i + 1 \quad (3.6.210)$$

so that

$$(m^a, J'm^b) = J'_{ab} = J'_{i,i+1} = 1. \quad (3.6.211)$$

3.6.16. Show that for a symplectic basis and a J of the form (2.10) there are the relations

$${}^r v^1 = Jv^2, {}^r v^2 = -Jv^1; {}^r v^3 = Jv^4, {}^r v^4 = -Jv^3; \text{ etc.} \quad (3.6.212)$$

3.6.17. Here is a curiosity: Given a set of $2n$ linearly independent vectors w^j , can one find a set of vectors ${}^{sr} w^i$ such that

$$({}^{sr} w^i, Jw^j) = J_{ij} ? \quad (3.6.213)$$

The answer is yes. Such a set will be called a *symplectic reciprocal* basis. Let the vectors ${}^r w^i$ denote the ordinary reciprocal basis to the w^j . See (6.18). Define a related basis \tilde{w}^i by the rule

$$\tilde{w}^i = J \cdot {}^r w^i. \quad (3.6.214)$$

This basis has the property

$$(\tilde{w}^i, Jw^j) = (J \cdot {}^r w^i, Jw^j) = ({}^r w^i, J^T Jw^j) = ({}^r w^i, w^j) = \delta_{ij}. \quad (3.6.215)$$

Now it is convenient to work with the J given by (2.10). With this choice in mind, define the ${}^{sr} w^i$ by the rule

$$\begin{aligned} {}^{sr} w^1 &= \tilde{w}^2, \\ {}^{sr} w^2 &= -\tilde{w}^1, \\ {}^{sr} w^3 &= \tilde{w}^4, \\ {}^{sr} w^4 &= -\tilde{w}^3, \\ &\text{etc.} \end{aligned} \quad (3.6.216)$$

Verify that these vectors satisfy (6.148).

3.6.18. Review the discussion of transformation groups at the end of Section 6.1. For each of the realizations of a group acting on itself introduce the notation

$$h \rightarrow T_g h = gh, \quad h \rightarrow T_g h = hg^{-1}, \quad h \rightarrow T_g h = ghg^{-1}; \quad g \in G, \quad h \in \mathcal{Z} = G. \quad (3.6.217)$$

Verify that in each realization there is the relation

$$T_{g_2} T_{g_1} = T_{g_2 g_1}, \quad (3.6.218)$$

which shows that in each realization the transformations T_g form a group.

3.7 Lie Algebraic Properties

3.7.1 Matrix Exponential and Logarithm

Let B be any matrix. The *exponential* of a matrix, written variously as e^B or $\exp(B)$, is defined by the exponential series

$$e^B = \exp(B) = I + B + B^2/2! + \dots = \sum_{n=0}^{\infty} B^n/n!. \quad (3.7.1)$$

(Here we adopt the usual convention that $B^0 = I$ for any matrix B .) Similarly, the *logarithm* of a matrix A (sufficiently near the identity) is defined by the logarithm series

$$\log(A) = \log[I - (I - A)] = - \sum_{n=1}^{\infty} (I - A)^n/n. \quad (3.7.2)$$

As might be expected, the exponential and logarithmic functions are related. Specifically, if one has

$$B = \log(A), \quad (3.7.3)$$

then it follows that

$$A = \exp(B), \quad (3.7.4)$$

and vice versa. Put another way, one has the relations

$$A = \exp[\log(A)] \text{ for } A \text{ sufficiently near the identity matrix,} \quad (3.7.5)$$

$$B = \log[\exp(B)] \text{ for } B \text{ sufficiently near the zero matrix.} \quad (3.7.6)$$

If a matrix A can be written in the form (7.4), we say that B generates A . It can be shown that there is the identity

$$[\exp(B/n)]^n = \exp(B) \quad (3.7.7)$$

for any integer n . See Exercise 7.5. Thus, if A is generated by B , we may also write

$$A = [\exp(B/n)]^n. \quad (3.7.8)$$

From (7.1) we see that

$$\exp(B/n) = I + B/n + O(1/n)^2.$$

Consequently, for sufficiently large n , B/n is near the zero matrix and $\exp(B/n)$ is near the identity matrix. Since B/n is near the zero matrix, we may regard it as an infinitesimal matrix. Correspondingly, in view of (7.8), we say that A is *infinitesimally generated* in that it can be written as the product of a large number of identical near identity matrices. Finally, like the ordinary exponential function, it can be verified that

$$\lim_{n \rightarrow \infty} (I + B/n)^n = \exp(B) \quad (3.7.9)$$

for any matrix B .

In some cases there may be several linearly independent matrices B_1, B_2, \dots, B_k and we consider elements of the form

$$A = \exp(s_1 B_1 + s_2 B_2 + \dots + s_k B_k).$$

Here the s_j are scalars. We would then say that the B_j generate such matrices A . Even more generally, we might consider matrices that are finite products of matrices of the form A ,

$$G = \exp(s_1 B_1 + s_2 B_2 + \dots + s_k B_k) \exp(t_1 B_1 + t_2 B_2 + \dots + t_k B_k) \dots .$$

We would again say that the B_j generate such matrices G . However, it might not be possible to write such matrices in the form

$$G = \exp(B) \quad (3.7.10)$$

where B is some linear combination of the B_j . Consequently, if (7.10) is not possible, we would say that G is generated by the B_j , but not infinitesimally generated as defined above. Alternatively, we might broaden our definition of “infinitesimally generated” to include finite products of matrices that are themselves infinitesimally generated.

Vector and Matrix Norms

To make our discussion more precise, it is useful to introduce the concepts of vector and matrix *norms*. We will use the same notation $\|\cdot\|$ to refer to either norm with the understanding that the exact meaning of the notation depends on whether it is being applied to a vector or a matrix.

A vector norm is a rule that assigns to any vector v a real non-negative number $\|v\|$, called the norm of v , in such a way that the following properties are satisfied:

$$\|v\| \geq 0, \text{ and } \|v\| = 0 \Leftrightarrow v = 0; \quad (3.7.11)$$

$$\|av\| = |a|\|v\|, \text{ } a \text{ any scalar}; \quad (3.7.12)$$

$$\|u + v\| \leq \|u\| + \|v\|. \quad (3.7.13)$$

Here the notation \Leftrightarrow is used to denote logical implication in both directions.

Similarly, a matrix norm is a rule that assigns to any matrix A a real non-negative number $\|A\|$, called the norm of A . The matrix norm is required to satisfy properties analogous to those for a vector norm plus a property associated with matrix multiplication:

$$\|A\| \geq 0, \text{ and } \|A\| = 0 \Leftrightarrow A = 0; \quad (3.7.14)$$

$$\|aA\| = |a|\|A\|, \text{ } a \text{ any scalar}; \quad (3.7.15)$$

$$\|A + B\| \leq \|A\| + \|B\|; \quad (3.7.16)$$

$$\|AB\| \leq \|A\|\|B\|. \quad (3.7.17)$$

Finally, a matrix norm is said to be *consistent* with a vector norm if the following condition is satisfied for any matrix A and vector v (assuming that A is $m \times m$ and v is m -dimensional):

$$\|Av\| \leq \|A\|\|v\|. \quad (3.7.18)$$

Note that the norm indicated in the left side of (7.18) is a vector norm since the quantity Av is a vector. By contrast, the norms on the right side of (7.18) are matrix and vector norms, respectively.

There are several ways of defining consistent matrix and vector norms. One of the more useful is to take for the matrix norm the *maximum column sum* norm. It is defined by the rule

$$\|A\| = \max_k \left(\sum_j |A_{jk}| \right). \quad (3.7.19)$$

[Sum over j while holding k fixed to add together the values of $|A_{jk}|$ for column k . Then, for the various columns (values of k), report the largest result found.] It can be shown that this norm satisfies the requirements (7.14) through (7.17). Furthermore, it can be shown that this norm is consistent with the *component moduli sum* vector norm defined by the rule

$$\|v\| = \sum_j |v_j|. \quad (3.7.20)$$

The *strongest* matrix norm is the *spectral* norm defined by

$$\|A\|_{\text{spct}} = +(\text{maximum eigenvalue of } A^\dagger A)^{1/2}. \quad (3.7.21)$$

Note that for any matrix A the eigenvalues of $A^\dagger A$ are guaranteed to be real and nonnegative so that (7.21) is well defined. The spectral norm is strongest in the sense that for any matrix A there is the inequality

$$\|A\|_{\text{spct}} \leq \|A\| \quad (3.7.22)$$

where $\|A\|_{\text{spct}}$ denotes the spectral norm and $\|A\|$ denotes any other matrix norm. However, to compute the spectral norm generally requires considerable work, and therefore it is sometimes more of theoretical value rather than suitable for frequent computation.

It can be shown that the matrix spectral norm is consistent with the *Euclidean* vector norm. Let v be a possibly complex m -dimensional vector and let $(*, *)$ denote the usual complex inner product. The Euclidean vector norm $\|v\|_E$ is defined by the rule

$$(\|v\|_E)^2 = (v, v) = \sum_j |v_j|^2. \quad (3.7.23)$$

There are also other ways of defining vector and matrix norms. See Exercise 7.1 for the definition of the Euclidean matrix norm.

Convergence of Series

With the aid of the concept of a matrix norm, it can be shown that the series (7.1) converges for *any* matrix B , and that one has the relations

$$\|\exp(B)\| \leq e^{\|B\|}, \quad (3.7.24)$$

$$\|[\exp(B) - I]\| \leq e^{\|B\|} - 1. \quad (3.7.25)$$

[There is a theorem to the effect that if a matrix power series $\sum_n c_n B^n$ converges in norm (*i.e.*, if the series converges with all coefficients c_n replaced by $|c_n|$ and with B^n replaced by $\|B\|^n$), then it also converges for each individual matrix element.] By contrast, it can be shown that in general the series (7.2) converges only when $\|(A - I)\| < 1$ for some norm, and that then one has the relation

$$\|\log(A)\| \leq -\log[1 - \|(A - I)\|]. \quad (3.7.26)$$

3.7.2 Application to Symplectic Matrices

With this background in mind, suppose that M is a real symplectic matrix near the identity. We start our analysis in a heuristic fashion by assuming that M can be written in the form

$$M = \exp(\epsilon B) \quad (3.7.27)$$

where ϵ is small so that ϵB is near the zero matrix. We then have the expansions

$$M = I + \epsilon B + O(\epsilon^2), \quad (3.7.28)$$

$$M^T = I + \epsilon B^T + O(\epsilon^2).$$

Upon inserting these expansions into the symplectic condition (1.2) and equating powers of ϵ , we find the result

$$B^T J + J B = 0. \quad (3.7.29)$$

The relation (7.29) is a key result that we will now prove rigorously for $\epsilon = 1$ provided B itself is sufficiently small. Specifically, assume that M and M^{-1} are sufficiently near the identity so that $\log(M)$ and $\log(M^{-1})$ can be computed using (7.2). That is, suppose the following two series converge:

$$B = \log(M) = - \sum_{n=1}^{\infty} (I - M)^n / n, \quad (3.7.30)$$

$$-B = \log(M^{-1}) = - \sum_{n=1}^{\infty} (I - M^{-1})^n / n. \quad (3.7.31)$$

Use the series (7.30) to compute the quantity $J^{-1}B^T J$. Doing so gives the result

$$\begin{aligned} J^{-1}B^T J &= - \sum_{n=1}^{\infty} (I - J^{-1}M^T J)^n / n \\ &= - \sum_{n=1}^{\infty} (I - M^{-1})^n / n \\ &= -B. \end{aligned} \quad (3.7.32)$$

Here use has also been made of (7.31) and (1.9). Now compare the beginning and end of (7.32) to get the equivalent results

$$J^{-1}B^T J = -B \text{ or } JB^T J^{-1} = -B \text{ or } B^T J + J B = 0 \text{ or } JB^T + B J = 0. \quad (3.7.33)$$

Note that (7.29) is among these equivalent results. A matrix B that satisfies (7.33) is sometimes called *Hamiltonian* or *infinitesimally symplectic*.²³

To understand the implications of the condition (7.33), suppose that B is written in the form

$$B = JS. \quad (3.7.34)$$

[Reader, verify that given any matrix B , because J is nonsingular, there is always a well-defined S such that (7.34) is satisfied.] Upon inserting (7.34) into (7.33), one finds the equivalent condition

$$-S^T JJ + JJS = 0 \text{ or } S^T = S. \quad (3.7.35)$$

That is, S must be a symmetric matrix. Parenthetically, we note that any Hamiltonian matrix (any matrix of the form JS with S symmetric) must be traceless. Verify this claim!

²³This usage of the adjective *Hamiltonian* should not be confused with its usage in Quantum Mechanics where a Hamiltonian matrix would be a matrix formed by taking the matrix elements of a Hamiltonian operator with respect to an orthonormal basis. Such a matrix would generally have complex entries and would be Hermitian.

It follows that any matrix M of the form $M = \exp(JS)$ must have unit determinant. See Exercise 7.10.

We have learned that any real symplectic matrix M sufficiently near the identity can be written in the form

$$M = e^B = e^{JS}, \quad (3.7.36)$$

with S small, real, and symmetric. Conversely, suppose that B is any matrix of the form (7.34) with S real and symmetric. Then, the matrix M given by (7.36) is symplectic. To verify this assertion, simply compute! One finds the results

$$M = \exp(JS), \quad (3.7.37)$$

$$M^T = \exp(-SJ),$$

$$\begin{aligned} M^TJM &= \exp(-SJ)J\exp(JS) \\ &= JJ^{-1}\exp(-SJ)J\exp(JS) \\ &= J\exp(-J^{-1}SJ^2)\exp(JS) \\ &= J\exp(-JS)\exp(JS) \\ &= J. \end{aligned} \quad (3.7.38)$$

What has been shown is that any symplectic matrix M sufficiently near the identity can be written in the form (7.36) with S small and symmetric, and vice versa.²⁴ Note that the symplectic condition as expressed by (1.2) is a set of *quadratic* relations among the matrix elements of M . By contrast, the conditions (7.33) or (7.35) are *linear* relations among the matrix elements of B or S , respectively. We see that the use of an exponential representation has converted a set of quadratic relations, which are generally more difficult to work with due to their nonlinearity, into a set of simple linear relations.

Finally, we remark that not every symplectic matrix can be written in single exponential form. See Exercise 7.12.

3.7.3 Matrix Lie Algebra and Lie Group: The Baker-Campbell-Hausdorff (BCH) Multiplication Theorem

The stage is now set for the introduction of a central discovery of Sophus Lie, the concept of a *Lie* algebra. We will first introduce this concept in a concrete matrix setting, and then place it in a more general abstract setting.

A set A of $m \times m$ matrices forms a Lie algebra if it satisfies the following properties:

- i. If the matrix A is in the Lie algebra, then so is the matrix aA where a is any scalar.
- ii. If two matrices A and B are in the Lie algebra, then so is their sum.

²⁴Here we see the beginning of a grand theme: There is a close relation between symplectic and symmetric matrices. This theme will be developed fully in Sections 3.11, 5.13, and 6.7. We also note that in the calculation (7.38) we have used the results of Exercises 7.5 and 7.11.

- iii. If two matrices A and B are in the Lie algebra, then so is their *commutator* $[A, B]$.
The *commutator* is defined by the relation

$$[A, B] = AB - BA. \quad (3.7.39)$$

Note that the commutator symbol $[,]$ is the same as that used earlier for a Poisson bracket. This is somewhat awkward, but unfortunately there are not always enough convenient symbols to go around. Later, when there is greater chance of confusion, we will use the symbols $\{, \}$ to denote a commutator.

At this point the reader should take pen in hand and verify that the set of matrices of the form JS with S symmetric is a Lie algebra. That is, *Hamiltonian matrices form a Lie algebra*.

That Hamiltonian matrices form a Lie algebra is no accident. It is a remarkable fact that there is a close connection between the concept of a Lie algebra and that of a group. The connection arises from a deep property of the exponential function that generally bears the names *Baker-Campbell-Hausdorff* (BCH). Their result, in a matrix setting, may be stated as follows: Let A and B be any two matrices (square and of the same dimension). Form the matrices $\exp(sA)$ and $\exp(tB)$ where s and t are parameters. Next form their product. Then, for s and t sufficiently small, it is possible to write

$$\exp(sA) \exp(tB) = \exp(C), \quad (3.7.40)$$

where C is some other matrix. The remarkable fact is that C is a member of the Lie algebra *generated* by A and B .²⁵ That is, C is a sum of elements formed *only* from A and B and their *multiple commutators*. Specifically, one has the relation

$$\begin{aligned} C(s, t) = sA &+ tB + (st/2)[A, B] + (s^2t/12)[A, [A, B]] \\ &+ (st^2/12)[B, [B, A]] \\ &- (s^2t^2/24)[A, [B, [A, B]]] + O(s^4t, s^3t^2, s^2t^3, st^4). \end{aligned} \quad (3.7.41)$$

No isolated terms of the form A^2 , B^2 , AB , $[A^2, B^2]$, etc. occur! Although stated in terms of matrices, this result can be extended to the case of linear operators.

In general, the series for C (called the BCH series) contains an infinite number of terms and may converge only for sufficiently small s and t . It may not converge at all if the Lie algebra generated by A and B is infinite dimensional and A and B are unbounded operators.²⁶

²⁵Here is yet another, and different, use of the word *generate*. Suppose one has a collection of $n \times n$ matrices B_i . Form their commutators to produce possibly new linearly independent matrices. Next, join the set of these matrices to the original set of the B_i . Now form the commutators of all these matrices, and join these matrices to the set already obtained. Repeat this process ad infinitum until no new linearly independent matrices are obtained. In the matrix case this process must terminate because there are only n^2 linearly independent $n \times n$ matrices. The net result of this procedure is a Lie algebra, which is referred to as the Lie algebra generated by the B_i . Although we have been talking about matrix Lie algebra, the same construction can be carried out for any collection of Lie elements drawn from some Lie algebra with the commutator replaced by the abstract Lie product.

²⁶The BCH series can be summed in the case of $sp(2, \mathbb{C}) = sl(2, \mathbb{C})$ which includes $su(2)$ and $sp(2, \mathbb{R})$. See Subsection 8.7.1. For an example of divergence in the infinite-dimensional case, see Section 38.7.

The proof of this theorem is difficult and is given in Appendix C.²⁷ For present purposes, it shows that given any Lie algebra L of matrices, there exists a corresponding *Lie group* G . Furthermore, the rules for multiplying any two group elements are contained within the Lie algebra. To see the truth of this assertion, consider all matrices of the form $g(s) = \exp(s\ell)$ with ℓ contained in L . According to the previous result, one has

$$\exp(s\ell) \exp(t\ell') = \exp \ell''$$

with ℓ'' given by a relation of the form (7.41) for s, t sufficiently small. Also

$$g(0) = I \text{ and } g^{-1}(s) = g(-s).$$

Thus these matrices, at least those sufficiently near the identity, form a group. Once the group has been obtained near the identity, it can be extended to a global group by successively multiplying the different g 's already obtained. We remark that if the Lie algebras of two sets of matrices are the same, it does not necessarily follow that the two corresponding groups constructed in this way are globally the same. They may only be related by a homomorphism. The groups $SU(2)$ and $SO(3, \mathbb{R})$ provide an example of this possibility.²⁸ Information about the matrices beyond their Lie algebra is needed to determine the global properties of the group.

It has already been shown that symplectic matrices form a group. Furthermore, it has been shown that symplectic matrices near the identity can be written as the exponentials of elements of a Lie algebra. It follows that $Sp(2n)$, the group of symplectic matrices, is a Lie group. The Lie algebra associated with $Sp(2n)$, the Lie algebra of Hamiltonian matrices, is denoted by $sp(2n)$. More specifically, the Lie algebra associated with $Sp(2n, \mathbb{R})$ is denoted by $sp(2n, \mathbb{R})$, and that associated with $Sp(2n, \mathbb{C})$ is denoted by $sp(2n, \mathbb{C})$. Where there is no possibility of confusion, we will use the notation $sp(2n)$ to mean $sp(2n, \mathbb{R})$.

Properties 1 and 2 of a Lie algebra indicate that the elements of a Lie algebra form a linear vector space. It is therefore natural to speak of the *dimension* of a Lie algebra. For the case of the symplectic group, elements of the Lie algebra are of the form (7.34) where S is any symmetric matrix. The dimension of the Lie algebra in this case, therefore, is just the dimensionality of the set of all $2n \times 2n$ symmetric matrices. This number is easily computed. There are $2n$ independent entries on the diagonal of a $2n \times 2n$ symmetric matrix, and $[(2n)^2 - 2n]/2$ independent entries above the diagonal. Finally, all the entries below the diagonal are given in terms of the entries above the diagonal by the symmetry condition. Therefore, the dimension of the symplectic group Lie algebra, which will be written as $\dim sp(2n)$, is given by the relation

$$\dim sp(2n) = 2n + [(2n)^2 - 2n]/2 = n(2n + 1). \quad (3.7.42)$$

For example, the dimensions of $sp(2)$, $sp(4)$, and $sp(6)$ are 3, 10, and 21, respectively. See Table 7.1 below.

Let M be some element of $Sp(2n)$ that can be written in the exponential form (7.36). To the extent that the elements of $Sp(2n)$ in some neighborhood of M can also be written

²⁷Also see Appendix C for a discussion of the converse Zassenhaus formula.

²⁸The groups $SU(n)$ will be defined in Subsection 7.6.

Table 3.7.1: Dimension of $sp(2n)$.

n	$2n$	$\dim sp(2n)$	n	$2n$	$\dim sp(2n)$
1	2	3	5	10	55
2	4	10	6	12	78
3	6	21	7	14	105
4	8	36	8	16	136

in exponential form, we may say that the dimension of this neighborhood is also given by (7.42). However, in a while we will see that not all elements of $Sp(2n)$ can be written in exponential form. See Exercise 7.12 and Subsection 8.7.2. What about the general case? In Subsection 8.2 it is shown that every symplectic matrix can be written as the product of two symplectic matrices, each of which can be written in exponential form. And the *total* dimension count of the two of them together is again given by (7.42). See Exercise 9.10. Consequently we may say that $su(2n)$ as a vector space and $Sp(2n)$ as a manifold have the same dimension.

3.7.4 Abstract Definition of a Lie Algebra

For future use, it is essential to put the concept of a Lie algebra, as just defined in a matrix context, into a more general setting. We will begin by defining the concept of an algebra.

Algebra

Naively speaking, algebra has to do with the concepts of addition and multiplication. The concept of addition can be generalized to yield the concept of a linear vector space. The concept of multiplication has several possible generalizations. Formally, an *algebra* A over a field of numbers F is defined as a linear vector space supplemented by a rule for multiplying two vectors to yield a third vector. This multiplication rule must satisfy certain conditions having to do jointly with vector space properties and multiplication properties. Indicating multiplication by the symbol \circ , we require that to every ordered pair of elements $x, y \in A$ there corresponds a third unique element of A , denoted by $x \circ y$, and called the *product* of x and y . The product should satisfy the following requirements:

$$1. (cx) \circ y = x \circ (cy) = c(x \circ y) \quad (3.7.43)$$

$$2. (x + y) \circ z = x \circ z + y \circ z \quad (\text{right distributive}) \quad (3.7.44)$$

$$3. x \circ (y + z) = x \circ y + x \circ z \quad (\text{left distributive}) \quad (3.7.45)$$

for any $x, y, z \in A$ and $c \in F$.

Associative Algebra

An example of an algebra is the set of all $m \times m$ matrices. The set of all $m \times m$ matrices forms an m^2 dimensional vector space. It also forms an algebra if we use for the \circ operation

ordinary matrix multiplication. Note that in this case multiplication is *associative*, that is, the multiplication rule satisfies the property

$$(x \circ y) \circ z = x \circ (y \circ z). \quad (3.7.46)$$

Lie Algebra

A second example of an algebra is the set of all 3-vectors with the multiplication rule given by the relation

$$\mathbf{a} \circ \mathbf{b} = \mathbf{a} \times \mathbf{b}. \quad (3.7.47)$$

Here \times denotes the usual cross product. This algebra is *not* associative,

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} \neq \mathbf{a} \times (\mathbf{b} \times \mathbf{c}).$$

A Lie algebra L is an algebra for which the multiplication rule (sometimes now called a Lie product) satisfies two *further* properties. For convenience, multiplication of x and y will now be denoted by the symbol $[x, y]$,

$$[x, y] = x \circ y.$$

In using this customary notation, however, it should be understood that the bracket $[,]$ does not necessarily refer to a commutator (or a Poisson bracket). Rather, in this context, it refers to the Lie product abstractly, and independently of any particular realization. The two additional properties for a Lie product are the following:

$$4. [x, y] = -[y, x] \quad (\text{antisymmetry}) \quad (3.7.48)$$

$$5. [x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0 \quad (\text{Jacobi condition or identity}) \quad (3.7.49)$$

We note that a Lie algebra is not associative. Instead, the associativity condition (7.46) has been replaced by the *Jacobi condition* (7.49). We also remark, because Lie algebras are often realized in terms of matrices, two elements in a Lie algebra are said to *commute* if their Lie product vanishes.

A subalgebra K of a Lie algebra L is a subset of L whose elements also satisfy the above properties 1 through 5. Let ℓ be any element of L . Then the set of all scalar multiples of ℓ , which by definition includes the zero element, evidently forms a subalgebra of L . Whether L has any other nontrivial subalgebras depends on the nature of L .

To settle these concepts into the mind, the reader is invited to verify that the set of all 3-vectors with the multiplication rule (7.47) forms a Lie algebra. Next, she or he should verify that the set of all $m \times m$ matrices forms a Lie algebra if the Lie product is taken to be the commutator. In both cases, it is necessary to verify that properties 1 through 5 above are satisfied for the particular Lie product involved.

3.7.5 Abstract Definition of a Lie Group

At this point we should make a side comment. We have defined, and will define, various Lie groups in the context of matrix groups. However, Lie groups can also be defined abstractly. Abstractly, a Lie group is a set G with the following properties:

1. G is a *manifold*. Roughly speaking this means that G , at and near each point, looks like Euclidean space of some fixed dimension m , and there are local coordinates described by m quantities x_1, \dots, x_m . For example, consider the set of all real 2×2 matrices. Since each such matrix has 4 entries, this set can be viewed as being identical to E^4 , 4-dimensional Euclidean space. Within this space is the set of 2×2 matrices M that satisfy (1.2), the set $Sp(2)$ of symplectic matrices. Since (1.2) constitutes a collection of algebraic equations among the entries in M , the set of symplectic matrices forms a manifold within E^4 . Elements of $Sp(2)$ sufficiently near the identity can be written in the form (7.37), and we know that the dimension m of the set of 2×2 matrices of the form JS is 3. See (7.42). Let B_1 through B_3 be a basis for this set. See (7.66) through (7.68) for one possibility. Then, near the identity, we may write $M = \exp(x_1 B_1 + x_2 B_2 + x_3 B_3)$.
2. G is also a group. See the definition of an abstract group in Section 3.6.1. Moreover, the multiplication and inversion operations are required to be *continuous*. Suppose M and N are any two group elements. Continuity means that the coordinates of the product MN are continuous functions of the coordinates of M and N , and the coordinates of M^{-1} are continuous functions of the coordinates of M .

From these assumptions it can be proved that the group operations can actually be made *analytic*.²⁹ That is, there is a choice of coordinates such that the coordinates of the product MN are analytic functions of the coordinates of M and N , and the coordinates of M^{-1} are analytic functions of the coordinates of M . Based on this analyticity, one can differentiate group elements with respect to their coordinates. Next, from the group elements and their derivatives, one can construct entities (vector fields) that can be shown to form a Lie algebra of dimension m . Also, the process can be turned around to reconstruct the group elements from the Lie algebra. Among other things, it can be shown that the Jacobi identity for the Lie algebra is a consequence of the associativity property assumed for the operation of group multiplication. See Appendix R.

3.7.6 Classification of Lie Algebras

Let us return to the main discussion. One of the key discoveries of modern physics is that Lie groups are important for the description of Nature. (Mathematicians already knew earlier that they were important on aesthetic grounds.) Since Lie groups are important, it would be nice to classify them. Because of the close connection between Lie groups and Lie algebras, a natural starting point is to try to classify Lie algebras. This classification has been substantially carried out, initially by *Wilhelm Killing* (1847-1923), and subsequently by *Élie Cartan* (1869-1951) and others. Once Lie algebras/Lie groups have been classified, a next important step is to find *representations* for them in terms of matrices or possibly nonlinear transformations acting on some space. For examples, matrix representations of $su(2)$ are familiar from the Quantum Mechanical theory of angular momentum, matrix representations of $su(3)$ are described in Section 5.8, and matrix representations of $sp(2)$

²⁹See Chapter 38 for a discussion of analyticity.

through $sp(6)$ are described in Chapter 27. Finally, Section 5.12 describes the nonlinear action of $Sp(2n)$ on Siegel space.

The first step in the classification, or even description, of Lie algebras is the introduction of the concept of *structure constants*. Suppose L is a Lie algebra. Since a Lie algebra is a vector space, it must have a basis. Suppose some basis is selected, and let the various basis elements be denoted as B_1, B_2, \dots, B_k where k is the dimension of L . Now consider the Lie product of any two basis elements. Since the Lie product is again an element in the Lie algebra, it must be expandable in the terms of the basis elements. Consequently, there must be a set of coefficients $c_{\alpha\beta}^\gamma$, called *structure constants*, such that one has the relations

$$[B_\alpha, B_\beta] = \sum_\gamma c_{\alpha\beta}^\gamma B_\gamma. \quad (3.7.50)$$

Note that once the Lie product has been specified for the basis elements as in (7.50), then the Lie product for all other elements in L follows from the right and left distributive properties 2 and 3.³⁰

Simple observation shows that, as a consequence of the antisymmetry condition (7.48), the structure constants must obey the relations

$$c_{\alpha\beta}^\gamma = -c_{\beta\alpha}^\gamma. \quad (3.7.51)$$

Somewhat lengthier analysis shows that, as a consequence of the Jacobi condition (7.49), the structure constants must also obey the relations

$$\sum_\sigma (c_{\alpha\beta}^\sigma c_{\gamma\sigma}^\tau + c_{\beta\gamma}^\sigma c_{\alpha\sigma}^\tau + c_{\gamma\alpha}^\sigma c_{\beta\sigma}^\tau) = 0. \quad (3.7.52)$$

Evidently, the problem of classifying all Lie algebras is equivalent to finding all sets of structure constants satisfying (7.51) and (7.52).

Of course, the structure constants depend on the choice of basis elements. Suppose $\tilde{B}_1, \tilde{B}_2, \dots$ is another set of basis elements. Associated with this basis set there will be a set of structure constants $\tilde{c}_{\alpha\beta}^\gamma$ with the property

$$[\tilde{B}_\alpha, \tilde{B}_\beta] = \sum_\gamma \tilde{c}_{\alpha\beta}^\gamma \tilde{B}_\gamma. \quad (3.7.53)$$

Also, since both the B_α and \tilde{B}_β are sets of basis elements, the B_α can be expanded in terms of the \tilde{B}_β , and vice versa. That is, there must be an *invertible* matrix T with the property

$$\tilde{B}_\alpha = \sum_\beta T_{\alpha\beta} B_\beta, \quad (3.7.54)$$

³⁰Strictly speaking, what has been defined in Subsection 7.4 is a *free* Lie algebra. That is, *no* restrictions have been placed on the Lie product save antisymmetry and the Jacobi condition. [As an example of this kind of reasoning/terminology, we may say that the BCH series (7.41) is a free Lie algebraic result because it holds for all Lie algebras no matter what the structure constants may be.] By contrast, once a basis and structure constants have been selected/determined, the Lie algebra is no longer “free” in that it is then completely specified.

$$B_\alpha = \sum_\beta (T^{-1})_{\alpha\beta} \tilde{B}_\beta. \quad (3.7.55)$$

By using (7.50), (7.53), (7.54), and (7.55), we find that the structure constants $c_{\alpha\beta}^\gamma$ and $\tilde{c}_{\alpha\beta}^\gamma$ are connected by the relations

$$\tilde{c}_{\alpha\beta}^\gamma = \sum_{\mu\sigma\tau} T_{\alpha\mu} T_{\beta\sigma} (T^{-1})_{\tau\gamma} c_{\mu\sigma}^\tau, \quad (3.7.56)$$

$$c_{\alpha\beta}^\gamma = \sum_{\mu\sigma\tau} (T^{-1})_{\alpha\mu} (T^{-1})_{\beta\sigma} (T)_{\tau\gamma} \tilde{c}_{\mu\sigma}^\tau. \quad (3.7.57)$$

Often two Lie algebras are deemed to be *equivalent* if their structure constants are related by a change of basis. Sometimes it is important to consider the field from which the entries of T are taken. For example, two Lie algebras may be equivalent if the entries of T are allowed to be complex, but may be inequivalent if T is required to be real. Finally we remark that, in the classification or description of a Lie algebra, it is often convenient to choose a basis in such a way that the structure constants become as neatly organized as possible. For example, one might like to arrange that all the structure constants be real (or purely imaginary). This is possible for all the so-called *simple* Lie algebras.³¹ One might also like to have as many of them vanish as possible, and to have those that do not vanish satisfy some geometric properties. As will be illustrated by examples in Section 5.8 and Chapter 27, so doing for the simple Lie algebras was one of the accomplishments of Killing and Cartan.

The classification of all Lie algebras and Lie groups is a difficult task that lies beyond the scope of our discussion. We shall be primarily interested in the symplectic group and, as will be seen in Chapters 5 and 6, the group of all symplectic maps. However, there are certain Lie groups that arise naturally as subgroups of the symplectic group, and are therefore of direct interest to us. We close this section with a brief discussion of these groups.

Consider the set of all *invertible* $n \times n$ matrices. It is easily verified that this set of matrices forms a group. This group is called the *general linear* group, and is denoted by the symbols $GL(n, \mathbb{R})$ or $GL(n, \mathbb{C})$ depending on the choice of the field to be employed (real or complex). We also use the notation $GL(n, \mathbb{R}, +)$ to indicate the subgroup of $GL(n, \mathbb{R})$ consisting of matrices with positive determinant. Next consider the set of all $n \times n$ matrices with determinant +1. This set of matrices also forms a group, called the *special* linear group. It is denoted by the symbols $SL(n, \mathbb{R})$ or $SL(n, \mathbb{C})$. Evidently, the special linear group is a subgroup of the linear group. The groups $GL(n, \mathbb{R})$, $GL(n, \mathbb{C})$, $SL(n, \mathbb{R})$, and $SL(n, \mathbb{C})$ are all Lie groups. Their associated Lie algebras are denoted by the symbols $gl(n, \mathbb{R})$, $gl(n, \mathbb{C})$, $sl(n, \mathbb{R})$, and $sl(n, \mathbb{C})$, respectively.³²

³¹Here is a wonderful definition: A Lie algebra is called *simple* if it has no *ideals*. See Section 8.9. A Lie algebra can have one or more ideals. It is called *semisimple* if none of the ideals are Abelian. It can be shown that a semisimple Lie algebra is the *direct sum* of simple Lie algebras. (For the purposes of this definition, these simple Lie algebras must have dimension greater than one.) By direct sum it is meant that linear combinations can be formed of the elements in the various Lie algebras, but the Lie products of elements in different Lie algebras are defined to be zero. For example, $su(2)$ is simple. And, because $so(4) = su(2) \oplus su(2)$, $so(4)$ is semisimple.

³²It is customary for *special* to be denoted by the symbols S or s where special means having determinant

We have already learned in Section 3.6 about the orthogonal group and its connected subgroups. The groups $SO(n, \mathbb{R})$ and $SO(n, \mathbb{C})$ are Lie groups. Their associated Lie algebras are denoted by the symbols $so(n, \mathbb{R})$ and $so(n, \mathbb{C})$.

An $n \times n$ matrix U that satisfies the condition

$$U^\dagger U = I \quad (3.7.58)$$

is called *unitary*. The set of all such matrices forms a group, called the unitary group, and is denoted by the symbol $U(n)$. (Here the field is naturally taken to be the complex field.) Next consider the subset of all $n \times n$ unitary matrices having determinant +1. This subset also forms a group [a subgroup of $U(n)$] called the *special* unitary group (or sometimes the unitary *unimodular* group), and is denoted by the symbols $SU(n)$. The groups $U(n)$ and $SU(n)$ are Lie groups. Their associated Lie algebras are denoted by the symbols $u(n)$ and $su(n)$, respectively.

The groups $SU(n)$, $Sp(2n)$, $SO(n)$ and their related Lie algebras $su(n)$, $sp(2n)$, $so(n)$ have been studied extensively. In the mathematics literature they are referred to as the *classical groups* and are given the symbols A_ℓ , B_ℓ , C_ℓ , D_ℓ .³³ To facilitate entrée to this literature, Table 7.2 below summarizes the notation and a few key properties for these groups.³⁴ [Contrary to what might be expected, the groups/algebras $so(2\ell+1)$ and $so(2\ell)$ have different structures, and hence are given the different symbols B_ℓ and D_ℓ .] These groups/algebras form infinite families since they exist for each integer value of $\ell = 1, 2, \dots$. Here, as in the Table, the subscript denotes the *rank* ℓ of the Lie algebra. The concept of rank is defined in Sections 5.8 and 27.4. It is the dimension of the so called *Cartan* subalgebra of the full Lie algebra, which is a particular subalgebra having ℓ mutually commuting elements

By their definitions, the classical Lie algebras/groups can be realized in terms of certain matrices. These realizations are called the *fundamental* or *defining representations*. What is meant by a *representation* in this context is described in Subsection 7.7. (See also Exercise 7.36.) For a given classical Lie algebra, the dimension of the vector space on which the matrices for the fundamental representation act is given within the parentheses associated with its name. For example, the fundamental representation of $sp(2\ell)$ employs $2\ell \times 2\ell$ matrices.

In addition there are a finite number, namely 5, *exceptional groups/algebras* called $G_2(14)$, $F_4(52)$, $E_6(78)$, $E_7(133)$, $E_8(248)$. [We remark that the exceptional Lie algebras are nested as subalgebras according to the relations $G_2(14) \subset F_4(52) \subset E_6(78) \subset E_7(133) \subset E_8(248)$.] Taken together, the classical and exceptional Lie algebras comprise *all* the *simple* Lie algebras.

The naming convention for the exceptional Lie algebras/groups is somewhat different. Here the number within the parentheses associated with the name of such a Lie algebra is its *dimension* and, as done for A_ℓ through D_ℓ , the subscript is its rank ℓ . For example, $E_6(78)$ has dimension 78 and rank 6.

+1; and G or g means *general*, i.e. having determinant possibly $\neq 1$ but nonvanishing. The exceptions to this convention are the notations Sp and sp , where S and s stand for *symplectic*.

³³The term *classical groups* is due to Weyl.

³⁴We note that some authors identify A_m with $sl(m+1, \mathbb{R})$. It can be shown that $su(n)$ and $sl(n, \mathbb{R})$ are equivalent over the complex field. See Exercise 7.29.

The exceptional Lie algebras/groups can also be realized in terms of matrices, and the smallest such matrices for any given exceptional Lie algebra/group provide its fundamental representation. The construction of these matrices is quite difficult and beyond the scope of our discussion. For examples, the fundamental representation of $G_2(14)$ involves 7×7 matrices, and the fundamental representation of $E_8(248)$ involves 248×248 matrices.

Finally, for the classical Lie algebras/groups, there is some redundancy for low values of ℓ . There are the equivalencies $su(2) = so(3) = sp(2)$, $sp(4) = so(5)$, and $su(4) = so(6)$.³⁵ In mathematical notation, these equivalencies are $A_1 = B_1 = C_1$, $B_2 = C_2$, and $A_3 = D_3$. Moreover, $so(2)$ is one dimensional; and $so(4)$ is not simple, but rather is the direct sum of two commuting $su(2)$ algebras: $so(4) = su(2) \oplus su(2)$.³⁶

It has been discovered that all Lie algebras can be constructed by putting together in various ways the simple and the so-called *solvable* and *nilpotent* Lie algebras. The solvable and nilpotent Lie algebras have more or less all been classified. And, as we have just seen, all the simple Lie algebras have been classified. Thus, after over a century of work since the time of Lie, all finite-dimensional Lie algebras and their associated Lie groups are reasonably well classified and their properties reasonably well understood. For our purposes, we are primarily interested in simple Lie algebras and the Lie algebras made out of them.

Table 3.7.2: Cartan Catalog of the Classical and Exceptional Lie Groups/Algebras.

Classical Lie Groups/Algebras, infinite families with an entry for each integer value of ℓ :

<u>Symbol</u>	<u>Lie Algebra</u>	<u>Dimension</u>	<u>Rank</u>
A_ℓ	$su(\ell + 1)$	$\ell(\ell + 2)$	ℓ
B_ℓ	$so(2\ell + 1)$	$\ell(2\ell + 1)$	ℓ
C_ℓ	$sp(2\ell)$	$\ell(2\ell + 1)$	ℓ
D_ℓ	$so(2\ell)$	$\ell(2\ell - 1)$	ℓ

Exceptional Lie Groups/Algebras:

$E_6(78)$, $E_7(133)$, $E_8(248)$, $F_4(52)$, $G_2(14)$

³⁵The equivalence $su(2) = so(3)$ is discussed in Exercise 3.7.31. The equivalence $sp(2) = su(2)$ is treated in Exercise 7.3.24. For the equivalences $sp(4) = so(5)$ and $su(4) = so(6)$ see Chapter 28.

³⁶See Exercises 4.3.19 and 4.3.20.

3.7.7 Adjoint Representation of a Lie Algebra

We close this section with a brief discussion of the subject of *representations* of Lie algebras and, in particular, the *adjoint* representation. Suppose we are given a Lie algebra L . That is, we are told that there are k basis elements (where k is the dimension of L) and we are given a set of structure constants satisfying (7.51) and (7.52). A *representation* of L is a set of $m \times m$ matrices \hat{B}_α that, in analogy to (7.50), obeys the rules

$$\{\hat{B}_\alpha, \hat{B}_\beta\} = \sum_\gamma c_{\alpha\beta}^\gamma \hat{B}_\gamma \quad (3.7.59)$$

where here, to be perfectly explicit, $\{\cdot, \cdot\}$ denotes the matrix commutator,

$$\{\hat{B}_\alpha, \hat{B}_\beta\} = \hat{B}_\alpha \hat{B}_\beta - \hat{B}_\beta \hat{B}_\alpha. \quad (3.7.60)$$

This representation is said to be of *dimension* m since the matrices \hat{B}_α act on an m -dimensional vector space.

At this point some clarifying comments are in order. The first comment concerns definitions. The classical Lie algebras are specified by certain matrix properties associated with their initial specifications. For example, the initially defining matrices for $sp(2n)$ obey the relation (7.29). Upon verifying that these matrices form a Lie algebra, a basis can be chosen and the structure constants associated with this basis can be found. Once the structure constants have been specified, one can search for other sets of matrices which also form a Lie algebra with the same structure constants. However, these other matrices need not satisfy the the matrix properties associated with the initial specification. For example, general representation matrices for $sp(2n)$ need not satisfy (7.29).

The second comment has to do with dimensionality. Note that the dimension m of a representation is not to be confused with the dimension k of the underlying Lie algebra. They may be different.³⁷ However, since the set of $m \times m$ matrices may be viewed as a vector space of dimension m^2 , there must be the relation $k \leq m^2$ if we require that the \hat{B}_α be linearly independent.³⁸

As already described, by their specification, the Classical Lie algebras $su(n)$, $so(n)$, and $sp(2n)$ have natural matrix representations which are called the fundamental or defining representations. We also remarked that the Exceptional Lie algebras have fundamental matrix representations, but that their construction is complicated. The existence of a matrix representation for an arbitrary Lie algebra is even less obvious. The purpose of the present discussion is to observe that every Lie algebra L has a matrix representation, called the *adjoint* representation, which is constructed from the structure constants. This construction turns out to be quite elementary. [According to a much more difficult theorem of *Ado*, which is far beyond the scope of our discussion, every Lie algebra over the complex field is *isomorphic* to some matrix Lie algebra. That is, every (finite-dimensional) abstract Lie

³⁷We also note that the word *representation* can have different meanings depending on context. Here, and in Section 5.8.5 and Chapter 27 and perhaps elsewhere, it means a set of matrices having some desired commutation rules. Another possibility, as in Sections 3.8 and 3.12, is that it may mean that some matrix may be written (represented) in a useful way as some function of some other matrix or matrices.

³⁸For example, the fundamental representation of $sp(2)$ involves 2×2 matrices, and the dimension of $sp(2)$ is 3.

algebra may be viewed as (is isomorphic to) a subalgebra of some $gl(n, \mathbb{C})$. However, it is not the case that every finite-dimensional Lie group is isomorphic to a subgroup of some $GL(n, \mathbb{C})$. The metaplectic group is a counter example.]

After some trial and error in the search for a representation, we hit upon the matrices \hat{B}_α defined in terms of the structure constants by the rules

$$(\hat{B}_\alpha)_{\mu\nu} = c_{\alpha\nu}^\mu. \quad (3.7.61)$$

Note that these matrices are $k \times k$ where k is the dimension of L . (They are also *real* if the structure constants are real.) So, in this case, we have $m = k$.³⁹ Let us verify that the prescription (7.61) works. Using (7.60) and the rules for matrix multiplication, we write

$$\{\hat{B}_\alpha, \hat{B}_\beta\}_{\mu\nu} = (\hat{B}_\alpha \hat{B}_\beta)_{\mu\nu} - (\hat{B}_\beta \hat{B}_\alpha)_{\mu\nu} = \sum_{\mu'} (\hat{B}_\alpha)_{\mu\mu'} (\hat{B}_\beta)_{\mu'\nu} - (\hat{B}_\beta)_{\mu\mu'} (\hat{B}_\alpha)_{\mu'\nu}. \quad (3.7.62)$$

Inserting the definition (7.61) into (7.62) gives the result

$$\begin{aligned} \{\hat{B}_\alpha, \hat{B}_\beta\}_{\mu\nu} &= \sum_{\mu'} (c_{\alpha\mu'}^\mu c_{\beta\nu}^{\mu'} - c_{\beta\mu'}^\mu c_{\alpha\nu}^{\mu'}) \\ &= \sum_{\mu'} (c_{\beta\nu}^{\mu'} c_{\alpha\mu'}^\mu - c_{\alpha\nu}^{\mu'} c_{\beta\mu'}^\mu) = \sum_{\mu'} (c_{\beta\nu}^{\mu'} c_{\alpha\mu'}^\mu + c_{\nu\alpha}^{\mu'} c_{\beta\mu'}^\mu). \end{aligned} \quad (3.7.63)$$

Here we have also used (7.51). But, from (7.52) with a change of indices, we find the relation

$$\begin{aligned} \sum_{\mu'} (c_{\beta\nu}^{\mu'} c_{\alpha\mu'}^\mu + c_{\nu\alpha}^{\mu'} c_{\beta\mu'}^\mu) &= - \sum_{\mu'} c_{\alpha\beta}^{\mu'} c_{\nu\mu'}^\mu \\ &= \sum_{\mu'} c_{\alpha\beta}^{\mu'} c_{\mu'\nu}^\mu = \sum_\gamma c_{\alpha\beta}^\gamma c_{\gamma\nu}^\mu. \end{aligned} \quad (3.7.64)$$

Here we have again used (7.51). Upon combining (7.63), (7.64), and (7.61) we find the final result

$$\{\hat{B}_\alpha, \hat{B}_\beta\}_{\mu\nu} = \sum_\gamma c_{\alpha\beta}^\gamma c_{\gamma\nu}^\mu = \sum_\gamma c_{\alpha\beta}^\gamma (\hat{B}_\gamma)_{\mu\nu}, \quad (3.7.65)$$

and hence (7.59) is satisfied.

As a concrete example, let us construct the adjoint representation of $sp(2, \mathbb{R})$. To begin, there is the 2×2 representation of $sp(2, \mathbb{R})$ which we have agreed to call the defining or fundamental representation. According to (7.34) the Lie algebra of $sp(2, \mathbb{R})$ in the defining representation consists of 2×2 matrices of the form JS with S real and symmetric. These

³⁹Look at Table 7.2. For the Classical Lie algebras compare the dimension m of the fundamental representation with the dimension k of the Lie algebra. For example, in the case of $su(\ell+1)$, compare $m = \ell+1$ with $k = \ell(\ell+2)$. One finds that $m < k$ save for the cases of $so(2)$ and $so(3)$. That is, with these two exceptions, for the Classical Lie algebras the dimension of the fundamental representation is less than the dimension of the adjoint representation. In the case of $so(2)$, the fundamental representation is two-dimensional, and the Lie algebra is one-dimensional. In the case of $so(3)$, $m = k = 3$, and it turns out that the fundamental representation is the adjoint representation. See Exercise 7.30. It can be shown that $m < k$ for the Exceptional Lie algebras as well save for $E_8(248)$. For $E_8(248)$ the fundamental representation is the adjoint representation.

matrices form a 3-dimensional vector space and therefore $k = 3$. See (7.42) evaluated at $n = 1$. A convenient basis for this vector space is provided by the matrices B_1 , B_2 , and B_3 given by the relations

$$\begin{aligned} B_1 &= (1/2)F \\ &= (1/2) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (1/2)\sigma^1, \end{aligned} \quad (3.7.66)$$

$$\begin{aligned} B_2 &= (1/2)B^0 \\ &= (1/2) \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = (i/2)\sigma^2, \end{aligned} \quad (3.7.67)$$

$$\begin{aligned} B_3 &= (1/2)G \\ &= (1/2) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = (1/2)\sigma^3. \end{aligned} \quad (3.7.68)$$

See Section 5.6 where the matrices B^0 , F , and G are constructed and their commutation rules are derived. (Here we have also referenced the Pauli matrices σ^α . See Exercise 3.7.31. This referencing will be useful later.) Note that B_1 and B_3 are Hermitian, and B_2 is anti-Hermitian.

From the commutation rules (5.6.18) through (5.6.20) and the definitions given by the first parts of (7.66) through (7.68) it follows that the B_α obey the commutation rules

$$\{B_1, B_2\} = -B_3, \quad (3.7.69)$$

$$\{B_2, B_3\} = -B_1, \quad (3.7.70)$$

$$\{B_3, B_1\} = B_2, \quad (3.7.71)$$

which are a variant of the commutation rules for $sp(2, \mathbb{R})$. From these rules we see that the only nonzero structure constants in this case are given by the relations

$$c_{12}^3 = -c_{21}^3 = -1, \quad (3.7.72)$$

$$c_{23}^1 = -c_{32}^1 = -1, \quad (3.7.73)$$

$$c_{31}^2 = -c_{13}^2 = 1. \quad (3.7.74)$$

Correspondingly, according to (7.61), the adjoint representation has the associated elements

$$\hat{B}_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}, \quad (3.7.75)$$

$$\hat{B}_2 = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad (3.7.76)$$

$$\hat{B}_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (3.7.77)$$

The reader should check that the \hat{B}_α do indeed satisfy (7.59), i.e., the “hatted” version of (7.69) through (7.71). Note, in accord with the comments made earlier, these matrices do *not* satisfy the original defining relation (7.29).

How could one have guessed the construction (7.61)? There is a way, which at first may also seem obscure, but which will ultimately prove to be very useful. Suppose A is some element in the Lie algebra L . We know that a Lie algebra is a vector space. We are going to associate with A a *linear* operator, denoted by the symbols $(\text{ad } A)$ and called the *adjoint* of A , that will send L into itself.⁴⁰ The action of this operator on any element C in L is defined by the rule

$$(\text{ad } A)C = [A, C]. \quad (3.7.78)$$

Since both C and $[A, C]$ are in L , the operator $(\text{ad } A)$ does indeed send L into itself. It is also obviously linear because of the left distributive property (7.45) of the Lie product.

What can be said about these operators? First, they too form a linear vector space. To see this, suppose $(\text{ad } B)$ is the operator associated with the element B in L . Then, from the definition (7.78) and the right distributive property (7.44), there is the relation

$$\{\text{ad } (A + B)\}C = [A + B, C] = [A, C] + [B, C] = (\text{ad } A)C + (\text{ad } B)C. \quad (3.7.79)$$

Since C is an arbitrary element in L , we may rewrite this relation in the operator form

$$\text{ad } A + \text{ad } B = \text{ad } (A + B), \quad (3.7.80)$$

and take this result to be the definition of operator addition.

Second, these operators also form a Lie algebra with the Lie product taken to be the commutator. This fact is partly obvious since we know that linear operators may be viewed as matrices and, as seen earlier, matrices do form a linear vector space and the matrix or linear operator commutator does satisfy the requirements for a Lie product. However, we have to verify that the linear operator commutator of two adjoint operators is again the adjoint operator for some element in L . Let us check. We find the results

$$(\text{ad } A)(\text{ad } B)C = (\text{ad } A)[B, C] = [A, [B, C]], \quad (3.7.81)$$

$$(\text{ad } B)(\text{ad } A)C = (\text{ad } B)[A, C] = [B, [A, C]]. \quad (3.7.82)$$

It follows that

$$\{(\text{ad } A), (\text{ad } B)\}C = [A, [B, C]] - [B, [A, C]]. \quad (3.7.83)$$

But, from the Jacobi identity and antisymmetry, we have the result

$$\begin{aligned} [A, [B, C]] - [B, [A, C]] &= [A, [B, C]] + [B, [C, A]] = -[C, [A, B]] \\ &= [[A, B], C] = \text{ad } ([A, B])C. \end{aligned} \quad (3.7.84)$$

⁴⁰Note that in this context the term *adjoint* is not to be confused with the concept of Hermitian conjugate.

[Note that (7.64) is also a result of the Jacobi identity.] Upon combining (7.83) and (7.84) and recalling that C is any element in L , we may write the operator identity

$$\{(\text{ad } A), (\text{ad } B)\} = \text{ad } ([A, B]), \quad (3.7.85)$$

which shows that the adjoint operators do indeed form a Lie algebra.

Moreover, this Lie algebra has the *same* structure constants as L . To see this, consider the operators $(\text{ad } B_\alpha)$ and compute:

$$\{(\text{ad } B_\alpha), (\text{ad } B_\beta)\} = \text{ad } ([B_\alpha, B_\beta]) = \text{ad } \left(\sum_\gamma c_{\alpha\beta}^\gamma B_\gamma \right) = \sum_\gamma c_{\alpha\beta}^\gamma (\text{ad } B_\gamma). \quad (3.7.86)$$

Finally, since the adjoint operators are linear operators, let us compute the matrix elements for their equivalent matrices. Suppose D is an arbitrary element in L . Since the B_α form a basis, D has an expansion of the form

$$D = \sum_\nu d_\nu B_\nu. \quad (3.7.87)$$

Let $(\text{ad } B_\alpha)$ act on D to produce a “transformed” D ,

$$D^{\text{tr}} = (\text{ad } B_\alpha)D. \quad (3.7.88)$$

The transformed element D^{tr} has an expansion of the form

$$D^{\text{tr}} = \sum_\mu d_\mu^{\text{tr}} B_\mu. \quad (3.7.89)$$

How are the components d_μ^{tr} related to the components d_ν ? From (7.87) through (7.89) we have the relations

$$\begin{aligned} \sum_\mu d_\mu^{\text{tr}} B_\mu &= D^{\text{tr}} = (\text{ad } B_\alpha)D = [B_\alpha, D] \\ &= [B_\alpha, \sum_\nu d_\nu B_\nu] = \sum_\nu [B_\alpha, B_\nu] d_\nu = \sum_{\nu\gamma} c_{\alpha\nu}^\gamma d_\nu B_\gamma. \end{aligned} \quad (3.7.90)$$

However, since the B_γ form a basis, the relation (7.90) is equivalent to the matrix relation

$$d_\mu^{\text{tr}} = \sum_\nu c_{\alpha\nu}^\mu d_\nu. \quad (3.7.91)$$

Consequently, (7.88) is logically equivalent to (7.91). Moreover, by using the definition (7.61), the relation (7.91) can be rewritten in the form

$$d_\mu^{\text{tr}} = \sum_\nu (\hat{B}_\alpha)_{\mu\nu} d_\nu. \quad (3.7.92)$$

We see that \hat{B}_α is simply the matrix corresponding to the linear operator $(\text{ad } B_\alpha)$; and the fact that these operators satisfy the commutation rules (7.86) implies that their matrix representatives must do so as well.

There is one last point to be made. Suppose it happens that there is some nonzero element in L , call it A , such that the Lie product of A with any element C in L vanishes,

$$[A, C] = 0 \text{ for all } C \in L. \quad (3.7.93)$$

Then we have the results

$$\text{ad } A = 0, \quad (3.7.94)$$

$$\hat{A} = 0. \quad (3.7.95)$$

Since the B_α form a basis, and A is nonzero, A must have an expansion of the form

$$A = \sum_{\alpha} a_{\alpha} B_{\alpha} \quad (3.7.96)$$

where at least some of the components a_{α} are nonzero. As a consequence of (7.80), (7.94), and (7.95) we find the result

$$\sum_{\alpha} a_{\alpha} \hat{B}_{\alpha} = \hat{A} = 0, \quad (3.7.97)$$

which shows that the \hat{B}_{α} in this case are *linearly dependent*. We see that while (by the definition of a basis) the B_{α} are linearly independent, it can happen that the \hat{B}_{α} are not. Therefore, it may happen that the adjoint representation of L provided by the \hat{B}_{α} is *not* isomorphic to L . If the Lie algebra provided by the matrices of a representation of a Lie algebra L is isomorphic to L , then this representation is said to be *faithful*. We have learned that the adjoint representation need not be faithful.

Exercises

3.7.1. Show that the Euclidean vector norm defined by (7.23) satisfies all the requirements for a vector norm. Show that the Euclidean matrix norm defined by the rule

$$(\|A\|_E)^2 = \text{tr}(A^\dagger A) = \sum_{jk} |A_{jk}|^2 \quad (3.7.98)$$

satisfies all the requirements for a matrix norm. (The Euclidean matrix norm is also sometimes called the *Frobenius* norm.) Note that the Euclidean vector and matrix norms are analogous in that both involve a sum of absolute values squared. Show that the Euclidean vector and matrix norms are consistent.

The Euclidean matrix norm is easy to compute, but is weaker than the maximum column sum norm, which is also easy to compute. Show that, for example in the $m \times m$ case,

$$\|I\| = \sqrt{m} \quad (3.7.99)$$

for the Euclidean norm, while

$$\|I\| = 1 \quad (3.7.100)$$

for the maximum column sum and spectral norms.

Show that

$$\|J\| = \sqrt{2n} \quad (3.7.101)$$

for the Euclidean norm, while

$$\|J\| = 1 \quad (3.7.102)$$

for the maximum column sum and spectral norms.

Suppose u and v are any two real vectors, and let (u, v) denote the usual real Euclidean inner product,

$$(u, v) = \sum_j u_j v_j. \quad (3.7.103)$$

Verify the *Schwarz inequality*

$$|(u, v)| \leq \|u\| \|v\| \quad (3.7.104)$$

where here the vector norm on the right side of (7.104) is the real Euclidean norm. Verify an analogous result for complex vectors when the Euclidean complex inner product is used, in which case

$$\langle u, v \rangle = \sum_j \bar{u}_j v_j = (\bar{u}, v). \quad (3.7.105)$$

Suppose u and v are two real $2n$ -dimensional vectors that are symplectically conjugate in the sense that

$$(u, Jv) = \pm 1. \quad (3.7.106)$$

Verify the chain of reasoning

$$1 = |(u, Jv)| \leq \|u\| \|Jv\| \leq \|u\| \|J\| \|v\| \leq \|u\| \|v\| \quad (3.7.107)$$

where here the matrix spectral norm (which is consistent with the Euclidean vector norm) has been used for $\|J\|$. Thus, (7.106) implies the inequality

$$\|u\| \|v\| \geq 1. \quad (3.7.108)$$

3.7.2. The Schwarz *inequality* (7.104) is a relation between the absolute value of an inner product and the norms of its ingredients. The *Lagrange identity* is an *equality* that reveals what terms have been omitted to make the Schwarz inequality a true inequality. Suppose $u = (u_1, u_2, \dots, u_n)$ and $v = (v_1, v_2, \dots, v_n)$ are any two n -component vectors, real or complex. Then, according to the Lagrange identity, they satisfy the relation

$$\left(\sum_j u_j^2 \right) \left(\sum_k v_k^2 \right) - \left(\sum_j u_j v_j \right)^2 = (1/2) \sum_{jk} (u_j v_k - u_k v_j)^2. \quad (3.7.109)$$

The relation (7.109) may also be written in the form

$$\left(\sum_j u_j v_j \right)^2 = \left(\sum_j u_j^2 \right) \left(\sum_k v_k^2 \right) - (1/2) \sum_{jk} (u_j v_k - u_k v_j)^2. \quad (3.7.110)$$

If the entries in u and v are real, then the last term on the right side of (7.110) can never be positive. In that case there is the inequality

$$\left(\sum_j u_j v_j \right)^2 \leq \left(\sum_j u_j^2 \right) \left(\sum_k v_k^2 \right), \quad (3.7.111)$$

which can be written in the more compact form

$$(u, v)^2 \leq \|u\|^2 \|v\|^2. \quad (3.7.112)$$

Evidently, the relations (7.104) and (7.112) are equivalent.

Suppose \mathbf{u} and \mathbf{v} are two real 3-component vectors. For this case, verify the identity

$$(\mathbf{u} \cdot \mathbf{v})^2 + (\mathbf{u} \times \mathbf{v}) \cdot (\mathbf{u} \times \mathbf{v}) = (\mathbf{u} \cdot \mathbf{u})(\mathbf{v} \cdot \mathbf{v}), \quad (3.7.113)$$

and show that this identity is the Lagrange identity for the instance $n = 3$. How can the Lagrange identity be verified for the case of general n ? Here is one way: Define the matrix A by the rule

$$A = |u|(v| - |v)(u| \quad (3.7.114)$$

where, in the formation of dyads, complex conjugation is *not* to be employed. Show that, by this definition, A is antisymmetric and has the matrix elements

$$A_{jk} = (e^j, Ae^k) = u_j v_k - u_k v_j. \quad (3.7.115)$$

Show, for any antisymmetric matrix A , that

$$\text{tr}(A^2) = -\text{tr}(A^T A) = -\sum_{jk} (A_{jk})^2. \quad (3.7.116)$$

Verify the dyadic relation

$$\begin{aligned} A^2 &= [|u](v| - |v)(u|) [|u](v| - |v)(u|] \\ &= |u)(v, u)(v| - |u)(v, v)(u| - |v)(u, u)(v| + |v)(u, v)(u|. \end{aligned} \quad (3.7.117)$$

Use this dyadic result to show that

$$\begin{aligned} \text{tr}(A^2) &= (v, u)^2 - (u, u)(v, v) - (u, u)(v, v) + (u, v)^2 \\ &= 2(u, v)^2 - 2(u, u)(v, v). \end{aligned} \quad (3.7.118)$$

By comparing (7.116) and (7.118), show that

$$(u, v)^2 - (u, u)(v, v) = -(1/2) \sum_{jk} (A_{jk})^2. \quad (3.7.119)$$

Verify that (7.110) and (7.119) agree.

Suppose that the usual complex inner product (7.105) is of interest rather than the usual real inner product (7.103). In this case there is the associated Lagrange identity

$$\left(\sum_j |u_j|^2 \right) \left(\sum_k |v_k|^2 \right) - \left| \sum_j \bar{u}_j v_j \right|^2 = (1/2) \sum_{jk} |u_j v_k - u_k v_j|^2. \quad (3.7.120)$$

The relation (7.120) can also be written in the form

$$\left| \sum_j \bar{u}_j v_j \right|^2 = \left(\sum_j |u_j|^2 \right) \left(\sum_k |v_k|^2 \right) - (1/2) \sum_{jk} |u_j v_k - u_k v_j|^2. \quad (3.7.121)$$

Prove this result as follows: Define A exactly as before using (7.114). Show that

$$\text{tr}(A\bar{A}) = -\text{tr}(AA^\dagger) = -\sum_{jk} |A_{jk}|^2. \quad (3.7.122)$$

Verify the dyadic relation

$$\begin{aligned} A\bar{A} &= [|u](v| - |v)(u|) [|\bar{u})(\bar{v}| - |\bar{v})(\bar{u}|] \\ &= |u)(v, \bar{u})(\bar{v}| - |u)(v, \bar{v})(\bar{u}| - |v)(u, \bar{u})(\bar{v}| + |v)(u, \bar{v})(\bar{u}|. \end{aligned} \quad (3.7.123)$$

Use this dyadic result to show that

$$\begin{aligned} \text{tr}(A\bar{A}) &= (\bar{v}, u)(v, \bar{u}) - (\bar{u}, u)(v, \bar{v}) - (u, \bar{u})(\bar{v}, v) + (u, \bar{v})(\bar{u}, v) \\ &= 2|(\bar{u}, v)|^2 - 2(\bar{u}, u)(\bar{v}, v). \end{aligned} \quad (3.7.124)$$

Compare (7.122) and (7.124) to show that

$$|(\bar{u}, v)|^2 - (\bar{u}, u)(\bar{v}, v) = -(1/2) \sum_{jk} |A_{jk}|^2. \quad (3.7.125)$$

Verify that (7.121) and (7.125) agree.

3.7.3. Show that the maximum column sum matrix norm defined by (7.19) satisfies the relation

$$|B_{jk}| \leq \|B\|. \quad (3.7.126)$$

Show that the series (7.1) converges for any matrix B . (Hint: Show that the set of partial sums forms a Cauchy sequence.) Consider the matrix function $F(s)$ defined by the equation

$$F(s) = \exp(sB). \quad (3.7.127)$$

Show, by term-by-term differentiation of the power series for $F(s)$, that $F(s)$ satisfies the differential equation

$$dF(s)/ds = BF(s) = F(s)B \quad (3.7.128)$$

with the initial condition

$$F(0) = I. \quad (3.7.129)$$

Justify the required interchange of the operations of (infinite) summation and differentiation.

3.7.4. Show that the series (7.2) converges for A sufficiently near the identity matrix I . Note that when A is near the identity, then $\log(A)$ is near the zero matrix. Thus, any matrix A sufficiently near the identity has a generator, namely $\log(A)$. Moreover, from the work of Section 3.7.1, we know that such an A is infinitesimally generated.

3.7.5. Suppose that B_i and B_j are any two $m \times m$ matrices that commute,

$$\{B_i, B_j\} = 0. \quad (3.7.130)$$

Equivalently, we may say that the Lie products of B_i and B_j vanish. Show from the power series definition (7.1) that in this case there is the relation

$$\exp(s_i B_i) \exp(s_j B_j) = \exp(s_i B_i + s_j B_j), \quad (3.7.131)$$

where s_i and s_j are any scalars. Verify (7.7) and (7.10). Suppose there are k linearly independent elements B_1, B_2, \dots, B_k all of which mutually commute (Lie products mutually vanish) as in (7.130). Show that these elements span a Lie algebra L . Show that all elements in L commute. A Lie algebra with this property is called *Abelian*. Consider all elements $G(s_1, \dots, s_k)$ of the form

$$G(s_1, \dots, s_k) = \exp(s_1 B_1 + s_2 B_2 + \dots + s_k B_k).$$

Show that these elements form a group with the property

$$G(0, \dots, 0) = I$$

and the group multiplication rule

$$G(s_1, \dots, s_k)G(t_1, \dots, t_k) = G(s_1 + t_1, \dots, s_k + t_k).$$

Show that all elements in this group commute with respect to group multiplication,

$$G(s_1, \dots, s_k)G(t_1, \dots, t_k) = G(t_1, \dots, t_k)G(s_1, \dots, s_k).$$

A group for which all elements commute is also called Abelian. We have shown, in the context of matrices, that exponentiating an Abelian Lie algebra produces an Abelian Lie group. Conversely, again in the matrix context, the Lie algebra of an Abelian Lie group is Abelian. The same can be shown to be true for all Lie algebras and their related Lie groups.

Finally, as a special case, consider all elements of the form

$$G(s) = \exp(sB)$$

where B is some matrix. They evidently form a one-parameter Abelian Lie subgroup of $GL(m)$.

3.7.6. Verify the relations (7.24) through (7.26).

3.7.7. Verify the relations given by (7.3) through (7.6) using the definitions (7.1) and (7.2).

3.7.8. Verify (7.29) using the expansions (7.28) and the symplectic condition (1.2).

3.7.9. The calculation leading from (7.30) to (7.32) involved interchanges of the operations of matrix multiplication and transposition, and the operation of summation. Verify that these interchanges do not affect the convergence of the infinite series involved.

3.7.10. Consider two matrices A and B related by (7.4). The purpose of this exercise is to show that the determinant of A is related to the trace of B . We will do so by setting up and solving a differential equation.

Suppose that ϵ is a small parameter and B is an arbitrary matrix. Verify the expansion

$$\det(I + \epsilon B) = 1 + \epsilon \operatorname{tr}(B) + O(\epsilon^2). \quad (3.7.132)$$

Let $f(\lambda)$ be the function

$$f(\lambda) = \det[\exp(\lambda B)].$$

Verify the expansion

$$\begin{aligned} f(\lambda + d\lambda) &= \det\{\exp[(\lambda + d\lambda)B]\} \\ &= \det[\exp(\lambda B) \exp(d\lambda B)] \\ &= \det[\exp(\lambda B)] \det[\exp(d\lambda B)] \\ &= f(\lambda) \det\{1 + d\lambda B + O[(d\lambda)^2]\} \\ &= f(\lambda)\{1 + d\lambda \operatorname{tr}(B) + O[(d\lambda)^2]\}. \end{aligned}$$

Show that $f(\lambda)$ obeys the differential equation

$$df/d\lambda = f(\lambda) \operatorname{tr}(B)$$

with the initial condition

$$f(0) = 1.$$

Show that this differential equation has the unique solution

$$f(\lambda) = \exp[\lambda \operatorname{tr}(B)].$$

Consequently show that if A and B are any two matrices related by (7.4), then

$$\det(A) = \det[\exp(B)] = f(1) = \exp[\operatorname{tr}(B)]. \quad (3.7.133)$$

The relation (7.133), sometimes also called *Liouville's formula*, is useful and memorable. Note that, in view of the differential equation obeyed by f , this formula is a special case of the Liouville-Ostrogradski formula derived in Exercise 1.4.6.

As an application, first verify that it is always possible to find a matrix S such that (7.34) is true. Next verify that the matrix JS is traceless if S is symmetric. Finally, show that any matrix of the form $\exp(JS)$ must have determinant +1.

3.7.11. Verify the details of the calculation described in (7.37) and (7.38) using the series definition of the exponential function as given by (7.1).

3.7.12. This exercise presents two challenges:

- Show that the matrix given by the relation

$$M = \begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix} \quad (3.7.134)$$

is symplectic, but cannot be written in the form (7.36).

- What happens if the -1 in the upper right corner of (7.134) is replaced by $+1$? See the matrix N below. Can the resulting symplectic matrix be written in the form (7.36)?

As preparatory observations, verify the symplectic conjugacy relation (see Exercises 5.7 and 8.13)

$$N = \begin{pmatrix} -1 & +1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}. \quad (3.7.135)$$

Note that the conjugating matrix A in this case [see (5.50)] is symplectic, but complex. Next, suppose M and N are *any* two $2n \times 2n$ symplectic matrices that are symplectically conjugate. Suppose also that M can be written in the form (7.36). Show that the same must then also be true for N .

Hint for meeting the challenges: Take them in opposite order. Now let N again be the matrix on the left side of (7.135). First prove that that N cannot be diagonalized by a similarity transformation. Next, suppose N is written in exponential form,

$$N = \exp(E). \quad (3.7.136)$$

This is possible because N is invertible. Verify this claim! The matrix E also cannot be diagonalized by a similarity transformation, because if it could be so diagonalized, then so could N . Therefore both eigenvalues of E must be identical. If we *assume* that E is of the form JS , then E must be traceless, and these eigenvalues must both be zero. Consequently, there is a similarity transformation B that brings E to the Jordan form,

$$BEB^{-1} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \quad (3.7.137)$$

It follows that N can be written in the form

$$\begin{aligned} N = \exp(E) &= \exp \left[B^{-1} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} B \right] = B^{-1} \left[\exp \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \right] B \\ &= B^{-1} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} B. \end{aligned} \quad (3.7.138)$$

But (7.138) is absurd because the eigenvalues of N are both -1 whereas the eigenvalues of the matrix on the right side of (7.138) are both $+1$. [Look ahead to Exercise 7.16. Compute the characteristic polynomial of the matrix product on the right side of (7.138).] Conclude that our assumption has produced a contradiction, and therefore neither N nor M can be written in the exponential form (7.36).

Show that the matrix $L = -M$, with M given by (7.134), is also symplectic, and can be written in the form (7.36). Find S for this case. See Exercise 5.6.7.

3.7.13. Verify that the set of matrices of the form JS (that is, the set of all Hamiltonian matrices) is indeed a Lie algebra by showing that properties i through iii are satisfied.

3.7.14. Let $B = JS$ be a *real* $2n \times 2n$ Hamiltonian matrix. Section 3.4 described the eigenvalue spectrum of real symplectic matrices. Derive related results for the eigenvalue spectrum of real Hamiltonian matrices. See (7.33). If $P(\lambda) = \det(JS - \lambda I)$ is the characteristic polynomial of a real Hamiltonian matrix, show that all its coefficients are real, and hence $\bar{P}(\lambda) = P(\bar{\lambda})$. Show that $P(\lambda) = P(-\lambda)$ and hence P contains only even powers of λ . Hint: Verify the chain of relations

$$\begin{aligned}\det(JS - \lambda I) &= \det[(JS - \lambda I)^T] = \det(-SJ - \lambda I) \\ &= \det[J(-SJ - \lambda I)J^{-1}] = \det(-JS - \lambda I) \\ &= \det[(-I)(JS + \lambda I)] = \det(-I)\det(JS + \lambda I) \\ &= \det(JS + \lambda I).\end{aligned}$$

Here we have used that fact that, because $-I$ is $2n \times 2n$, $\det(-I) = 1$. Show that if λ is an eigenvalue, so are $\bar{\lambda}$ and $-\lambda$. Show that if $\lambda = 0$ is an eigenvalue, it must have even multiplicity. Thus, if α is a real eigenvalue, $-\alpha$ must also be an eigenvalue, and real eigenvalues must come in $\pm\alpha$ pairs. Similarly, if $i\beta$ is a pure imaginary eigenvalue, $-i\beta$ must also be an eigenvalue, and pure imaginary eigenvalues must come in $\pm i\beta$ pairs. Finally, if $\alpha + i\beta$ is a complex eigenvalue, there must be a quartet of complex eigenvalues $\pm\alpha \pm i\beta$ with all signs taken independently. Section 3.4.4 showed that the problem of finding the eigenvalues of a $2n \times 2n$ symplectic matrix can be simplified to that of finding the roots of a polynomial of degree n followed by the solution of a quadratic equation. Show that the same is true for a Hamiltonian matrix. Show that the eigenvalues can be found in terms of radicals for the cases $n \leq 4$.

Finally we remark that if S is positive (or negative) definite, then JS is what we call a “special” real Hamiltonian matrix, and *must* have purely imaginary eigenvalues coming in complex conjugate pairs. There are no other possibilities. Moreover, Hamiltonian matrices of this special kind can always be diagonalized by a real symplectic similarity transformation. See Section 34.6.4.

3.7.15. Let B be a Hamiltonian matrix. See (7.33). Show that B obeys the relation

$$KB = -B^T K \tag{3.7.139}$$

with K given by (5.3). Using the angular inner product (5.2), study the eigenvector structure of real Hamiltonian matrices in a manner similar to that done for symplectic matrices in Section 3.5. You will need the eigenvalue spectrum results of Exercise 7.14.

3.7.16. The *characteristic polynomial* $P(\lambda)$ of a matrix A is defined by the equation

$$P(\lambda) = \det(A - \lambda I). \tag{3.7.140}$$

The solutions of the equation $P(\lambda) = 0$ are the eigenvalues of A . Show that the matrices A and $A' = SAS^{-1}$, where S is any invertible matrix, have the same characteristic polynomial and hence the same eigenvalues. You have verified that the set of eigenvalues is invariant under *similarity* transformations.

3.7.17. Suppose A is $m \times m$ and let $P(\lambda)$ be its characteristic polynomial (7.140).

- a) Verify that $P(\lambda)$ has the expansion

$$P(\lambda) = \sum_{\ell=0}^m a_\ell(-\lambda)^\ell \quad (3.7.141)$$

with

$$a_0 = \det(A) \quad (3.7.142)$$

and

$$a_m = 1. \quad (3.7.143)$$

What can be said about the other a_ℓ ? Using the results (7.1) through (7.4) and the result of Exercise (7.7), we may write the relations

$$\begin{aligned} P(\lambda) &= \det(A - \lambda I) = \det[(-\lambda I)(I - A/\lambda)] \\ &= (-\lambda)^m \det(I - A/\lambda) \\ &= (-\lambda)^m \det\{\exp[\log(I - A/\lambda)]\} \\ &= (-\lambda)^m \det\{\exp[-\sum_{\ell=1}^{\infty} (A/\lambda)^\ell / \ell]\} \\ &= (-\lambda)^m \exp[-\sum_{\ell=1}^{\infty} (1/\lambda)^\ell (1/\ell) \text{tr}(A^\ell)]. \end{aligned} \quad (3.7.144)$$

- b) Verify the statement made above. Note that the series employed will certainly converge for λ large enough, because $\|A/\lambda\|$ is then small.
c) Now expand out the exponential function, and collect powers of λ to get an expression of the form

$$P(\lambda) = \sum_{\ell=-\infty}^m a_\ell(-\lambda)^\ell. \quad (3.7.145)$$

It follows from (7.141) that $a_\ell = 0$ for $\ell < 0$. Show that this fact gives an infinite collection of identities. Find the first few coefficients a_{m-1}, a_{m-2}, \dots and verify the results

$$a_{m-1} = \text{tr}(A), \quad (3.7.146)$$

$$a_{m-2} = \{[\text{tr}(A)]^2 - \text{tr}(A^2)\}/2, \quad (3.7.147)$$

$$a_{m-3} = (1/3) \text{tr}(A^3) - (1/2)[\text{tr}(A)][\text{tr}(A^2)] + (1/6)[\text{tr}(A)]^3, \quad (3.7.148)$$

$$\begin{aligned} a_{m-4} &= (1/24)\{[\text{tr}(A)]^4 - 6[\text{tr}(A)]^2[\text{tr}(A^2)] + 3[\text{tr}(A^2)]^2 \\ &\quad + 8[\text{tr}(A)][\text{tr}(A^3)] - 6[\text{tr}(A^4)]\}. \end{aligned} \quad (3.7.149)$$

Show that all the a_ℓ are functions of $[\text{tr}(A^j)]^k$ for various values of j and k . Verify (4.25) and (4.26). Make a similar study of $[1/P(\lambda)]$.

d) Show that $\det(A)$ is also expressible in terms of $[\text{tr}(A^j)]^k$. Verify the results

$$\det(A) = \{[\text{tr}(A)]^2 - \text{tr}(A^2)\}/2 \text{ when } m = 2, \quad (3.7.150)$$

$$\det(A) = [\text{tr}(A^3)]/3 - [\text{tr}(A)][\text{tr}(A^2)]/2 + [\text{tr}(A)]^3/6 \text{ when } m = 3, \quad (3.7.151)$$

$$\begin{aligned} \det(A) &= (1/24)\{[\text{tr}(A)]^4 - 6[\text{tr}(A)]^2[\text{tr}(A^2)] + 3[\text{tr}(A^2)]^2 \\ &\quad + 8[\text{tr}(A)][\text{tr}(A^3)] - 6[\text{tr}(A^4)]\} \text{ when } m = 4, \text{ etc.} \end{aligned} \quad (3.7.152)$$

e) As in Exercise 7.10, let ϵ be a small parameter and C an arbitrary matrix. Verify the results

$$(I + \epsilon C) = \exp[\log(I + \epsilon C)] = \exp[-\sum_{n=1}^{\infty} (-\epsilon C)^n/n], \quad (3.7.153)$$

$$\begin{aligned} \det(I + \epsilon C) &= \exp[-\sum_{n=1}^{\infty} (1/n)(-\epsilon)^n \text{tr}(C^n)] \\ &= 1 + \epsilon \text{tr}(C) + (\epsilon^2/2)\{[\text{tr}(C)]^2 - \text{tr}(C^2)\} + \dots \end{aligned} \quad (3.7.154)$$

3.7.18. Let A and B be any two $m \times m$ matrices. Define matrices C and D by the equations $C = AB, D = BA$. Show that C and D have the same eigenvalue spectrum. Hint: Use Exercise (7.17) to show that they have the same characteristic polynomial.

3.7.19. Given an $n \times n$ matrix A , equation (7.140) gives a polynomial $P(\lambda)$. Show that $P(\lambda)$ has the leading term $(-\lambda)^n$. Consider the inverse problem: given a polynomial $P(\lambda)$ with leading term $(-\lambda)^n$, can one find a matrix A such that $P(\lambda)$ is the characteristic polynomial for A ? Suppose the roots of $P(\lambda)$ are known. Call them λ_j . Find a *diagonal* A such that (7.140) holds. Remarkably, one does not need to know the roots to find an A that works. Show that the matrix A , called the *companion matrix* for $P(\lambda)$, given by

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & & & & 1 \\ b_1 & b_2 & \cdots & & b_n \end{pmatrix} \quad (3.7.155)$$

has the characteristic polynomial

$$P(\lambda) = (-1)^n(\lambda^n - b_n\lambda^{n-1} - b_{n-1}\lambda^{n-2} - \cdots - b_1). \quad (3.7.156)$$

Show that this A may not always be diagonalizable. Hint: Study the 2×2 case. Show that a general eigenvector is of the form $(1, \lambda)^T$, and show that there is only one (linearly independent) eigenvector if the eigenvalues are degenerate. Generalize to the $n \times n$ case and show that there are as many linearly independent eigenvectors as there are distinct eigenvalues.

3.7.20. Verify that the cross-product algebra given by (7.47) is not associative. Verify that it is a Lie algebra. In particular, check the Jacobi condition.

There is a theorem in plane Euclidean geometry to the effect that the three altitudes of a triangle intersect in a common point (called the *orthocenter*). (This point is in the interior of the triangle if the triangle is *acute*. Recall that a triangle is called acute if all its angles are less than 90 degrees.) It can be shown that the existence of a common intersection is a consequence of the Jacobi identity for the cross-product Lie algebra. Google “Jacobi identity altitudes of a triangle”.

3.7.21. Verify that the set of all $m \times m$ matrices with the multiplication rule defined by $[A, B] = AB - BA$ forms a Lie algebra. In particular, check the Jacobi condition.

3.7.22. Given any algebra, define the *associator* $A(x, y, z)$ of any 3 elements x, y, z by the rule $A(x, y, z) = (x \circ y) \circ z - x \circ (y \circ z)$. An algebra is called associative if the associator vanishes. Show that a Lie algebra is generally not associative. Hint: Use the Jacobi condition (and antisymmetry) to compute the associator. From this perspective, the Jacobi condition (along with antisymmetry) may be viewed as a rule that specifies the associator.

3.7.23. Suppose the vectors $\mathbf{e}_1 = \mathbf{e}_x$, $\mathbf{e}_2 = \mathbf{e}_y$, and $\mathbf{e}_3 = \mathbf{e}_z$ form a right-handed orthonormal triad in three-dimensional Euclidean space. Use them to form a basis for the Lie algebra (7.47). Find the structure constants $c_{\alpha\beta}^{\gamma}$ for this Lie algebra and basis. Show that these structure constants are related to the *Levi-Civita* tensor $\epsilon_{\alpha\beta\gamma}$. Consider a complex “spherical” basis $\mathbf{e}_{-1}, \mathbf{e}_0, \mathbf{e}_{+1}$ defined by the relations

$$\mathbf{e}_{+1} = -(1/\sqrt{2})(\mathbf{e}_x + i\mathbf{e}_y), \quad (3.7.157)$$

$$\mathbf{e}_0 = i\mathbf{e}_z,$$

$$\mathbf{e}_{-1} = (1/\sqrt{2})(\mathbf{e}_x - i\mathbf{e}_y).$$

Find the structure constants for this choice of basis.

3.7.24. Verify the relations (7.51), (7.52), (7.56), and (7.57). Verify that if (7.51) and (7.52) hold for some basis set, then the Lie algebraic properties (7.48) and (7.49) (antisymmetry and Jacobi condition) are satisfied.

3.7.25. Classify all two- and three-dimensional Lie algebras.

3.7.26. Verify that $GL(n, \mathbb{R})$, $GL(n, \mathbb{C})$, $SL(n, \mathbb{R})$, and $SL(n, \mathbb{C})$ are indeed groups. Characterize the Lie algebras $gl(n, \mathbb{R})$, $gl(n, \mathbb{C})$, $sl(n, \mathbb{R})$, and $sl(n, \mathbb{C})$. That is, what properties are satisfied by such matrices? Find the dimensions of these Lie algebras. In the complex case, find the dimension over both the real and complex fields. [Hint: Use the relation (7.133).]

3.7.27. Verify that $O(n, \mathbb{R})$, $O(n, \mathbb{C})$, $SO(n, \mathbb{R})$, and $SO(n, \mathbb{C})$ are indeed groups. Characterize the Lie algebras $so(n, \mathbb{R})$ and $so(n, \mathbb{C})$. That is, what properties are satisfied by such matrices? Find the dimensions of these Lie algebras. In the complex case, find the dimension over both the real and complex fields. [Hint: Use the relation (7.133).]

3.7.28. Show from (7.58) that $|\det(U)| = 1$. Verify that $U(n)$ and $SU(n)$ are indeed groups. Show that the set of $n \times n$ anti-Hermitian matrices forms a Lie algebra. This set is $u(n)$, the Lie algebra of $U(n)$. Find its dimension. Show that the set of all traceless matrices in $u(n)$ forms a sub Lie algebra. This sub Lie algebra is called $su(n)$. Find its dimension. Show that $su(n)$ is the Lie algebra of the group $SU(n)$. [Hint: Use the relation (7.133).]

3.7.29. By construction, because only real matrices are involved in their definitions, there are basis choices for the Lie algebras $so(n, \mathbb{R})$ and $sp(2n, \mathbb{R})$ for which the structure constants are *real*. What about the case of $su(n)$? Let \mathcal{A} be the set of all real $n \times n$ antisymmetric matrices. Since all the diagonal entries in all antisymmetric matrices vanish, every antisymmetric matrix is traceless. Show that the dimension of \mathcal{A} is $(n^2 - n)/2$. Let \mathcal{S} be the set of all real $n \times n$ symmetric and traceless matrices. Show that the dimension of \mathcal{S} is $[(n^2 - n)/2 + n - 1]$. Let the matrices A_j and S_k form bases for the sets \mathcal{A} and \mathcal{S} , respectively. Show that the matrices A_j and iS_k form a basis for $su(n)$. See Exercise 7.28. Verify that the commutator of any two matrices is traceless. Verify that the commutator of any two antisymmetric matrices is antisymmetric. Verify that the commutator of an antisymmetric matrix and a symmetric matrix is symmetric. Verify that the commutator of any two symmetric matrices is antisymmetric. We can write these relations symbolically in the form

$$\{A, A'\} \propto A'', \quad (3.7.158)$$

$$\{A, S\} \propto S', \quad (3.7.159)$$

$$\{S, S'\} \propto A. \quad (3.7.160)$$

Correspondingly, verify that there are also the relations

$$\{A, A'\} \propto A'', \quad (3.7.161)$$

$$\{A, (iS)\} \propto (iS'), \quad (3.7.162)$$

$$\{(iS), (iS')\} \propto A. \quad (3.7.163)$$

Thus, show that in this basis, which may be viewed as a natural basis for $su(n)$, the structure constants are all real.

Consider next the Lie algebra $sl(n, \mathbb{R})$. Show that the matrices A_j and S_k form a basis for this Lie algebra. See Exercise 7.26. Thus show that $su(n)$ and $sl(n, \mathbb{R})$ have the same dimension. Show, in fact, that $su(n)$ and $sl(n, \mathbb{R})$ are *equivalent* over the complex field.

3.7.30. Verify that the \hat{B}_j given by (7.75) through (7.77) for the $sp(2, \mathbb{R})$ case satisfy commutation rules analogous to (7.69) through (7.71). Verify that in this case the \hat{B}_j are linearly independent.

3.7.31. This exercise explores the relations between the Lie algebras $su(2)$, $so(3, \mathbb{R})$, and the cross-product Lie algebra. It also explores the relations between $SU(2)$ and $SO(3, \mathbb{R})$. It presumes that you have worked, or at least read, Exercises 7.23, 7.27, 7.28, and 7.29. The *Euclidean* group in three dimensions is defined to be the rotation group $SO(3, \mathbb{R})$ augmented by *translations*. It might also be called $ISO(3, \mathbb{R})$, the *inhomogeneous* rotation group in three dimensions. The Euclidean group in three dimensions is studied as part of Chapter 26.

Define the *Pauli* matrices σ^α for $\alpha = 1, 2, 3$ by the rules

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (3.7.164)$$

$$\sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \Leftrightarrow \sigma^2 = -iJ_2 \Leftrightarrow J_2 = i\sigma^2, \quad (3.7.165)$$

$$\sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (3.7.166)$$

[We remark that the Pauli matrices were discovered by Klein some 50 years prior to the time of Pauli. They came to bear Pauli's name because he was the first to use them to describe electron spin. Pauli (1900-1958) died in room 137 $\simeq 1/\alpha$ of the Red-Cross hospital at Zurich.] Verify that the Pauli matrices are traceless, Hermitian,

$$(\sigma^\alpha)^\dagger = \sigma^\alpha, \quad (3.7.167)$$

and satisfy the relations

$$(1/2)\text{tr}(\sigma^\alpha \sigma^\beta) = \delta_{\alpha\beta}. \quad (3.7.168)$$

Verify also that the Pauli matrices are unitary, $(\sigma^\alpha)^\dagger \sigma^\alpha = I$. (For further properties of the Pauli matrices, see Section 5.7 and Exercises 5.7.2 and 5.7.7.)

Let K^1 through K^3 be the traceless anti-Hermitian matrices

$$K^1 = (-i/2)\sigma^1 = (-i/2) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (3.7.169)$$

$$K^2 = (-i/2)\sigma^2 = (-i/2) \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad (3.7.170)$$

$$K^3 = (-i/2)\sigma^3 = (-i/2) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (3.7.171)$$

In Exercise 7.27 you should have found that the Lie algebra $su(n)$ consists of all $n \times n$ traceless anti-Hermitian matrices. Show that K^1 through K^3 form a basis for the Lie algebra $su(2)$. Show that they obey the multiplication and commutation rules

$$K^\alpha K^\beta = (1/2)K^\gamma, \quad (3.7.172)$$

$$\{K^\alpha, K^\beta\} = K^\gamma, \quad (3.7.173)$$

where α, β, γ is any cyclic permutation of 1, 2, 3. Thus, with this choice of basis, the structure constants for $su(2)$ are the components of the Levi-Civita tensor,

$$c_{\alpha\beta}^\gamma = \epsilon_{\alpha\beta\gamma}. \quad (3.7.174)$$

We remark that we have been following what might be called a mathematician's approach to $su(2)$ [and $so(3, \mathbb{R})$] in which basis elements are chosen so that the structure constants are all *real*. For a quantum physicist's approach for which a basis is chosen to make all the structure constants *pure imaginary*, see Exercise 7.43.

Verify also that the K^α obey the anticommutation rules

$$\{K^\alpha, K^\beta\}_+ = K^\alpha K^\beta + K^\beta K^\alpha = -(1/2)\delta_{\alpha\beta}I. \quad (3.7.175)$$

Let \mathbf{a} be a three-component vector with entries (a_1, a_2, a_3) . Introduce the notation

$$\mathbf{a} \cdot \mathbf{K} = \sum_\alpha a_\alpha K^\alpha. \quad (3.7.176)$$

Show that there is the multiplication rule

$$(\mathbf{a} \cdot \mathbf{K})(\mathbf{b} \cdot \mathbf{K}) = -(1/4)(\mathbf{a} \cdot \mathbf{b})I + (1/2)(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{K}. \quad (3.7.177)$$

Compute the matrices $(K^\alpha)^2$. Observe that they are diagonal, and hence mutually commute. Show that they sum to $-(3/4)I$.

Let L^1 through L^3 be the matrices

$$L^1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad (3.7.178)$$

$$L^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad (3.7.179)$$

$$L^3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (3.7.180)$$

In Exercise 7.27 you should have found that the Lie algebra $so(n, \mathbb{R})$ consists of all $n \times n$ real antisymmetric matrices. Show that L^1 through L^3 form a basis for the Lie algebra $so(3, \mathbb{R})$. Show that they also obey the commutation rules

$$\{L^\alpha, L^\beta\} = L^\gamma \quad (3.7.181)$$

where α, β, γ is any cyclic permutation of 1, 2, 3. Evidently, according to (7.173) and (7.181), the Lie algebras $su(2)$ and $so(3, \mathbb{R})$ have the same structure constants, and are therefore the same. Compute the matrices $(L^\alpha)^2$. Observe that they are diagonal, and hence mutually commute. Show that they sum to $-2I$.

Let the matrices \hat{K}^α be the adjoint representation matrices associated with the K^α . See (7.61). Verify the relations

$$(L^\alpha)_{\beta\gamma} = -\epsilon_{\alpha\beta\gamma} \text{ and therefore } \hat{K}^\alpha = L^\alpha. \quad (3.7.182)$$

You have shown that the L^α matrices are those for the adjoint representation of $su(2)$. Since $su(2)$ and $so(3, \mathbb{R})$ have the same structure constants, show that the adjoint representation of $so(3, \mathbb{R})$ is the fundamental representation.

Show from (7.177) that there is the relation

$$\{\mathbf{a} \cdot \mathbf{K}, \mathbf{b} \cdot \mathbf{K}\} = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{K}. \quad (3.7.183)$$

Define $\mathbf{a} \cdot \mathbf{L}$ in an analogous way to (7.176) and show that there is also the relation

$$\{\mathbf{a} \cdot \mathbf{L}, \mathbf{b} \cdot \mathbf{L}\} = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{L}. \quad (3.7.184)$$

You have shown that the cross-product Lie algebra (7.47) is intimately related to the Lie algebra for $su(2)$ and $so(3, \mathbb{R})$. Indeed, let \mathbf{e}_1 through \mathbf{e}_3 be the unit vectors of Exercise 7.22. Make the correspondences

$$\mathbf{e}_\alpha \leftrightarrow K^\alpha \leftrightarrow L^\alpha. \quad (3.7.185)$$

Then, you have shown that there are also the correspondences

$$(\mathbf{e}_\alpha \times \mathbf{e}_\beta) \leftrightarrow \{K^\alpha, K^\beta\} \leftrightarrow \{L^\alpha, L^\beta\}. \quad (3.7.186)$$

Finally, in Exercise 7.23 you should have found that the structure constants for the cross-product Lie algebra are the same as those for $su(2)$ and $so(3, \mathbb{R})$. Therefore the cross-product Lie algebra is the same as that of $su(2)$ and $so(3, \mathbb{R})$.

We have studied the Lie algebras $su(2)$, $so(3, \mathbb{R})$ and the cross-product Lie algebra. We now explore the relation between the groups $SU(2)$ and $SO(3, \mathbb{R})$. Begin with the case of $SU(2)$. Let \mathbf{n} be a unit vector. Define $SU(2)$ matrices $v(\theta, \mathbf{n})$ by the rule

$$v(\theta, \mathbf{n}) = \exp(\theta \mathbf{n} \cdot \mathbf{K}). \quad (3.7.187)$$

Show, in accord with Section 3.8.1, that any $SU(2)$ matrix can be written in the form (7.187). Show that

$$(\mathbf{n} \cdot \mathbf{K})^2 = -(1/4)I. \quad (3.7.188)$$

Use this relation to sum the series implied by (7.187) to find the explicit result

$$v(\theta, \mathbf{n}) = I \cos(\theta/2) + 2(\mathbf{n} \cdot \mathbf{K}) \sin(\theta/2). \quad (3.7.189)$$

Show that

$$v(2\pi, \mathbf{n}) = -I, \quad (3.7.190)$$

$$v(4\pi, \mathbf{n}) = +I. \quad (3.7.191)$$

Show that $SU(2)$ is covered once and only once when $\theta \in [0, 2\pi]$ and \mathbf{n} is allowed to be any unit vector.

As special cases of (7.187), verify the relations

$$v(\theta, \mathbf{e}_1) = \exp(\theta K^1) = \exp[(-i/2)\theta \sigma^1] = \begin{pmatrix} \cos(\theta/2) & -i \sin(\theta/2) \\ -i \sin(\theta/2) & \cos(\theta/2) \end{pmatrix}, \quad (3.7.192)$$

$$v(\theta, \mathbf{e}_2) = \exp(\theta K^2) = \exp[(-i/2)\theta \sigma^2] = \begin{pmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{pmatrix}, \quad (3.7.193)$$

$$v(\theta, \mathbf{e}_3) = \exp(\theta K^3) = \exp[(-i/2)\theta \sigma^3] = \begin{pmatrix} \exp(-i\theta/2) & 0 \\ 0 & \exp(i\theta/2) \end{pmatrix}. \quad (3.7.194)$$

The Euler-angle parameterization of $SU(2)$ is defined by the rule

$$v(\phi, \theta, \psi) = \exp(\phi K^3) \exp(\theta K^2) \exp(\psi K^3). \quad (3.7.195)$$

Note that K^1 does not appear in the formula. Verify that every element in $SU(2)$ can be written in Euler form. Verify that $SU(2)$ is covered once, and only once, if the Euler angles lie in the ranges $\phi \in [0, 2\pi]$, $\theta \in [0, \pi]$, $\psi \in [0, 4\pi]$. By carrying out the matrix multiplications implied by (7.195), verify that $v(\phi, \theta, \psi)$ has the explicit form

$$v(\phi, \theta, \psi) = \begin{pmatrix} \cos(\theta/2) \exp[-(i/2)(\phi + \psi)] & -\sin(\theta/2) \exp[(i/2)(-\phi + \psi)] \\ \sin(\theta/2) \exp[-(i/2)(-\phi + \psi)] & \cos(\theta/2) \exp[(i/2)(\phi + \psi)] \end{pmatrix}. \quad (3.7.196)$$

The relation (7.189) is a formula for computing the 2×2 $SU(2)$ matrix v given θ and \mathbf{n} . Suppose, instead, that one is given $v \in SU(2)$ and wants to know θ and \mathbf{n} . Show that there are the formulas

$$2 \cos(\theta/2) = \text{tr}(v) \quad (3.7.197)$$

and

$$4(\mathbf{n} \cdot \mathbf{K}) \sin(\theta/2) = v - v^\dagger, \quad (3.7.198)$$

from which it follows that

$$n_\alpha \sin(\theta/2) = (i/4) \text{tr}[\sigma^\alpha(v - v^\dagger)]. \quad (3.7.199)$$

Consider next the case of $SO(3, \mathbb{R})$. Define $SO(3, \mathbb{R})$ matrices $R(\theta, \mathbf{n})$ by the rule

$$R(\theta, \mathbf{n}) = \exp(\theta \mathbf{n} \cdot \mathbf{L}). \quad (3.7.200)$$

Show, in accord with Section 3.8.1, that any $SO(3, \mathbb{R})$ matrix can be written in this form. For any two vectors \mathbf{a} and \mathbf{b} , verify the relation

$$(\mathbf{a} \cdot \mathbf{L})\mathbf{b} = (\mathbf{a} \times \mathbf{b}). \quad (3.7.201)$$

Use this result to show that $R(\theta, \mathbf{n})$ produces a rotation by angle θ about the axis \mathbf{n} .

Verify the result

$$(\mathbf{n} \cdot \mathbf{L})^3 = -\mathbf{n} \cdot \mathbf{L}. \quad (3.7.202)$$

Use this relation to sum the series implied by (7.200) to find the explicit results

$$R(\theta, \mathbf{n}) = I + (\mathbf{n} \cdot \mathbf{L}) \sin \theta + (\mathbf{n} \cdot \mathbf{L})^2 (1 - \cos \theta), \quad (3.7.203)$$

$$R(2\pi, \mathbf{n}) = I. \quad (3.7.204)$$

Verify that $SO(3, \mathbb{R})$ is covered once and only once when $\theta \in [0, \pi]$ and \mathbf{n} is allowed to be any unit vector.

As special cases of (7.203), verify the relations

$$R(\theta, \mathbf{e}_1) = \exp(\theta L^1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix}, \quad (3.7.205)$$

$$R(\theta, \mathbf{e}_2) = \exp(\theta L^2) = \begin{pmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{pmatrix}, \quad (3.7.206)$$

$$R(\theta, \mathbf{e}_3) = \exp(\theta L^3) = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (3.7.207)$$

The Euler-angle parameterization of $SO(3, \mathbb{R})$ is defined by the rule

$$R(\phi, \theta, \psi) = \exp(\phi L^3) \exp(\theta L^2) \exp(\psi L^3). \quad (3.7.208)$$

Note that L^1 does not appear in the formula. Verify that every element in $SO(3, \mathbb{R})$ can be written in Euler form. Verify that $SO(3, \mathbb{R})$ is covered once, and only once, if the Euler angles lie in the ranges $\phi \in [0, 2\pi]$, $\theta \in [0, \pi]$, $\psi \in [0, 2\pi]$. By carrying out the matrix multiplications implied by (7.208), verify that $R(\phi, \theta, \psi)$ has the explicit form

$$R(\phi, \theta, \psi) = \begin{pmatrix} \cos \phi \cos \theta \cos \psi - \sin \phi \sin \psi & -\cos \phi \cos \theta \sin \psi - \sin \phi \cos \psi & \cos \phi \sin \theta \\ \sin \phi \cos \theta \cos \psi + \cos \phi \sin \psi & -\sin \phi \cos \theta \sin \psi + \cos \phi \cos \psi & \sin \phi \sin \theta \\ -\sin \theta \cos \psi & \sin \theta \sin \psi & \cos \theta \end{pmatrix}. \quad (3.7.209)$$

The relation (7.203) is a formula, sometimes called *Rodrigues' rotation formula*, for computing the 3×3 rotation matrix R given the rotation angle θ and the axis of rotation \mathbf{n} . Suppose, instead, that one is given R and wants to know the rotation angle θ and the axis \mathbf{n} . First verify the relation

$$\text{tr}[(\mathbf{a} \cdot \mathbf{L})^2] = -2\mathbf{a} \cdot \mathbf{a}. \quad (3.7.210)$$

for any vector \mathbf{a} . Now show that there are the formulas

$$1 + 2 \cos \theta = \text{tr}(R), \quad (3.7.211)$$

$$2(\mathbf{n} \cdot \mathbf{L}) \sin \theta = R - R^T. \quad (3.7.212)$$

Let R be any element of $SO(3, \mathbb{R})$. Show that R must have $+1$ as an eigenvalue and that \mathbf{n} is the associated eigenvector.⁴¹ Show that the other two eigenvalues of R are $\exp(\pm i\theta)$.

Since the Lie algebras $su(2)$ and $so(3, \mathbb{R})$ are the same, we may expect a close relation between the groups $SU(2)$ and $SO(3, \mathbb{R})$. Are they perhaps the same? The answer is *no* as can be seen by comparing (7.190) and (7.204). Although the groups have the same Lie algebra, they are not the same globally. Exercises 8.2.10 and 8.2.11 show that there is a two-to-one homomorphism between $SU(2)$ and $SO(3, \mathbb{R})$. In fact these exercises show that, given $v \in SU(2)$, there is an $R(v) \in SO(3, \mathbb{R})$ specified by the two-to-one homomorphic map

$$R_{\alpha\beta}(v) = (1/2)\text{tr}(v^\dagger \sigma^\alpha v \sigma^\beta). \quad (3.7.213)$$

There is another way to highlight the distinction between $su(2)$ and $so(3, \mathbb{R})$. For a Lie algebra realized in terms of matrices, it is useful, when possible, to form the matrix for the second-order Casimir operator. The second-order Casimir matrix is defined in terms of the

⁴¹This result is sometimes called Euler's theorem for rigid body motion.

structure constants and the basis matrices for the Lie algebra. See Section 27.11 for details. In the case of the K^α verify that there is the matrix relation

$$(K^1)^2 + (K^2)^2 + (K^3)^2 = -(3/4)I. \quad (3.7.214)$$

In the case of the L^α verify that there is the matrix relation

$$(L^1)^2 + (L^2)^2 + (L^3)^2 = -(2)I. \quad (3.7.215)$$

In our case the structure constants are given by (7.174), and it can be shown that the quantities on the left sides of (7.214) and (7.215) are the Casimir matrices formed from the K^α and the L^α , respectively. The coefficients in parentheses on the right sides of (7.214) and (7.215) are of the form $j(j+1)$ with $j = 1/2$ in the case of the K^α and $j = 1$ in the case of the L^α . Analogous relations may be familiar to the reader from the quantum theory of spin/angular momentum. We see that $su(2)$ corresponds to the case $j = 1/2$ and $so(3, \mathbb{R})$ corresponds to the case $j = 1$.

3.7.32. Suppose that \mathcal{R} is a map that sends three-dimensional Euclidean space into itself and has a cyclic action on the points e_1, e_2, e_3 :

$$\mathcal{R}e_1 = e_2, \mathcal{R}e_2 = e_3, \mathcal{R}e_3 = e_1. \quad (3.7.216)$$

Extend \mathcal{R} to all of three-dimensional Euclidean space by linearity so that its action can be represented by an associated matrix R . Show that $R \in SO(3, \mathbb{R})$. Use the results of Exercise 7.31 to find the axis \mathbf{n} and angle θ for R .

3.7.33. Show that the groups $Sp(2, \mathbb{R})$ and $SL(2, \mathbb{R})$ are the same, and therefore their Lie algebras are the same: $Sp(2, \mathbb{R}) = SL(2, \mathbb{R})$ and $sp(2, \mathbb{R}) = sl(2, \mathbb{R})$. Show that the groups $Sp(2, \mathbb{C})$ and $SL(2, \mathbb{C})$ are the same, and therefore their Lie algebras are the same: $Sp(2, \mathbb{C}) = SL(2, \mathbb{C})$ and $sp(2, \mathbb{C}) = sl(2, \mathbb{C})$. See Exercises 1.2 and 1.3. [Subsequently it will be shown that the Lie algebra of the Lorentz group is the same as the Lie algebras $sp(2, \mathbb{C}) = sl(2, \mathbb{C})$. See Exercises 7.3.30 and 8.2.14.] Show that the Lie algebras $sp(2, \mathbb{R})$, $sl(2, \mathbb{R})$, $su(2)$, and $so(3, \mathbb{R})$ have the same dimension. Show that these Lie algebras are in fact the same (equivalent) over the complex field. Which of these Lie algebras are equivalent over the real field?

3.7.34. Show that the Lie algebras $sp(4, \mathbb{R})$ and $so(5, \mathbb{R})$ have the same dimension. See Exercise 27.5.4 for a demonstration that these Lie algebras are in fact the same (equivalent) over the complex field, but not the real field.

3.7.35. Show that the Lie algebras $su(4)$ and $so(6, \mathbb{R})$ have the same dimension. In fact, these Lie algebras are the same (equivalent) over the real field. Moreover, as shown in Exercise 8.2.12, there is a corresponding two-to-one homomorphism between the groups $SU(4)$ and $SO(6, \mathbb{R})$ just as there is a two-to-one homomorphism between the groups $SU(2)$ and $SO(3, \mathbb{R})$. See Exercises 7.29, 8.2.11, and 8.2.12.

3.7.36. Let $c_{\alpha\beta}^\gamma$ be a set of structure constants for some Lie algebra L as in (7.50). Let B_α be a set of $m \times m$ matrices that forms a basis for L thereby providing a *representation* of L . That is, the matrices satisfy the commutation rules

$$\{B_\alpha, B_\beta\} = \sum_\gamma c_{\alpha\beta}^\gamma B_\gamma. \quad (3.7.217)$$

Let E be any $m \times m$ invertible matrix, and define matrices B'_α by the rule (similarity transformation)

$$B'_\alpha = EB_\alpha E^{-1}. \quad (3.7.218)$$

View the B'_α as basis elements and show that the B'_α also form a representation of L . That is, the B'_α obey commutation rules identical to those of the B_α in (7.217) with the *same* structure constants. The representations provided by the B'_α and the B_α are called *equivalent*, and for many purposes may be viewed as being essentially the same.⁴² Conversely, given two sets of $m \times m$ representation matrices B_α and B'_α that obey the same commutation rules, one can inquire whether there is an invertible matrix E such that (7.218) holds. If there is, the two representations are said to be equivalent.

Given the B_α , suppose we define *conjugate* matrices \tilde{B}_α by the “tilde” rule

$$\tilde{B}_\alpha = -B_\alpha^T. \quad (3.7.219)$$

Note that this tilde rule is an *involution*. That is, let $\tilde{\mathcal{C}}$ denote the tilde conjugacy operator defined by

$$\tilde{\mathcal{C}}(B_\alpha) = \tilde{B}_\alpha. \quad (3.7.220)$$

Verify that $\tilde{\mathcal{C}}^2$ has the property

$$\tilde{\mathcal{C}}^2(B_\alpha) = B_\alpha \quad (3.7.221)$$

so that $\tilde{\mathcal{C}}^2 = \mathcal{I}$ on every element on which it acts.

Show that the \tilde{B}_α also form a representation of L . This representation is called a conjugate representation. That is, the \tilde{B}_α obey commutation rules identical to those of the B_α in (7.217) with the *same* structure constants. Put another way, show that there is the result

$$\tilde{\mathcal{C}}(\{B_\alpha, B_\beta\}) = \{\tilde{\mathcal{C}}(B_\alpha), \tilde{\mathcal{C}}(B_\beta)\}, \quad (3.7.222)$$

which displays that $\tilde{\mathcal{C}}$ is a homomorphism for the Lie product provided by the commutator $\{*, *\}$.

Whether this conjugate representation is equivalent (in the sense defined three paragraphs above) to that provided by the B_α depends on the Lie algebra L and the representation provided by the B_α . If a representation and its conjugate are equivalent in the sense

⁴²Observe that this definition of *equivalent* need not be the same as that given in Subsection 7.6. There *arbitrary* linear invertible transformations on the basis were considered with the aim of modifying the structure constants, and the structure constants were found to change according to the rules (7.56) and (7.57). But the new basis elements, being linear combinations of the B_α , are still elements of L . In the present case the structure constants are required to remain unchanged even though the basis is changed. And in this case the B_α have to be matrices or linear operators or some such things for (7.218) to make sense. Finally, unless the B_α form a basis for the set of $m \times m$ matrices, the B'_α need not be in L . For example, if we consider the fundamental representation of $sp(2n)$, the B_α are of the form JS . But the same need not be true of the B'_α .

(7.218), the representations are said to be *self conjugate*. (We remark that if the B_α are the matrices for the *adjoint* representation of L , then the matrices \tilde{B}_α are sometimes referred to as the *coadjoint* representation of L .)

There is a second “conjugacy” possibility. Suppose that for some choice of basis elements [see (7.56) and (7.57)] the structure constants $c_{\alpha\beta}^\gamma$ can all be made *real*. (This can be shown to be the case, for example, for all the classical and exceptional Lie algebras in Table 7.2.) Also allow the possibility that at least some of the B_α may have some complex entries. In this case, define matrices \check{B}_α , again called conjugate matrices, by the “accent breve” rule

$$\check{B}_\alpha = \bar{B}_\alpha \quad (3.7.223)$$

where a bar indicates complex conjugation. Call the breve conjugacy operator $\check{\mathcal{C}}$ so that we may write

$$\check{\mathcal{C}}(B_\alpha) = \check{B}_\alpha = \bar{B}_\alpha. \quad (3.7.224)$$

Verify that $\check{\mathcal{C}}$ is also an involution. Show that the \check{B}_α also form a representation of L or, equivalently, $\check{\mathcal{C}}$ is also a commutator Lie product homomorphism. Whether this conjugate representation is equivalent to that provided by the B_α also depends on the Lie algebra L and the representation provided by the B_α . A representation involving complex matrices that is equivalent to itself under the breve operation (7.222) is called *pseudoreal*. In analogy with the tilde operation case, it may also be called self conjugate.

If the structure constants cannot be made real, show that the \check{B}_α still form a Lie algebra. Whether or not this Lie algebra is different from the original one, or can be brought to the same form as the original Lie algebra by a suitable new choice of basis elements as in (7.56) and (7.57), then requires further investigation.

There is a third conjugacy operator possibility, which we will call the “accent grave” rule, that is sometimes of use. Denote it by $\grave{\mathcal{C}}$. Given basis matrices B_α for a Lie algebra L with real structure constants, define matrices \grave{B}_α , again called conjugate matrices, by the rule

$$\grave{\mathcal{C}}(B_\alpha) = \grave{B}_\alpha = -B_\alpha^\dagger. \quad (3.7.225)$$

Verify that the grave rule is also an involution. Show that the \grave{B}_α also form a representation of L so that $\grave{\mathcal{C}}$ is also a commutator Lie product homomorphism. Show that $\grave{\mathcal{C}}$ consists of combining the operations of the first two conjugation rules by verifying that

$$\grave{\mathcal{C}} = \check{\mathcal{C}}\tilde{\mathcal{C}} = \tilde{\mathcal{C}}\check{\mathcal{C}}. \quad (3.7.226)$$

In summary, we have defined three possible conjugacy rules: tilde \sim , breve $\check{}$, and grave $\grave{}$. Note that, in all cases and for all three conjugacy rules, a representation and its conjugate have the same dimension.

Let us now explore conjugacy relations for some of the familiar Lie algebras: Consider the *fundamental/defining* representation of $sp(2, \mathbb{R})$ provided by the matrices (7.66) through (7.68). Find the associated tilde representation given by (7.219) above. Verify that this representation is equivalent to the fundamental representation using

$$E = J. \quad (3.7.227)$$

Consider the fundamental representation of $sp(2n, \mathbb{R})$ for *any* n . Using (7.34) show that

$$\tilde{B} = SJ. \quad (3.7.228)$$

Verify that

$$J\tilde{B}J^{-1} = JS = B, \quad (3.7.229)$$

and therefore the tilde representation is equivalent to the fundamental representation, again using (7.227), for all n . Put another way, under the tilde operation and for all n , the fundamental representation of $sp(2n, \mathbb{R})$ is *self conjugate*.

Suppose that either the breve or grave conjugacy operations are used instead. Show that, since the fundamental representation of $sp(2n, \mathbb{R})$ consists of real matrices, it is also self conjugate under both the breve and grave operation for all n .

Consider the adjoint representation of $sp(2, \mathbb{R})$. It is provided by the matrices (7.75) through (7.77). Find the associated tilde representation given by (7.219) above. Verify that this representation is equivalent to the adjoint representation using

$$E = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \quad (3.7.230)$$

In fact, it can be shown that *all* representations of $sp(2n, \mathbb{R})$ are self conjugate for all n under all three conjugacy relations tilde, breve, and grave.

Consider the fundamental representation of $su(2)$ provided by the matrices (7.169) through (7.171). Observe that some of them are complex. Show that the associated breve representation given by (7.223) above yields the result

$$\check{K}^1 = -K^1, \quad (3.7.231)$$

$$\check{K}^2 = K^2, \quad (3.7.232)$$

$$\check{K}^3 = -K^3. \quad (3.7.233)$$

Verify that this representation is equivalent to the original representation using

$$E = \exp(\pi K^2) = -J_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \quad (3.7.234)$$

[Look ahead to see (8.2.338) or (8.2.374) through (8.2.376) if you need help.] Thus the representation of $su(2)$ provided by the matrices (7.169) through (7.171) is pseudoreal. Finally, it can be shown that *all* representations of $su(2)$ are self conjugate under the breve operation because any representation of $su(2)$ can be obtained by taking suitable real linear combinations of tensors formed from vectors (actually, *spinors*) that transform according to the fundamental representation.

Suppose that instead the tilde operation (7.219) is employed. Show that in this case there is the result

$$\tilde{K}^1 = -K^1, \quad (3.7.235)$$

$$\tilde{K}^2 = K^2, \quad (3.7.236)$$

$$\tilde{K}^3 = -K^3. \quad (3.7.237)$$

The tilde and breve operations (7.219) and (7.223), when acting on the fundamental representation of $su(2)$, give the same result. Therefore for $su(2)$ the tilde representation is also equivalent to the fundamental representation again using the E given by (7.234).

Show that the reason the tilde and breve operations give the same result in the case of $su(2)$ fundamental representation is that the K^α are anti-Hermitian, and that this “same result” conclusion will also hold for all cases in which the B_α are anti-Hermitian, including all $su(n)$ cases.

Suppose the grave operation acts on the fundamental representation of $su(n)$. Show that in this case there is the “no effect” result

$$\grave{B}_\alpha = B_\alpha \quad (3.7.238)$$

because the B_α are anti-Hermitian.

We remark that the case of $su(2)$ is special. Subsequently we will find that the fundamental representation of $su(n)$ for any $n > 2$ is not equivalent to its complex conjugate (breve) representation. Therefore these fundamental representations are not self-conjugate/pseudoreal.

Consider the orthogonal group Lie algebras $so(n, \mathbb{R})$. In this case the Lie algebra for the fundamental representation consists of all real antisymmetric matrices A . Correspondingly, the breve operation has no effect on elements in the fundamental representation. Show that, because A is real and antisymmetric, the tilde operation (7.219) and the grave operation (7.225) also have no effect for the fundamental representation. We conclude that the fundamental representation of $so(n, \mathbb{R})$ is self conjugate for all three conjugacy definitions. Since all representations can be obtained from the fundamental representation by suitable linear combinations of tensor products, all three conjugacy operations will also have no effect on any representation, and they are therefore all self conjugate for all conjugacy definitions.

In summary, the representations of $sp(2n)$, $so(n)$, and $su(2)$ are all self conjugate for all three conjugacy operations. And for $su(n)$ the tilde and breve conjugacy operations yield the same result. Finally, for the case of $su(n)$ with $n > 2$, it can be shown that there are representations that are not self conjugate. See, for example, Exercise 5.8.29. In particular, the representations 3 and $\bar{3}$ for $su(3)$ are not equivalent.⁴³

Subsequently we will also be interested in the Lorentz group Lie algebra and in $s\ell(2, \mathbb{C})$, which will be found to be the Lie algebra for the covering group of the Lorentz group. For these Lie algebras the three conjugacy operations will again be useful. See Exercises 7.3.27, 7.3.29 through 7.3.31, and 8.2.14.

3.7.37. Review Exercise 7.36 which described tilde, breve, and grave rules for defining conjugate matrices in Lie algebras. The purpose of this exercise is to extend these rules to the associated matrix groups. Suppose the Lie algebra L has dimension n and basis elements B_α . Let $b = (b_1, b_2, \dots, b_n)$ be a collection of n real parameters, and consider the group element

$$G(b) = \exp\left(\sum_{\alpha=1}^n b_\alpha B_\alpha\right). \quad (3.7.239)$$

⁴³Here the “bar” notation, which is customary, is used to refer to the use of what we have called the breve conjugacy operator \check{C} because this operator involves complex conjugation. Recall (7.223).

In the case of the tilde rule, define the group element $\tilde{G}(b)$ by the rule

$$\tilde{G}(b) = \exp\left(\sum_{\alpha=1}^n b_\alpha \tilde{B}_\alpha\right) = \exp\left(-\sum_{\alpha=1}^n b_\alpha B_\alpha^T\right). \quad (3.7.240)$$

Show that

$$\tilde{G}(b) = [G^T(b)]^{-1}. \quad (3.7.241)$$

[See (7.266).] Show that the tilde operation when acting as defined on group elements is again an involution. That is, for group elements define a tilde operator, which we will call $\tilde{\mathcal{D}}$, by the rule

$$\tilde{\mathcal{D}}G(b) = \tilde{G}(b) = [G^T(b)]^{-1}, \quad (3.7.242)$$

and show that

$$\tilde{\mathcal{D}}^2 G(b) = G(b). \quad (3.7.243)$$

Moreover, verify that it follows from (7.242) and (7.243) that

$$\tilde{\mathcal{D}}\tilde{G}(b) = G(b) = [\tilde{G}^T(b)]^{-1}, \quad (3.7.244)$$

which is the counterpart to (7.242). Show that $\tilde{\mathcal{D}}$ is also a group homomorphism. That is, if G and G' are any two group elements, then

$$\tilde{\mathcal{D}}(GG') = \tilde{\mathcal{D}}(G)\tilde{\mathcal{D}}(G'). \quad (3.7.245)$$

Comparison of (7.222) and (7.245) shows that we have converted a Lie product (commutator) homomorphism into a Lie group homomorphism.

In the cases of the breve and grave rules make definitions of their actions on group elements analogous that for the tilde operation. For these rules show that

$$\breve{G}(b) = \bar{G}(b) \quad (3.7.246)$$

and

$$\grave{G}(b) = [G^\dagger(b)]^{-1}. \quad (3.7.247)$$

[See (7.267).] Let $\breve{\mathcal{D}}$ and $\grave{\mathcal{D}}$ be the breve and grave operators which act on group elements so that we may write

$$\breve{\mathcal{D}}G(b) = \breve{G}(b) = \bar{G}(b) \quad (3.7.248)$$

and

$$\grave{\mathcal{D}}G(b) = \grave{G}(b) = [G^\dagger(b)]^{-1}. \quad (3.7.249)$$

Show that $\breve{\mathcal{D}}$ and $\grave{\mathcal{D}}$ are also involutions and group homomorphisms. Observe that if G is a unitary matrix ($G^\dagger = G^{-1}$) it follows from (7.249) that

$$\grave{G}(b) = [G^\dagger(b)]^{-1} = G(b). \quad (3.7.250)$$

Finally, review Exercise 1.6.18. Verify that the relation between K and Λ given by (1.6.287) is the same as the relation between $\tilde{G}(b)$ and $G(b)$ given by (7.241). According to Exercise 7.36 the B_α and the \tilde{B}_α obey the same Lie algebra. This fact is consistent with the result that K must be a Lorentz transformation matrix if Λ is a Lorentz transformation matrix, and vice versa.

3.7.38. Suppose g is a $(m+n) \times (m+n)$ diagonal matrix with m diagonal entries having value $+1$ followed by n diagonal entries having value -1 . Show that the set of all $(m+n) \times (m+n)$ matrices O that satisfy the relation

$$O^T g O = g \quad (3.7.251)$$

forms a group. This group is called the *indefinite* orthogonal group, and is sometimes denoted by the symbol $O(m, n)$, or by the symbols $O(m, n, \mathbb{R})$ and $O(m, n, \mathbb{C})$ if the choice of field needs to be explicit. Note that (7.251) continues to hold if g is replaced by $-g$, and therefore there is no distinction between $O(m, n)$ and $O(n, m)$. Show from (7.251) that there is the result

$$\det(O) = \pm 1. \quad (3.7.252)$$

Show that indefinite orthogonal matrices with determinant $+1$ form a subgroup, called $SO(m, n)$. Find the Lie algebra $so(m, n, \mathbb{R})$, and show that it is equivalent to the Lie algebra $so(m + n, \mathbb{R})$ when working over the complex field.

3.7.39. Review Exercise 7.38 above. Let g be the matrix defined there. Show that the set of all complex $(m + n) \times (m + n)$ matrices U that satisfy the relation

$$U^\dagger g U = g \quad (3.7.253)$$

forms a group. This group is called the *indefinite* unitary group, and is denoted by the symbol $U(m, n)$. Here the field is naturally \mathbb{C} . Note that (7.253) continues to hold if g is replaced by $-g$, and therefore there is no distinction between $U(m, n)$ and $U(n, m)$. Show from (7.253) that there is the result

$$|\det(U)| = 1. \quad (3.7.254)$$

Verify that indefinite unitary matrices with determinant $+1$ form a subgroup. It is called $SU(m, n)$. Find the Lie algebra $su(m, n)$, and show that it is equivalent to the Lie algebra $su(m + n)$ when working over the complex field.

3.7.40. Review Exercise 7.38 above. Also look ahead and review Exercise 6.2.6. The 4×4 real matrices Λ defined there form a group, generally called the Lorentz group. From (6.2.20) we see that the Lorentz group is analogous to the rotation group $SO(4, \mathbb{R})$ except the 4×4 identity matrix I has been replaced by the 4×4 diagonal matrix g . Note that g has three equal diagonal entries with the same sign, and one with the opposite sign. For this reason, the Lorentz group is also referred to as $SO(3, 1, \mathbb{R})$. Find the Lie algebra $so(3, 1, \mathbb{R})$ and show that it is equivalent to the Lie algebra $so(4, \mathbb{R})$ when working over the complex field, and hence also equivalent to $su(2) \oplus su(2)$. Since the representations of $su(2)$ are well known, the finite-dimensional (and nonunitary) representations of the Lorentz Lie algebra and Lie group are also well understood. See Exercises 4.3.19, 4.3.20, and 7.3.28.

3.7.41. Suppose f and g are two elements of some group G . Using group multiplication, form the group element h defined by the rule

$$h = (gf)^{-1} fg = f^{-1} g^{-1} fg. \quad (3.7.255)$$

This element is called the *group commutator* of f and g . Note that if f and g commute, $fg = gf$, then $h = (gf)^{-1}fg = (fg)^{-1}fg = I$.

Suppose that G is a matrix Lie group and consider elements $f(s)$ and $g(s)$ of the form

$$f(s) = \exp(sa), \quad (3.7.256)$$

$$g(s) = \exp(sb), \quad (3.7.257)$$

where a and b are in the Lie algebra of G . Let $h(s)$ be the group commutator of $f(s)$ and $g(s)$,

$$h(s) = f^{-1}(s)g^{-1}(s)f(s)g(s) = \exp(-sa)\exp(-sb)\exp(sa)\exp(sb). \quad (3.7.258)$$

Show, using the BCH formula (7.41), that there is the relation

$$h(s) = \exp\left(s^2\{a,b\} + O(s^3)\right). \quad (3.7.259)$$

Thus, for a matrix Lie group, there is a relation between the group commutator and the Lie algebra commutator. It can be shown that an analogous relation holds for abstract Lie groups: There is a relation between the group commutator and the Lie product of associated elements in the Lie algebra.

For extra credit, use through third order the BCH formula (7.41) to show that

$$h(s) = \exp\left(s^2\{a,b\} - (s^3/2)\{(a+b), \{a,b\}\} + O(s^4)\right). \quad (3.7.260)$$

Moreover verify that if $\{a,b\} = 0$ then f and g commute, and vice versa.

3.7.42. Suppose A and B are two $n \times n$ matrices that satisfy the relation

$$AB = I.$$

Show it follows that

$$BA = I,$$

and therefore A and B commute.

Suppose C and D are two commuting $n \times n$ matrices and that C is invertible. Show that then C^{-1} and D also commute. Show that C^m and D also commute for all integer values (positive, zero, and negative) of m .

Suppose that a matrix E is a function of C , and specifically is defined in terms of C as some convergent power series in C (and possibly also powers of C^{-1} if C is invertible). Show that then E and C also commute.

3.7.43. Review Exercise 7.31. There it is shown that the generators K^α for $su(2)$ and the generators L^α for $so(3, \mathbb{R})$ can be selected to be *anti-Hermitian* and, writing the generators generically as J^α , satisfy the same commutation rules

$$\{J^\alpha, J^\beta\} = \sum_\gamma \epsilon_{\alpha\beta\gamma} J^\gamma \quad (3.7.261)$$

with *real* structure constants $\epsilon_{\alpha\beta\gamma}$. Verify that the commutation rules (7.248) are consistent with the J^α being anti-Hermitian. That is, verify that the commutator of two anti-Hermitian generators is again anti-Hermitian.

Correspondingly, the associated group elements U are of the form

$$U = \exp\left(\sum_\gamma \lambda_\gamma J^\gamma\right) \quad (3.7.262)$$

with *real* parameters λ_γ . Verify that U as given by (7.262) is unitary when the λ_γ are real.

We might say this treatment of $su(2)$ and $so(3, \mathbb{R})$ is the mathematicians' approach. By contrast, in the quantum treatment of angular momentum, physicists work with *Hermitian* generators, call them \tilde{J}^α , that are required to satisfy the commutation rules

$$\{\tilde{J}^\alpha, \tilde{J}^\beta\} = \sum_\gamma i\epsilon_{\alpha\beta\gamma} \tilde{J}^\gamma \quad (3.7.263)$$

with *purely imaginary* structure constants $i\epsilon_{\alpha\beta\gamma}$. Verify that the commutation rules (7.263) are consistent with the \tilde{J}^α being Hermitian. That is, verify that the commutator of two Hermitian generators is anti-Hermitian.

Correspondingly, the associated group elements U are of the form

$$U = \exp\left(\sum_\gamma -i\lambda_\gamma \tilde{J}^\gamma\right) \quad (3.7.264)$$

with *real* parameters λ_γ . Verify that U as given by (7.264) is unitary when the λ_γ are real.

Verify that the mathematicians' and quantum physicists' approaches are connected by the (*complex*) change of basis

$$J^\alpha = -i\tilde{J}^\alpha \Leftrightarrow \tilde{J}^\alpha = iJ^\alpha. \quad (3.7.265)$$

That is, the Lie algebras defined by (2.61) and (2.63) are equivalent over the complex field. Verify also that U as given by (2.62) and (2.64) are unaffected by this change of basis.

Why do quantum physicists insert what would appear to mathematicians to be superfluous factors of i ? They do so because they wish to associate physical observables with Hermitian operators in order to ensure that the expectation values and eigenvalues of physical observables are *real* numbers.

3.7.44. Suppose A is *any* $n \times n$ matrix. Verify the relations

$$[\exp(A)]^T = \exp(A^T), \quad (3.7.266)$$

$$[\exp(A)]^\dagger = \exp(A^\dagger). \quad (3.7.267)$$

Suppose H is a *Hermitian* $n \times n$ matrix. Verify that then $\exp(H)$ is also Hermitian,

$$[\exp(H)]^\dagger = \exp(H). \quad (3.7.268)$$

Verify the line of reasoning below to show that $\exp(H)$ is positive definite: Let v be any nonzero n -component vector. Then there is the result

$$\begin{aligned} (v, \exp(H)v) &= (v, \exp(H/2)\exp(H/2)v) = (\exp(H/2)v, \exp(H/2)v) \\ &= (w, w) > 0 \end{aligned} \quad (3.7.269)$$

where we have employed the usual complex scalar product and

$$w = \exp(H/2)v. \quad (3.7.270)$$

3.7.45. Review Table 7.2 that provides the dimensions of the Classical Lie Algebras for each integer value of ℓ . Verify that, consistent with the Table, we may define functions \mathcal{A} through \mathcal{D} by the rules

$$\mathcal{A}(n) = \dim[su(n)] = n^2 - 1, \quad (3.7.271)$$

$$\mathcal{B}(n) = \dim[so(n)] = (1/2)n(n - 1), \text{ } n \text{ odd}, \quad (3.7.272)$$

$$\mathcal{C}(n) = \dim[sp(n)] = (1/2)n(n + 1), \text{ } n \text{ even}, \quad (3.7.273)$$

$$\mathcal{D}(n) = \dim[so(n)] = (1/2)n(n - 1), \text{ } n \text{ even}, \quad (3.7.274)$$

where \dim stands for *dimension*. Note that $\mathcal{B}(n)$ and $\mathcal{D}(n)$ may be regarded as odd n and even n evaluations of a common formula. Recall Exercises 7.27 and 7.28. Suppose this common formula is evaluated for *negative* even values of n . Show that

$$\mathcal{D}(-n) = (1/2)(-n)(-n - 1) = (1/2)n(n + 1) = \mathcal{C}(n), \text{ } n \text{ even}. \quad (3.7.275)$$

We also observe that if the right side of (7.271) is taken to define $\mathcal{A}(n)$ for all values of n , then there is the relation

$$\mathcal{A}(-n) = \mathcal{A}(n). \quad (3.7.276)$$

For a discussion of what to make of these results, see the book of P. Cvitanović cited at the end of this chapter.

3.8 Exponential Representations of Group Elements

Lie group elements that are sufficiently near the identity can be written as exponentials of elements in the corresponding Lie algebra. This rule which sends a Lie algebra element into a group element is called the *exponential map*. Can this be done globally? That is, can *every* Lie group element be written as the exponential of some element in the associated Lie algebra? In this section we will answer this question for the Lie groups $SO(n, \mathbb{R})$, $SO(n, \mathbb{C})$, $U(n)$, $SU(n)$, and $Sp(2n, \mathbb{R})$.⁴⁴ The answer for all these groups is *yes* save for $Sp(2n, \mathbb{R})$ where the matter is more complicated. Finally, we note that being global is not the same as being a bijection. As is evident from (7.187) for $SU(2)$, for example, $v(\theta, \mathbf{n}) = v(\theta', \mathbf{n})$ whenever θ and θ' differ by a multiple of 4π . In general, exponentials of various elements in the Lie algebra may result in the same group element.

⁴⁴Exercise 8.2.16 shows that for the case of $SL(2, \mathbb{C})$, the covering group of the Lorentz group, not every element can be written in single exponential form. Nevertheless, every element of the Lorentz group can be written in single exponential form.

3.8.1 Exponential Representation of Orthogonal and Unitary Matrices

Suppose O is an orthogonal matrix with $\det(O) = 1$. Then it can be shown that there is an antisymmetric matrix A such that

$$O = \exp(A). \quad (3.8.1)$$

Conversely, if A is an antisymmetric matrix, then the O given by (8.1) will be orthogonal and have unit determinant. We conclude that every element $O(n) \in SO(n)$ can be written as the exponent of an element $A \in so(n)$.⁴⁵ This is true when working over either the real or the complex field. In particular, if O is real, then there is a real antisymmetric A satisfying (8.1), and A will be complex if O is complex. Finally, consider all elements of the form

$$O(s) = \exp(sA).$$

We see that all elements of $SO(n)$ lie on some one-parameter subgroup of $SO(n)$.

Similarly, suppose U is a unitary matrix. Then it can be shown that there is an anti-Hermitian matrix A such that

$$U = \exp(A). \quad (3.8.2)$$

And, if U has unit determinant, then there is a traceless anti-Hermitian matrix A such that (8.2) holds. Conversely, if A is anti-Hermitian, then the U given by (8.2) will be unitary; and if A is also traceless, then U will have unit determinant as well. We conclude that every element $U \in U(n)$ can be written as the exponent of an element $A \in u(n)$, and every element $U \in SU(n)$ can be written as the exponent of an element $A \in su(n)$. Finally, consider all elements of the form

$$U(s) = \exp(sA).$$

We see that all elements of $U(n)$ lie on some one-parameter subgroup of $U(n)$.

In summary, the exponential maps (8.1) and (8.2) for the orthogonal and unitary groups are *global*. That is, every orthogonal and every unitary matrix can be written as the exponential of some element in the associated Lie algebra.

3.8.2 Exponential Representation of Symplectic Matrices

The case of $Sp(2n, \mathbb{R})$ is more complicated. Again the discussion so far has shown that symplectic matrices sufficiently near the identity element can be written as exponentials of elements in the symplectic group Lie algebra. But what can be said about representing symplectic matrices in general? Thanks to the work of Exercise 7.12 we know that not every symplectic matrix can be written in single exponential form. The purpose of this subsection is to study what can be accomplished.

To proceed, it is useful to employ polar decomposition. See Subsection 6.4 and Section 4.2. Any real nonsingular matrix M can be written uniquely in the form

$$M = PO, \quad (3.8.3)$$

⁴⁵If $\det(O) = -1$, which is the other possibility, define an $n \times n$ diagonal and orthogonal matrix K by the rule $K_{11} = -1$ and all other $K_{jj} = 1$. Then O can be written in the form $O = KO'$ where O' is in $SO(n)$.

where P is a real positive-definite symmetric matrix and O is a real orthogonal matrix. Now suppose that M is symplectic. Using (1.9), the symplectic condition can be written in the form

$$M = J^{-1}(M^T)^{-1}J. \quad (3.8.4)$$

Then, upon inserting the polar decomposition (8.3) into (8.4), one finds the relation

$$PO = J^{-1}P^{-1}JJ^{-1}OJ. \quad (3.8.5)$$

Next, observe that the matrix $J^{-1}P^{-1}J$ is real, symmetric, and positive definite; and observe that the matrix $J^{-1}OJ$ is real and orthogonal. Consequently, because polar decomposition is unique, (8.5) implies the relations

$$P = J^{-1}P^{-1}J, \quad (3.8.6)$$

$$O = J^{-1}OJ. \quad (3.8.7)$$

Using the fact that P is symmetric and O is orthogonal, (8.6) and (8.7) can also be written in form

$$P = J^{-1}(P^T)^{-1}J, \quad (3.8.8)$$

$$O = J^{-1}(O^T)^{-1}J. \quad (3.8.9)$$

It follows that each of the matrices P and O are themselves symplectic.

The next thing to do is to work with the matrices O and P . Consider first the matrix O . Since O is real orthogonal and has determinant +1 (O is symplectic), it can be written in the form

$$O = \exp(F), \quad (3.8.10)$$

where F is a real antisymmetric matrix,

$$F^T = -F. \quad (3.8.11)$$

Upon inserting the representation (8.10) into the condition (8.7), we find the condition

$$O = \exp(F) = \exp(J^{-1}FJ). \quad (3.8.12)$$

Note that the matrix $(J^{-1}FJ)$ is real antisymmetric if the matrix F is. Therefore, in view of (8.12), it is tempting to assume that F has the property

$$F = J^{-1}FJ \text{ or } JF = FJ. \quad (3.8.13)$$

In general this assumption need not be correct because the logarithm of an orthogonal matrix is not unique. However, it will be shown in the next section that we may indeed require (8.13) for the present problem. Using (8.11), the condition (8.13) can also be written in the form

$$F^TJ + JF = 0. \quad (3.8.14)$$

Now compare (8.14) with (7.29) or (7.33). According to the argument employed earlier, the matrix F can be written in the form

$$F = JS^c, \quad (3.8.15)$$

where S^c is a real symmetric matrix. Furthermore, since F commutes with J , see (8.13), it follows that S^c commutes with J ,

$$S^c J = JS^c. \quad (3.8.16)$$

In summary, it has been shown that O can be written in the form

$$O = \exp(JS^c), \quad (3.8.17)$$

where S^c is a real symmetric matrix that commutes with J .

It remains to see what can be said about the matrix P . Since P is real, symmetric, and positive definite, it can be written in the form

$$P = \exp(G), \quad (3.8.18)$$

where G is real and symmetric,

$$G^T = G. \quad (3.8.19)$$

Moreover, it can be shown that the real and symmetric logarithm of a real symmetric positive definite matrix is unique. Now insert the representation (8.18) into the condition (8.8) to obtain the result

$$P = \exp(G) = \exp(-J^{-1}GJ). \quad (3.8.20)$$

Since the matrix $(-J^{-1}GJ)$ is real symmetric if the matrix G is, and since G is unique, it follows from (8.20) that G has the property

$$(-J^{-1}GJ) = G \text{ or } GJ + JG = 0. \quad (3.8.21)$$

Using (8.19), the condition (8.21) can be re-expressed in the form

$$G^T J + JG = 0. \quad (3.8.22)$$

Consequently, G can also be written in the form

$$G = JS^a, \quad (3.8.23)$$

where S^a is a real symmetric matrix. However, in this case (8.19) implies the condition

$$JS^a + S^aJ = 0. \quad (3.8.24)$$

That is, S^a anticommutes with J . In summary, it has been shown that P can be written in the form

$$P = \exp(JS^a), \quad (3.8.25)$$

where S^a is a real symmetric matrix that anticommutes with J .

Now combine (8.3), (8.17), and (8.25). The result is that any real symplectic matrix can be written in the form

$$M = \exp(JS^a) \exp(JS^c). \quad (3.8.26)$$

It has been shown that the most general real symplectic matrix can be written as the product of two exponentials of elements in the real symplectic group Lie algebra, and each of the elements is of a special type. See Exercise 5.10.16 for analogous results for the complex case.

It is interesting to examine the properties of commuting and anticommuting with J in a bit more detail. Let S be any symmetric matrix. Form the matrices S^a and S^c by the rules

$$\begin{aligned} S^a &= (S - J^{-1}SJ)/2, \\ S^c &= (S + J^{-1}SJ)/2. \end{aligned} \quad (3.8.27)$$

It is easily verified that S^a and S^c are symmetric and anticommute and commute respectively with J as the notation suggests. And, if we wish, we may express the properties of anticommuting and commuting by the relations

$$\begin{aligned} JS^aJ^{-1} &= -S^a, \\ JS^cJ^{-1} &= S^c. \end{aligned} \quad (3.8.28)$$

Also, it is obvious by construction that

$$S = S^a + S^c. \quad (3.8.29)$$

That is, any symmetric matrix can be uniquely decomposed into a sum of two symmetric matrices that anticommute and commute with J respectively.

We have seen, according to (8.26), that any real symplectic matrix can be written as the product of two symplectic matrices, each itself written in exponential form with the exponent being a real Hamiltonian matrix. We also know, according to (7.36), that any symplectic matrix sufficiently near the identity can be written in single exponential form. Moreover, the two exponentials appearing in (8.26) can, in principle, be combined into a single exponential using the Baker-Campbell-Hausdorff formula (7.41) providing the series converges. Finally, we know from Exercise 7.12 that not every symplectic matrix can be written as the exponential of a Hamiltonian matrix. Consequently, for $Sp(2n, \mathbb{R})$, there must be cases in which the Baker-Campbell-Hausdorff series diverges.

$$G - I = 0 + O(\epsilon) \quad (3.8.30)$$

We have learned that in the case of $Sp(2n, \mathbb{R})$ the exponential map is *not* global. It follows, unlike the case for $SO(n)$ and $SU(n)$, that not every element of $Sp(2n, \mathbb{R})$ lies on a one-parameter subgroup of $Sp(2n, \mathbb{R})$. Instead, as (8.26) shows, to reach some elements in $Sp(2n, \mathbb{R})$ from the identity requires taking a *dogleg* path.

Since the exponential map is *not* global in the case of $Sp(2n, \mathbb{R})$, it is natural to ask under what conditions a symplectic matrix can be written in single exponential form. It is known that any matrix that is invertible (has nonzero determinant, or, equivalently, all its eigenvalues are nonzero) has a logarithm. But, like the case of numbers, this logarithm may be complex even if the matrix is real. Since any symplectic matrix has determinant $+1$, it must have a logarithm. But this logarithm may be complex. It is known that a sufficient, but not necessary condition, for a real invertible matrix to have a real logarithm is that none of its eigenvalues be negative. It is also known that if a real invertible matrix has a real square root, then it has a real logarithm and vice versa. What we are interested in for real symplectic matrices is the possibility of the logarithm being real and Hamiltonian. The analysis required to answer this question is complicated, and beyond the scope of our

present discussion. It is known that if a real symplectic matrix has a real logarithm, then this logarithm will be Hamiltonian. For an analysis of the 2×2 case, see Section 8.7.2. Basically, as one might guess from Exercise 7.12, problems can occur when -1 appears as a repeated eigenvalue in Jordan blocks. Further information may be found in the references listed at the end of this chapter.

Exercises

3.8.1. Prove the statements made in Section 3.8.1 about orthogonal matrices.

3.8.2. Prove the statements made in Section 3.8.1 about unitary matrices.

3.8.3. Show that

$$\exp(\theta J) = I \cos \theta + J \sin \theta. \quad (3.8.31)$$

Note that this result is a generalization of Euler's formula

$$\exp(i\theta) = \cos \theta + i \sin \theta.$$

3.8.4. Show that the matrices J , $-I$, $-J$, and $(1/\sqrt{2})(I \pm J)$ can be written in the form $\exp(JS^c)$. Find S^c in each case.

3.8.5. Verify (8.5).

3.8.6. Verify that $J^{-1}P^{-1}J$ is real, symmetric, and positive definite. Verify that $J^{-1}OJ$ is real and orthogonal.

3.8.7. Verify (8.8) and (8.9), and the claim that O and P are symplectic.

3.8.8. Verify (8.12) using (8.9) and the definition (7.1).

3.8.9. Verify that $(J^{-1}FJ)$ is real antisymmetric if F is.

3.8.10. Verify that $(-J^{-1}GJ)$ is real symmetric if G is.

3.8.11. Verify (8.27) through (8.29).

3.8.12. Show that every matrix of the form $M = \exp(JS^a)$ is symplectic and has all its eigenvalues on the positive real axis. Show that every matrix of the form $M = \exp(JS^c)$ is symplectic, diagonalizable, and has all its eigenvalues on the unit circle. To prove the "diagonalizable" claim you may have to read the next section, Section 9.

3.8.13. Let M and A be the symplectic matrices

$$M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad A = \begin{pmatrix} \tau & 0 \\ 0 & 1/\tau \end{pmatrix}. \quad (3.8.32)$$

Show that there is the symplectic conjugacy relation

$$AMA^{-1} = \begin{pmatrix} 1 & \tau^2 \\ 0 & 1 \end{pmatrix}. \quad (3.8.33)$$

3.8.14. Show that

$$\exp(JS^c) = \cosh(JS^c) + \sinh(JS^c), \quad (3.8.34)$$

$$\exp(JS^a) = \cosh(JS^a) + \sinh(JS^a). \quad (3.8.35)$$

Using the property that J commutes with S^c , show that

$$\begin{aligned} \cosh(JS^c) &= I + (JS^c)^2/2! + (JS^c)^4/4! + \dots \\ &= I + J^2(S^c)^2/2! + J^4(S^c)^4/4! + \dots \\ &= I - (S^c)^2/2! + (S^c)^4/4! + \dots = \cos(S^c), \end{aligned} \quad (3.8.36)$$

$$\begin{aligned} \sinh(JS^c) &= JS^c + (JS^c)^3/3! + \dots \\ &= JS^c + J^3(S^c)^3/3! + \dots \\ &= J[S^c - (S^c)^3/3! + \dots] = J \sin(S^c). \end{aligned} \quad (3.8.37)$$

Thus, show that

$$\exp(JS^c) = \cos(S^c) + J \sin(S^c). \quad (3.8.38)$$

This relation may be viewed as a symplectic Euler formula in which J plays the role of i .

Using the property that J anticommutes with S^a , show that

$$\begin{aligned} \cosh(JS^a) &= I + (JS^a)^2/2! + (JS^a)^4/4! + \dots \\ &= I - J^2(S^a)^2/2! + J^4(S^a)^4/4! + \dots \\ &= I + (S^a)^2/2! + (S^a)^4/4! + \dots = \cosh(S^a), \end{aligned} \quad (3.8.39)$$

$$\begin{aligned} \sinh(JS^a) &= JS^a + (JS^a)^3/3! + \dots \\ &= JS^a - J^3(S^a)^3/3! + \dots \\ &= J[S^a + (S^a)^3/3! + \dots] = J \sinh(S^a). \end{aligned} \quad (3.8.40)$$

Thus, show that

$$\exp(JS^a) = \cosh(S^a) + J \sinh(S^a). \quad (3.8.41)$$

3.8.15. Show that N as given by (5.60) and (5.61) is symplectic. Let E^ℓ be the matrix defined by the relation

$$E^\ell = \begin{pmatrix} 0_1 & & & & \\ & 0_2 & & & \\ & & \ddots & & \\ & & & E_\ell^{[2]} & \\ & & & & \ddots \\ & & & & & 0_{n-1} & \\ & & & & & & 0_n \end{pmatrix}. \quad (3.8.42)$$

Here each 0_ℓ is a 2×2 null matrix, $E_\ell^{[2]}$ is the 2×2 identity matrix,

$$E_\ell^{[2]} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad (3.8.43)$$

and all other entries are zero so that the $2n \times 2n$ identity matrix has the decomposition

$$I = \sum_{\ell=1}^n E^\ell. \quad (3.8.44)$$

Show that N can be written in the exponential form

$$N = \exp(JS^c) \quad (3.8.45)$$

with J given by (2.10) and S^c given by

$$S^c = \sum_{\ell=1}^n \phi_\ell E^\ell. \quad (3.8.46)$$

Note, as the notation is meant to indicate, that S^c commutes with J . Suppose that (5.53) is solved for M to give the result

$$M = ANA^{-1}. \quad (3.8.47)$$

Use this result and (8.44) to show that

$$M = ANA^{-1} = A[\exp(JS^c)]A^{-1} = \exp(AJS^cA^{-1}). \quad (3.8.48)$$

Next show that

$$AJ = J(A^{-1})^T. \quad (3.8.49)$$

Finally, show that M can be written in the form $M = \exp(JS)$ with

$$S = (A^{-1})^T S^c A^{-1}. \quad (3.8.50)$$

3.8.16. This exercise presumes that you have read Exercise 8.15 above. Its purpose is to describe various invariant quadratic forms.

Suppose a real $2n \times 2n$ symplectic matrix M can be written in the exponential form (7.30) with S real and symmetric. Let $z = (z_1, z_2, \dots, z_{2n})$ be the vector formed from the $2n$ variables z_1 through z_{2n} . Define the quadratic form $Q(z)$ by the rule

$$Q(z) = (z, Sz) \quad (3.8.51)$$

where $(*, *)$ denotes the usual real vector inner product. Suppose that \bar{z} is defined in terms of M and z by the rule

$$\bar{z} = Mz. \quad (3.8.52)$$

Show that Q is *invariant* under this action. That is, show that

$$Q(\bar{z}) = Q(Mz) = Q(z). \quad (3.8.53)$$

Begin by verifying that

$$Q(\bar{z}) = (\bar{z}, S\bar{z}) = (Mz, SMz) = (z, M^T SMz). \quad (3.8.54)$$

Next verify that

$$M^T S M = -M^T J J S M = -M^T J M J S = -J J S = S, \quad (3.8.55)$$

from which it follows that

$$Q(\bar{z}) = (\bar{z}, S\bar{z}) = (z, Sz) = Q(z). \quad (3.8.56)$$

Consider the quadratic forms

$$Q_k(z) = (z, J[JS]^k z) \text{ for } k = 1, 2, \dots \quad (3.8.57)$$

and

$$Q'_k(z) = (z, [SJ]^k J z) \text{ for } k = 1, 2, \dots. \quad (3.8.58)$$

Let a and b be any two possibly non-commuting entities. For all integers $k > 0$ verify the identity

$$(ab)^k = a(ba)^{k-1}b. \quad (3.8.59)$$

Use this identity to verify that

$$Q_k(z) = Q'_k(z). \quad (3.8.60)$$

Show that the $Q_k(z)$ are invariant and vanish for *even* k .

Consider the quadratic forms

$$\tilde{Q}_k(z) = (z, JM^k z) \text{ for } k = 1, 2, \dots \quad (3.8.61)$$

and

$$\tilde{Q}'_k(z) = (z, [M^T]^k J z) \text{ for } k = 1, 2, \dots. \quad (3.8.62)$$

Show from the symplectic condition that $\tilde{Q}'_k(z)$ can also be written in the form

$$\tilde{Q}'_k(z) = (z, JM^{-k} z) \text{ for } k = 1, 2, \dots \quad (3.8.63)$$

so that \tilde{Q}_k and \tilde{Q}'_k are analogous. Show that \tilde{Q}_k and \tilde{Q}'_k are also invariant. Note that the construction of these invariants does not require that M can be written in exponential form.

Suppose f is some function that sends the $2n \times 2n$ matrix JS to some other $2n \times 2n$ matrix $f(JS)$, and suppose that JS and $f(JS)$ commute. Suppose also that (7.30) holds. Show that Q_f defined by

$$Q_f(z) = (z, Jf(JS)z) \quad (3.8.64)$$

is then an invariant function. Similarly, suppose g is some function that sends the $2n \times 2n$ matrix M to some other $2n \times 2n$ matrix $g(M)$, and suppose that M and $g(M)$ commute. Show that Q_g defined by

$$Q_g(z) = (z, Jg(M)z) \quad (3.8.65)$$

is then an invariant function. See Exercise 11.4 for an example of such an invariant.

3.8.17. Work Exercises 8.15 and 8.16 above if you have not already done so. One might wonder whether/how all the invariants found in Exercise 8.16 are related. With what we know so far, we can study this question for the case in which all the eigenvalues of M lie on the unit circle and are distinct. In that case we can use the normal-form results of Section 3.5.

Let us begin with the case of $Q(z)$. From (8.49) there is the result

$$Q(z) = (z, Sz) = (z, [A^{-1}]^T S^c A^{-1} z) = (A^{-1} z, S^c A^{-1} z). \quad (3.8.66)$$

Consider the *normalized* quadratic form $Q^{\text{norm}}(z)$ defined by writing

$$Q^{\text{norm}}(z) = (z, S^c z). \quad (3.8.67)$$

Show from (8.41) and (8.45) that there is the result

$$Q^{\text{norm}}(z) = \sum_{\ell=1}^n \phi_\ell(p_\ell^2 + q_\ell^2). \quad (3.8.68)$$

Define transformed variables \hat{z} by writing

$$\hat{z} = A^{-1} z. \quad (3.8.69)$$

With this definition, show that (8.65) can be rewritten in the form

$$Q(z) = (A^{-1} z, S^c A^{-1} z) = (\hat{z}, S^c \hat{z}) = Q^{\text{norm}}(\hat{z}) = \sum_{\ell=1}^n \phi_\ell(\hat{p}_\ell^2 + \hat{q}_\ell^2). \quad (3.8.70)$$

Let us next consider the Q_k . Show from Exercise 8.15 that

$$JS = AJ S^c A^{-1}, \quad (3.8.71)$$

from which it follows that

$$(JS)^k = A(J S^c)^k A^{-1} \quad (3.8.72)$$

and

$$J(JS)^k = JA(J S^c)^k A^{-1} = (A^{-1})^T J(J S^c)^k A^{-1}. \quad (3.8.73)$$

Consequently, show that

$$Q_k(z) = (z, J(JS)^k z) = (z, (A^{-1})^T J(J S^c)^k A^{-1} z) = (A^{-1} z, J(J S^c)^k A^{-1} z). \quad (3.8.74)$$

Define $Q_k^{\text{norm}}(z)$ by writing

$$Q_k^{\text{norm}}(z) = (z, J(J S^c)^k z). \quad (3.8.75)$$

Then, again using (8.68), show that

$$Q_k(z) = Q_k^{\text{norm}}(\hat{z}). \quad (3.8.76)$$

We still have to evaluate Q_k^{norm} . Since J and S^c commute, we may write

$$J(J S^c)^k = J^{k+1}(S^c)^k \quad (3.8.77)$$

and therefore

$$Q_k^{\text{norm}}(z) = (z, J^{k+1}(S^c)^k z). \quad (3.8.78)$$

We already know that we only have to deal with the odd k case, in which case

$$J^{k+1} = (-1)^{(k+1)/2} I. \quad (3.8.79)$$

Also, show from (8.45) that

$$(S^c)^k = \sum_{\ell=1}^n (\phi_\ell)^k E^\ell. \quad (3.8.80)$$

Show, therefore, that for odd k there is the result

$$Q_k^{\text{norm}}(z) = (-1)^{(k+1)/2} \sum_{\ell=1}^n (\phi_\ell)^k (p_\ell^2 + q_\ell^2). \quad (3.8.81)$$

It follows that

$$Q_k(z) = (-1)^{(k+1)/2} \sum_{\ell=1}^n (\phi_\ell)^k (\hat{p}_\ell^2 + \hat{q}_\ell^2). \quad (3.8.82)$$

The last task is to consider \tilde{Q}_k and \tilde{Q}'_k . From the representation (8.46) show that

$$M^k = (ANA^{-1})^k = AN^k A^{-1} \quad (3.8.83)$$

and consequently

$$\begin{aligned} \tilde{Q}_k(z) &= (z, JM^k z) = (z, JAN^k A^{-1} z) = (z, [A^{-1}]^T J N^k A^{-1} z) \\ &= (A^{-1} z, J N^k A^{-1} z) = (\hat{z}, J N^k \hat{z}) = \tilde{Q}_k^{\text{norm}}(\hat{z}) \end{aligned} \quad (3.8.84)$$

where

$$\tilde{Q}_k^{\text{norm}}(z) = (z, J N^k z). \quad (3.8.85)$$

Let us write N as given by (5.60) in the more explicit form

$$N(\phi_1, \phi_2, \dots, \phi_n) = \begin{pmatrix} R_1(\phi_1) & & & \\ & R_2(\phi_2) & & \\ & & \ddots & \\ & & & R_n(\phi_n) \end{pmatrix} \quad (3.8.86)$$

to emphasize that it depends on n angles ϕ_1 through ϕ_n . Verify that

$$[N(\phi_1, \phi_2, \dots, \phi_n)]^k = N(k\phi_1, k\phi_2, \dots, k\phi_n) \quad (3.8.87)$$

so that

$$\tilde{Q}_k^{\text{norm}}(z) = (z, J N^k z) = (z, J N(k\phi_1, k\phi_2, \dots, k\phi_n) z). \quad (3.8.88)$$

Show from (5.60) and (5.61) that there is the result

$$(z, J N z) = - \sum_{\ell=1}^n (\sin \phi_\ell)(p_\ell^2 + q_\ell^2), \quad (3.8.89)$$

and therefore

$$(z, JN(k\phi_1, k\phi_2, \dots, k\phi_n)z) = - \sum_{\ell=1}^n (\sin k\phi_\ell)(p_\ell^2 + q_\ell^2). \quad (3.8.90)$$

You have shown that

$$\tilde{Q}_k(z) = - \sum_{\ell=1}^n (\sin k\phi_\ell)(\hat{p}_\ell^2 + \hat{q}_\ell^2). \quad (3.8.91)$$

Verify also, in view of (8.62), that

$$\tilde{Q}'_k(z) = \sum_{\ell=1}^n (\sin k\phi_\ell)(\hat{p}_\ell^2 + \hat{q}_\ell^2). \quad (3.8.92)$$

At this point it is evident that all the invariant quadratic forms found above involve the n quantities $(\hat{p}_\ell^2 + \hat{q}_\ell^2)$. You are now to show that these n quantities themselves are also invariant under the action of M . In particular, define quadratic forms $I_\ell(z)$ by the rule

$$I_\ell(z) = (z, [A^{-1}]^T E_\ell A^{-1} z). \quad (3.8.93)$$

Your task is to show that these n quadratic forms are equal to the $(\hat{p}_\ell^2 + \hat{q}_\ell^2)$, and are also invariant under the action of M . In so doing, you will have shown that all the (infinite in number) invariant quadratic forms found above are functions of the n functionally independent invariants I_ℓ .

Begin by verifying that

$$I_\ell(z) = (z, [A^{-1}]^T E_\ell A^{-1} z) = (A^{-1}z, E_\ell A^{-1} z) = (\hat{z}, E_\ell \hat{z}) = \hat{p}_\ell^2 + \hat{q}_\ell^2. \quad (3.8.94)$$

Next, as preparatory steps, show that N commutes with each E_ℓ and that N is orthogonal. Finally, verify that the I_ℓ are invariant under the action of M by checking that

$$\begin{aligned} I_\ell(Mz) &= (Mz, [A^{-1}]^T E_\ell A^{-1} Mz) = (AN A^{-1} z, [A^{-1}]^T E_\ell A^{-1} AN A^{-1} z) \\ &= (A^{-1} z, N^T A^T [A^{-1}]^T E_\ell N A^{-1} z) = (A^{-1} z, N^T E_\ell N A^{-1} z) \\ &= (A^{-1} z, N^T N E_\ell A^{-1} z) = (A^{-1} z, E_\ell A^{-1} z) = (z, [A^{-1}]^T E_\ell A^{-1} z) \\ &= I_\ell(z). \end{aligned} \quad (3.8.95)$$

Here again you will need to use the representation (8.46).

3.8.18. This exercise is devoted to the Krein-Moser theorem. It presumes that you have worked Exercises 8.16 and 8.17 above.

Let M be a real symplectic matrix all of whose eigenvalues are distinct, lie on the unit circle, and are different from ± 1 . Suppose we compute the quadratic form $\tilde{Q}_1(z)$ as given by (8.60). For notational convenience we will simply call it Q . Then we know from (8.90) that it has the representation

$$Q(z) = - \sum_{\ell=1}^n (\sin \phi_\ell)(\hat{p}_\ell^2 + \hat{q}_\ell^2). \quad (3.8.96)$$

We have agreed to employ the range $\phi_\ell \in (-\pi, \pi)$ and to exclude the possibilities $\phi_\ell = 0$ and $\phi_\ell = \pm\pi$. Show that as a consequence there is the relation

$$\text{sign}(\sin \phi_\ell) = \text{sign}(\phi_\ell). \quad (3.8.97)$$

It follows that $Q(z)$ is a negative-definite quadratic form if all phase advances are positive, and a positive-definite quadratic form if all phase advances are negative. If some phase advances are positive and some are negative, then $Q(z)$ is an indefinite quadratic form.

Now suppose, for example, that all phase advances are positive, and that two of them, say ϕ_1 and ϕ_2 , are nearly equal. Then the eigenvalues λ_1 and λ_2 associated with each of them, as given by (5.39), are very nearly equal so that they are likely to collide if M is perturbed. According to the discussion in Section 3.5, these two eigenvalues will have the same signature, namely +1. There will be another pair λ_{-1} and λ_{-2} given by (5.37). They will also have the same signature, namely -1, and they will also collide if the first pair collides. Suppose that each pair does collide under perturbation of M , and afterward, contrary to the Krein-Moser theorem, each pair leaves the unit circle to form a Krein quartet. See Figure 5.1. Then there will be two eigenvalues, call them λ_+ and $\bar{\lambda}_+$, such that

$$|\lambda_+| = |\bar{\lambda}_+| > 1. \quad (3.8.98)$$

Show that, correspondingly, there will then be an initial condition z^0 such that the distance from the origin of the points z^k given by

$$z^k = M^k z^0 \quad (3.8.99)$$

grows without bound as $k \rightarrow \infty$.

Another more delicate situation that we need to consider is that M becomes undiagonalizable when some eigenvalues coincide so that they are no longer distinct. Then the best that can be achieved for M is that it can be brought to Jordan normal form with some +1's above the diagonal. Show that there will again be an initial condition z^0 such that the distance from the origin of the points z^k given by (8.98) grows without bound as $k \rightarrow \infty$.

Under the assumptions made, Q is negative definite before M is perturbed. By continuity, it will remain negative definite under perturbation of M . See Appendix O. But now we have reached a contradiction. Show that if z^k grows without bound as $k \rightarrow \infty$ and Q is negative definite, then $Q(z^k)$ must become ever more negative as $k \rightarrow \infty$. Indeed, suppose that after M is perturbed we use the representation (O.7) for Q .⁴⁶ We know that all the σ_j will be negative because Q is negative definite. Let s_{\min} be the minimum of the quantities $-\sigma_j$. Show that

$$-Q(z) \geq s_{\min} \|z\|^2. \quad (3.8.100)$$

But, because Q is invariant, we must also have the relation

$$Q(z^k) = Q(z^0) \quad (3.8.101)$$

so that $Q(z^k)$ must, in fact, remain constant as $k \rightarrow \infty$. It follows that the eigenvalues associated with ϕ_1 and ϕ_2 cannot leave the unit circle as M is perturbed, nor can M become undiagonalizable.

⁴⁶The representation (8.95) cannot be employed in this case because its construction required that all the eigenvalues be on the unit circle and distinct.

Carry out similar reasoning for the case where all phase advances are negative. Finally, show that Q is indefinite if some phase advances are positive and some are negative so that the above reasoning cannot be applied in that case. In fact, Exercise 25.2.9 provides an example for which two pairs of eigenvalues of opposite signature do indeed collide and then leave the unit circle to become a Krein quartet.

Suppose that Q is indefinite before M is perturbed. Suppose also that two pairs of eigenvalues do come off the unit circle when M is perturbed. Show that Q must then remain indefinite after M is perturbed. Indeed, show that there is a contradiction if Q becomes definite.

We have used the invariant quadratic form $\tilde{Q}_1(z)$ in all our analysis above. Show that $\tilde{Q}'_1(z)$, and the $Q_c(z)$ described in Exercise 11.4 of Section 3.11, could also have been used. Note that we need an invariant form that is defined in terms of M itself since, at least without further work, we cannot presume that a suitable S , as employed to construct the $Q_k(z)$ in Exercise 8.16, can be found after M is perturbed. Recall that in that exercise our construction of S itself assumed that the eigenvalues of M were on the unit circle and distinct. What is needed, if we are not sure this is the case, is some other way of constructing S from M , say by proving the existence of $\log M$ and verifying that it has various desired properties.

3.9 Unitary Subgroup Structure

It is easily verified that the commutator of any two matrices of the form JS^c is again a matrix of the form JS^c . Consequently, matrices of the form JS^c constitute a Lie algebra all by themselves. By contrast, the commutator of a matrix of the form JS^c with that of the form JS^a is again a matrix of the form JS^a . Finally, the commutator of two matrices of the form JS^a is a matrix of the form JS^c . We summarize these results by writing the relations

$$\{JS^c, JS^{c'}\} \propto JS^{c''}, \quad (3.9.1)$$

$$\{JS^c, JS^a\} \propto JS^{a'}, \quad (3.9.2)$$

$$\{JS^a, JS^{a'}\} \propto JS^c. \quad (3.9.3)$$

Since matrices of the form JS^c form a Lie algebra, their exponentials must form a group, and this group will be a subgroup of the full symplectic group. Let us call this subgroup H . We know that it is symplectic and, since it arose from polar decomposition [see (8.15)], it is also orthogonal.⁴⁷ Therefore H is in the intersection of the orthogonal and symplectic groups,

$$H = O(2n, \mathbb{R}) \cap Sp(2n, \mathbb{R}). \quad (3.9.4)$$

The purpose of this section is to study H . For this study it is useful to employ the form (1.1) for J . We will find that H is isomorphic to the unitary group $U(n)$.

The most general $2n \times 2n$ real symmetric matrix S can be written in the block form

$$S = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}, \quad (3.9.5)$$

⁴⁷Note also that matrices of the form JS^c are antisymmetric and therefore, when exponentiated, must produce orthogonal matrices.

where the matrices A , B , and C are $n \times n$ and real, and the matrices A and C are themselves symmetric,

$$A^T = A, \quad (3.9.6)$$

$$C^T = C. \quad (3.9.7)$$

Requiring that J commute with S gives the restrictions

$$B^T = -B \quad (3.9.8)$$

$$C = A. \quad (3.9.9)$$

Thus, the most general S^c is of the form

$$S^c = \begin{pmatrix} A & B \\ -B & A \end{pmatrix} \quad (3.9.10)$$

with the restrictions (9.6) and (9.8). Correspondingly, JS^c is of the form

$$JS^c = \begin{pmatrix} -B & A \\ -A & -B \end{pmatrix}. \quad (3.9.11)$$

Let W be the unitary and (complex) symplectic matrix

$$W = \frac{1}{\sqrt{2}} \begin{pmatrix} I & iI \\ iI & I \end{pmatrix}. \quad (3.9.12)$$

Here each block in W is $n \times n$. Then it is easily verified that the similarity transformation produced by W brings matrices of the form JS^c to block diagonal form. From (9.11) and (9.12) we find the result

$$W^{-1}(JS^c)W = \begin{pmatrix} -B + iA & 0 \\ 0 & -B - iA \end{pmatrix}. \quad (3.9.13)$$

Here each block is again $n \times n$. Now observe that matrices of the form $-B + iA$ with A and B real and obeying (9.6) and (9.8) span the space of all $n \times n$ anti-Hermitian matrices. Consequently, upon exponentiation, matrices of the form $-B + iA$ generate the unitary group $U(n)$. Correspondingly, the matrices $-B - iA$ generate the complex conjugate representation for which we employ the abusive notation $\bar{U}(n)$. Therefore, the Lie algebra spanned by the matrices JS^c is reducible, and is a variant of $u(n)$, the Lie algebra of $U(n)$.

To see how this works in more detail, exponentiate both sides of (9.13) to get the result

$$\exp[W^{-1}(JS^c)W] = \begin{pmatrix} \exp(-B + iA) & 0 \\ 0 & \exp(-B - iA) \end{pmatrix}. \quad (3.9.14)$$

Define a matrix v by the rule

$$v = \exp(-B + iA). \quad (3.9.15)$$

As described earlier, v is unitary as a result of (9.6) and (9.8),

$$v^\dagger = v^{-1}. \quad (3.9.16)$$

Also, any unitary matrix can be written in the form (9.15). Next, observe that the left side of (9.14) can be written in the form

$$\exp[W^{-1}(JS^c)W] = W^{-1} \exp(JS^c)W = W^{-1}MW. \quad (3.9.17)$$

Finally, solving (9.17) and (9.14) for M gives the result

$$M(v) = W \begin{pmatrix} v & 0 \\ 0 & \bar{v} \end{pmatrix} W^{-1}. \quad (3.9.18)$$

Suppose m is an arbitrary $n \times n$ matrix with possibly complex entries. Define an associated $2n \times 2n$ matrix $M(m)$ by the rule

$$M(m) = W \begin{pmatrix} m & 0 \\ 0 & \bar{m} \end{pmatrix} W^{-1}. \quad (3.9.19)$$

Then it is easily verified that there are the relations

$$M(I_n) = I_{2n}, \quad (3.9.20)$$

$$M(m_1 m_2) = M(m_1)M(m_2), \quad (3.9.21)$$

$$M(m^{-1}) = M^{-1}(m), \quad (3.9.22)$$

$$M^\dagger(m) = M(m^\dagger). \quad (3.9.23)$$

Here I_n denotes the $n \times n$ identity matrix. Also, if (9.19) is multiplied out explicitly, we find the result

$$M(m) = \begin{pmatrix} \operatorname{Re}(m) & \operatorname{Im}(m) \\ -\operatorname{Im}(m) & \operatorname{Re}(m) \end{pmatrix}. \quad (3.9.24)$$

It follows that $M(m)$ is real for any m . Consequently, we also have the relation

$$M^T(m) = M^\dagger(m) = M(m^\dagger). \quad (3.9.25)$$

Use of (9.24) for the case $m = iI_n$ gives the result

$$M(iI_n) = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix} = J. \quad (3.9.26)$$

[Note that the matrix (iI_n) is unitary. Note also that (9.26) is consistent with J providing an almost complex structure. See Exercise 2.6.] Suppose we compute M^TJM . By using (9.21), (9.25), and (9.26), we find the result

$$\begin{aligned} M^T(m)JM(m) &= M(m^\dagger)M(iI_n)M(m) = M[m^\dagger(iI_n)m] \\ &= M[m^\dagger m(iI_n)] = M(m^\dagger m)M(iI_n) \\ &= M(m^\dagger m)J. \end{aligned} \quad (3.9.27)$$

However, from (9.24) we also have the result

$$M(m^\dagger m) = \begin{pmatrix} \operatorname{Re}(m^\dagger m) & \operatorname{Im}(m^\dagger m) \\ -\operatorname{Im}(m^\dagger m) & \operatorname{Re}(m^\dagger m) \end{pmatrix}. \quad (3.9.28)$$

Consequently, inspection of (9.27) and (9.28) shows that a necessary and sufficient condition for $M(m)$ to be symplectic is that m be unitary,

$$m^\dagger m = I. \quad (3.9.29)$$

Also, if m is unitary, then use of (9.25) and (9.22) gives the result

$$M^T(m) = M^\dagger(m) = M(m^\dagger) = M(m^{-1}) = M^{-1}(m). \quad (3.9.30)$$

Thus $M(m)$ is also orthogonal if m is unitary.

Conversely, suppose M is a real symplectic matrix that is also orthogonal. Write M in the form

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad (3.9.31)$$

where the matrices a, b, c , and d are real and $n \times n$. Impose on M the condition (8.7), which is equivalent to M being both symplectic and orthogonal. See (1.9). Doing so gives the results

$$c = -b, \quad d = a. \quad (3.9.32)$$

Consequently, a real symplectic orthogonal M must be of the form

$$M = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}. \quad (3.9.33)$$

Next, define the $n \times n$ matrix m by the relation

$$m = a + ib. \quad (3.9.34)$$

As a result of (9.33) and (9.34), M can be written in the form

$$M = M(m) \quad (3.9.35)$$

with $M(m)$ defined by (9.24). Finally, as has been seen, use of the symplectic condition for M implies that m is unitary, so (9.30) also holds. Thus, we conclude that (9.24), (9.34), and (9.35) give a one-to-one correspondence between $2n \times 2n$ real symplectic orthogonal matrices and $n \times n$ unitary matrices. Moreover, the relations (9.20) through (9.22) show that this correspondence is an isomorphism. More precisely, the set of $2n \times 2n$ real symplectic orthogonal matrices forms a group that is the representation $U(n) \oplus \bar{U}(n)$ of $U(n)$. We will sometimes refer to these matrices, which form the subgroup we have called H , as the $U(n)$ subgroup of $Sp(2n, \mathbb{R})$.

At this point we can also provide another proof of the fact that symplectic matrices must have determinant ± 1 . For simplicity, we will restrict our discussion to the case of real symplectic matrices. Suppose M is written in the polar form (8.3). Then we know from (8.8) and (8.9) that both factors P and O are symplectic and hence, according to (1.8), must satisfy the relations

$$\det P = \pm 1, \quad (3.9.36)$$

$$\det O = \pm 1. \quad (3.9.37)$$

But since P is real positive definite and symmetric, its eigenvalues must be real and positive, and hence its determinant must be positive. Thus we must have the relation

$$\det P = +1. \quad (3.9.38)$$

Next consider the matrix O , which is symplectic and real orthogonal. According to (9.35) and (9.19), O can be written in the form

$$O = W \begin{pmatrix} m & 0 \\ 0 & \bar{m} \end{pmatrix} W^{-1}. \quad (3.9.39)$$

Now take the determinant of both sides of (9.39). Doing so gives the result

$$\begin{aligned} \det(O) &= [\det(W)][\det(m)][\det(\bar{m})][\det(W^{-1})] \\ &= [\det(m)][\det(\bar{m})] = |\det(m)|^2 \geq 0. \end{aligned} \quad (3.9.40)$$

Comparison of (9.37) and (9.40) gives the result

$$\det O = +1. \quad (3.9.41)$$

Note that (9.40) and (9.41) are consistent with the fact that $|\det(m)|^2 = 1$ for any unitary matrix m . And from (8.3), (9.38), and (9.41) we conclude that

$$\det M = +1. \quad (3.9.42)$$

Finally it remains to be shown, as promised, that a real symplectic orthogonal matrix M can be written in the form (8.10) with F real and satisfying (8.11) and (8.13). As has been seen, such an M can be written in the form (9.19) with m unitary. Since m is unitary, there exist real matrices A and B satisfying (9.6) and (9.8) such that m can be written in the form

$$m = \exp(-B + iA). \quad (3.9.43)$$

Correspondingly, using (9.43) and (9.19), M can be written in the form

$$M = W \begin{pmatrix} \exp(-B + iA) & 0 \\ 0 & \exp(-B - iA) \end{pmatrix} W^{-1}. \quad (3.9.44)$$

However, the right side of (9.44) can be manipulated to take the form

$$W \begin{pmatrix} \exp(-B + iA) & 0 \\ 0 & \exp(-B - iA) \end{pmatrix} W^{-1} = W \exp(H) W^{-1} = \exp(W H W^{-1}), \quad (3.9.45)$$

where here H is a matrix defined by the relation

$$H = \begin{pmatrix} -B + iA & 0 \\ 0 & -B - iA \end{pmatrix}. \quad (3.9.46)$$

Define a matrix F by the relation

$$F = W H W^{-1}. \quad (3.9.47)$$

Then, use of (9.44) through (9.47) gives the result

$$M = \exp(F). \quad (3.9.48)$$

Also, explicit calculation using (9.46), (9.47), and (9.12) gives the result

$$F = \begin{pmatrix} -B & A \\ -A & -B \end{pmatrix}. \quad (3.9.49)$$

It is readily verified from (9.6) and (9.8) that F satisfies (8.11) and (8.13). Finally, if this F is used in (8.15) to solve for S^c , one finds the result (9.10).

We close this section with one last observation. Consider the $n \times n$ diagonal unitary matrix v given by the relation

$$v(\phi_1, \phi_2, \dots, \phi_n) = \begin{pmatrix} \exp(i\phi_1) & & & \\ & \exp(i\phi_2) & & \\ & & \ddots & \\ & & & \exp(i\phi_n) \end{pmatrix}. \quad (3.9.50)$$

Let $V(\phi_1, \phi_2, \dots, \phi_n)$ be the associated real symplectic and orthogonal matrix given by the relation

$$V = M(v). \quad (3.9.51)$$

Explicit calculation gives the result

$$V = \begin{pmatrix} \operatorname{Re}(v) & \operatorname{Im}(v) \\ -\operatorname{Im}(v) & \operatorname{Re}(v) \end{pmatrix} = \begin{pmatrix} C & S \\ -S & C \end{pmatrix}. \quad (3.9.52)$$

Here C and S are $n \times n$ diagonal matrices given by the relations

$$C = \begin{pmatrix} \cos(\phi_1) & & & \\ & \cos(\phi_2) & & \\ & & \ddots & \\ & & & \cos(\phi_n) \end{pmatrix}, \quad (3.9.53)$$

$$S = \begin{pmatrix} \sin(\phi_1) & & & \\ & \sin(\phi_2) & & \\ & & \ddots & \\ & & & \sin(\phi_n) \end{pmatrix}. \quad (3.9.54)$$

Let us seek to write V in exponential form. Since V belongs to the $U(n)$ subgroup of $Sp(2n, \mathbb{R})$, there must be a matrix \hat{S}^c such that

$$V = \exp(J\hat{S}^c). \quad (3.9.55)$$

From Exercise 3.8.14 we know that

$$V = \exp(J\hat{S}^c) = \cos(\hat{S}^c) + J \sin(\hat{S}^c). \quad (3.9.56)$$

But we also see from (9.52) that there is the relation

$$V = \begin{pmatrix} C & S \\ -S & C \end{pmatrix} = \begin{pmatrix} C & 0 \\ 0 & C \end{pmatrix} + \begin{pmatrix} 0 & S \\ -S & 0 \end{pmatrix} = \begin{pmatrix} C & 0 \\ 0 & C \end{pmatrix} + J \begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix}. \quad (3.9.57)$$

Upon comparing (9.56) and (9.57) we find that

$$\cos(\hat{S}^c) = \begin{pmatrix} C & 0 \\ 0 & C \end{pmatrix} \quad (3.9.58)$$

and

$$\sin(\hat{S}^c) = \begin{pmatrix} S & 0 \\ 0 & S \end{pmatrix}. \quad (3.9.59)$$

It follows that

$$\hat{S}^c = \begin{pmatrix} \phi & 0 \\ 0 & \phi \end{pmatrix} \quad (3.9.60)$$

where ϕ is the diagonal matrix

$$\phi = \begin{pmatrix} \phi_1 & & & \\ & \phi_2 & & \\ & & \ddots & \\ & & & \phi_n \end{pmatrix}. \quad (3.9.61)$$

Evidently incrementing any of the angles ϕ_ℓ by 2π brings V (or v) back to itself. Thus these elements form an *n-torus* within $Sp(2n, \mathbb{R})$. An *n-torus* is the topological product of *n* circles, has dimension *n*, and will be denoted by the symbol T^n . The *n-torus* $V(\phi_1, \phi_2, \dots, \phi_n)$, with each ϕ_ℓ ranging over $[0, 2\pi]$, is called a *maximal* torus within $Sp(2n, R)$ because there is no torus within $Sp(2n, \mathbb{R})$ having a dimension larger than *n*.

By construction, V is symplectic with respect to the J given by (1.1). Let us find the corresponding V' that is symplectic with respect to the J' given by (2.10). According to (2.15), it is given by the relation

$$V' = PVP^T \quad (3.9.62)$$

where, here, P is the permutation matrix of Section 3.2. Note that, since P is orthogonal, V' will also be orthogonal. It is easily verified that carrying out the calculation (9.62) gives the result

$$V'(\phi_1, \phi_2, \dots, \phi_n) = N(\phi_1, \phi_2, \dots, \phi_n) \quad (3.9.63)$$

where N is given by (8.85). Observe that the normal form N given by (8.85), or by (5.60) and (5.61), is orthogonal and real symplectic for the J' given by (2.10). Moreover, from the work of Section 3.5, we know that any real symplectic M with all eigenvalues distinct and on the unit circle is conjugate to such an N by a real symplectic similarity transformation. See also Exercises (8.9) and (8.12). We conclude that all these matrices are related to the maximal *n-torus* $V(\phi_1, \phi_2, \dots, \phi_n)$.

Exercises

3.9.1. Verify the relations (9.1) through (9.3).

3.9.2. Consider the matrix M written in the form (8.26). Show that it can also be written in the form

$$M = \exp(JS^c) \exp(JS^{a'}). \quad (3.9.64)$$

Find the matrix $S^{a'}$.

Answer: $S^{a'} = [\exp(-JS^c)]S^a[\exp(JS^c)]$. Show that $S^{a'}$ is symmetric and anticommutes with J .

3.9.3. Verify that the requirement that J commute with S does indeed give the restrictions (9.8) and (9.9).

3.9.4. Verify that W as given by (9.12) is unitary. That is, $W^\dagger W = I$. Show also that W is (complex) symplectic. That is, show that W belongs to $Sp(2n, C)$.

3.9.5. Verify (9.13).

3.9.6. Verify (9.17).

3.9.7. Verify (9.20) through (9.23).

3.9.8. Verify (9.24).

3.9.9. Verify (9.25) and (9.30).

3.9.10. Find the dimension of the Lie algebra generated by all $2n \times 2n$ matrices of the form JS^c . Verify that this dimension is the same as that of $u(n)$. See Exercise 7.27. Find the dimension of the vector space spanned by all $2n \times 2n$ matrices of the form JS^a . You should have found the dimensions n^2 and $(n^2 + n)$, respectively. Verify, in accord with (8.29), that their sum is $\dim sp(2n)$ as given by (7.42).

3.9.11. Verify (9.49) starting with (9.43) and (9.12). Verify that F satisfies (8.11) and (8.13).

3.9.12. Use the methods of this section to show that all (real) symplectic matrices of the form $\exp(JS^c)$, i.e. all real symplectic orthogonal matrices, can be brought to the normal form (5.60) and (5.61) even if the eigenvalues are not necessarily distinct. In addition, show that the transforming matrix A can be taken to be both real symplectic and orthogonal. Hint: Use the fact that any unitary matrix can be brought to diagonal form by a unitary similarity transformation.

3.9.13. Suppose M is real orthogonal and symplectic with respect to the J of (1.1). Show that then M' as given by (2.15) is real orthogonal and symplectic with respect to the J' of (2.10), and vice versa.

3.9.14. Show that J belongs to the $U(n)$ subgroup of $Sp(2n, \mathbb{R})$, and also commutes with all matrices in $U(n)$.

3.9.15. Verify the relations (9.36) through (9.42).

3.9.16. Was the condition $\det M = +1$ used to derive (9.48) and (9.49)? Show that (9.48), (9.49), (9.8), and (7.104) imply the relation $\det(M) = +1$.

3.9.17. Verify (9.63).

3.9.18. Refer to Exercise 2.6. Given the real $2n$ -vector z in (2.18), let $w(z)$ denote the complex n -vector given by (2.19). Suppose m is a (possibly complex) $n \times n$ matrix. Let m act on w to get the result

$$\begin{aligned} mw &= [\operatorname{Re}(m) + i\operatorname{Im}(m)][x + iy] \\ &= [\operatorname{Re}(m)x - \operatorname{Im}(m)y] + i[\operatorname{Im}(m)x + \operatorname{Re}(m)y]. \end{aligned} \quad (3.9.65)$$

Define a $2n \times 2n$ real matrix $N(m)$ by the rule

$$N(m) = \begin{pmatrix} \operatorname{Re}(m) & -\operatorname{Im}(m) \\ \operatorname{Im}(m) & \operatorname{Re}(m) \end{pmatrix}. \quad (3.9.66)$$

Prove the relations

$$mw(z) = w(N(m)z), \quad (3.9.67)$$

$$(mw, mw') = (Nz, Nz') + i(Nz, JNz'). \quad (3.9.68)$$

Suppose m is unitary. Show that N is then both orthogonal and symplectic. Refer to (9.23). Show that

$$N(m) = M(\overline{m}), \quad (3.9.69)$$

where an overbar denotes the operation of complex conjugation. Show that if m is unitary, then so is \overline{m} .

3.9.19. The purpose of this exercise is to understand more about the correspondence relation (9.19). Consider the set of all matrices $g \in GL(2n, \mathbb{R})$. Next consider the subset of such matrices that also commute with J . Show that these matrices form a subgroup H of $GL(2n, \mathbb{R})$. But, what is this subgroup H ?

Write g in the block form

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad (3.9.70)$$

where the matrices a, b, c , and d are real and $n \times n$. Show that there are the results

$$Jg = \begin{pmatrix} c & d \\ -a & -b \end{pmatrix}, \quad (3.9.71)$$

and

$$gJ = \begin{pmatrix} -b & a \\ -d & c \end{pmatrix}. \quad (3.9.72)$$

Show that requiring that g commute with J yields the restrictions

$$c = -b \quad (3.9.73)$$

and

$$d = a. \quad (3.9.74)$$

Thus, g is of the form

$$g = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}. \quad (3.9.75)$$

Show that the dimension of H is $2n^2$.

Next define matrices A and B by the rules

$$A = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}, \quad (3.9.76)$$

and

$$B = \begin{pmatrix} b & 0 \\ 0 & b \end{pmatrix}. \quad (3.9.77)$$

Verify that both A and B commute with J . Show that there is also the relation

$$JB = \begin{pmatrix} 0 & b \\ -b & 0 \end{pmatrix}. \quad (3.9.78)$$

Therefore, we may also write

$$g = A + JB. \quad (3.9.79)$$

Suppose g_1 and g_2 are two matrices that commute with J and we use the representation (9.79) to write

$$g_k = A_k + JB_k. \quad (3.9.80)$$

Then, recalling that the A_k and B_k commute with J and that $J^2 = -I$, show that there is the product relation

$$g_1 g_2 = (A_1 A_2 - B_1 B_2) + J(A_1 B_2 + B_1 A_2). \quad (3.9.81)$$

We see that, in (9.80) and (9.81), the matrix J plays a role analogous to the imaginary number i . Recall Exercise 2.6 that dealt with almost complex structure.

This analogy can be made explicit using the machinery of this section. An arbitrary $n \times n$ matrix m with possibly complex entries can be written in the form (9.34) where a and b are real $n \times n$ matrices. Let us multiply two such matrices together. Show that so doing gives the result

$$m_1 m_2 = (a_1 a_2 - b_1 b_2) + i(a_1 b_2 + b_1 a_2). \quad (3.9.82)$$

Note the resemblance between the pairs (9.79), (9.34) and (9.81), (9.82). To pursue the analogy further, verify that there is the relation

$$g = M(m). \quad (3.9.83)$$

Next take the determinant of both sides of (9.83). Show that doing so gives the result

$$\begin{aligned} \det(g) &= [\det(W)][\det(m)][\det(\bar{m})][\det(W^{-1})] \\ &= [\det(m)][\det(\bar{m})] = |\det(m)|^2 \geq 0. \end{aligned} \quad (3.9.84)$$

Matrices of the form (9.34) constitute the group $GL(n, \mathbb{C})$ provided we add the condition

$$\det(m) \neq 0. \quad (3.9.85)$$

In view of (9.20) through (9.22), you have shown that the set of matrices $g \in GL(2n, \mathbb{R}, +)$ that also commute with J constitutes a group that is the representation $GL(n, \mathbb{C}) \oplus GL(n, \mathbb{C})$ of $GL(n, \mathbb{C})$; and (9.19) is the relation that provides the isomorphism between them. Note, as a sanity check, that the dimension of $GL(n, \mathbb{C})$ is $2n^2$, which you have already shown is also the dimension of H .

Suppose we impose the further condition

$$\det(m) = 1. \quad (3.9.86)$$

Show that then

$$\det(g) = 1. \quad (3.9.87)$$

Matrices of the form (9.34) subject to the further condition (9.86) constitute the group $SL(n, \mathbb{C})$. In view of (9.20) through (9.22) and (9.87), you have shown that the set of matrices $g \in SL(2n, \mathbb{R})$ that also commute with J constitutes a group that is isomorphic to $SL(n, \mathbb{C})$. More precisely, the set of matrices $g \in SL(2n, \mathbb{R})$ that also commute with J constitutes a group that is the representation $SL(n, \mathbb{C}) \oplus \overline{SL(n, \mathbb{C})}$ of $SL(n, \mathbb{C})$.

3.9.20. This exercise explores some further properties of the matrix W given by (9.12). To begin, review Exercise 9.4. Next, show that from (9.19) and (9.26) that there is the relation

$$WJW^{-1} = \begin{pmatrix} iI & 0 \\ 0 & -iI \end{pmatrix}. \quad (3.9.88)$$

Thus, W provides a similarity transformation that diagonalizes J .⁴⁸

Suppose a vector w is defined by the rule

$$w = Wz. \quad (3.9.89)$$

Show that if z is given by (1.7.9), then w has the entries

$$w = (1/\sqrt{2})(q_1 + ip_1, \dots, q_n + ip_n; iq_1 + p_1, \dots, iq_n + p_n). \quad (3.9.90)$$

Suppose instead a vector w is defined by the rule

$$w = W^{-1}z. \quad (3.9.91)$$

Show that then

$$w = (1/\sqrt{2})(q_1 - ip_1, \dots, q_n - ip_n; -iq_1 + p_1, \dots, -iq_n + p_n). \quad (3.9.92)$$

Let S^a be the symmetric matrix defined by the rule

$$S^a = \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix}. \quad (3.9.93)$$

⁴⁸In Chapter 27 it will be found that the Lie transformation realization of W diagonalizes all the Lie operators : $(p_j^2 + q_j^2)/2$:

Verify that the matrix JS^a is given by the relation

$$JS^a = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}. \quad (3.9.94)$$

Verify, as the notation indicates, that S^a anticommutes with J . Let $U(\theta)$ be the matrix defined by the relation

$$U(\theta) = \exp(i\theta JS^a). \quad (3.9.95)$$

Verify that U is (complex) symplectic because S^a is symmetric and U is unitary because JS^a is Hermitian. Verify that there is the relation

$$(JS^a)^2 = I. \quad (3.9.96)$$

Use this relation to sum the series implied by (9.95) to find the relation

$$U(\theta) = I \cos(\theta) + iJS^a \sin(\theta) = \begin{pmatrix} \cos(\theta) & i \sin(\theta) \\ i \sin(\theta) & \cos(\theta) \end{pmatrix}. \quad (3.9.97)$$

Show that there is the relation

$$U(\pi/4) = W. \quad (3.9.98)$$

Verify that there is the relation

$$W^4 = U(\pi) = -I. \quad (3.9.99)$$

Suppose w is defined by the rule

$$w = U(\theta)z. \quad (3.9.100)$$

Show that

$$w = (w_1, \dots, w_n; w_{n+1}, \dots, w_{2n}) \quad (3.9.101)$$

with

$$w_a = q_a \cos(\theta) + ip_a \sin(\theta) \quad \text{for } a = 1, n \quad (3.9.102)$$

and

$$w_{n+a} = iq_a \sin(\theta) + p_a \cos(\theta) \quad \text{for } a = 1, n. \quad (3.9.103)$$

3.10 Other Subgroup Structure

Consider symplectic matrices of the form (3.9) through (3.11). We have seen that they generate all symplectic matrices. We will now see that, when taken individually, they generate subgroups.

Consider first matrices of the form (3.9). If M and M' are two such matrices, we find the multiplication rule

$$M'M = \begin{pmatrix} I & B' \\ 0 & I \end{pmatrix} \begin{pmatrix} I & B \\ 0 & I \end{pmatrix} = \begin{pmatrix} I & B' + B \\ 0 & I \end{pmatrix}. \quad (3.10.1)$$

It follows, if the matrices B are taken to be arbitrary $n \times n$ matrices, then matrices of the form (3.9) comprise a group. Moreover, (10.1) shows that the elements of this group commute. That is, the group is Abelian. (See Exercise 7.5 for the definition of *Abelian*). Further thought reveals that this group is isomorphic to the translation group in n^2 dimensions. Finally, if B' and B satisfy (3.12), so does their sum $B' + B$. We conclude that symplectic matrices of the form (3.9) comprise a subgroup of the symplectic group. Moreover, this subgroup is isomorphic to the translation group in $n(n+1)/2$ dimensions.

Suppose B satisfies (3.12). Then the matrix S defined by the equation

$$S = \begin{pmatrix} 0 & 0 \\ 0 & B \end{pmatrix} \quad (3.10.2)$$

is symmetric and satisfies the relation

$$JS = \begin{pmatrix} 0 & B \\ 0 & 0 \end{pmatrix}. \quad (3.10.3)$$

Furthermore, the matrix JS is *nilpotent*. That is, JS satisfies the relation

$$(JS)^2 = \begin{pmatrix} 0 & B \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & B \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0. \quad (3.10.4)$$

Consequently, the exponential of JS is given by the simple relation

$$\exp(JS) = I + JS = \begin{pmatrix} I & B \\ 0 & I \end{pmatrix} = M. \quad (3.10.5)$$

We conclude that symplectic matrices of the form (3.9) can be written in the exponential form (10.5) with S given by (10.2).

Similar statements can be made about matrices of the form (3.10). They also form an Abelian subgroup. They can be written in the form

$$M = \exp(JS) \quad (3.10.6)$$

with S given by the relation

$$S = \begin{pmatrix} -C & 0 \\ 0 & 0 \end{pmatrix}. \quad (3.10.7)$$

Moreover, matrices of the form (3.10) are *conjugate* to matrices of the form (3.9) under the action of J . Compute the matrix $J^{-1}MJ$ with M given by (3.9). Matrix multiplication gives the result

$$J^{-1}MJ = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} \begin{pmatrix} I & B \\ 0 & I \end{pmatrix} \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} = \begin{pmatrix} I & 0 \\ -B & I \end{pmatrix}. \quad (3.10.8)$$

Consider matrices of the form (3.11). Let M and M' be two such matrices. We find the multiplication rule

$$M'M = \begin{pmatrix} A' & 0 \\ 0 & D' \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix} = \begin{pmatrix} A'A & 0 \\ 0 & D'D \end{pmatrix}. \quad (3.10.9)$$

Also, if D and D' satisfy (3.13), we have the result

$$D'D = [(A')^T]^{-1}[(A)^T]^{-1} = [(A'A)^T]^{-1}. \quad (3.10.10)$$

Consequently, matrices of the form (3.11) also form a subgroup. Note that the condition (3.13) places no restrictions on the matrices A save that they be invertible. Also, once A is given, D is completely specified by (3.13). This observation, when combined with the multiplication rule (10.9), shows that the subgroup is isomorphic to $GL(n, \mathbb{R})$, the *general linear* group of $n \times n$ invertible matrices over the real field.

Suppose the (real) matrix A is sufficiently near the identity. Then there is real matrix a such that A can be written in the form

$$A = \exp(a). \quad (3.10.11)$$

From (3.13) we find that D can be written in the form

$$D = \exp(-a^T). \quad (3.10.12)$$

Let S be the symmetric matrix defined by the relation

$$S = \begin{pmatrix} 0 & a^T \\ a & 0 \end{pmatrix}. \quad (3.10.13)$$

Then the matrix JS is given by the relation

$$JS = \begin{pmatrix} a & 0 \\ 0 & -a^T \end{pmatrix}. \quad (3.10.14)$$

Evidently, matrices M of the form (3.11) with A sufficiently near the identity can be written as

$$M = \exp(JS) \quad (3.10.15)$$

with S given by (10.13).

Let M be a symplectic matrix of the form

$$M = \begin{pmatrix} A & B \\ 0 & D \end{pmatrix}, \quad (3.10.16)$$

and let M' be another such matrix. Then we find the multiplication rule

$$M'M = \begin{pmatrix} A' & B' \\ 0 & D' \end{pmatrix} \begin{pmatrix} A & B \\ 0 & D \end{pmatrix} = \begin{pmatrix} A'A & A'B + B'D \\ 0 & D'D \end{pmatrix}. \quad (3.10.17)$$

We conclude that such matrices also form a subgroup. This subgroup is the *semi-direct* product of the subgroups of matrices of the forms (3.11) and (3.9). Note that as far as the subgroup of matrices of the form (3.11) is concerned, the multiplication rule (10.17) is the same as (10.9). That is, the diagonal blocks of (10.9) and (10.17) are the same. However, the upper right block of (10.17) is not the same as that of (10.1), but instead involves A' and D . We see that the subgroup of matrices of the form (3.9) is *transformed* under the

action of the subgroup of matrices of the form (3.11). For this reason, the subgroup product is said to be semi-direct rather than simply *direct*. Sometimes it is convenient to use the matrix identity

$$\begin{pmatrix} A' & 0 \\ 0 & D' \end{pmatrix} \begin{pmatrix} I & B' \\ 0 & I \end{pmatrix} = \begin{pmatrix} A' & A'B' \\ 0 & D' \end{pmatrix}. \quad (3.10.18)$$

[Observe that the right side of (10.18) has the desired subgroup block form (10.16), and the matrices on the left have the subgroup block forms (3.11) and (3.9).] When this done, the only conditions that need to be enforced to ensure symplecticity are of the forms (3.12) and (3.13).

In a similar fashion it can be shown that symplectic matrices of the form

$$M = \begin{pmatrix} A & 0 \\ C & D \end{pmatrix} \quad (3.10.19)$$

also constitute a subgroup. This subgroup is the semi-direct product of the subgroups of matrices of the forms (3.11) and (3.10). For this subgroup it is sometimes convenient to use the matrix identity

$$\begin{pmatrix} A' & 0 \\ 0 & D' \end{pmatrix} \begin{pmatrix} I & 0 \\ C' & I \end{pmatrix} = \begin{pmatrix} A' & 0 \\ D'C' & D' \end{pmatrix}. \quad (3.10.20)$$

Exercises

3.10.1. Strictly speaking, (10.1) shows only that the set of matrices of the form (3.9) is closed under multiplication. Show that the other requirements for a (sub)group are also satisfied. See Section 3.6. Verify that matrices of the form (3.10) also constitute a subgroup. Verify (10.6) through (10.8).

3.10.2. Verify the relations (10.2) through (10.5).

3.10.3. Verify the relations (10.6) through (10.8). Also verify that the requirements for a subgroup are met. [See Exercise (10.1) above.]

3.10.4. Verify the relations (10.9) through (10.15). Also verify that the requirements for a subgroup are met. [See Exercise (10.1) above.]

3.10.5. Verify that symplectic matrices of the form (10.16) constitute a subgroup. [See Exercise (10.1) above.] Also verify that symplectic matrices of the form (10.19) constitute a subgroup.

3.11 Other Factorizations/Decompositions

Sections 3.3.1 and 3.10 demonstrated that usually a symplectic matrix can be written as a product of three symplectic matrices of the form (3.9) through (3.11), and a product of six such factors always suffices. Section 3.8 showed that any symplectic matrix has a polar decomposition, and hence can be written as a product of two symplectic matrices in the form (8.26). The purpose of this section is to describe other possible factorizations/decompositions of symplectic matrices that may be of subsequent use.

3.12 Cayley Representation of Symplectic Matrices

In Sections 3.7 and 3.8 we saw that there is a connection between symplectic matrices and symmetric matrices, namely the relations (7.36) and (8.26). In this section we will find another connection, and in Section 5.13 we will see that this connection is but one of a whole *infinite* family of such connections.⁴⁹ The connection to be described here is based on the *Cayley* representation/transformation.⁵⁰ It is a matrix generalization of the hyperbolic function identity

$$\exp(z) = \cosh(z) + \sinh(z) = [1 + \tanh(z/2)]/[1 - \tanh(z/2)].$$

Let M be a (real) symplectic matrix sufficiently near the identity. Then, according to (7.36), M can be written in the form

$$M = \exp(JS) \quad (3.12.1)$$

with S real and symmetric. Now watch closely. By algebraic manipulation involving properties of the exponential function, we may write the following chain of relations:

$$\begin{aligned} M &= \exp(JS) = [\exp(\frac{1}{2}JS)][\exp(-\frac{1}{2}JS)]^{-1} \\ &= [\cosh(\frac{1}{2}JS) + \sinh(\frac{1}{2}JS)][\cosh(\frac{1}{2}JS) - \sinh(\frac{1}{2}JS)]^{-1} \\ &= [I + \tanh(JS/2)][I - \tanh(JS/2)]^{-1}. \end{aligned} \quad (3.12.2)$$

Next, define a matrix W by the equation

$$W = -J \tanh(JS/2). \quad (3.12.3)$$

Then, we also have the relation

$$JW = \tanh(JS/2). \quad (3.12.4)$$

Consequently, using (12.2) and (12.4), M can be written in the form

$$M = (I + JW)(I - JW)^{-1} = (I - JW)^{-1}(I + JW). \quad (3.12.5)$$

We will call this form the *Cayley* representation of M .⁵¹

The alert reader will have observed that, in going from (12.1) to (12.5), no use was made of the symplectic condition. We now show that M being *symplectic* implies that W is *symmetric*, and vice versa,

$$W = W^T \Leftrightarrow M^TJM = J. \quad (3.12.6)$$

⁴⁹And, in Section 6.7, we will see that there is an analogous connection between symplectic maps and gradient maps.

⁵⁰Arthur Cayley (1821-1895), in his 1858 “A Memoir on the Theory of Matrices”, was the first to define matrices abstractly and to describe general matrix algebra including matrix inversion.

⁵¹The terminology *Cayley transform* or *Cayley trivialization* is also used in the literature.

First, suppose W is symmetric. Then taking the transpose of (12.5) gives the representation

$$\begin{aligned} M^T &= [(I - JW)^T]^{-1}[(I + JW)^T] \\ &= (I + WJ)^{-1}(I - WJ). \end{aligned} \quad (3.12.7)$$

Next use the representations (12.5) and (12.7) to compute the quantity M^TJM . Doing so gives the result

$$M^TJM = (I + WJ)^{-1}(I - WJ)J(I + JW)(I - JW)^{-1}. \quad (3.12.8)$$

Insert judicious factors of $I = J^{-1}J$ into part of (12.8) to get the simplification

$$\begin{aligned} J(I + JW)(I - JW)^{-1} &= J(I + JW)J^{-1}J(I - JW)^{-1}J^{-1}J \\ &= (I + WJ)(I - WJ)^{-1}J. \end{aligned} \quad (3.12.9)$$

Here use has been made of (1.3). Correspondingly, (12.8) now simplifies to the form

$$M^TJM = (I + WJ)^{-1}(I - WJ)(I + WJ)(I - WJ)^{-1}J. \quad (3.12.10)$$

Observe that the second and third factors in the right side of (12.10) commute. Thus, we also have the relation

$$\begin{aligned} M^TJM &= (I + WJ)^{-1}(I + WJ)(I - WJ)(I - WJ)^{-1}J \\ &= J, \end{aligned} \quad (3.12.11)$$

which is what we wanted to prove.

Conversely, suppose that M is symplectic. Solve (12.5) for the quantity JW to get the relation

$$JW = (M + I)^{-1}(M - I) = (M - I)(M + I)^{-1}. \quad (3.12.12)$$

Now take the transpose of (12.12) to get the result

$$-W^TJ = (M^T - I)(M^T + I)^{-1}. \quad (3.12.13)$$

The symplectic condition can be written in the form

$$M^T = JM^{-1}J^{-1}. \quad (3.12.14)$$

See (1.9). Consequently, (12.13) can also be written in the form

$$\begin{aligned} -W^TJ &= (JM^{-1}J^{-1} - I)(JM^{-1}J^{-1} + I)^{-1} \\ &= J(M^{-1} - I)(M^{-1} + I)^{-1}J^{-1} \\ &= J(I - M)M^{-1}[(I + M)M^{-1}]^{-1}J^{-1} \\ &= J(I - M)(I + M)^{-1}J^{-1} \\ &= J(-JW)J^{-1} = -WJ. \end{aligned} \quad (3.12.15)$$

It follows from (12.15) that W is symmetric,

$$W^T = W. \quad (3.12.16)$$

Now consider matrices of the form JW . We know they are Hamiltonian, that is, they belong to the Lie algebra $sp(2n, \mathbb{R})$ [or, more generally, $sp(2n, \mathbb{C})$] if W is symmetric. Since we have seen that W is symmetric if M is symplectic, we conclude that for M sufficiently near the identity matrix and for JW sufficiently near the zero matrix there is the relation

$$M \in Sp(2n, \mathbb{R}) \Leftrightarrow JW \in sp(2n, \mathbb{R}), \quad (3.12.17)$$

or, more generally,

$$M \in Sp(2n, \mathbb{C}) \Leftrightarrow JW \in sp(2n, \mathbb{C}). \quad (3.12.18)$$

Thus, near the identity in group space and near the origin in Lie-algebra space, the Cayley representation, like the exponential map, provides a local bijection between group elements and Lie-algebra elements.

Again we note that the symplectic condition as expressed by (1.2) is a set of quadratic relations, and the use of the Cayley representation converts these quadratic relations into the simple linear relations (12.16).

We also need to make an important observation. It is easily checked that $-I$ is a symplectic matrix. However, $(M + I)$ is singular for $M = -I$. Indeed, $(M + I)$ is singular whenever M has -1 as an eigenvalue. It follows that JW does not exist in these cases. Consequently, unlike the two-exponentials product representation (8.26), the Cayley representation is not global.

We note for future use that (12.12) can be solved for W to give the relation

$$W = (-JM + J)(M + I)^{-1}. \quad (3.12.19)$$

Finally, we observe that (12.5) and the inverse relation (12.19) stand on their own without any need of the motivational assumption (12.1).

We close this section by noting that there are also Cayley representations for all the so-called *quadratic* matrix groups including orthogonal, unitary, and Lorentz transformation matrices. See Exercises 12.5 and 12.6.

Exercises

3.12.1. Show that (12.3) and (12.4) have the expansions

$$\begin{aligned} W &= -J \tanh(JS/2) = -J[(JS/2) - (1/3)(JS/2)^3 + (2/15)(JS/2)^5 - \dots] \\ &= S/2 - SJSJS/24 + SJSJSJSJS/240 - \dots. \end{aligned} \quad (3.12.20)$$

$$\begin{aligned} JW &= \tanh(JS/2) = (JS/2) - (1/3)(JS/2)^3 + (2/15)(JS/2)^5 - \dots \\ &= JS/2 - (JS)^3/24 + (JS)^5/240 - \dots. \end{aligned} \quad (3.12.21)$$

Show directly from (12.20) that W is symmetric if S is. Show from (12.21) that the matrices JW and JS commute,

$$\{JW, JS\} = 0. \quad (3.12.22)$$

Show that (12.20) and (12.21) can be inverted to give the relations

$$JS/2 = \tanh^{-1}(JW) = [JW + (1/3)(JW)^3 + (1/5)(JW)^5 + \dots], \quad (3.12.23)$$

$$JS = 2 \tanh^{-1}(JW) = 2[JW + (1/3)(JW)^3 + (1/5)(JW)^5 + \dots], \quad (3.12.24)$$

$$\begin{aligned} S &= -2J \tanh^{-1}(JW) = -2J[JW + (1/3)(JW)^3 + (1/5)(JW)^5 + \dots] \\ &= 2W + (2/3)WJWJW + (2/5)WJWJWJWJW + \dots. \end{aligned} \quad (3.12.25)$$

Show from (12.25) that, conversely, S is symmetric if W is.

3.12.2. Derive (12.12) from (12.5).

3.12.3. Find the Cayley representation for the matrix N given by (5.60) and (5.61). That is, find the matrix W in this case. Show explicitly that the representation is not global, i.e., does not hold for all values of ϕ_ℓ .

3.12.4. Read Exercise 8.13. Let W and M be the matrices appearing in the Cayley relations (12.5) and (12.19). Define what we will call the Cayley quadratic form Q_c by the relation

$$Q_c(z) = (z, Wz). \quad (3.12.26)$$

Verify that W is of the form

$$W = Jg(M) \quad (3.12.27)$$

with

$$g(M) = (-M + I)(M + I)^{-1} \quad (3.12.28)$$

and that M commutes with $g(M)$. Show that Q_c is invariant under the action of M .

Show, for the case described in Exercise 8.14, that Q_c is given by the relation

$$Q_c(z) = \sum_{\ell=1}^n [\tan(\phi_\ell/2)](\hat{p}_\ell^2 + \hat{q}_\ell^2). \quad (3.12.29)$$

3.12.5. Section 3.12 described the Cayley representation of symplectic matrices. The purpose of this exercise is to explore Cayley representations for other kinds of matrices including orthogonal, unitary, and Lorentz transformation matrices. Let L be any fixed real nonsingular $m \times m$ matrix, and consider all $m \times m$ matrices M such that

$$M^T LM = L. \quad (3.12.30)$$

Show that

$$\det M = \pm 1, \quad (3.12.31)$$

and therefore all such matrices are invertible. Indeed, show from (12.30) that

$$M^{-1} = L^{-1} M^T L. \quad (3.12.32)$$

Note that, while matrix inversion is usually a computationally intensive task, all that is required in this case is the inversion of L , which can be done once and for all, the transposing of M , and two matrix multiplications.

Verify that all matrices that satisfy (12.30) form a group, call it G . Since the relation (12.30), is an algebraic one among the entries in M , G is an algebraic group. Indeed, since (12.30) is a quadratic relation, G is also sometimes called a *quadratic* group. Thus, for example, the orthogonal, symplectic, and Lorentz groups are quadratic groups.

Let $(*, *)$ denote the usual real inner product. Define an angular inner product, more accurately a bilinear form, $\langle *, * \rangle$ by the rule

$$\langle u, v \rangle = (u, Lv). \quad (3.12.33)$$

Verify that

$$\langle Mu, Mv \rangle = (Mu, LMv) = (u, M^T LMv) = (u, Lv) = \langle u, v \rangle. \quad (3.12.34)$$

That is, G preserves the bilinear form $\langle *, * \rangle$.

Show that G consists of two disconnected components comprised of elements with determinant $+1$ and elements with determinant -1 . [Actually, it can happen, as is the case for $Sp(2n, \mathbb{R})$, that the component with determinant -1 is empty.] Show that the matrices $M \in G$ such that $\det M = 1$ form a subgroup, call it SG .

Consider matrices in SG that are sufficiently close to the identity so that they can be written in the exponential form

$$M = \exp(\epsilon A) \quad (3.12.35)$$

where ϵ is a sufficiently small parameter. Show, by equating powers of ϵ , that (12.30) and (12.34) require that A obey the relation

$$A^T L + LA = 0 \quad (3.12.36)$$

or, equivalently,

$$L^{-1} A^T L = -A. \quad (3.12.37)$$

Conversely, show that if A satisfies the relation (12.36), then any M given by (12.34) satisfies (12.30), and therefore belongs to G . Verify that matrices A that satisfy (12.36) form a Lie algebra, and therefore G is a Lie group. Show that (12.36) implies the relation

$$\text{tr } A = 0, \quad (3.12.38)$$

and therefore any M of the form (12.34) belongs to SG . Correspondingly, following our usual nomenclature, we may define sg to be the Lie algebra of all matrices A that satisfy (12.36).

Set $\epsilon = 1$ in (12.34). Following the logic of Section 3.12, show that matrices M sufficiently near the identity can be written in the form

$$M = (I + V)/(I - V) \quad (3.12.39)$$

where

$$V = \tanh(A/2) = (A/2) - (1/3)(A/2)^3 + (2/15)(A/2)^5 + \dots \quad (3.12.40)$$

and

$$A = 2 \tanh^{-1} V = 2[V + (1/3)V^3 + (1/5)V^5 + \dots]. \quad (3.12.41)$$

Show that if A satisfies (12.36), then so does V , and vice versa,

$$L^{-1}A^T L = -A \Leftrightarrow L^{-1}V^T L = -V. \quad (3.12.42)$$

[Here it assumed that A is sufficiently small for the series (12.39) and its inverse relation (12.40) to be convergent.] We conclude that $V \in sg$.

Verify that (12.38) can be solved for V to yield the inverse relation

$$V = (M - I)/(M + I). \quad (3.12.43)$$

Verify that, for M sufficiently near the identity matrix and for V sufficiently near the zero matrix, there is the relation

$$M \in SG \Leftrightarrow V \in sg. \quad (3.12.44)$$

Consequently, for quadratic groups, (12.38) and (12.42) provide a mapping between SG and sg , which is a bijection between elements in SG sufficiently near the identity and elements in sg sufficiently near the origin.

Sometimes it is convenient to define a function that is a variant of the relation (12.38). Given any matrix X that does not have -1 as an eigenvalue, define a matrix function cay by the rule

$$cay(X) = (I - X)/(I + X). \quad (3.12.45)$$

With this definition, (12.38) becomes

$$M = cay(-V) \quad (3.12.46)$$

and (12.42) becomes

$$V = -cay(M). \quad (3.12.47)$$

Show that we may also write

$$cay(V) = M^{-1} \quad (3.12.48)$$

and

$$cay(M^{-1}) = V. \quad (3.12.49)$$

Note that $V \in sg$ implies that $-V \in sg$, and vice versa; and $M \in SG$ implies that $M^{-1} \in SG$, and vice versa. Therefore, in our context, the function cay also provides a bijection between elements in SG sufficiently near the identity and elements in sg sufficiently near the origin. A map or operator whose square is the identity is often called an *involution*. Show that the map cay is an *involution*. That is, show that

$$(cay)^2(X) = cay[cay(X)] = X. \quad (3.12.50)$$

Verify that in the case of $SO(m)$, for which $L = I$, the matrices A and V are antisymmetric.

Scan Exercise 6.2.6. Be aware that in this exercise the symbol g is used to denote the metric tensor. Verify that the Lorentz group is a quadratic group, and therefore has a Cayley representation.

Reapply, with necessary modifications, the arguments made so far to the case of matrices M that satisfy the relation

$$M^\dagger LM = L. \quad (3.12.51)$$

Show that such matrices form a group G and that there is a Cayley representation that provides a map between G and its Lie algebra g . Apply your results to the case of $U(m)$, for which $L = I$, and show that in this case the matrices A and V are anti-Hermitian. That is, $A \in u(m)$ and $V \in u(m)$. Verify that in the quantum-mechanical theory of scattering, for which M is the unitary scattering matrix S , there is the relation

$$S = M = (I + iK)/(I - iK) \quad (3.12.52)$$

where the so called K matrix given by $K = -iV$ is Hermitian. Verify also the inverse relation

$$K = -i(S - I)/(S + I). \quad (3.12.53)$$

The relations (12.51) and (12.52) provide a map between unitary matrices S and Hermitian matrices K . We remark that if the scattering process is time symmetric, then it can be shown that K is real, and therefore also symmetric.

Are there familiar examples of groups for which there is no Cayley representation? It depends what one means by a Cayley representation. If one means that the Cayley relations are required to supply a bijection between the group and its Lie algebra, then there are groups for which there is no Cayley representation in the sense that the Cayley relations do not provide a bijection between the group and its Lie algebra. The groups $SL(m, \mathbb{R})$ for $m > 2$ are examples. In working Exercise 7.25 you should have found that $sl(m, \mathbb{R})$, the Lie algebra of $SL(m, \mathbb{R})$, consists of all real $m \times m$ matrices A that are traceless. Consider the case of $sl(3, \mathbb{R})$ and the Lie algebraic element

$$A = \epsilon \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \quad (3.12.54)$$

where ϵ is small. Verify that V as given by (12.39) is not traceless and therefore does not belong to $sl(3, \mathbb{R})$. Thus, there is no Cayley representation for $SL(3, \mathbb{R})$. That is, although the relations (12.38) and (12.42) continue to hold, there are group elements $M \in SL(3, \mathbb{R})$, and arbitrarily near the identity, for which the corresponding V given by (12.42) is not in $sl(3, \mathbb{R})$.

Verify that $SL(3, \mathbb{R})$ is a subgroup of $SL(m, \mathbb{R})$ for $m > 3$ and therefore there is no Cayley representation for $SL(m, \mathbb{R})$ when $m > 2$. Is $SL(m, \mathbb{R})$ an algebraic group?

3.12.6. The aim of this exercise is to explore in some detail the use of Cayley parameterizations for the cases of $SO(3, \mathbb{R})$ and $SU(2)$.

Assume, for the case of $SO(3, \mathbb{R})$, that group elements are parameterized in *exponential* form by the relation

$$R_e(\boldsymbol{\lambda}) = \exp(\boldsymbol{\lambda} \cdot \mathbf{L}). \quad (3.12.55)$$

Show that they have an associated *Cayley* parameterization of the form

$$R_c(\boldsymbol{\mu}) = (I + \boldsymbol{\mu} \cdot \mathbf{L})/(I - \boldsymbol{\mu} \cdot \mathbf{L}). \quad (3.12.56)$$

Show that if $R_e(\boldsymbol{\lambda}) = R_c(\boldsymbol{\mu}) = R$, then the parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are interconnected by the relations

$$\boldsymbol{\mu} \cdot \mathbf{L} = \tanh(\boldsymbol{\lambda} \cdot \mathbf{L}/2) \quad (3.12.57)$$

and

$$\boldsymbol{\lambda} \cdot \mathbf{L} = 2 \tanh^{-1}(\boldsymbol{\mu} \cdot \mathbf{L}). \quad (3.12.58)$$

Also verify that (12.55) can be inverted to give the relation

$$\boldsymbol{\mu} \cdot \mathbf{L} = [R - I]/[R + I]. \quad (3.12.59)$$

Verify, for the case of $so(3, \mathbb{R})$, that there is the relation

$$(\boldsymbol{\nu} \cdot \mathbf{L})^3 = (i|\boldsymbol{\nu}|)^2 \boldsymbol{\nu} \cdot \mathbf{L} \quad (3.12.60)$$

for any 3-vector $\boldsymbol{\nu}$. See (7.201). Use this relation to show that (12.56) and (12.57) can be rewritten in the forms

$$\boldsymbol{\mu} \cdot \mathbf{L} = (\boldsymbol{\lambda} \cdot \mathbf{L}/|\boldsymbol{\lambda}|) \tan(|\boldsymbol{\lambda}|/2) \quad (3.12.61)$$

and

$$\boldsymbol{\lambda} \cdot \mathbf{L} = 2(\boldsymbol{\mu} \cdot \mathbf{L}/|\boldsymbol{\mu}|) \tan^{-1}(|\boldsymbol{\mu}|). \quad (3.12.62)$$

Hint: Expand (12.56) and (12.57) in Taylor series, use (12.59) in these series, and then sum the transformed series to get the advertised results. Show it follows from (12.60) and (12.61) that

$$\boldsymbol{\mu} = (\boldsymbol{\lambda}/|\boldsymbol{\lambda}|) \tan(|\boldsymbol{\lambda}|/2) \quad (3.12.63)$$

and

$$\boldsymbol{\lambda} = 2(\boldsymbol{\mu}/|\boldsymbol{\mu}|) \tan^{-1}(|\boldsymbol{\mu}|). \quad (3.12.64)$$

Verify that both (12.62) and (12.63) imply the relation

$$|\boldsymbol{\mu}| = \tan(|\boldsymbol{\lambda}|/2). \quad (3.12.65)$$

Observe that (12.64) is singular when $|\boldsymbol{\lambda}| = \pi$. Verify that this singularity is to be expected because then R has -1 as an eigenvalue, from which it follows that the factor $[R + I]^{-1}$ in (12.58) is singular.

Assume, for the case of $SU(2)$, that group elements are parameterized in exponential form by the relation

$$u_e(\boldsymbol{\lambda}) = \exp(\boldsymbol{\lambda} \cdot \mathbf{K}). \quad (3.12.66)$$

In the notation of Exercise 12.5, group elements near the identity have an associated parameterization of the form

$$u_c = (I + V)/(I - V). \quad (3.12.67)$$

Show that setting $u_e = u_c = u$ yields the result

$$V = \tanh(\boldsymbol{\lambda} \cdot \mathbf{K}/2). \quad (3.12.68)$$

In order for this parameterization to be a Cayley parameterization we must verify that $V \in su(2)$. Check, for the case of $su(2)$, that there is the relation

$$(\boldsymbol{\nu} \cdot \mathbf{K})^3 = (i|\boldsymbol{\nu}|/2)^2 \boldsymbol{\nu} \cdot \mathbf{K} \quad (3.12.69)$$

for any 3-vector $\boldsymbol{\nu}$. See (7.187). Use this relation in the Taylor series for the right side of (12.67) to show that (12.67) can be rewritten in the form

$$V = (\boldsymbol{\lambda} \cdot \mathbf{K})(2/|\boldsymbol{\lambda}|) \tan(|\boldsymbol{\lambda}|/4), \quad (3.12.70)$$

from which it follows, in particular, that $V \in su(2)$, and $SU(2)$ has a Cayley parameterization.⁵² Therefore, we may write (12.66) in the Cayley form

$$u_c(\boldsymbol{\mu}) = (I + \boldsymbol{\mu} \cdot \mathbf{K})/(I - \boldsymbol{\mu} \cdot \mathbf{K}) \quad (3.12.71)$$

with

$$\boldsymbol{\mu} \cdot \mathbf{K} = V = (\boldsymbol{\lambda} \cdot \mathbf{K})(2/|\boldsymbol{\lambda}|) \tan(|\boldsymbol{\lambda}|/4). \quad (3.12.72)$$

From (12.71) show that

$$\boldsymbol{\mu} = 2(\boldsymbol{\lambda}/|\boldsymbol{\lambda}|) \tan(|\boldsymbol{\lambda}|/4) \quad (3.12.73)$$

and

$$|\boldsymbol{\mu}| = 2 \tan(|\boldsymbol{\lambda}|/4). \quad (3.12.74)$$

Show also that (12.70) can be inverted to give the relation

$$\boldsymbol{\mu} \cdot \mathbf{K} = [u - I]/[u + I]. \quad (3.12.75)$$

Note that (12.73) is singular when $|\boldsymbol{\lambda}| = 2\pi$. Verify that this singularity is to be expected because then $u = -I$, see (7.189), from which it follows that the factor $[u + I]^{-1}$ in (12.74) is singular. Finally, show that (12.72) and (12.73) can be solved for $\boldsymbol{\lambda}$ to give the inverse relation

$$\boldsymbol{\lambda} = 4(\boldsymbol{\mu}/|\boldsymbol{\mu}|) \tan^{-1}(|\boldsymbol{\mu}|/2). \quad (3.12.76)$$

Note that both (12.62) and (12.72) yield the relation $\boldsymbol{\mu} \simeq \boldsymbol{\lambda}/2$ for small $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$. But they differ in higher order.

3.13 General Symplectic Forms, Darboux Transformations, Pfaffians, and Variant Symplectic Groups

3.13.1 General Symplectic Forms

According to Exercise 2.7, the symplectic group consists of all linear transformations that preserve the fundamental symplectic 2-form (2.3). The matrix J in this 2-form has the property that it is real, antisymmetric, and nonsingular. We will now see that there is an endless supply of matrices K with this property; and we will call each (w, Kz) a *generalized symplectic 2-form*.

First we assert that such a matrix must be $2n \times 2n$ for some choice of n . For suppose K is $m \times m$ with m odd. Then we find that

$$\det K = \det K^T = \det(-K) = (-1)^m \det K = -\det K, \quad (3.13.1)$$

⁵²We remark that although $U(n)$ has a Cayley parameterization, $SU(n)$ does not when $n > 2$.

from which it follows that $\det K = 0$ and therefore K is singular, contrary to one of our stipulations about K .

Next, let N be any matrix in $GL(2n, \mathbb{R})$. Define an associated matrix K by the rule

$$K = NJN^T. \quad (3.13.2)$$

That is, K and J are congruent under the action of N . Evidently K is real. We also find by direct calculation that

$$K^T = (NJN^T)^T = NJ^TN^T = -NJJ^TN^T = -K. \quad (3.13.3)$$

Moreover,

$$\det K = (\det N)(\det J)(\det N^T) = (\det N)^2 > 0. \quad (3.13.4)$$

Therefore K is nonsingular.

The converse is also true. Given any real, antisymmetric, and nonsingular $2n \times 2n$ matrix K , there is a matrix $N \in GL(2n, \mathbb{R})$ such that (13.2) holds.

We begin the demonstration of this claim by showing that there is a set of (real) basis vectors v^1, v^2, \dots, v^{2n} such that

$$(v^i, Kv^j) = J'_{ij}, \quad (3.13.5)$$

where J' is the matrix given by (2.10). The construction of the v^i is very similar to that used for Darboux symplectification. Let w^1, \dots, w^{2n} be any set of $2n$ real and linearly independent vectors. For convenience, they might be taken to be the unit vectors e^1, \dots, e^{2n} given by (6.4). Now follow this algorithm:

1. Define v^1 by the simple rule

$$v^1 = w^1. \quad (3.13.6)$$

2. Starting with w^2 , search through the w^j with $j \geq 2$ to find the first j , call it k , with the property

$$(v^1, Kw^j) \neq 0. \quad (3.13.7)$$

[Better yet, if one is working numerically and therefore only to finite precision, select j so that $|(v^1, Kw^j)|$ is maximized. The analogous choices should also be made in steps 6, 10, etc. below.] Renumber the vectors $w^2 \cdots w^{2n}$ so that w^k becomes w^2 .

3. Define v^2 by the rule

$$v^2 = w^2 / [(v^1, Kw^2)]. \quad (3.13.8)$$

We then have the result

$$(v^1, Kv^2) = 1 = J'_{12}. \quad (3.13.9)$$

And, since K is antisymmetric, at this stage we have the result

$$(v^i, Kv^j) = J'_{ij} \text{ for } i, j = 1 \text{ to } 2. \quad (3.13.10)$$

4. Using the remaining vectors $w^3 \dots w^{2n}$, define new vectors ${}^1w^j$ for $j \geq 3$ by the rule

$${}^1w^j = w^j + (v^2, Kw^j)v^1 - (v^1, Kw^j)v^2. \quad (3.13.11)$$

As a result of this rule there are the relations

$$(v^i, K {}^1w^j) = 0 \text{ for } i = 1, 2 \text{ and } j = 3, 4, \dots, 2n. \quad (3.13.12)$$

5. Define v^3 by the rule

$$v^3 = {}^1w^3. \quad (3.13.13)$$

6. Starting with ${}^1w^4$, search through the ${}^1w^j$ with $j \geq 4$ to find the first j , call it k , with the property

$$(v^3, K {}^1w^j) \neq 0. \quad (3.13.14)$$

Renumber the vectors ${}^1w^4 \dots {}^1w^{2n}$ so that ${}^1w^k$ becomes ${}^1w^4$.

7. Define v^4 by the rule

$$v^4 = {}^1w^4 / [(v^3, K {}^1w^4)]. \quad (3.13.15)$$

At this stage we have the results

$$(v^i, Kv^j) = J'_{ij} \text{ for } i, j = 1 \text{ to } 4. \quad (3.13.16)$$

8. Using the remaining vectors ${}^1w^5 \dots {}^1w^{2n}$, define new vectors ${}^2w^j$ for $j \geq 5$ by the rule

$${}^2w^j = {}^1w^j + (v^4, K {}^1w^j)v^3 - (v^3, K {}^1w^j)v^4. \quad (3.13.17)$$

Now we have the relations

$$(v^i, K {}^2w^j) = 0 \text{ for } i = 1 \text{ to } 4 \text{ and } j = 5, 6, \dots, 2n. \quad (3.13.18)$$

9. Define v^5 by the rule

$$v^5 = {}^2w^5. \quad (3.13.19)$$

10. Starting with ${}^2w^6$, search through the ${}^2w^j$ with $j \geq 6$ to find the first j , call it k , with the property

$$(v^5, K {}^2w^j) \neq 0. \quad (3.13.20)$$

Renumber the vectors ${}^2w^6 \dots {}^2w^{2n}$ so that ${}^2w^k$ becomes ${}^2w^6$.

11. Define v^6 by the rule

$$v^6 = {}^2w^6 / [(v^5, K {}^2w^6)]. \quad (3.13.21)$$

At this stage we have the results

$$(v^i, Kv^j) = J'_{ij} \text{ for } i, j = 1 \text{ to } 6. \quad (3.13.22)$$

12. Proceed with the obvious extension of the above process to construct $v^7, v^8, \dots, v^{2n-2}$. Then at the last stage we have

$$v^{2n-1} = {}^m w^{2n-1}, \quad (3.13.23)$$

$$v^{2n} = {}^m w^{2n}/[(v^{2n-1}, K {}^m w^{2n})], \quad (3.13.24)$$

with

$$m = n - 1. \quad (3.13.25)$$

As was the case with Darboux symplectification, how does one know that the required vectors ${}^m w^k$ described in steps 2, 6, 10, etc. exist? And how does one know that the vectors $v^3, v^5, \dots, v^{2n-1}$ given in steps 5, 9, etc. are nonzero? Again difficulties do not arise because the w^i are assumed to be linearly independent and K is assumed to be invertible. See Exercise 13.1.

Next, let e^j denote the unit column vector with 1 in its j th entry and zeroes elsewhere. See (6.4). Then (13.5) can be written in the form

$$(v^i, Kv^j) = (e^i, J'e^j). \quad (3.13.26)$$

Define a linear transformation L by the rule

$$v^j = Le^j. \quad (3.13.27)$$

It has the matrix elements

$$L_{ij} = (e^i, Le^j) = (e^i, v^j). \quad (3.13.28)$$

Upon inserting (13.27) into (13.5) we find the relation

$$J'_{ij} = (e^i, J'e^j) = (Le^i, KLe^j) = (e^i, L^T KLe^j), \quad (3.13.29)$$

which is equivalent to the matrix relation

$$J' = L^T KL. \quad (3.13.30)$$

We also observe that L is invertible. Indeed, taking the determinant of both sides of (13.30) yields the relation

$$\det J' = (\det L^T)(\det K)(\det L) = (\det K)(\det L)^2, \quad (3.13.31)$$

from which we find the result

$$(\det L)^2 = (\det J')/(\det K) = 1/(\det K) \neq 0 \text{ or } \infty \quad (3.13.32)$$

since K is assumed to be invertible.

We are almost done. Since L is invertible, we may also write (13.30) in the form

$$K = (L^{-1})^T J' L^{-1}. \quad (3.13.33)$$

Now make use of (2.14) to find the result

$$K = (L^{-1})^T PJP^T L^{-1}. \quad (3.13.34)$$

Finally, define N by the rule

$$N = (L^{-1})^T P. \quad (3.13.35)$$

This N is evidently real and in $GL(2n, \mathbb{R})$, and direct calculation shows that it has the desired property

$$NJN^T = [(L^{-1})^T P]J[(L^{-1})^T P]^T = (L^{-1})^T PJP^T L^{-1} = K. \quad (3.13.36)$$

3.13.2 Darboux Transformations

We have found that K and J are congruent under the action of the intertwining transformation N . An intertwining congruence transformation, such as N above, that relates two different antisymmetric matrices is sometimes called a *Darboux* transformation because he was the first to study such transformations in the context of Classical Mechanics, and N can be called a Darboux matrix.⁵³ Darboux transformations and matrices will be essential for the work of Sections 5.13 and 6.7 and Chapter 34. We note in passing that the relations (2.14), (6.118), and (6.119) are Darboux relations.

Suppose K and \hat{K} are two symplectic 2-form matrices of the same dimension. Then we know that there are Darboux matrices N and \hat{N} that connect them to the J of this same dimension by the relations (13.2) and

$$\hat{K} = \hat{N}J\hat{N}^T. \quad (3.13.37)$$

Upon combining (13.2) and (13.37), we see that

$$\hat{K} = (\hat{N}N^{-1})K(\hat{N}N^{-1})^T. \quad (3.13.38)$$

Thus, \hat{K} and K are connected by the Darboux matrix $(\hat{N}N^{-1})$.

Given K , what can be said about the N that satisfy (13.2)? Suppose N' is another matrix that satisfies (13.2),

$$K = N'J(N')^T. \quad (3.13.39)$$

Combining (13.2) and (13.39) yields the relation

$$N J N^T = N' J (N')^T, \quad (3.13.40)$$

from which we conclude that

$$[N^{-1}N']J[N^{-1}N']^T = J. \quad (3.13.41)$$

Therefore, if we make the definition

$$M = N^{-1}N', \quad (3.13.42)$$

we see that M is a symplectic matrix. Moreover, (13.42) can be rewritten in the form

$$N' = NM. \quad (3.13.43)$$

That is, N' and N are related by multiplication on the right by a symplectic matrix. Finally, suppose that M is any symplectic matrix, and use (13.43) to define N' . Then we find the result

$$N'J(N')^T = NMJM^TN^T = NJN^T = K. \quad (3.13.44)$$

Thus, all Darboux matrices (for any fixed K) are related by multiplication on the right by symplectic matrices, and this symplectic matrix can be any symplectic matrix.⁵⁴ It follows

⁵³The reader is warned that the words *Darboux transformation* are also employed, with a different meaning, in the context of differential equations.

⁵⁴Note that (13.43) can be rewritten in the form $N = N'M^{-1}$. We know that M^{-1} is symplectic if M is. Thus N and N' are also related by multiplication on the right by a symplectic matrix.

that the dimensionality of the space of Darboux matrices (for any fixed K) is the same as that of $Sp(2n, \mathbb{R})$, namely $n(2n + 1)$.

Matrices N' and N in $GL(2n, \mathbb{R})$ that are related by an equation of the form (13.43) are said to be in the same (left) *coset* of $GL(2n, \mathbb{R})$ relative to the subgroup $Sp(2n, \mathbb{R})$. The collection of these cosets is denoted by the symbols $GL(2n, \mathbb{R})/Sp(2n, \mathbb{R})$. See Section 5.12 for a discussion of cosets. We conclude that the coset space $GL(2n, \mathbb{R})/Sp(2n, \mathbb{R})$ is in one-to-one correspondence with the set of all symplectic 2-forms on $2n$ -dimensional space,

$$GL(2n, \mathbb{R})/Sp(2n, \mathbb{R}) \leftrightarrow \{K \mid K \text{ is real, } 2n \times 2n, \text{ antisymmetric, and nonsingular}\}. \quad (3.13.45)$$

Observe, as a sanity check, that the dimension of $GL(2n, \mathbb{R})$ is $(2n)^2$, the dimension of $Sp(2n, \mathbb{R})$ is $n(2n + 1)$, and the dimension of the space of all real $2n \times 2n$ antisymmetric matrices is $(1/2)[(2n)^2 - 2n]$. But there is the relation

$$(1/2)[(2n)^2 - 2n] = (2n)^2 - n(2n + 1), \quad (3.13.46)$$

which verifies that the dimensionality count works out properly. In Section 5.12 we will learn that the set of all symplectic 2-forms on $2n$ -dimensional space constitutes a *homogeneous* space under the action of $GL(2n, \mathbb{R})$.

Can we restrict our attention to Darboux matrices N that have unit determinant so that $N \in SL(2n, \mathbb{R})$? Then the 2-form matrices K will also have unit determinant, which we might like. The answer is *no*. Consider the 2×2 case and suppose

$$K = -J. \quad (3.13.47)$$

According to (1.7) there is the relation

$$N J N^T = [\det(N^T)] J = [\det(N)] J. \quad (3.13.48)$$

Thus in this case, for (13.2) and (13.47) to hold, we must have the relation

$$\det N = -1. \quad (3.13.49)$$

Note also that if we instead impose the condition

$$\det N = \pm 1, \quad (3.13.50)$$

then, according to (13.4), we will still have the result

$$\det K = 1. \quad (3.13.51)$$

There is another interesting feature of Darboux transformations. Let us use the representation (13.2) to compute K^2 . Doing so gives the result

$$K^2 = N J N^T N J N^T = N J (N^T N) J N^T \quad (3.13.52)$$

Suppose N is orthogonal. Then we find that

$$K^2 = N J^2 N^T = -N N^T = -I. \quad (3.13.53)$$

Thus symplectic 2-form matrices K that are related to J by orthogonal Darboux transformations are those that are most analogous to J . We have already seen an instance of this fact in the case of the symplectic form matrix J' given by (2.10). Since orthogonal matrices from a group, we see from (13.38) that symplectic-form matrices that are related to J by orthogonal Darboux matrices are also related to each other by orthogonal Darboux matrices.

Suppose, conversely, that

$$K^2 = -I. \quad (3.13.54)$$

Then we find from (13.52) and (13.54) that

$$(N^T N) J (N^T N)^T = J, \quad (3.13.55)$$

from which it follows that M defined by

$$M = N^T N \quad (3.13.56)$$

is a symplectic matrix. Also, we see from (13.56) that M is symmetric and positive definite. Therefore we know, from the work of Section 3.8, that there is a unique symmetric matrix S^a such that

$$N^T N = \exp(JS^a). \quad (3.13.57)$$

Suppose we make a polar decomposition for N^T by writing

$$N^T = PO. \quad (3.13.58)$$

See Section 4.2 for information about polar decomposition. Then we find that

$$N^T N = P^2, \quad (3.13.59)$$

from which we conclude that

$$P = \exp(JS^a/2) \quad (3.13.60)$$

and

$$N^T = \exp(JS^a/2)O. \quad (3.13.61)$$

Taking the transpose of both sides of (13.61) gives the result

$$N = O' \exp(JS^a/2) \quad (3.13.62)$$

where $O' = O^T$ is also an orthogonal matrix. This is the most general form for N when (13.54) holds. All such N belong to cosets $GL(2n, \mathbb{R})/Sp(2n, \mathbb{R})$ that contain an orthogonal matrix. In the case that N is orthogonal, we see that $S^a = 0$.

Finally, suppose we replace O' in (13.62) by O'' where

$$O'' = O' \exp(JS^c). \quad (3.13.63)$$

We know that all matrices of the form $\exp(JS^c)$ are orthogonal and therefore O'' is also orthogonal. They are also symplectic, and form a subgroup H of the symplectic group. Recall the work of Section 3.9. Therefore N' defined by

$$N' = O'' \exp(JS^a/2) = O'[\exp(JS^c) \exp(JS^a/2)] \quad (3.13.64)$$

produces the same K as N does when used in (13.2). We conclude that what matters in (13.62) is the coset $O(2n, \mathbb{R})/H$ to which O' belongs.

3.13.3 Symplectic Forms and Pfaffians

Let A be a $2n \times 2n$ antisymmetric matrix. The *Pfaffian* of A , denoted by $\text{Pf}(A)$, is a certain polynomial of degree n in the entries of A (with real coefficients). For our present purposes we need not know all about Pfaffians, but only that they have certain remarkable properties.

The first of these is that

$$\det A = [\text{Pf}(A)]^2. \quad (3.13.65)$$

From (13.65) we see that any real nonsingular antisymmetric matrix must have a positive determinant. This result can also be proved without the use of Pfaffians. See Exercise 13.2.

Other remarkable Pfaffian properties are given by the relations

$$\text{Pf}(NAN^T) = [\det(N)]\text{Pf}(A), \quad (3.13.66)$$

$$\text{Pf}(\lambda A) = \lambda^n \text{Pf}(A), \quad (3.13.67)$$

$$\text{Pf}(J') = 1. \quad (3.13.68)$$

From (2.14), (13.66), and (13.68) we deduce the relation

$$\text{Pf}(J) = (-1)^{n(n-1)/2}. \quad (3.13.69)$$

As special cases of (13.54) we find the results

$$\text{Pf}(J) = 1, -1, -1 \quad (3.13.70)$$

for $n = 1, 2, 3$, respectively.

Upon employing (13.66) in (13.2) and using (13.69), we find the result

$$\text{Pf}(K) = [\det(N)](-1)^{n(n-1)/2}. \quad (3.13.71)$$

We see that symplectic forms can be classified according to the signs of their Pfaffians. Suppose K and K' are two symplectic forms. Then, from (2.38), we know that they are related by an equation of the form

$$K' = MKM^T \quad (3.13.72)$$

with $M \in GL(2n, \mathbb{R})$. If their Pfaffians have the same sign, then $M \in GL(2n, \mathbb{R}, +)$. Here $GL(2n, \mathbb{R}, +)$ denotes the set of real $2n \times 2n$ matrices with *positive* determinant. Such matrices evidently form a subgroup of $GL(2n, \mathbb{R})$. If the Pfaffians of K and K' have different signs, then $M \in GL(2n, \mathbb{R}, -)$. Here $GL(2n, \mathbb{R}, -)$ denotes the set of real $2n \times 2n$ matrices with *negative* determinant. They evidently do not form a subgroup of $GL(2n, \mathbb{R})$, but rather are in a disconnected piece of $GL(2n, \mathbb{R})$ that does not contain the identity matrix I . Given any element $F \in GL(2n, \mathbb{R}, -)$, all elements of $GL(2n, \mathbb{R}, -)$ can be obtained by multiplying F (either on the left or right) by all elements of $GL(2n, \mathbb{R}, +)$.

3.13.4 Variant Symplectic Groups?

Consider the general symplectic 2-form (w, Kz) , and suppose that R is a real matrix that preserves this 2-form. Then it follows that R must satisfy the generalized symplectic relation

$$R^T KR = K. \quad (3.13.73)$$

It is easily verified that all such matrices form a group. One might wonder if this group is something new or is merely $Sp(2n, \mathbb{R})$ in disguise. We will see that the latter is true. It follows that the group $Sp(2n, \mathbb{R})$ is as general as might be desired.

Suppose we employ (13.2) in (13.73). Doing so gives the relation

$$R^T N J N^T R = N J N^T \quad (3.13.74)$$

from which it follows that

$$[N^{-1} R^T N] J [N^{-1} R^T N]^T = J. \quad (3.13.75)$$

We conclude that the M now defined by the relation

$$M^T = N^{-1} R^T N \quad (3.13.76)$$

is a symplectic matrix. Upon solving (13.76) for R we find the result

$$R = N^T M (N^T)^{-1}. \quad (3.13.77)$$

Thus, we see that the group of matrices R is related to the group $Sp(2n, \mathbb{R})$ simply by the similarity transformation (13.77).

Exercises

3.13.1. Review Exercise 6.12. Show that the steps 1 through 12 in Section 3.13.1 can always be executed. Alternatively, verify by induction on n that the construction of the desired v^j is always possible.

3.13.2. Verify that any real nonsingular antisymmetric matrix must have a positive determinant. Hint: Use (13.4).

3.13.3. In the 2×2 case verify that using

$$N = B_3 \quad (3.13.78)$$

in (13.2), with B_3 given by (7.61), yields (13.47). Note also that (13.49) holds in this case as it should.

3.13.4. Verify that (13.52) and (13.54) together imply (13.55).

3.13.5. Verify that the matrices R that satisfy (13.73) form a group.

3.13.6. Take the Pfaffian of both sides of (1.10) or (1.2), and use (13.66) and (13.69), to show that symplectic matrices always have determinant +1.

3.13.7. Let N_2 be the matrix

$$N_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (3.13.79)$$

Verify that

$$N_2 J_2 N_2^T = -J_2, \quad (3.13.80)$$

and therefore N_2 is a Darboux matrix relating J_2 and $-J_2$. Recall (2.11). Use these results to show that N' defined by

$$N' = \begin{pmatrix} N_2 & & & \\ & N_2 & & \\ & & \ddots & \\ & & & N_2 \end{pmatrix} \quad (3.13.81)$$

has the property

$$N' J' (N')^T = -J'. \quad (3.13.82)$$

Therefore N' is a Darboux matrix relating J' and $-J'$. Define the matrix N by the rule

$$N = P^T N' P. \quad (3.13.83)$$

Verify the relation

$$N J N^T = P^T N' P J P^T (N')^T P = P^T N' J' (N')^T P = -P^T J' P = -J, \quad (3.13.84)$$

which demonstrates that N is a Darboux matrix relating J and $-J$. Verify that

$$\det N = \det(P^T N' P) = \det N' = (-1)^n. \quad (3.13.85)$$

Show that in fact N has the simple block form

$$N = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \quad (3.13.86)$$

so that (13.84) and (13.85) follow immediately.

3.13.8. Suppose a real $2n \times 2n$ matrix M satisfies the condition

$$M^T J M = -J. \quad (3.13.87)$$

Such a matrix is said to be *antisymplectic*. (In Section 31.1 it will be shown that antisymplectic matrices arise in the study of *reversal symmetry*.) Show that if M is antisymplectic, then

$$\det M = \pm 1, \quad (3.13.88)$$

and therefore M is invertible. Show that if M is antisymplectic, then so are $-M$, M^T , and M^{-1} . Show that the product of two antisymplectic matrices is symplectic, and the product (in either order) of a symplectic and an antisymplectic matrix is antisymplectic. Thus, antisymplectic matrices do not form a group. For example, the identity matrix is symplectic,

but not antisymplectic. Show, when taken together, that symplectic and antisymplectic matrices do form a group. This group does not seem to have have a name, but might be called the *complete* symplectic group.

Since, as you have proved, (13.87) implies the relation

$$MJM^T = -J, \quad (3.13.89)$$

it follows that M is a Darboux matrix connecting J and $-J$. Take the Pfaffian of both sides of (13.89) and use (13.67) to conclude that

$$\text{Pf}(MJM^T) = \text{Pf}(-J) = (-1)^n \text{Pf}(J). \quad (3.13.90)$$

But, by (13.66), we also have the relation

$$\text{Pf}(MJM^T) = [\det(M)]\text{Pf}(J). \quad (3.13.91)$$

Upon comparing (13.90) and (13.91) you have shown that, in fact,

$$\det M = (-1)^n. \quad (3.13.92)$$

Let N be the Darboux matrix given by (13.83) or (13.86) so that

$$N J N^T = -J. \quad (3.13.93)$$

Evidently N is antisymplectic. Show that any antisymplectic M can be written in the form

$$M = LN = NL' \quad (3.13.94)$$

where L and L' are symplectic. Show that (13.92) also follows from (13.94) and (13.85). Show that the set of antisymmetric matrices is *connected*. (See Section 5.9.1.) Show that what we have called the complete symplectic group consists of two disconnected pieces in $GL(2n, \mathbb{R})$, each of which itself is connected.

As in (1.7.9), write

$$z = (q_1, \dots, q_n; p_1, \dots, p_n). \quad (3.13.95)$$

Define \bar{z} by the rule

$$\bar{z} = Nz. \quad (3.13.96)$$

Show that

$$\bar{z} = (q_1, \dots, q_n; -p_1, \dots, -p_n) \quad (3.13.97)$$

so that N leaves the q_j in peace and changes the signs of all the p_j . Verify that this same result holds when the N' given by (13.81) is used, for which z has the form

$$z = (q_1, p_1, q_2, p_2, \dots, q_n, p_n). \quad (3.13.98)$$

Bibliography

Matrix Theory

- [1] A.C. Aitken, *Determinants and Matrices*, Eighth Edition, Oliver and Boyd Ltd. (1954).
- [2] R.E. Bellman, *Introduction to Matrix Analysis*, Second Edition, Society for Industrial and Applied Mathematics (1997).
- [3] G. Hadley, *Linear Algebra*, Addison-Wesley (1961).
- [4] P. Lax, *Linear Algebra and its Applications*, Second Edition, John Wiley (2007).
- [5] C.G. Cullen, *Matrices and Linear Transformations*, Second Edition, Dover (1990).
- [6] J.N. Franklin, *Matrix Theory*, Dover (2000).
- [7] F.R. Gantmacher, *The Theory of Matrices, Vols. One and Two*, Chelsea (1959).
- [8] F.R. Gantmacher and M. G. Krein, *Oscillation Matrices and Kernels and Small Vibrations of Mechanical Systems*, AMS Chelsea (2002).
- [9] F. Zhang, *Matrix Theory*, Springer Verlag (1999).
- [10] C. Cullen, *Matrices and Linear Transformations*, Second Edition, Dover (1990).
- [11] M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities*, Dover (1992).
- [12] A. E. Fekete, *Real Linear Algebra*, Marcel Dekker (1985).

Linear Stability Analysis and Stability Diagrams

- [13] J.E. Howard and R.S. MacKay, “Linear Stability of Symplectic Maps”, *J. Math. Phys.* **28**, 1036 (1987).
- [14] J.E. Howard and R.S. MacKay, “Calculation of Linear Stability Boundaries for Equilibria of Hamiltonian Systems”, *Phys. Let. A* **122**, 331 (1987).
- [15] J.E. Howard, *Celestial Mechanics and Dynamical Astronomy* **48**, 267 (1990).

Normal Forms

- [16] J. Moser, *Comm. Pure and Appl. Math.* **11**, 81 (1958).
- [17] A. Weinstein, *Bull. Am. Math. Soc.* **77**, 814 (1971).
- [18] J. Williamson, *Am. J. Math.* **58**, 141 (1936); **59**, 599 (1937).
- [19] A. Wintner, *Ann. di. Mat.* **13**, 105 (1934).
- [20] N. Burgoyne and R. Cushman, *Celest. Mech.* **8**, 435 (1974).
- [21] N. Burgoyne and R. Cushman, “Normal forms in linear Hamiltonian systems”, published in the *1976 Ames Research Center (NASA) conference on geometric control theory*, C. Martin and R. Hermann, eds. Math. Sci. Press (Brookline, Mass., 1977).
- [22] A.J. Laub and K. Meyer, *Celest. Mech* **9**, 213 (1974).
- [23] R. Abraham and J.E. Marsden, *Foundations of Mechanics*, American Mathematical Society (2008).
- [24] V.I. Arnold, *Mathematical Methods of Classical Mechanics*, Second Edition, Springer-Verlag (1989).
- [25] M. Moshinsky and P. Winternitz, *J. Math Phys.* **21**, 1667 (1980).
- [26] A.J. Dragt *et al.*, *Phys. Rev. A* **45**, 2572 (1992).
- [27] S-N. Chow, C. Li, and D. Wang, *Normal Forms and Bifurcation of Planar Vector Fields*, Cambridge University Press (1994).
- [28] R. Churchill and M. Kummer, “A Unified Approach to Linear and Nonlinear Normal Forms for Hamiltonian systems”, *J. Symbolic Computation* **27**, p. 49, (1999).
- [29] J. Murdock, *Normal Forms and Unfoldings for Local Dynamical Systems*, Springer-Verlag (2003).
- [30] H. Hofer and E. Zehnder, *Symplectic Invariants and Hamiltonian Dynamics*, Birkhäuser Verlag (1994).
- [31] Y. Long, *Index Theory for Symplectic Paths with Applications*, Progress in Mathematics, Vol. 207, Birkhäuser Verlag (2002).
- [32] K. Weierstrass, *Mathematische Werke* Band I: 233-246, Band II: 19-44, Nachtrag: 139-148, Berlin (1858).

Krein-Moser Theory and Periodic Linear Systems

- [33] M. Krein, “A Generalization of some Investigations on Linear Differential Equations with Periodic Coefficients”, *Doklady Akad. Nauk. SSSR N.S.*, Vol. 73, 445-448 (1950).
- [34] M. Krein, “On the Application of an Algebraic Proposition in the Theory of Monodromy Matrices”, *Uspekhi Math. Nauk.* **6**, 171-177 (1951).

- [35] M. Krein, “On the Theory of Entire Matrix-Functions of Exponential Type”, *Ukrainian Math. Journal* **3**, 164-173 (1951).
- [36] M. Krein, “On Some Maximum and Minimum Problems for Characteristic Numbers and Liapunov Stability Zones”, *Prikl. Math. Mekh.* **15**, 323-348 (1951).
- [37] M. Krein, “On Criteria for Stability and Boundedness of Solutions of Periodic Canonical Systems, *Prikl. Math. Mekh.* **19**, 641-680 (1955).
- [38] M. Krein and G. Lyubarski, “On Analytical Properties of Multipliers of Periodic Canonical Differential Systems of Positive Type, *Izv. Ak. Nauk. SSSR* **26**, 542-572 (1962).
- [39] J. Moser, “New Aspects in the Theory of Stability of Hamiltonian Systems”, *Communications on Pure and Applied Mathematics*, Vol. XI, 81-114 (1958).
- [40] I. M. Gelfand and L. D. Lidskii, “On the structure of stability of linear Hamiltonian systems of differential equations with periodic coefficients”, *Uspekhi Mat Nauk* **10**, AMS Translation **2** 8 pp. 143-181 (1958).
- [41] N. Erugin, *Linear Systems of Ordinary Differential Equations with Periodic and Quasi-Periodic Coefficients*, Academic Press (1966).
- [42] V. Yakubovich and V. Starzhinskii, *Linear Differential Equations with Periodic Coefficients*, Vols. 1 and 2, Wiley (1975).
- [43] F. M. Arscott, *Periodic Differential Equations: An Introduction to Mathieu, Lamé, and Allied Functions*, MacMillan (1964).
- [44] I. Gohberg, P. Lancaster, and L. Rodman, *Matrices and Indefinite Scalar Products*, Birkhäuser Verlag (1983).
- [45] I. Ekeland, *Convexity Methods in Hamiltonian Mechanics*, Springer-Verlag (1990).
- [46] J. Meiss, *Differential Dynamical Systems*, Section 9.11, SIAM (2007).
- [47] E. Forest, *Beam Dynamics, a New Attitude and Framework*, Section 4.5.1, Harwood (1998)
- [48] M. Kuwamura and E. Yanagida, “Krein’s formula for indefinite multipliers in linear periodic Hamiltonian systems”, *J. Differential Equations* **230**, 446-464 (2006).
- [49] R. Cordeiro and R. Vieira Martins, “Krein Stability in the Disturbed Two-Body Problem”, *Chaos, Resonance, and Collective Dynamical Phenomena in the Solar System*, F. Ferraz-Mello, Ed., pp. 369-374, International Astronomical Union (1992).
- [50] T. Bridges and J. Furter, *Singularity Theory and Equivariant Symplectic Maps*, Springer-Verlag (1993).
- [51] A. Abbondandolo, *Morse theory for Hamiltonian systems*, Chapman & Hall/CRC (2001).

Vector and Matrix Norms

- [52] L. Collatz, *Functional Analysis and Numerical Mathematics*, Section 9, Academic Press (1966).
- [53] I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products*, Section 15, Academic Press (1980).

Linear Algebra, Polar Decomposition, Orthogonalization, and Symplectic Bases

- [54] N. Jacobson, *Lectures in Abstract Algebra*, Vol. II - *Linear Algebra*, D. Van Nostrand (1953). See Section 10, beginning on page 159, for a discussion of symplectic forms. See page 188 for a description of polar decomposition. See also the book of F.R. Gantmacher cited in reference 5 above.
- [55] P.R. Halmos, *Finite-Dimensional Vector Spaces*, D. Van Nostrand (1958).
- [56] P.R. Halmos, *Linear Algebra Problem Book*, Mathematical Association of America (1995).
- [57] V. Moretti, *Multi-Linear Algebra, Tensors and Spinors in Mathematical Physics*. See the Web site <http://www.science.unitn.it/~moretti/tensori.pdf>.
- [58] N. Higham, *Functions of Matrices: Theory and Computation*, SIAM (2008).
- [59] J.B. Keller, “Closest Unitary, Orthogonal and Hermitian Operators to a Given Operator”, *Mathematics Magazine* **48**, p. 192 (1975).
- [60] H.C. Schweinler and E.P. Wigner, *J. Math. Phys.* **11**, p. 1693 (1970).
- [61] R. Simon, S. Chaturvedi, and V. Srinivasan, *J. Math. Phys.* **40**, p. 3632 (1999).
- [62] P. Libermann and C-M. Marle, *Symplectic Geometry and Analytical Mechanics*, D. Reidel (1987).
- [63] A. Baker, *Matrix Groups: An Introduction to Lie Group Theory*, Springer (2006). See Section 8.8 for Darboux symplectification.

Modified Darboux Symplectification

- [64] F. Neri invented what we have called modified Darboux symplectification, and incorporated it in the code MaryLie circa 1986.

Generating Function Symplectification

- [65] For a description of the mixed-variable generating functions F_1 through F_4 , see H. Goldstein, *Classical Mechanics*, Addison-Wesley (1980). See also H.D. Block, *J. of Mathematics and Physics* **32**, p. 207 (1953-54).

- [66] For a description of how mixed-variable generating functions can be used to symplectify matrices arising in the context of Accelerator Physics, see D.R. Douglas, “Interpolation of Off-Energy Matrices: Symplectic and Otherwise (A Comparison of Methods)”, SSC Central Design Group Report SSC-TM-4003 (1985).

Exponential Representations and Logarithms

See also the Matrix Theory and Polar Decomposition references at the beginning of this bibliography.

- [67] J. Williamson, *Am. J. Math* **61**, 897 (1939).
- [68] V. Yakubovich and V. Starzhinskii, *Linear Differential Equations with Periodic Coefficients*, Vols. 1 and 2, Wiley (1975).
- [69] Y. Sibuya, “Note on Real Matrices and Linear Dynamical Systems with Periodic Coefficients”, *J. Math. Anal. Appl.* **1**, 363-372 (1960).
- [70] K. Meyer, G. Hall, and D. Offin, *Introduction to Hamiltonian Dynamical Systems and the N-Body Problem*, Second Edition, Springer (2009).
- [71] W. Culver, “On the Existence and Uniqueness of the Real Logarithm of a Matrix”, *Proceedings of the American Mathematical Society* **17**, p. 1146 (1966).
- [72] J. Gallier, “Logarithms and Square Roots of Real Matrices”, available on the Web at ScholarlyCommons (2008): http://repository.upenn.edu/cis_reports/876/.
- [73] N. Higham, *Functions of Matrices: Theory and Computation*, SIAM (2008).
- [74] Zhong-Qi Ma, *Group Theory for Physicists*, World Scientific (2007).
- [75] A. Dooley and N. Wildberger, “Harmonic analysis and the global exponential map for compact Lie groups”, *Functional Analysis and Its Applications* **27**, p. 21 (1993).

Numerical Methods for Symplectic and Hamiltonian Matrices

- [76] H. Fassbender, *Symplectic Methods for the Symplectic Eigenproblem*, Kluwer/Plenum (2000).
- [77] P. Benner, R. Byers, and E. Barth, “Algorithm 800: Fortran 77 Subroutines for Computing the Eigenvalues of Hamiltonian Matrices I: The Square-Reduced Method”, *ACM Transactions on Mathematical Software* **26**, p. 49-77 (2000).
- [78] L. Dieci, “Considerations on computing real logarithms of matrices, Hamiltonian logarithms, and skew-symmetric logarithms”, *Linear Algebra Appl.* **244**, p. 35-54 (1996).
- [79] L. Dieci, “Real Hamiltonian logarithm of a symplectic matrix”, *Linear Algebra Appl.* **281**, p. 227-246 (1998).

See also the Group/Lie Algebra Theory sections of the Bibliographies for Chapters 5 and 27.

- [80] M. Hamermesh, *Group Theory and its Application to Physical Problems*, Addison-Wesley (1962).
- [81] K. Tapp, *Matrix Groups for Undergraduates*, American Mathematical Society (2005). For a supplementary 10th chapter, see the Web site <http://people.sju.edu/~ktapp/>.
- [82] A.O. Barut and R. Raczka, *Theory of Group Representations and Applications*, World Scientific (1986).
- [83] N. Bourbaki, *Lie Groups and Lie Algebras, Elements of Mathematics, Chapters 1-3*, Springer-Verlag (1989).
- [84] T. Brocker and T.T. Dieck, *Representations of Compact Lie Groups*, Springer-Verlag (1985).
- [85] D. Bump, *Lie Groups*, Springer (2004).
- [86] R. Cahn, *Semi-Simple Lie Algebras and their Representations*, Dover (2006).
- [87] J-Q. Chen, *Group Representation Theory for Physicists*, World Scientific (1989).
- [88] W. Fulton and J. Harris, *Representation Theory, A First Course*, Corrected third printing, Springer-Verlag (1996).
- [89] H. Georgi, *Lie Algebras in Particle Physics*, Perseus Books (1999).
- [90] R. Goodman and N.R. Wallach, *Representations and Invariants of the Classical Groups*, Cambridge University Press (1998).
- [91] R. Goodman and N.R. Wallach, *Symmetry, Representations, and Invariants*, Springer (2009).
- [92] J.E. Humphreys, *Introduction to Lie Algebras and Representation Theory*, Springer-Verlag (1972).
- [93] N. Jacobson, *Lectures in Abstract Algebra*, Vol. I - *Basic Concepts*, D. Van Nostrand (Princeton, 1951).
- [94] N. Jacobson, *Lie Algebras*, Interscience Publishers (1962).
- [95] A.W. Knapp, *Lie Groups Beyond an Introduction*, Second Edition, Birkhäuser (2005).
- [96] A.W. Knapp, *Representation Theory of Semisimple Groups, An Overview Based on Examples*, Princeton (1986).
- [97] C. Procesi, *Lie Groups, An Approach through Invariants and Representations*, Springer (2007).

- [98] V.S. Varadarajan, Review of “Lie Groups, An Approach through Invariants and Representations, by C. Procesi”, *Bulletin of the American Mathematical Society* **45**, p. 661 (2008).
- [99] V.S. Varadarajan, *Lie Groups, Lie Algebras, and Their Representations*, Springer-Verlag (1984).
- [100] D.H. Sattinger and O.L. Weaver, *Lie Groups and Algebras with Applications to Physics, Geometry, and Mechanics*, Springer-Verlag (1986).
- [101] J.E. Campbell, *Introductory Treatise on Lie's Theory of Finite Continuous Transformation Groups*, Chelsea Publishing (1903 and 1966).
- [102] H. Weyl, *The Classical Groups: Their Invariants and Representations*, Princeton University Press (1946).
- [103] E.P. Wigner, *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra*, Academic Press (1959).
- [104] B.G. Wybourne, *Classical Groups for Physicists*, John Wiley and Sons (1974).
- [105] A. Baker, *Matrix Groups: An Introduction to Lie Group Theory*, Springer (2006).
- [106] J.G.F. Belinfante and B. Kolman, *A Survey of Lie Groups and Lie Algebras with Applications and Computational Methods*, Society for Industrial and Applied Mathematics (1972).
- [107] D. Montgomery and L. Zippin, *Topological Transformation Groups*, Interscience (1955).
- [108] P. Tondeur, *Introduction to Lie Groups and Transformation Groups*, Lecture Notes in Mathematics **7**, Springer-Verlag (1965).
- [109] J. Stillwell, *Naive Lie Theory*, Springer (2008).
- [110] P. Szekeres, *A Course in Modern Mathematical Physics: Groups, Hilbert Space, and Differential Geometry*, Cambridge University Press (2005).
- [111] M. Curtis, *Matrix Groups*, Second Edition, Springer (1984).
- [112] R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications*, Dover (2006).
- [113] Arvind, B. Dutta, N. Mukunda, and R. Simon, “The Real Symplectic Groups in Quantum Mechanics and Optics”, arXiv:quant-ph/9509002v3 24 Nov 1995, (2008).
- [114] P. Ramond, *Group Theory, A Physicist's Survey*, Cambridge University Press (2010).
- [115] W. Miller, *Symmetry Groups and Their Applications*, Academic Press (1972).
- [116] J. Gallier, *Geometric Methods and Applications: For Computer Science and Engineering*, Springer-Verlag (2001).

- [117] P. Winternitz, Edit., *Group Theory and Numerical Analysis*, American Mathematical Society (2005).
- [118] N. Ibragimov, *Transformation Groups Applied to Mathematical Physics*, D. Reidel (1985).
- [119] L. P. Eisenhart, *Continuous Groups of Transformations*, Dover (1961).
- [120] A. Henderson, *Representations of Lie algebras: An Introduction Through $gl(n)$* , Cambridge (2012).
- [121] G. Bredon, *Introduction to Compact Transformation Groups*, Academic Press (1972).
- [122] N. Ibragimov, *Transformation Groups and Lie Algebras*, World Scientific (2013).
- [123] W. Pfeifer, *The Lie Algebras $su(n)$: An Introduction*, Birkhäuser Verlag (2003).
- [124] P. Teodorescu and N.-A. Nicorovici, *Applications of the Theory of Groups in Mechanics and Physics*, Kluwer (2004).
- [125] R. Carter, G. Segal, and I. Macdonald, *Lectures on Lie Groups and Lie Algebras*, Cambridge University Press (1995).
- [126] S. Sternberg, *Lie Algebras*, (2004). See the Web site http://www.math.harvard.edu/~shlomo/docs/lie_algebras.pdf.
- [127] P. Cvitanović, *Group Theory: Birdtracks, Lie's, and Exceptional Groups*, Princeton University Press (2008).
- [128] J. Arthur, *The Endoscopic Classification of Representations: Orthogonal and Symplectic Groups*, American Mathematical Society (2013).
- [129] J. Talman, *Special Functions: A Group Theoretic Approach Based on Lectures by E. Wigner*, Benjamin (1968).
- [130] N. Vilenkin, *Special Functions and the Theory of Group Representations*, American Mathematical Society (1968).
- [131] F. Iachello, *Lie Algebras and Applications*, 2nd edition, Springer (2015).
- [132] A. Zee, *Group Theory in a Nutshell for Physicists*, Princeton University Press (2016).
- [133] B. Hall, *Lie Groups, Lie Algebras, and Representations: an Elementary Introduction*, 2nd edition, Springer (2015).
- [134] A. Baker, *Matrix Groups: An Introduction to Lie Group Theory*, Springer (2002).
- [135] K. Erdmann and M. Wildon, *Introduction to Lie Algebras*, Springer (2011).
- [136] H. Pollatsek, *Lie Groups: A Problem-Oriented Introduction via Matrix Groups*, Mathematical Association of America (2009).

- [137] P. Woit, *Quantum Theory, Groups and Representations: An Introduction*, Springer (2017). See also the Web site <http://www.math.columbia.edu/~woit/QM/qmbook.pdf>.
- [138] A. Kirillov Jr., *An Introduction to Lie Groups and Lie Algebras*, Cambridge University Press (2017). You can try Googling Lie Group PDF or the url <https://www.math.stonybrook.edu/~krillov/mat552/liegroups.pdf>
- [139] Adam Marsh, *Mathematics for Physics: An Illustrated Handbook*, World Scientific (2018).
- [140] Gregory W. Moore, *Applied Group Theory*, <http://www.physics.rutgers.edu/~gmoore/618Spring2018/GroupTheory-Spring2018.html> (2018).
- [141] J. D. Vergados, *Group and Representation Theory*, World Scientific (2017).
- [142] S. Garibaldi, “ E_8 , The Most Exceptional Group”, *Bulletin of the American Mathematical Society*, Volume 53, Number 4, (October 2016).

Pfaffians

- [143] R. Vein and P. Dale *Determinants and Their Applications in Mathematical Physics*, Springer (1999).
- [144] Google “Pfaffian” and look at the *Wikipedia* and *PlanetMath* entries.

Chapter 4

Matrix Exponentiation, Polar Decompositions, and Symplectifications

Overview

Matrix Exponentiation

We have learned in Section 3.8 and elsewhere that we need to compute matrices of the form $\exp(JS)$. That is, we need to *exponentiate* the matrix JS . Sometimes, as will be seen in later chapters, this exponentiation can be done analytically. However in many cases numerical methods are required. We may also be interested in exponentiating other matrices. When numerical methods are used, it is desirable that these methods be fast and accurate. No completely satisfactory method is known for this purpose, but one of the better methods available will be described in Section 4.1. This description begins with the problem of computing the ordinary exponential function, and then moves on to the computation of the matrix exponential function.

Polar Decompositions

Suppose z is a complex number. It can then be written/decomposed in the factored product form

$$z = re^{i\phi} \quad (4.0.1)$$

where r is its magnitude and ϕ is its phase. Polar decompositions are a matrix generalization of this decomposition.

(Orthogonal) Polar Decomposition

Suppose M is a real $m \times m$ matrix. Then there exists a real positive definite matrix P and an orthogonal matrix O such that

$$M = PO. \quad (4.0.2)$$

The matrix P is unique, and the matrix O is unique if M is invertible. See Section *.

If M is real and symplectic (in which case $m = 2n$), then P and O are symplectic as well as being positive definite and orthogonal, respectively. They are also unique. And, since O is both orthogonal and symplectic, such matrices are isomorphic to $U(n)$. For example, in the case of $Sp(6, \mathbb{R})$, the matrices O provide a real representation of $U(3)$. See *,

(Unitary) Polar Decomposition

If M is complex, then there is a positive definite Hermitian matrix H and a unitary matrix U such that

$$M = HU. \quad (4.0.3)$$

The matrix H is unique, and the matrix U is unique if M is invertible. See Section *.

(Symplectic) Polar Decomposition

The polar decompositions described above are all part of standard matrix mathematics lore. What is new, and still under development, is what we call *symplectic* polar decomposition. Suppose G is a real $2n \times 2n$ invertible matrix, $G \in GL(2n, \mathbb{R})$, sufficiently *near* the identity matrix I . Then there is an antisymmetric matrix A and a symplectic matrix R such that G can be written in the form

$$G = \exp(JA)R. \quad (4.0.4)$$

(Like the previous decompositions, this decomposition is also named after the group properties of its second factor.) See Section * for a presentation of what is known about symplectic polar decomposition for matrices not necessarily near the identity.

Note that to the extent that symplectic polar decomposition can be uniquely achieved for some collection of matrices G , it provides a rule for assigning a unique symplectic matrix R to any matrix G in the collection. In this case we will call R a *symplectification* of G based on symplectic polar decomposition. In addition to symplectification based on symplectic polar decomposition, there are other possible rules for doing so. Recall the discussion in Subsection .

Symplectification

Roughly speaking, matrix symplectification is a procedure that takes a nearly symplectic matrix and produces a *nearby* matrix that is exactly symplectic. There are several circumstances in which matrix symplectification may be useful. Four come to mind:

First, attempted numerical exponentiation of JS may produce a matrix that is not as symplectic as desired. In that case we may take the numerical result, symplectify it, and accept the symplectified matrix as the desired result. Second, suppose that over the course of a numerical calculation we have multiplied together several symplectic matrices. For example, we will learn in Chapter 8 [see (8.4.20)] that such multiplication is required if we wish to *concatenate* a large number of maps. Then the net matrix result may not be exactly symplectic due to round-off error. Although we cannot recover an exact result, we can at least produce a result that is exactly symplectic (to machine precision) and also near the exact result.

Third, we will see in Section 9.3 that in the treatment of translations it is necessary to evaluate linear transformations of the form $\exp(: k_2 :)$ where k_2 arises solely from nonlinear feed-down effects. Since all calculations are carried out within the quotient algebra L^0/L^ℓ , k_2 in this case is only known up to some order in the size of the translation, and it seems pointless to evaluate $\exp(: k_2 :)$ to any order higher than what is known for k_2 . However, we may very well wish to have a result that is exactly symplectic. The meaning of the new notation just employed and the concepts alluded to will become evident in Chapter 9. Suffice it to say that there are cases where JS , while being exactly Hamiltonian, is yet only known approximately. In these cases we are content with a correspondingly approximate (but, we hope, rapidly computable) result for $\exp(JS)$ which is, nevertheless, exactly symplectic.

Fourth, there are occasions in which we may wish to factor a map into symplectic and nonsymplectic parts. See Section 29.1. The first step in this process is to factor, in some standard way, a matrix into symplectic and nonsymplectic parts.

Section 4.2 provides an initial background by describing the completely understood subject of orthogonal polar decomposition. Then Sections 4.3 and 4.4 provide a theoretical background for the more complicated subject of matrix symplectification and symplectic polar decomposition. They also give information concerning how the symplectic group lies within the general linear group. This information is useful when one considers non-Hamiltonian perturbations of Hamiltonian dynamics. Again see Section 29.1. Finally, Sections 4.5 through 4.8 describe four known methods for matrix symplectification.

4.1 Exponentiation by Scaling and Squaring

4.1.1 The Ordinary Exponential Function

The ordinary exponential function $\exp(z)$, where z is a complex variable, is defined by the Taylor series

$$\exp(z) = \sum_{\ell=0}^{\infty} z^\ell / \ell!. \quad (4.1.1)$$

This series converges everywhere, but is useful for computation only for small z . Consider computing, for example, $\exp(20)$. For $z = 20$, we find the numerical result

$$(20)^{60} / 60! = 1.4 \times 10^{-4}. \quad (4.1.2)$$

Consequently, when $z = 20$, at least 60 terms must be retained in (1.1) to even begin to get convergence. And at this stage the convergence is still quite slow since the ratio of successive terms is only about

$$(20)/(60) = 1/3. \quad (4.1.3)$$

Finally, if we want to compute $\exp(-20)$ using the Taylor series, we would have to use very high precision arithmetic to take into account the high degree of cancellation that in this case must occur between very large terms.

There is a better way to compute the exponential function based on the observation that it satisfies the functional *scaling* equation

$$\exp(z) = [\exp(z/m)]^m. \quad (4.1.4)$$

Suppose we set m to an integer power of 2,

$$m = 2^n. \quad (4.1.5)$$

Then the right side of (1.4) can be calculated by n successive *squarings*,

$$\exp(z) = \{\exp[z/(2^n)]\}^{2^n} = \{\cdots \{\{\exp[z/(2^n)]\}^2\}^2 \cdots\}^2 \text{ (n squarings).} \quad (4.1.6)$$

Next we observe that if $[z/(2^n)]$ is small enough, the quantity $\exp[z/(2^n)]$ can be computed to good accuracy using a Taylor series truncated at relatively low order,

$$\exp[z/(2^n)] \sim \sum_0^N [z/(2^n)]^\ell / \ell!. \quad (4.1.7)$$

Let $t\text{Nexp}(z)$ denote the *truncated* exponential function defined by the relation

$$t\text{Nexp}(z) = \sum_0^N z^\ell / \ell!. \quad (4.1.8)$$

Suppose we define *my* exponential function by the rule

$$\text{myexp}(z) = \{\cdots \{\{t\text{Nexp}[z/(2^n)]\}^2\}^2 \cdots\}^2 \text{ (n squarings).} \quad (4.1.9)$$

Then we might hope that for a suitable value of n (which depends on z) we would have to good accuracy the relation

$$\exp(z) \sim \text{myexp}(z). \quad (4.1.10)$$

In fact, using Taylor's formula with remainder, we find the result

$$\text{myexp}(z) = \exp(z) - \exp(z)z[z/(2^n)]^N / (N+1)! + h.o.t. \quad (4.1.11)$$

where "h.o.t." denotes still higher order error terms. Thus, the magnitude of the *relative* estimated error is given by the relation

$$\text{estimated error} \sim |z| [|z|/(2^n)]^N / (N+1)!. \quad (4.1.12)$$

We see that to achieve good accuracy what we must do is make N sufficiently large and $[z/(2^n)]$ sufficiently small that the error term above is small. Given moderate values of z , this can be done with quite small values of N and n . We also observe that the required value of n only grows as $\log(|z|)$, and that for a given $|z|$ and modest N the accuracy increases very rapidly with increasing n . The tables below show results for $N = 6$ and 9 , $-20 < z < 20$, and n selected so that $||z/(2^n)|| < (1/10)$. The error is also shown, and is consistent with the estimates (1.11) and (1.12). Note that (with $N = 9$) at most 16 ($9 - 1 + 8 = 16$) multiplications are required to achieve full (64 bit) machine precision. Indeed, the errors listed in Table 1.2 fluctuate in sign, and are mostly the result of working with only 64 bit arithmetic.

Table 4.1.1: $N = 6$; scaling n values chosen to make $|[z/(2^n)]| < (1/10)$.

z	n	$\text{myexp}(z)$	error	relative error
-20	8	0.206E-08	0.199E-17	0.966E-09
-19	8	0.560E-08	0.377E-17	0.672E-09
-18	8	0.152E-07	0.699E-17	0.459E-09
-17	8	0.414E-07	0.127E-16	0.307E-09
-16	8	0.113E-06	0.225E-16	0.200E-09
-15	8	0.306E-06	0.388E-16	0.127E-09
-14	8	0.832E-06	0.648E-16	0.779E-10
-13	8	0.226E-05	0.105E-15	0.463E-10
-12	7	0.614E-05	0.108E-13	0.175E-08
-11	7	0.167E-04	0.158E-13	0.948E-09
-10	7	0.454E-04	0.219E-13	0.483E-09
-9	7	0.123E-03	0.283E-13	0.229E-09
-8	7	0.335E-03	0.335E-13	0.999E-10
-7	7	0.912E-03	0.355E-13	0.390E-10
-6	6	0.248E-02	0.217E-11	0.877E-09
-5	6	0.674E-02	0.163E-11	0.242E-09
-4	6	0.183E-01	0.915E-12	0.500E-10
-3	5	0.498E-01	0.218E-10	0.439E-09
-2	5	0.135E+00	0.338E-11	0.250E-10
-1	4	0.368E+00	0.460E-11	0.125E-10
0	0	0.100E+01	0.000E+00	0.000E+00
1	4	0.272E+01	-0.304E-10	-0.112E-10
2	5	0.739E+01	-0.165E-09	-0.224E-10
3	5	0.201E+02	-0.748E-08	-0.372E-09
4	6	0.546E+02	-0.245E-08	-0.448E-10
5	6	0.148E+03	-0.313E-07	-0.211E-09
6	6	0.403E+03	-0.300E-06	-0.745E-09
7	7	0.110E+04	-0.388E-07	-0.354E-10
8	7	0.298E+04	-0.267E-06	-0.896E-10
9	7	0.810E+04	-0.164E-05	-0.203E-09

Table 4.1.1 continued

z	n	$\text{myexp}(z)$	error	relative error
10	7	0.220E+05	-0.928E-05	-0.421E-09
11	7	0.599E+05	-0.488E-04	-0.815E-09
12	7	0.163E+06	-0.242E-03	-0.149E-08
13	8	0.442E+06	-0.187E-04	-0.423E-10
14	8	0.120E+07	-0.852E-04	-0.708E-10
15	8	0.327E+07	-0.374E-03	-0.114E-09
16	8	0.889E+07	-0.159E-02	-0.179E-09
17	8	0.242E+08	-0.659E-02	-0.273E-09
18	8	0.657E+08	-0.266E-01	-0.406E-09
19	8	0.178E+09	-0.105E+00	-0.590E-09
20	8	0.485E+09	-0.409E+00	-0.843E-09

Table 4.1.2: $N = 9$; scaling n values chosen to make $|[z/(2^n)]| < (1/10)$.

z	n	$\text{myexp}(z)$	error	relative error
-20	8	0.206E-08	-0.773E-22	-0.375E-13
-19	8	0.560E-08	0.108E-21	0.193E-13
-18	8	0.152E-07	-0.586E-21	-0.385E-13
-17	8	0.414E-07	-0.242E-20	-0.585E-13
-16	8	0.113E-06	0.132E-20	0.118E-13
-15	8	0.306E-06	-0.116E-20	-0.381E-14
-14	8	0.832E-06	-0.392E-20	-0.471E-14
-13	8	0.226E-05	0.775E-19	0.343E-13
-12	7	0.614E-05	0.110E-19	0.179E-14
-11	7	0.167E-04	0.146E-18	0.872E-14
-10	7	0.454E-04	-0.854E-18	-0.188E-13
-9	7	0.123E-03	-0.239E-17	-0.193E-13
-8	7	0.335E-03	0.195E-17	0.582E-14
-7	7	0.912E-03	-0.217E-17	-0.238E-14
-6	6	0.248E-02	0.217E-17	0.875E-15
-5	6	0.674E-02	-0.633E-16	-0.940E-14
-4	6	0.183E-01	0.555E-16	0.303E-14
-3	5	0.498E-01	0.208E-16	0.418E-15
-2	5	0.135E+00	0.194E-15	0.144E-14
-1	4	0.368E+00	0.278E-15	0.754E-15
0	0	0.100E+01	0.000E+00	0.000E+00

Table 4.1.2 continued

z	n	$\text{myexp}(z)$	error	relative error
1	4	0.272E+01	0.488E-14	0.180E-14
2	5	0.739E+01	0.258E-13	0.349E-14
3	5	0.201E+02	0.355E-13	0.177E-14
4	6	0.546E+02	0.384E-12	0.703E-14
5	6	0.148E+03	-0.568E-13	-0.383E-15
6	6	0.403E+03	0.142E-11	0.352E-14
7	7	0.110E+04	-0.432E-11	-0.394E-14
8	7	0.298E+04	0.414E-10	0.139E-13
9	7	0.810E+04	-0.236E-10	-0.292E-14
10	7	0.220E+05	-0.182E-10	-0.826E-15
11	7	0.599E+05	0.800E-10	0.134E-14
12	7	0.163E+06	0.114E-08	0.697E-14
13	8	0.442E+06	0.827E-08	0.187E-13
14	8	0.120E+07	-0.978E-08	-0.813E-14
15	8	0.327E+07	0.118E-06	0.362E-13
16	8	0.889E+07	0.248E-06	0.279E-13
17	8	0.242E+08	-0.110E-05	-0.455E-13
18	8	0.657E+08	-0.380E-06	-0.579E-14
19	8	0.178E+09	0.149E-04	0.833E-13
20	8	0.485E+09	-0.715E-06	-0.147E-14

4.1.2 The Matrix Exponential Function

So far we have been discussing the ordinary exponential function. The matrix exponential function (3.7.1) has similar properties. Again its Taylor series may be only very slowly convergent, and again scaling and squaring can be used to good advantage. Let s be a parameter and Z any $m \times m$ matrix. Consider the matrix function $F(s)$ defined by the equation

$$F(s) = \exp(-sZ). \quad (4.1.13)$$

The function F satisfies the relations

$$F(0) = I, \quad (4.1.14)$$

$$F(1) = \exp(-Z), \quad (4.1.15)$$

$$(d/ds)F(s) = -Z \exp(-sZ). \quad (4.1.16)$$

See Exercise 3.7.1. Integrate both sides of (1.16) to get the result

$$\int_0^1 (d/ds)F(s)ds = F(s)|_0^1 = \exp(-Z) - I. \quad (4.1.17)$$

By combining (1.16) and (1.17) we find the integral formula

$$\exp(-Z) - I = -Z \int_0^1 \exp(-sZ) ds. \quad (4.1.18)$$

Now multiply both sides of (1.18) by $\exp(Z)$ to get the result

$$\exp(Z) = I + Z \exp(Z) \int_0^1 \exp(-sZ) ds. \quad (4.1.19)$$

Integration by parts yields the general formula

$$\int_0^1 \exp(-sZ) s^n ds = [1/(n+1)] \exp(-Z) + [1/(n+1)] Z \int_0^1 \exp(-sZ) s^{n+1} ds. \quad (4.1.20)$$

Now use (1.20) repeatedly in (1.19) to get the truncated Taylor series with remainder result

$$\exp(Z) = \sum_{\ell=0}^N Z^\ell / \ell! + (Z^{N+1} / N!) \exp(Z) \int_0^1 \exp(-sZ) s^N ds. \quad (4.1.21)$$

As before, we define a truncated exponential function by the formula

$$\text{tNexp}(Z) = \sum_{\ell=0}^N Z^\ell / \ell!. \quad (4.1.22)$$

Then from (1.21) and (1.22) we get the result

$$\text{tNexp}(Z) = \exp(Z) [I - (Z^{N+1} / N!) \int_0^1 \exp(-sZ) s^N ds]. \quad (4.1.23)$$

In analogy to (1.9) we define $\text{myexp}(Z)$ by the rule

$$\text{myexp}(Z) = \{\text{tNexp}[Z/(2^n)]\}^{2^n} = \{\cdots \{\{\text{tNexp}[z/(2^n)]\}^2\}^2 \cdots\}^2 \quad (n \text{ squarings}). \quad (4.1.24)$$

Now scale and square both sides of (1.23). Upon combining (1.23) and (1.24) we find the final result

$$\text{myexp}(Z) = \exp(Z) \{I - (1/N!) [Z/(2^n)]^{N+1} \int_0^1 \exp[-sZ/(2^n)] s^N ds\}^{2^n}. \quad (4.1.25)$$

Suppose we decide to make the approximation

$$\exp(Z) \sim \text{myexp}(Z). \quad (4.1.26)$$

It is easily checked that the *relative* error made in doing so has an estimated *norm* given by the relation

$$\text{estimated error} = \left\| 2^n \{(1/N!) [Z/(2^n)]^{N+1} \int_0^1 \exp[-sZ/(2^n)] s^N ds\} \right\|. \quad (4.1.27)$$

By using the properties (3.7.10) through (3.7.13) for a norm the expression (1.27) can be simplified to the form

$$\text{estimated error} \sim \{[1/(N+1)!] \|Z\| \|Z/(2^n)\|^N \exp[u \|Z/(2^n)\|]\}, \quad (4.1.28)$$

where u is a number in the range $0 < u < 1$.

For purposes of illustration, suppose we set $N = 10$ and select n so that $\|Z/(2^n)\| < (1/20)$. Then we find the estimates

$$\exp[u \|Z/(2^n)\|] < \exp(1/20) \sim 1.05, \quad (4.1.29)$$

$$\|Z/(2^n)\|^N < (1/20)^{10} \sim 9.8 \times 10^{-14}, \quad (4.1.30)$$

$$1/(N+1)! = 1/(11!) \sim 2.5 \times 10^{-8}. \quad (4.1.31)$$

Correspondingly, the error estimate becomes

$$\text{estimated error} \sim (2.6 \times 10^{-21}) \|Z\|, \quad (4.1.32)$$

which for reasonable values of $\|Z\|$ is well below round-off error for 64 bit arithmetic. We conclude that the error committed in using (1.26) can be made quite small by using modest values for N and n . Consequently, the computation of $\exp(Z)$ by scaling and squaring can be both very fast and very accurate. See Exercise 1.2.

We close this section by remarking that there are alternatives to using the truncated exponential series (1.22) to evaluate the exponential of the scaled exponent. These alternatives, which include the use of Padé approximants, give even better numerical performance at the expense of more elaborate programming. For further detail, see the references at the end of this chapter.

Exercises

4.1.1. Verify (1.6). Verify (1.17) through (1.25). Verify (1.27) through (1.32).

4.1.2. Suppose n is selected so that

$$\|Z/(2^n)\| = \|Z\|/(2^n) < (1/20). \quad (4.1.33)$$

Verify that n grows with increasing $\|Z\|$ like

$$n \sim [\log(20)]/\log(2) + [\log(\|Z\|)]/\log(2). \quad (4.1.34)$$

Consequently, for reasonable values of $\|Z\|$, the number of squarings required to evaluate (1.24) is quite modest, and the computation of $\exp(Z)$ by scaling and squaring is both accurate and remarkably fast. Show that for a given $\|Z\|$ and N , the relative error (1.28) decreases *exponentially* with increasing n .

4.2 (Orthogonal and Unitary) Polar Decompositions

4.2.1 Real Matrix Case

Consider the set of all *real* $n \times n$ matrices. This set obviously forms a Lie algebra, with the commutator as a Lie product, and this Lie algebra is $gl(n, \mathbb{R})$. Any matrix B in $gl(n, \mathbb{R})$ can be written in the form

$$B = S + A, \quad (4.2.1)$$

where S is (real) symmetric and A is (real) antisymmetric, and both are unique. Next we observe that the antisymmetric matrices form a Lie subalgebra by themselves,

$$\{A, A'\} = A'', \quad (4.2.2)$$

and this Lie algebra is $so(n, \mathbb{R})$. Finally, we observe that the remaining commutation rules for $gl(n, \mathbb{R})$ can be written in the form

$$\{A, S\} = S', \quad (4.2.3)$$

$$\{S, S'\} = A. \quad (4.2.4)$$

That is, the commutator of an antisymmetric and a symmetric matrix is a symmetric matrix, and the commutator of two symmetric matrices is an antisymmetric matrix.

If M is a matrix in $GL(n, \mathbb{R})$ sufficiently near the identity, it can be written in the exponential form

$$M = \exp(B). \quad (4.2.5)$$

See Section 3.7. Correspondingly, it can also be written in the form

$$M = \exp(S') \exp(A') \quad (4.2.6)$$

where S' is symmetric and A' is antisymmetric. Indeed, near the identity in $GL(n, \mathbb{R})$, which corresponds to being near the origin in $gl(n, \mathbb{R})$, one can in principle pass back and forth between the representations (2.5) and (2.6) by means of the BCH formula and an appropriate *Zassenhaus* formula. See Sections 3.7 and 8.8.

We observe that matrices of the form $\exp(S)$ are positive-definite symmetric, and matrices of the form $\exp(A)$ are orthogonal. See Exercise 2.2. Thus, any M sufficiently near the identity has the polar decomposition

$$M = PO \quad (4.2.7)$$

where P is positive-definite symmetric and O is orthogonal. To be more precise, we might call (2.7) an *orthogonal* polar decomposition to emphasize that the second factor in (2.7) is orthogonal.

So far we have examined matrices near the identity. In fact, the decomposition (2.7) can be made globally and is unique. It is easy to check that the matrix (MM^T) is positive symmetric, and consequently has a unique positive symmetric square root. See Exercise 2.3. Let us therefore define P by the rule

$$P = (MM^T)^{1/2}, \quad (4.2.8)$$

with the corresponding result

$$P^2 = MM^T. \quad (4.2.9)$$

Next assume M is invertible, in which case P is also invertible. Define a matrix O by the rule

$$O = P^{-1}M. \quad (4.2.10)$$

Calculation reveals that O is orthogonal,

$$\begin{aligned} OO^T &= (P^{-1}M)(P^{-1}M)^T = P^{-1}MM^T(P^{-1})^T \\ &= P^{-1}P^2(P^T)^{-1} = P^{-1}P^2P^{-1} = I. \end{aligned} \quad (4.2.11)$$

Thus (2.10) is equivalent to (2.7).

It can be shown that if M is invertible, then both P and O are unique, and P is positive-definite symmetric. See also Exercise 2.3. Moreover, the decomposition (2.7) is still possible if M is not invertible, and P in this case is still unique. However, O is no longer uniquely defined.

We close this section by noting that there is another way of looking at the decomposition (2.7) that deserves emphasis. We have been dealing with the group $GL(n, \mathbb{R})$ and its subgroup $O(n, \mathbb{R})$. Form the coset space $GL(n, \mathbb{R})/O(n, \mathbb{R})$ consisting of the left cosets of $GL(n, \mathbb{R})$ with respect to $O(n, \mathbb{R})$. See Section 5.12 for a detailed description of cosets. Equation (2.7) indicates that the elements of this coset space can be labeled by positive-definite symmetric matrices P . Moreover, any positive-definite symmetric matrix P can be written in the form

$$P = \exp(S) \quad (4.2.12)$$

where S is symmetric, and conversely. Symmetric matrices that are $n \times n$ form, in turn, a linear vector space whose dimension m is given by the relation

$$m = \dim(S) = (1/2)n(n + 1). \quad (4.2.13)$$

It follows that matrices P of the form (2.12) have the topology of E^m , m -dimensional Euclidean space, with m given by (2.13). Correspondingly, $GL(n, \mathbb{R})$ has the topology of $E^m \times O(n, \mathbb{R})$.

++++++

4.2.2 Application to the Symplectic Group

We begin with the simpler case of orthogonal polar decomposition. For the symplectic group we know that any real symplectic matrix $M \in Sp(2n, \mathbb{R})$ can be written in the product form

$$M = \exp(JS^a) \exp(JS^c), \quad (4.2.14)$$

each of the two factors in the product is *uniquely* defined in terms of M , and each is of a special type. The matrices $\exp(JS^c)$ are real, symplectic, orthogonal, and form a matrix group, which a subgroup of $M \in Sp(2n, \mathbb{R})$. This matrix group is a real representation of $U(n)$. In fact, there is the intersection result $U(n) = Sp(2n, \mathbb{R}) \cap SO(2n, \mathbb{R})$. The matrices

$\exp(JS^a)$ are real, symplectic, symmetric, and positive definite, but (taken together) do not form a group.

Here are some sanity checks: We know that all matrices of the form JS comprise the Lie algebra $sp(2n, \mathbb{R})$. Show that the Lie algebra generated by all $2n \times 2n$ matrices of the form JS^c comprises a subalgebra of $sp(2n, \mathbb{R})$, and find its dimension. Verify that this dimension is the same as that of $u(n)$. See Exercise 7.27. Find the dimension of the vector space spanned by all $2n \times 2n$ matrices of the form JS^a . You should have found the dimensions n^2 and $(n^2 + n)$, respectively. Verify, in accord with (8.29), that their sum is $\dim sp(2n\mathbb{R})$ as given by (7.42).

We know that all matrices of the form $\exp(JS)$ are elements of $Sp(2n, \mathbb{R})$. Moreover, any $Sp(2n, \mathbb{R})$ element sufficiently near the identity I can be written in single exponent form. Also, using the BCH series, the product of any two such matrices can be written in the single-exponent form $\exp(JS)$ provided the individual matrices are sufficiently near the identity I . Finally, there are elements in $Sp(2n, \mathbb{R})$ that *cannot* be written in single exponent form. It follows that the two exponents appearing in (5.55) cannot always be combined. For $Sp(2n, \mathbb{R})$, there must be cases for which the BCH series diverges.

4.2.3 Complex Matrix Case

Finally we remark that there is an analogous result for *complex* matrices. In that case a factorization of the form (2.7) still holds but now M is complex, P is Hermitian and positive definite, and O is unitary. This result is also called a polar decomposition. See Exercise 2.5.

Exercises

4.2.1. The purpose of this exercise is to verify that the decomposition (2.1) is unique. To begin, define matrices S and A by the explicit formulas

$$S = (1/2)(B + B^T), \quad (4.2.15)$$

$$A = (1/2)(B - B^T). \quad (4.2.16)$$

Verify that S is symmetric, A is antisymmetric, and (2.1) is satisfied. Next assume that there are symmetric and antisymmetric matrices S' and A' such that

$$B = S' + A'. \quad (4.2.17)$$

Verify that there must be the relations

$$S' = S, \quad (4.2.18)$$

$$A' = A. \quad (4.2.19)$$

4.2.2. Verify the commutation rules (2.2) through (2.4).

4.2.3. This exercise examines some of the properties of $\exp(A)$ and $\exp(S)$ where A and S are arbitrary real antisymmetric and symmetric matrices, respectively.

- a) Define a matrix O by the rule

$$O = \exp(A). \quad (4.2.20)$$

Show that

$$O^T = \exp(A^T) = \exp(-A), \quad (4.2.21)$$

and therefore

$$O^T O = O O^T = I. \quad (4.2.22)$$

Show, using (3.7.129), that

$$\det O = 1, \quad (4.2.23)$$

and therefore $O \in SO(n, \mathbb{R})$ if A is $n \times n$.

- b) Define a matrix P by the rule

$$P = \exp(S). \quad (4.2.24)$$

Show that P is real and symmetric. Use (3.7.129) to prove that P is nonsingular. Since S is real and symmetric, show that there is a real orthogonal matrix O such that

$$S = ODO^{-1} \quad (4.2.25)$$

where D is diagonal and real. Show that $\exp(D)$ is diagonal, real, and positive definite.

Show that

$$P = \exp(S) = O \exp(D) O^{-1} = O \exp(D) O^T. \quad (4.2.26)$$

Show, based on the representation (2.25), that P is positive definite.

There is also a more direct proof of this fact that does not involve matrix diagonalization. Show that the matrix $P^{1/2}$ defined by

$$P^{1/2} = \exp(S/2) \quad (4.2.27)$$

is real, symmetric, and nonsingular, and has the property

$$P^{1/2} P^{1/2} = P. \quad (4.2.28)$$

As a result, show that for any vector v there is the relation

$$(v, Pv) = (v, P^{1/2} P^{1/2} v) = (P^{1/2} v, P^{1/2} v) = \|P^{1/2} v\|^2 \geq 0. \quad (4.2.29)$$

Finally, demonstrate that $(v, Pv) = 0$ implies that $v = 0$.

- c) What about the converse? Suppose P is a real positive-definite symmetric matrix. Show that there is a real orthogonal matrix O such that

$$P = ODO^{-1} \quad (4.2.30)$$

where D is diagonal and has all positive entries on the diagonal. Define a symmetric matrix S' by the rule

$$S' = \log D \quad (4.2.31)$$

where $\log D$ is defined to be the diagonal matrix whose diagonal entries are the logarithms of the corresponding diagonal entries in D . Verify that, by this definition, S' is real and symmetric and has the feature

$$D = \exp(S'). \quad (4.2.32)$$

Define the matrix S by the rule

$$S = OS'O^{-1} = OS'O^T. \quad (4.2.33)$$

Verify that S is real and symmetric and has the property

$$P = \exp(S). \quad (4.2.34)$$

- d) There is more that can be said about the relation between real symmetric matrices S and real positive-definite symmetric matrices P . According to (2.23), P is a real analytic function of S . That is, each entry (matrix element) of P is an analytic function of the various entries in S , and each entry is real when the entries in S are real. (See Section 35.2 for a discussion of analyticity in several complex variables.) This result follows because all powers of S are analytic functions of S and the exponential series converges in norm for all S .

We will see the converse is also true. Namely, S is a real analytic function of P . This fact is not obvious from the work so far because the construction of S as given by (2.32) involved O and D , and their analytic properties are not evident. Indeed, as described in Section 26.13, the eigenvalues of a matrix M (and eigenvalues of P are involved in the construction of both O and D) need not be analytic functions of the entries in M . What is required is an alternate procedure for constructing S in terms of P that is manifestly analytic. Begin with the matrix Q defined in terms of P by the rule

$$Q = [1/\text{tr}(P)]P. \quad (4.2.35)$$

Evidently $[1/\text{tr}(P)]$ is a real analytic function of P as long as $\text{tr}(P) \neq 0$, and correspondingly Q is also an analytic function of P under this same proviso. Moreover, since P is positive definite, all its eigenvalues λ_j are real and positive, and therefore

$$\text{tr}(P) = (\sum \lambda_j) > 0. \quad (4.2.36)$$

The eigenvalues of Q , call them μ_j , are given by the relation

$$\mu_j = \lambda_j / (\sum \lambda_i), \quad (4.2.37)$$

and therefore are real and satisfy the conditions $0 < \mu_j < 1$. Next consider the matrix $I - Q$. Its eigenvalues, call them ν_j , satisfy the conditions $0 < \nu_j < 1$. It therefore follows that

$$\|I - Q\| < 1 \quad (4.2.38)$$

when the spectral norm is used. See Section 3.7.1. Moreover, suppose P is not exactly real, symmetric, and positive definite, but is in some sufficiently small and possibly

complex neighborhood of such a matrix. Show that (2.37) will continue to hold for such a P . [Use the norm property (3.7.12) and the fact that although the eigenvalues of M need not be analytic functions of M , they are continuous functions of M .] Now define S in terms of P by the rule

$$S = \{\log[\text{tr}(P)]\}I - \sum_{k=1}^{\infty} (1/k)(I - Q)^k = \{\log[\text{tr}(P)]\}I + \log(Q). \quad (4.2.39)$$

Evidently the infinite sum in (2.38) converges in norm because of (2.37), and therefore is an analytic function of P as long as P is in a sufficiently small neighborhood of a real positive-definite symmetric matrix. Also, the first term in (2.38) is an analytic function of P under the same proviso. Therefore S is an analytic function of P . Moreover, S is manifestly real and symmetric when P is real, positive definite, and symmetric. Finally, we find that

$$\begin{aligned} \exp(S) &= \exp\{\{\log[\text{tr}(P)]\}I + \log(Q)\} \\ &= I \exp\{\log[\text{tr}(P)]\} \exp[\log(Q)] = [\text{tr}(P)]Q = P. \end{aligned} \quad (4.2.40)$$

4.2.4. This exercise further explores polar decomposition. Let M be any real matrix. Show that the matrix Q defined by

$$Q = MM^T \quad (4.2.41)$$

is positive symmetric. Show that it has a positive symmetric square root. Hint: Since Q is symmetric, show that there exists a real orthogonal matrix O' that *diagonalizes* it,

$$O'Q(O')^{-1} = D. \quad (4.2.42)$$

Show that the entries in D are real and positive or zero. Define $D^{1/2}$ to be a diagonal matrix with entries equal to the positive square roots of the corresponding entries in D . Now construct P by the rule

$$P = (O')^{-1}D^{1/2}O'. \quad (4.2.43)$$

Show that P is positive symmetric and satisfies

$$P^2 = Q. \quad (4.2.44)$$

Show that if M is invertible, then so is P , and P is then positive-definite symmetric.

A bit more can be said in the way of analyticity if M is invertible. Review Exercise 2.3. From the definition (2.40) show that Q is real, symmetric, and positive definite if M is real and invertible. Verify that Q is analytic in M . From part d of Exercise 2.3 we know there is a real symmetric matrix S such that

$$Q = \exp(S) \quad (4.2.45)$$

and S is analytic in Q , and hence also in M . Now define P by the rule

$$P = \exp(S/2). \quad (4.2.46)$$

Evidently P is real, symmetric, and positive definite. Verify that it also satisfies (2.43). Also, we see that P is analytic in Q , and hence also in M . Finally, if we make a polar decomposition of M , we see from (2.10) that O is also analytic in M . This follows because P^{-1} is analytic in M if P is.

4.2.5. The purpose of this exercise is to define and study polar decomposition for complex matrices. Review the work of Subsection 4.2.1. We will follow an analogous path in the complex case.

Consider the set of all possibly complex $n \times n$ matrices. This set obviously forms a Lie algebra, with the commutator as a Lie product, and this Lie algebra is $g\ell(n, \mathbb{C})$. Verify that any matrix B in $g\ell(n, \mathbb{C})$ can be written uniquely in the form

$$B = H + A \quad (4.2.47)$$

where

$$H = (B + B^\dagger)/2 \quad (4.2.48)$$

and

$$A = (B - B^\dagger)/2. \quad (4.2.49)$$

Show that H is *Hermitian*,

$$H^\dagger = H, \quad (4.2.50)$$

and A is *anti-Hermitian*,

$$A^\dagger = -A. \quad (4.2.51)$$

Next we observe that anti-Hermitian matrices form a Lie subalgebra by themselves,

$$\{A, A'\} = A'', \quad (4.2.52)$$

and this Lie algebra is $u(n)$. Finally, we observe that the remaining commutation rules for $g\ell(n, \mathbb{C})$ can be written in the form

$$\{A, H\} = H', \quad (4.2.53)$$

$$\{H, H'\} = A. \quad (4.2.54)$$

That is, the commutator of an anti-Hermitian and a Hermitian matrix is a Hermitian matrix, and the commutator of two Hermitian matrices is an anti-Hermitian matrix. Verify these claims.

If M is a matrix in $GL(n, \mathbb{R})$ sufficiently near the identity, it can be written in the exponential form

$$M = \exp(B). \quad (4.2.55)$$

See Section 3.7. Correspondingly, it can also be written in the form

$$M = \exp(H') \exp(A') \quad (4.2.56)$$

where H' is Hermitian and A' is anti-Hermitian. Indeed, near the identity in $GL(n, \mathbb{C})$, which corresponds to being near the origin in $g\ell(n, \mathbb{C})$, one can in principle pass back and forth between the representations (2.54) and (2.55) by means of the BCH formula and an appropriate *Zassenhaus* formula. See Sections 3.7 and 8.8.

Verify that matrices of the form $\exp(H)$ are positive-definite Hermitian (see Exercise 3.7.44), and matrices of the form $\exp(A)$ are unitary. Thus, any M sufficiently near the identity has the polar decomposition

$$M = PU \quad (4.2.57)$$

where $P = \exp(H')$ is positive-definite Hermitian and $U = \exp(A')$ is unitary. To be more precise, we might call (2.56) a *unitary* polar decomposition to emphasize that the second factor in (2.56) is unitary.

So far we have examined matrices near the identity. In fact, the decomposition (2.56) can be made globally and is unique. Assume, for a moment, the correctness of (2.56). Verify that then there is the relation

$$MM^\dagger = PUU^\dagger P^\dagger = PP^\dagger = P^2. \quad (4.2.58)$$

Verify that the matrix (MM^\dagger) is positive Hermitian, and consequently has a unique positive Hermitian square root. Let us therefore *define* P by the rule

$$P = (MM^\dagger)^{1/2}, \quad (4.2.59)$$

with the corresponding result

$$P^2 = MM^\dagger. \quad (4.2.60)$$

Next assume M is invertible, in which case prove that P is also invertible. Define a matrix U by the rule

$$U = P^{-1}M. \quad (4.2.61)$$

Verify by calculation that U is unitary,

$$\begin{aligned} UU^\dagger &= (P^{-1}M)(P^{-1}M)^\dagger = P^{-1}MM^\dagger(P^{-1})^\dagger \\ &= P^{-1}P^2(P^\dagger)^{-1} = P^{-1}P^2P^{-1} = I. \end{aligned} \quad (4.2.62)$$

Thus the decomposition (2.56) can be made globally.

It can be shown that if M is invertible, then both P and U are unique, and P is positive-definite Hermitian. Moreover, the decomposition (2.56) is still possible if M is not invertible, and P in this case is still unique. However, U is no longer uniquely defined.

4.3 Matrix Symplectification

Introduction

In Subsection 3.6.5 we studied the construction of symplectic basis. There we considered the problem of, given a set of $2n$ linearly independent vectors w^i , how one may construct a set of vectors v^j such that they form a symplectic basis. Evidently we may view the vectors w^i as being the columns of some matrix $2n \times 2n$ matrix W , and we may view the vectors v^j as being the columns of some related symplectic matrix V . Thus given W whose columns are linearly independent, which is equivalent to the requirement $\det(W) \neq 0 \Leftrightarrow W \in G\ell(2n, \mathbb{R})$, we described various ways of finding a related symplectic matrix matrix V .

We have also defined and studied *Orthogonal* polar decomposition. It can be shown that if M is any matrix, then the *orthogonal* matrix O that is closest to M , in the sense of minimizing $\|M - O\|_E$, is given by the orthogonal matrix appearing in the polar decomposition (2.7). Thus there is a *group-theoretic* construction, which is what orthogonal polar

decomposition is, that leads from an arbitrary matrix $M \in G\ell(2n, \mathbb{R})$ to a geometrically related and uniquely defined orthogonal matrix O .

What we will now seek, under the rubric of Matrix Symplectification and in analogy to orthogonal polar decomposition, are various procedures that may be more demanding than just constructing symplectic bases: Given a matrix $W \in GL(2n, \mathbb{R})$ that is *nearly symplectic*, produce a symplectic matrix in $V \in Sp(2n, \mathbb{R})$ that is *close* to W . In so doing we shall have to define what is meant by the terms *nearly symplectic* and *close* as well as the procedure for constructing V from W . We will then call the matrix V a *symplectification* of the matrix W .

Symplectic Polar Decomposition

We will begin with a procedure we will call *Symplectic* polar decomposition and is in some ways analogous to Orthogonal polar decomposition. We will be working with real $2n \times 2n$ matrices. Consider all matrices of the form JS and JA where S and A are symmetric and antisymmetric, respectively. Verify that these matrices obey the commutation rules

$$\{JS, JS'\} = JS'', \quad (4.3.1)$$

$$\{JS, JA\} = JA', \quad (4.3.2)$$

$$\{JA, JA'\} = JS, \quad (4.3.3)$$

and constitute a basis for $g\ell(2n, \mathbb{R})$. Also recall that the JS constitute a basis for the subalgebra $sp(2n, \mathbb{R})$. Let G be an element in $GL(2n, \mathbb{R})$ that is sufficiently *near* the identity. Then we may write G in the form.

$$G = \exp(JA) \exp(JS). \quad (4.3.4)$$

In writing (5.59) we have employed a hybrid of Lie coordinates of the first and second kinds. See Section 7.9.

Any matrix R of the form

$$R = \exp(JS) \quad (4.3.5)$$

is symplectic and therefore, like all symplectic matrices, satisfies the relation

$$JR^T J^{-1} = R^{-1}. \quad (4.3.6)$$

See Sections 3.1 and 3.7. What can be said about the first factor, $\exp(JA)$, in (3.4)? Let Q be a matrix of the form

$$Q = \exp(JA). \quad (4.3.7)$$

It satisfies the relation

$$JQ^T J^{-1} = J[\exp(JA)]^T J^{-1} = J \exp[(JA)^T] J^{-1} = J \exp(AJ) J^{-1} = \exp(JA) = Q. \quad (4.3.8)$$

We will now make the *definition* that *any* matrix Q satisfying

$$JQ^T J^{-1} = Q \quad (4.3.9)$$

is a J -symmetric matrix.

Finally, we say that G has a symplectic polar decomposition if it can be written in the form

$$G = QR \quad (4.3.10)$$

where Q is J -symmetric and R is symplectic. Here, in this definition, Q may be an arbitrary J -symmetric matrix not necessarily expressible in the single-exponent form $\exp(JA)$ and R may be an arbitrary symplectic matrix not necessarily expressible in the single-exponent form $\exp(JS)$.

We have shown that G has a symplectic polar decomposition if it is sufficiently near the identity. We next seek to find elements in $GL(2n, \mathbb{R})$, not necessary near the identity I , which have a symplectic polar decomposition. To do so, given G , we have found it useful to define the matrix $N(G)$ by the rule

$$N(G) = GJG^T J^T. \quad (4.3.11)$$

The matrix $N(G)$ has two important properties. First, suppose G is symplectic. Then we find that

$$N(G) = GJG^T J^T = (GJG^T)J^T = JJ^T = I \quad (4.3.12)$$

and

$$\|N(G) - I\| = 0. \quad (4.3.13)$$

Since $N(G) = I$ when G is symplectic, we may define a measure $f(G)$ of the *failure* of G to be symplectic by the rule

$$f(G) = \|N(G) - I\|. \quad (4.3.14)$$

Suppose $f(G)$ is small. Then we say that G is nearly symplectic, and we might hope to find a matrix R that is both near G and exactly symplectic.

Second, suppose that Q and R in (3.10) are J -symmetric and symplectic, respectively. Then we find for G as given by (3.10) the result

$$\begin{aligned} N(G) &= GJG^T J^T = (QR)J(QR)^T J^T = Q(RJR^T)(Q^T J^T) \\ &= Q(JQ^T J^{-1}) = Q^2. \end{aligned} \quad (4.3.15)$$

We have learned that establishing the factorization (3.10) is equivalent to finding a J -symmetric matrix Q that satisfies (3.15).

4.3.1 Properties of J -Symmetric Matrices

Thanks to previous work we pretty much know the properties of symplectic matrices. For future use, let us work out, by a series of lemmas, some of the properties of J -symmetric matrices.

Lemma 3.1 If Q is J -symmetric, then Q can be written in the form

$$Q = J\hat{A} \quad (4.3.16)$$

where \hat{A} is antisymmetric, and conversely. To see this, solve (3.16) for \hat{A} to find the result

$$\hat{A} = J^T Q. \quad (4.3.17)$$

Then we see that

$$\hat{A}^T = (J^T Q)^T = Q^T J = J J^{-1} Q^T J = J J Q^T J^{-1} = J Q = -J^T Q = -\hat{A}. \quad (4.3.18)$$

Conversely, if \hat{A} is antisymmetric, we have from (3.15) the result

$$J Q^T J^{-1} = J(J\hat{A})^T J^{-1} = J\hat{A}^T J^T J^{-1} = -J\hat{A}J^T J^{-1} = J\hat{A} = Q. \quad (4.3.19)$$

Lemma 3.2 Any matrix Q that is symmetric and commutes with J is J -symmetric. Evidently with these two assumptions about Q we have the result

$$J Q^T J^{-1} = J Q J^{-1} = Q J J^{-1} = Q. \quad (4.3.20)$$

We conclude that all the matrices S^c are J -symmetric. See Section 3.8. In particular, the zero and identity matrices are J -symmetric.

Lemma 3.3 J -symmetric matrices form a linear vector space. We have already seen that the zero matrix is J -symmetric. Suppose Q_1 , and Q_2 are J -symmetric. Let a_1 and a_2 be any two scalars. Then we find the result

$$J(a_1 Q_1 + a_2 Q_2)^T J^{-1} = a_1 J Q_1^T J^{-1} + a_2 J Q_2^T J^{-1} = a_1 Q_1 + a_2 Q_2. \quad (4.3.21)$$

Lemma 3.4 Suppose Q is J -symmetric and has an inverse. Then Q^{-1} is J -symmetric:

$$J(Q^{-1})^T J^{-1} = J(Q^T)^{-1} J^{-1} = (J Q^T J^{-1})^{-1} = Q^{-1}. \quad (4.3.22)$$

Lemma 3.5 Suppose Q_1 and Q_2 are J -symmetric and commute. Then the product $Q_1 Q_2$ is J -symmetric:

$$J(Q_1 Q_2)^T J^{-1} = J(Q_2 Q_1)^T J^{-1} = J Q_1^T Q_2^T J^{-1} = J Q_1^T J^{-1} J Q_2^T J^{-1} = Q_1 Q_2. \quad (4.3.23)$$

Lemma 3.6 If Q is J -symmetric, then so are all powers of Q including, if Q is invertible, all negative powers. This result follows from Lemmas 3.3 and 3.4. We also note that (by definition) $Q^0 = I$ and (by Lemma 3.2) I is J -symmetric.

Lemma 3.7 If Q is J -symmetric and nonsingular, then

$$\det Q > 0. \quad (4.3.24)$$

Moreover, if G is nonsingular and has the symplectic polar decomposition (3.10), then

$$\det G > 0. \quad (4.3.25)$$

Thus, G is orientation preserving. To verify the first claim, take the determinant of both sides of (3.16) to find the relation

$$\det Q = (\det J)(\det \hat{A}) = \det \hat{A}. \quad (4.3.26)$$

We see that Q being nonsingular implies that \hat{A} is nonsingular. But, according to Exercise 3.12.2, it follows that $\det \hat{A} > 0$ and therefore (3.19) holds. To verify the second claim, take the determinant of both sides of (3.10) to find the relations

$$\det G = (\det Q)(\det R) = \det Q. \quad (4.3.27)$$

We see that Q is nonsingular if G is nonsingular. Therefore, if G is nonsingular and has a symplectic polar decomposition, (3.20) follows from the first claim.

Lemma 3.8 A J -symmetric matrix remains J -symmetric under the action of any symplectic similarity transformation. In other words, if Q is J -symmetric, if R is symplectic, and if we define the *transformed* matrix Q^{tr} by the rule

$$Q^{\text{tr}} = R^{-1}QR, \quad (4.3.28)$$

then Q^{tr} is J -symmetric. To check this claim, we carry out the computation

$$\begin{aligned} J(Q^{\text{tr}})^T J^{-1} &= J(R^{-1}QR)^T J^{-1} = JR^T Q^T (R^{-1})^T J^{-1} \\ &= JR^T J^{-1} J Q^T J^{-1} J (R^{-1})^T J^{-1} = R^{-1}QR = Q^{\text{tr}}. \end{aligned} \quad (4.3.29)$$

Note that if Q is written in the form (3.15), and Q^{tr} is written in the form

$$Q^{\text{tr}} = JA^{\text{tr}}, \quad (4.3.30)$$

then A and A^{tr} are related by the equation

$$A^{\text{tr}} = J^T Q^{\text{tr}} = J^T R^{-1}QR = J^T R^{-1}JAR = JR^{-1}J^{-1}AR = R^T AR. \quad (4.3.31)$$

Also note that if the matrix G has a symplectic polar decomposition, then so does the matrix $R'GR''$ where R' and R'' are any two symplectic matrices. To see this, use (3.10) to write

$$R'GR'' = R'QRR'' = R'Q(R')^{-1}R'RR'' = Q^{\text{tr}}R''' \quad (4.3.32)$$

where now

$$Q^{\text{tr}} = R'Q(R')^{-1} \quad (4.3.33)$$

and

$$R''' = R'RR''. \quad (4.3.34)$$

By the Lemma 3.8, Q^{tr} is J -symmetric. And, by the group property, R''' is symplectic. Therefore, the right side of (3.27) is a symplectic polar decomposition.

Conversely, suppose that a matrix M does not have a symplectic polar decomposition. (We will see in Subsection 4.3.5 that there are such matrices.) Again consider the matrix $R'MR''$ where R' and R'' are any two symplectic matrices. Then it is easy to verify, by *reductio ad absurdum*, that $R'MR''$ also does not have a symplectic polar decomposition. We conclude that the spaces of matrices that do and do not have symplectic polar decompositions are invariant under left and right translations/multiplications by elements in the symplectic group.

Lemma 3.9 Given any matrix G , form the matrix $N(G)$ by the rule

$$N(G) = GJG^T J^T. \quad (4.3.35)$$

Then N is J -symmetric. To see this, simply compute. We find the result

$$\begin{aligned} JN^T J^{-1} &= J(GJG^T J^T)^T J^{-1} = JJJGJ^T G^T J^{-1} \\ &= GJG^T J^T = N. \end{aligned} \quad (4.3.36)$$

We remark that if G is symplectic, then $N(G) = I$. Also, suppose we take the determinant of both sides of (3.30). Doing so gives the result

$$\det N = (\det J)^2 (\det G)^2 = (\det G)^2. \quad (4.3.37)$$

We see that if G is nonsingular, than so is N . Moreover, consistent with Lemma 3.7, N has a positive determinant.

Lemma 3.10 Suppose G is any matrix and R' and R'' are any two symplectic matrices. Then we have the relation

$$N(R'GR'') = R'N(G)(R')^{-1}. \quad (4.3.38)$$

Again we simply compute and use (3.7) to find the result

$$\begin{aligned} N(R'GR'') &= (R'R'')J(R'GR'')^T J^T \\ &= R'G[R''J(R'')^T]G^T(R')^T J^T \\ &= R'GJG^T J^T J(R')^T J^T \\ &= R'GJG^T J^T (R')^{-1} = R'N(G)(R')^{-1}. \end{aligned} \quad (4.3.39)$$

As a special case of (3.33) we have the relation

$$N(GR'') = N(G), \quad (4.3.40)$$

which shows that $N(G)$ depends only on the coset $GL(2n, \mathbb{R})/Sp(2n, \mathbb{R})$ to which G belongs.

Lemma 3.11 If the matrix G is transformed by a symplectic similarity transformation, then so is the matrix $N(G)$. Suppose G is any matrix and R is a symplectic matrix. Define the transformed matrix G^{tr} by the rule

$$G^{\text{tr}} = R^{-1}GR. \quad (4.3.41)$$

Let us compute the matrix N^{tr} associated with G^{tr} by the rule (3.30). As a special case of (3.33) we have the result

$$N^{\text{tr}} = N(G^{\text{tr}}) = N(R^{-1}GR) = R^{-1}N(G)R. \quad (4.3.42)$$

We call (3.10) a *symplectic* polar decomposition to emphasize that the second factor in (3.10) is symplectic. Of course, simply demanding that the second factor R in (3.10) be symplectic is not enough for a definition. We must also put requirements on Q because otherwise we could write $R = R'R''$, with R' and R'' symplectic, and then make the factorization $M = (QR')R''$ and claim R'' as the second factor. In what follows we will require that Q be J -symmetric. We might instead require that Q have a representation of the form (3.8). However, in all the cases for which we have been able to establish the existence of a symplectic polar decomposition, with Q being J -symmetric, we have then also been able to establish that Q has a representation of the form (3.8).

4.3.2 Initial Result on Symplectic Polar Decomposition

With these lemmas in hand, we are prepared to say more about the possibility of achieving the factorization (3.10) for general matrices G . Suppose G is invertible, and suppose there exists a J -symmetric matrix Q such that

$$N(G) = Q^2 \quad (4.3.43)$$

with N defined by (3.30). Then M has the factorization (3.10) with R symplectic.

To prove this result, we first observe that Q is invertible: we know from (3.32) that N is invertible if M is, and from (3.38) we see that Q is invertible if N is. Next, since Q is invertible, we define R by the rule

$$R = Q^{-1}M. \quad (4.3.44)$$

Then the computation

$$\begin{aligned} R J R^T &= (Q^{-1}M)J(Q^{-1}M)^T = Q^{-1}M J M^T (Q^{-1})^T \\ &= Q^{-1}M J M^T J^T J(Q^{-1})^T J^{-1}J = Q^{-1}N Q^{-1}J \\ &= Q^{-1}Q^2 Q^{-1}J = J \end{aligned} \quad (4.3.45)$$

shows that R is symplectic. Conversely, suppose that Q and R in (3.10) are J -symmetric and symplectic, respectively. Then we find the result

$$\begin{aligned} N &= M J M^T J^T = Q R J (Q R)^T J^T = Q R J R^T Q^T J^T \\ &= Q J Q^T J^{-1} = Q^2. \end{aligned} \quad (4.3.46)$$

We have learned that establishing the factorization (3.10) is equivalent to finding a J -symmetric matrix Q that satisfies (3.38).

Put another way, we seek a solution of the matrix equation

$$Q = [N(M)]^{1/2}, \quad (4.3.47)$$

provided a solution can be found, and further require that Q be J -symmetric.

***** and have shown that its properties can be used to infer in certain important cases the existence of a symplectic polar decomposition. Our next task, given G , is to determine, if possible, explicit formulas for the factors Q and R in its symplectic polar decomposition.

Suppose G is invertible, and suppose there exists a J -symmetric matrix Q such that

$$N(G) = Q^2 \quad (4.3.48)$$

with N defined by (5.65). We will see that the factorization (5.64) can be achieved if $f(G) < 1$.

To prove this result, we first observe that Q is invertible: we know from (3.32) that N is invertible if G is, and from (3.38) we see that Q is invertible if N is. Next, since Q is invertible, (5.64) can be solved for R to give

$$R = Q^{-1}G. \quad (4.3.49)$$

Finally the computation

$$\begin{aligned} RJR^T &= (Q^{-1}G)J(Q^{-1}G)^T = Q^{-1}GJG^T(Q^{-1})^T \\ &= Q^{-1}GJG^TJ^TJ(Q^{-1})^TJ^{-1}J = Q^{-1}NQ^{-1}J \\ &= Q^{-1}Q^2Q^{-1}J = J \end{aligned} \quad (4.3.50)$$

shows that R is symplectic, and we will say that this R is the *symplectification* of G ,

Conversely, suppose that Q and R in (3.10) are J -symmetric and symplectic, respectively. Then we find the result

$$\begin{aligned} N &= GJG^TJ^T = QRJ(QR)^TJ^T = QRJR^TQ^TJ^T \\ &= QJQ^TJ^{-1} = Q^2. \end{aligned} \quad (4.3.51)$$

We have learned that establishing the factorization (5.64) is equivalent to finding a J -symmetric matrix Q that satisfies (5.69).

We now turn to the task of given G , find Q . According to (5.69) we may hope to find a relation of the form

$$Q = [N(G)]^{1/2}, \quad (4.3.52)$$

with the further requirement that Q be J -symmetric. We will see that this is possible if $f(G) < 1$.

Begin by writing the identity

$$N = I + (N - I) \quad (4.3.53)$$

and, inspired by the binomial theorem, write

$$N^{1/2} = [I + (N - I)]^{1/2} = I + (1/2)(N - I) + \cdots = I + \sum_{\ell=1}^{\infty} c_{\ell}(N - I)^{\ell} \quad (4.3.54)$$

where the c_{ℓ} are the binomial coefficients

$$c_{\ell} = \binom{1/2}{\ell}. \quad (4.3.55)$$

Note that we may also write

$$N^{-1/2} = N^{-1}N^{1/2} = N^{-1}[I + (1/2)(N - I) + \cdots] = N^{-1} + N^{-1}\sum_{\ell=1}^{\infty} c_{\ell}(N - I)^{\ell}. \quad (4.3.56)$$

Note that (by Lemmas 3.1, 3.2, 3.5, and 3.9) all the terms I , N , and $[(I - N)^{\ell}/\ell]$ are J -symmetric matrices. Consequently, if the series (5.74) converges, then (by Lemma 3.2) Q will be a J -symmetric matrix.

Assuming these relations make sense, we have the result

$$Q = N^{1/2} = I + \sum_{\ell=1}^{\infty} c_{\ell}(N - I)^{\ell}. \quad (4.3.57)$$

The infinite sum on the right side of (5.78) is defined if the sum

$$\sum_{\ell=1}^{\infty} |c_{\ell}| \|N - I\|^{\ell} \quad (4.3.58)$$

converges which, in view of (5.68), is equivalent to the requirement

$$\sum_{\ell=1}^{\infty} |c_{\ell}| f^{\ell} < \infty. \quad (4.3.59)$$

It is easily checked that this requirement is met if $f < 1$.

We now move on to the calculation of R . In view of (5.70), we see see that what is also needed now is Q^{-1} . And according to (5.73), or (5.72), we may hope for a relation of the form

$$Q^{-1} = [N(G)]^{-1/2}, \quad (4.3.60)$$

and we further require that Q^{-1} be J -symmetric. We will see that this is possible if $f(G) < 1$. Again write (5.74) and, inspired by the binomial theorem, write

$$N^{-1/2} = [I + (N - I)]^{-1/2} = I - (1/2(N - I) + \dots) = I + \sum_{\ell=1}^{\infty} d_{\ell}(N - I)^{\ell} \quad (4.3.61)$$

where the d_{ℓ} are the binomial coefficients

$$d_{\ell} = \binom{-1/2}{\ell}. \quad (4.3.62)$$

This result for $N^{-1/2}$ is to be compared with the equally valid result (5.77). We also note that for $N^{1/2}$ there is the result

$$N^{1/2} = NN^{-1/2} = N[I - (1/2(N - I) + \dots)] = N + N \sum_{\ell=1}^{\infty} d_{\ell}(N - I)^{\ell}, \quad (4.3.63)$$

which is to be compared with (5.75).

Assuming these relations make sense, we have the result

$$Q^{-1} = N^{-1/2} = I + \sum_{\ell=1}^{\infty} d_{\ell}(N - I)^{\ell}. \quad (4.3.64)$$

The infinite sum on the right side of (5.85) is defined if the sum

$$\sum_{\ell=1}^{\infty} |d_{\ell}| \|N - I\|^{\ell} \quad (4.3.65)$$

converges which, in view of (5.68), is equivalent to the requirement

$$\sum_{\ell=1}^{\infty} |d_{\ell}| f^{\ell} < \infty. \quad (4.3.66)$$

It is easily checked that this requirement is met if $f < 1$. Also, again by Lemmas 3.1, 3.2, 3.5, and 3.9, all the terms I , N , and $[(I - N)^\ell/\ell]$ are J -symmetric matrices. Consequently, if the series (5.81) converges, then (by Lemma 3.2) Q^{-1} will be a J -symmetric matrix.

With Q^{-1} in hand we are able to compute R using (5.70) to find

$$R = Q^{-1}G = G + \left[\sum_{\ell=1}^{\infty} d_\ell (N - I)^\ell \right] G. \quad (4.3.67)$$

In view of (5.65), (5.78), and (5.88) we have now found, in terms of G , both the symplectic polar decomposition factors Q and R .

A standard general method for finding roots of a general matrix N is to first find, if possible, its logarithm.¹ Therefore, let us first try to compute $\log(N)$. From (3.7.2) we find the result

$$\log(N) = - \sum_{\ell=1}^{\infty} (I - N)^\ell / \ell. \quad (4.3.68)$$

Note that (by Lemmas 3.1, 3.2, 3.5, and 3.9) all the terms $[(I - N)^\ell/\ell]$ are J -symmetric matrices. Consequently, if the series (3.43) converges, then (by Lemma 3.2) $\log(N)$ will be a J -symmetric matrix. If the series does converge, let us define a matrix Q by the rule

$$Q = \exp[(1/2) \log(N)]. \quad (4.3.69)$$

The matrix Q will also be J -symmetric. [Apply to the series (3.7.1) arguments similar to those just made for $\log(N)$.] Moreover, Q will satisfy (3.38),

$$Q^2 = \{\exp[(1/2) \log(N)]\}^2 = \exp[\log(N)] = N. \quad (4.3.70)$$

Therefore we can achieve the factorization (3.10) if the series (3.43) converges.

The series (3.43) will converge if N is sufficiently near I . Specifically, the series will converge if $\|N - I\| < 1$ for some choice of matrix norm. But, according to the remark made in Lemma 3.9, $N = I$ if M is symplectic. Consequently, the series will converge if M is sufficiently near a symplectic matrix.

Start here.

From a counter example we know that there are matrices G that do not have symplectic polar decompositions. We also know from various theorems that there are broad conditions that guarantee the existence of a symplectic polar decomposition.

Because we will be working with real symplectic matrices, consider the set of all real $2n \times 2n$ matrices. Since the matrix J is invertible, any matrix $B \in gl(2n, \mathbb{R})$ can be written in the form

$$B = JS + JA, \quad (4.3.71)$$

where S and A are symmetric and antisymmetric, respectively. We also know that matrices of the form JS constitute a Lie subalgebra,

$$\{JS, JS'\} = JS'', \quad (4.3.72)$$

¹Recall from Exercise 2.3 that there are special methods for finding matrix roots if the matrix is positive symmetric.

and that this Lie algebra is $sp(2n, \mathbb{R})$. We next observe that the remaining matrices in $gl(2n, \mathbb{R})$ obey commutation rules of the form

$$\{JS, JA\} = JA', \quad (4.3.73)$$

$$\{JA, JA'\} = JS. \quad (4.3.74)$$

Finally, by arguments identical to those of the previous section, we conclude that if M is a matrix in $GL(2n, \mathbb{R})$ sufficiently near the identity, then it can be written in the form

$$M = \exp(JA') \exp(JS') \quad (4.3.75)$$

where S' is symmetric and A' is antisymmetric.

Any matrix R of the form

$$R = \exp(JS) \quad (4.3.76)$$

is symplectic, and therefore satisfies the relation

$$JR^T J^{-1} = R^{-1}. \quad (4.3.77)$$

See Sections 3.1 and 3.7. Let Q be any matrix of the form

$$Q = \exp(JA). \quad (4.3.78)$$

It is easily verified that Q satisfies the relation

$$JQ^T J^{-1} = Q. \quad (4.3.79)$$

We define a *J-symmetric* matrix to be *any* matrix Q that satisfies (3.9). See (3.1.12) and also note the similarity of (3.9) and (3.7.26). With these ideas in mind, we see from (3.5) that any M in $GL(2n, \mathbb{R})$ sufficiently near the identity has the decomposition

$$M = QR \quad (4.3.80)$$

where Q is *J-symmetric* and R is symplectic.

We have seen that orthogonal polar decomposition is possible globally and is unique. Is symplectic polar decomposition also possible globally and unique? To begin to answer these questions, we need to explore some of the properties of *J-symmetric* matrices. We will do so by proving a series of lemmas.

4.3.3 Extended Result on Symplectic Polar Decomposition

But still more can be said. Consider the matrix $N(\lambda M)$ which, according to (3.30), has the form

$$N(\lambda M) = (\lambda M)J(\lambda M)^T J^T = \lambda^2 M J M^T J^T = \lambda^2 N(M) \quad (4.3.81)$$

where λ is any real scalar in the range $0 < \lambda < \infty$. Also, let us view the set of all $2n \times 2n$ matrices as a linear vector space. This space is shown schematically in Figure 3.1. There we have depicted the zero matrix as the origin, and have also displayed the identity matrix

I. In addition we have depicted the various matrices $N(\lambda M)$ for fixed M as a ray (half line) emanating from the origin. These matrices do in fact lie on a ray because, according to (3.46), they are the λ^2 multiple of a fixed matrix. Finally, we have depicted the unit ball about the identity I . It is the set of matrices C that satisfy the requirement

$$\| C - I \| < 1 \quad (4.3.82)$$

for some choice of matrix norm. In drawing the unit ball about I we have assumed that the norm has the property $\| I \| = 1$ so that the zero matrix lies on the ball's surface. We are now ready to state an extended result in the form of a theorem.

Theorem 3.1 Suppose that for some value λ_0 the matrix $N(\lambda_0 M)$ lies *within* the unit ball around I . (This is the situation depicted in Figure 3.1.) Then the matrix M is invertible, has positive determinant, and has the symplectic polar decomposition (3.10).

Proof: Consider the matrix M_0 defined by the relation

$$M_0 = \lambda_0 M. \quad (4.3.83)$$

According to (3.46) the matrix N_0 associated with M_0 is given by the relation

$$N_0 = N(\lambda_0 M) = \lambda_0^2 N(M) = \lambda_0^2 N. \quad (4.3.84)$$

By hypothesis, we have the relation

$$\| N_0 - I \| = \| N(\lambda_0 M) - I \| < 1. \quad (4.3.85)$$

It follows that the series (3.43) for $\log(N_0)$ converges, and we can define a J -symmetric matrix Q_0 by the rule

$$Q_0 = \exp[(1/2) \log(N_0)]. \quad (4.3.86)$$

Moreover, according to Lemma 3.2, the matrix Q defined by

$$Q = (1/\lambda_0) Q_0 \quad (4.3.87)$$

is also J -symmetric. By (3.49), (3.51), and (3.52), it satisfies the relation

$$Q^2 = (1/\lambda_0)^2 Q_0^2 = (1/\lambda_0)^2 N_0 = N. \quad (4.3.88)$$

Consequently, M has the symplectic polar decomposition (3.10) with Q given by (3.52).

We also note that Q can be written in exponential form: We already know that $[(1/2) \log(N_0)]$ is J -symmetric. Consequently, according to Lemma 3.6, there exists an antisymmetric matrix A_0 such that

$$(1/2) \log(N_0) = JA_0, \quad (4.3.89)$$

and (3.51) can therefore be written in the form

$$Q_0 = \exp(JA_0). \quad (4.3.90)$$

Now use (3.52) and (3.55) to write Q in the form

$$\begin{aligned} Q &= (1/\lambda_0)Q_0 = \exp\{-[\log(\lambda_0)]I\} \exp(JA_0) \\ &= \exp\{JA_0 - [\log(\lambda_0)]I\} = \exp(JA) \end{aligned} \quad (4.3.91)$$

with A given by the relation

$$A = A_0 + [\log(\lambda_0)]J. \quad (4.3.92)$$

Finally, it follows from (3.56) and (3.7.129) that $\det Q > 0$ and hence, from (3.22), $\det M > 0$.

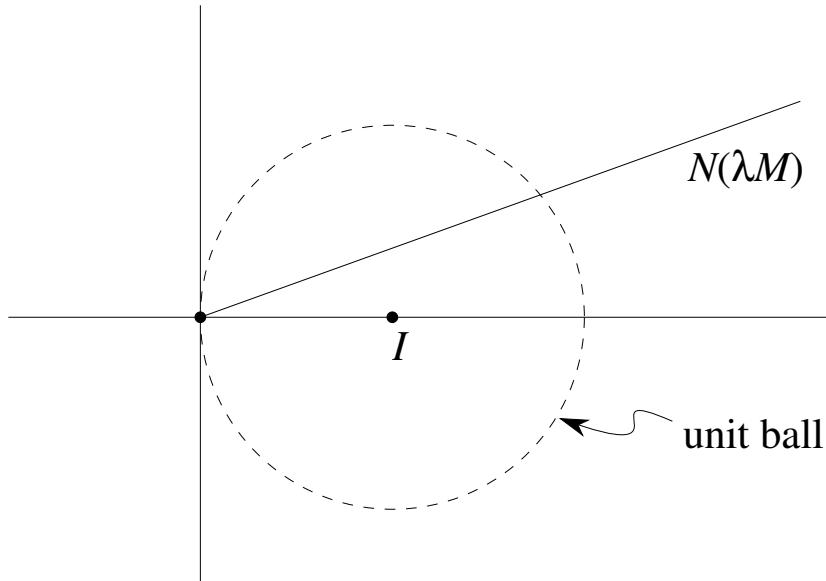


Figure 4.3.1: Schematic depiction of matrix space showing the zero matrix, the identity matrix I , the ray $N(\lambda M)$, and the unit ball around the identity matrix.

One might wonder if the condition of Theorem 3.1 is necessary. We can easily see that it is for the case of $GL(2, \mathbb{R})$. However, it is not necessary for some examples in the case of $GL(4, \mathbb{R})$, and presumably not for some examples in any $GL(2n, \mathbb{R})$ with $n \geq 2$. See Exercise 3.19.

In the $GL(2, \mathbb{R})$ case, for any 2×2 matrix M , we have the result

$$N(\lambda M) = \lambda^2 M J M^T J^T = \lambda^2 [\det(M)] I. \quad (4.3.93)$$

(See Exercise 3.1.2.) Thus we have the relation

$$\| N(\lambda M) - I \| = \| [\lambda^2 \det(M) - 1] I \| = |\lambda^2 \det(M) - 1| \| I \| . \quad (4.3.94)$$

Evidently, if $\det(M) > 0$, we can find a λ such that the right side of (3.59) is less than 1. Also, we can write M in the form (3.10) with Q and R given by the relation

$$Q = +[\det(M)]^{1/2} I, \quad (4.3.95)$$

$$R = +\{1/[\det(M)]^{1/2}\}M. \quad (4.3.96)$$

On the other hand, if $\det(M) < 0$, no choice of (real) λ will make the right side of (3.59) less than 1. This is consistent with Lemma 3.7 which states that $\det M > 0$ is a necessary condition for M to have a symplectic polar decomposition.

We also observe that we could replace the + signs in (3.60) and (3.61) by - signs and also obtain a (different) symplectic polar decomposition. Exercise 3.13 shows that the use of the Theorem 3.1 procedure, which is always possible in the 2×2 case when $\det M > 0$, produces the + signs choice.

Let us summarize our results in the language of cosets. (Again, see Section 5.12, if necessary, for a detailed discussion of cosets.) We have been dealing with the group $GL(2n, \mathbb{R})$ and its subgroup $Sp(2n, \mathbb{R})$. Form the coset space $GL(2n, \mathbb{R})/Sp(2n, \mathbb{R})$ consisting of the left cosets of $GL(2n, \mathbb{R})$ with respect to $Sp(2n, \mathbb{R})$. Equations (3.10), (3.56), and (3.57) indicate that the left cosets $GL(2n, \mathbb{R})/Sp(2n, \mathbb{R})$, for those M which satisfy the requirement of Theorem 3.1, can be put into one-to-one correspondence with $2n \times 2n$ (real) nonsingular antisymmetric matrices A . Such matrices form a linear vector space whose dimension m is given by the relation

$$m = \dim(A) = n(2n - 1). \quad (4.3.97)$$

Thus, the portion of $GL(2n, \mathbb{R})$ that satisfies the requirement of Theorem 3.1 has the topology of $E^m \times Sp(2n, \mathbb{R})$ with m given by (3.62).

4.3.4 Symplectic Polar Decomposition Not Globally Possible

We have already seen that symplectic polar decomposition is not possible for M in the cosets with $\det M < 0$. Are there other cosets as well for which symplectic decomposition is impossible? We will see that there are. Therefore symplectic polar decomposition is not possible globally even with the restriction $\det M > 0$.

Consider, as a possible 4×4 counter example, the diagonal matrix M given by

$$M = \begin{pmatrix} \mu_1 & 0 & 0 & 0 \\ 0 & \mu_2 & 0 & 0 \\ 0 & 0 & \nu_1 & 0 \\ 0 & 0 & 0 & \nu_2 \end{pmatrix} \quad (4.3.98)$$

where the μ_j and ν_j are real and nonzero. Its determinant satisfies the condition

$$\det(M) = \mu_1\mu_2\nu_1\nu_2. \quad (4.3.99)$$

Take for J the matrix

$$J = \begin{pmatrix} J_2 & 0 \\ 0 & J_2 \end{pmatrix}. \quad (4.3.100)$$

See (3.2.10). Then we find, using (3.30) and the results of Exercise 3.1.2, the relations

$$N(M) = \begin{pmatrix} \mu_1\mu_2 & 0 & 0 & 0 \\ 0 & \mu_1\mu_2 & 0 & 0 \\ 0 & 0 & \nu_1\nu_2 & 0 \\ 0 & 0 & 0 & \nu_1\nu_2 \end{pmatrix} \quad (4.3.101)$$

and

$$N(\lambda M) = \begin{pmatrix} \lambda^2 \mu_1 \mu_2 & 0 & 0 & 0 \\ 0 & \lambda^2 \mu_1 \mu_2 & 0 & 0 \\ 0 & 0 & \lambda^2 \nu_1 \nu_2 & 0 \\ 0 & 0 & 0 & \lambda^2 \nu_1 \nu_2 \end{pmatrix}. \quad (4.3.102)$$

It follows that

$$N(\lambda M) - I = - \begin{pmatrix} 1 - \lambda^2 \mu_1 \mu_2 & 0 & 0 & 0 \\ 0 & 1 - \lambda^2 \mu_1 \mu_2 & 0 & 0 \\ 0 & 0 & 1 - \lambda^2 \nu_1 \nu_2 & 0 \\ 0 & 0 & 0 & 1 - \lambda^2 \nu_1 \nu_2 \end{pmatrix}. \quad (4.3.103)$$

Therefore, using the spectral norm, which is the strongest, we find the result

$$\|N(\lambda M) - I\| = \max[|(1 - \lambda^2 \mu_1 \mu_2)|, |(1 - \lambda^2 \nu_1 \nu_2)|]. \quad (4.3.104)$$

We see that the ray $N(\lambda M)$ does not pass through the interior of the unit ball about the identity if

$$\mu_1 \mu_2 < 0 \text{ or } \nu_1 \nu_2 < 0. \quad (4.3.105)$$

Consequently the series (3.43) used to construct $\log(N)$, and hence Q , might be expected to diverge in these cases.²

We will show that, in fact, for these cases there is no J -symmetric matrix Q such that

$$Q^2 = N(\lambda M). \quad (4.3.106)$$

Suppose that such a Q exists. By Lemma 3.6 there is an antisymmetric matrix A such that $Q = JA$. Write A in the 2×2 block form

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \quad (4.3.107)$$

Then we find for Q the result

$$Q = JA = \begin{pmatrix} J_2 & 0 \\ 0 & J_2 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} J_2 a & J_2 b \\ J_2 c & J_2 d \end{pmatrix}. \quad (4.3.108)$$

And for Q^2 we find the result

$$Q^2 = \begin{pmatrix} J_2 a & J_2 b \\ J_2 c & J_2 d \end{pmatrix} \begin{pmatrix} J_2 a & J_2 b \\ J_2 c & J_2 d \end{pmatrix} = \begin{pmatrix} (J_2 a)^2 + J_2 b J_2 c & J_2 a J_2 b + J_2 b J_2 d \\ J_2 c J_2 a + J_2 d J_2 c & J_2 c J_2 b + (J_2 d)^2 \end{pmatrix}. \quad (4.3.109)$$

Let us work out the properties of the entries in Q^2 . Since A is antisymmetric, the matrices a and d are antisymmetric and therefore have the form

$$a = \begin{pmatrix} 0 & \alpha \\ -\alpha & 0 \end{pmatrix}, \quad (4.3.110)$$

²In fact, in this case because of the diagonal form of $[N(\lambda M) - I]$, it easily verified that the series (3.43) does diverge.

$$d = \begin{pmatrix} 0 & \delta \\ -\delta & 0 \end{pmatrix}. \quad (4.3.111)$$

Consequently, we have the relations

$$J_2a = -\alpha I, \quad (4.3.112)$$

$$J_2d = -\delta I, \quad (4.3.113)$$

from which it follows that

$$(J_2a)^2 = \alpha^2 I, \quad (4.3.114)$$

$$(J_2d)^2 = \delta^2 I. \quad (4.3.115)$$

Next we find, again using the results of Exercise 3.1.2, that

$$J_2bJ_2c = -J_2bJ_2b^T = \det(b)I \quad (4.3.116)$$

and

$$J_2cJ_2b = -J_2b^T J_2b = \det(b)I. \quad (4.3.117)$$

Here we have used the relation

$$c = -b^T, \quad (4.3.118)$$

which also follows from the fact that A is antisymmetric. Finally, we have the results

$$J_2aJ_2b + J_2bJ_2d = -(\alpha + \delta)J_2b, \quad (4.3.119)$$

$$J_2cJ_2a + J_2dJ_2c = -(\alpha + \delta)J_2c = (\alpha + \delta)J_2b^T. \quad (4.3.120)$$

Now require that (3.71) hold. So doing yields the relations

$$\alpha^2 + \det(b) = \lambda^2 \mu_1 \mu_2, \quad (4.3.121)$$

$$\delta^2 + \det(b) = \lambda^2 \nu_1 \nu_2, \quad (4.3.122)$$

$$-(\alpha + \delta)J_2b = 0, \quad (4.3.123)$$

$$(\alpha + \delta)J_2b^T = 0. \quad (4.3.124)$$

Note that the relations (3.88) and (3.89) are equivalent, and yield the two possibilities

$$\alpha = -\delta \quad (4.3.125)$$

or

$$J_2b = 0 \text{ which implies } b = 0. \quad (4.3.126)$$

If (3.90) holds, the relations (3.86) and (3.87) become

$$\alpha^2 + \det(b) = \lambda^2 \mu_1 \mu_2, \quad (4.3.127)$$

$$\alpha^2 + \det(b) = \lambda^2 \nu_1 \nu_2, \quad (4.3.128)$$

and they are contradictory if $\mu_1\mu_2 \neq \nu_1\nu_2$. If (3.91) holds, the relations (3.86) and (3.87) become

$$\alpha^2 = \lambda^2 \mu_1 \mu_2, \quad (4.3.129)$$

$$\delta^2 = \lambda^2 \nu_1 \nu_2, \quad (4.3.130)$$

and at least one of them is an impossibility in the cases (3.70).

We conclude that no J -symmetric matrix Q exists that satisfies (3.71) when M is of the form (3.63) and (3.70) holds. Therefore, symplectic polar decomposition for such M is impossible. The same is true for any matrices M' that are in the same cosets as such M . Finally we note from (3.64) that if both the cases (3.70) hold, then it is still possible to have $\det(M) > 0$.

From Theorem 3.1 we know that a *sufficient* condition for M to have a symplectic polar decomposition is that the ray $N(\lambda M)$ intersect the unit ball about I . From the example of this section one might be tempted to conjecture that this intersection condition is also a *necessary* condition. However, as already mentioned earlier, Exercise 3.19 shows that the intersection condition is not necessary for a particular $GL(4, \mathbb{R})$ example.

4.3.5 Uniqueness of Symplectic Polar Decomposition

There remains the question of uniqueness. Suppose that M has the two symplectic polar decompositions

$$M = QR \quad (4.3.131)$$

and

$$M = Q'R'. \quad (4.3.132)$$

Then we see that

$$Q' = QR(R')^{-1}. \quad (4.3.133)$$

But, since symplectic matrices form a group, the matrix $R(R')^{-1}$ is symplectic, and therefore Q' and Q are in the same $GL(2n, \mathbb{R})/Sp(2n, \mathbb{R})$ coset. By Lemma 3.2, $-Q$ is a J -symplectic matrix if $+Q$ is, and they are related under multiplication by the symplectic matrix $-I$, and are therefore in the same coset. Thus, if symplectic polar decomposition is possible at all, there are always at least two possibilities. In the case of Theorem 3.1 we imposed the unit ball condition of Figure 3.1, and were able to make the choice specified by (3.51) and (3.52). In the 2×2 case this choice dictates the $+$ sign in (3.60). We will see that an analogous choice can be made in the general case.

As in Theorem 3.1, let Q_0 be the matrix associated with $N_0 = N(\lambda_0 M)$. Write the matrix identity

$$-2I = (-Q_0 - I) + (Q_0 - I) \quad (4.3.134)$$

and use the triangle inequality (3.7.12) to deduce the inequality

$$\| -2I \| \leq \| -Q_0 - I \| + \| Q_0 - I \| . \quad (4.3.135)$$

With an appropriate norm, such as the spectral norm, we have the relation

$$\| 2I \| = 2, \quad (4.3.136)$$

and we conclude from (3.100) that

$$\| -Q_0 - I \| \geq 2 - \| Q_0 - I \| . \quad (4.3.137)$$

We will now seek an estimate for the quantity $\| Q_0 - I \|$.

Consider the function $g(x)$ defined by the equation

$$(1-x)^{1/2} = 1 - g(x). \quad (4.3.138)$$

It is readily verified that this function has the expansion

$$g(x) = \sum_{\ell=1}^{\infty} d_{\ell} x^{\ell} = x/2 + x^2/8 + \dots \quad (4.3.139)$$

where all the coefficients d_{ℓ} are *positive*. Moreover, $g(x)$ satisfies the inequality

$$(x/2) \leq g(x) \leq x \text{ for } x \in [0, 1]. \quad (4.3.140)$$

Instead of the method of Theorem 3.1, let us use a more direct (but equivalent) way of defining Q_0 by writing

$$Q_0 = (N_0)^{1/2} = [I - (I - N_0)]^{1/2} = I - g(I - N_0). \quad (4.3.141)$$

As in the proof of Theorem 3.1, let us also make the assumption that

$$\| N_0 - I \| < 1. \quad (4.3.142)$$

Under this hypothesis the relation (3.106) yields the inequality

$$\begin{aligned} \| Q_0 - I \| &= \| -g(I - N_0) \| = \| g(I - N_0) \| \\ &= \| \sum_{\ell=1}^{\infty} d_{\ell} (I - N_0)^{\ell} \| \leq \sum_{\ell=1}^{\infty} d_{\ell} \| I - N_0 \|^{\ell} \\ &= g(\| I - N_0 \|) = g(\| N_0 - I \|) \leq \| N_0 - I \| < 1. \end{aligned} \quad (4.3.143)$$

Here we have made use of the positivity of the d_{ℓ} and the relation (3.105).

Finally, combine (3.102) and (3.108) to get the result

$$\| -Q_0 - I \| > 1. \quad (4.3.144)$$

We see from (3.108) that the use of (3.106) or, equivalently, the use of the method of Theorem 3.1, produces a Q_0 that is inside the unit ball shown in Figure 3.1. And, correspondingly, (3.109) shows that $-Q_0$ is outside this unit ball. Thus, the method of Theorem 3.1 assures that the J -symplectic factor Q_0 is as close to the identity I as possible.

4.3.6 Concluding Summary

Let us summarize what has been learned. We have seen that symplectic polar decomposition is possible and unique if M is sufficiently near the symplectic group so that $N(M)$ lies within the unit ball about I . We have extended this result to show that symplectic polar decomposition is possible and unique if the ray $N(\lambda M)$ passes through the unit ball about I . Also, we have found a family of counter examples that show that symplectic polar decomposition is not possible globally. Naturally, for any counter example, the ray $N(\lambda M)$ cannot pass through the unit ball about I . However, as illustrated in Exercise 3.21, there are examples where the ray $N(\lambda M)$ does not pass through the unit ball about I and symplectic polar decomposition is still possible and unique. See also Exercises 3.22 through 3.24 for further examples of when symplectic polar decomposition is and is not possible.

Exercises

4.3.1. Verify the commutation rules (3.2) through (3.4).

4.3.2. Suppose M is a $2n \times 2n$ matrix near the identity. Then M can be written in the form

$$M = \exp[\epsilon(JS + JA)] \quad (4.3.145)$$

where ϵ is a small parameter. Show that M can also be written in the form

$$M = \exp(\epsilon JA') \exp(\epsilon JS'), \quad (4.3.146)$$

and determine the first few terms in A' and S' when expressed as a power series expansion in ϵ .

4.3.3. Show that any matrix of the form (3.8) satisfies (3.9), and hence is J -symmetric.

4.3.4. Review Exercise 3.1.9. Employing a slightly different notation for the symplectic transpose, define the matrix M' by the rule

$$M' = M^S = JM^T J^{-1}. \quad (4.3.147)$$

Show that any matrix of the form JA is symmetric under this priming operation,

$$(JA)' = JA, \quad (4.3.148)$$

and any matrix of the form JS is antisymmetric,

$$(JS)' = -JS. \quad (4.3.149)$$

Thus, verify that (3.1) is a decomposition of B into symmetric and antisymmetric parts with respect to the symplectic transpose operation.

4.3.5. Verify the calculations associated with Lemmas 3.1 through 3.11.

4.3.6. Refer to Lemma 3.1. Show that any two of the following three properties implies the third: (i) symmetric, (ii) commutes with J , (iii) J -symmetric.

4.3.7. Suppose M_1 and M_2 are two commuting matrices, and suppose M_2 is invertible. Verify that M_1 and M_2^{-1} also commute. Show that the set of all commuting J -symmetric matrices in $GL(2n, \mathbb{R})$ forms a group.

4.3.8. Review Lemma 3.6. Show that if A is any antisymmetric matrix, there exists another antisymmetric matrix A' such that

$$JA = A'J. \quad (4.3.150)$$

4.3.9. Given any factorization of the form (3.10), use Lemma 3.8 to show that M also has the factorization

$$M = QR = RQ^{\text{tr}}. \quad (4.3.151)$$

If Q is of the form (3.8), find the A^{tr} associated with Q^{tr} .

4.3.10. If M is symplectic, verify that N as given by (3.30) satisfies $N = I$.

4.3.11. If you have not already done so in Exercise 3.5, verify (3.40) and (3.41).

4.3.12. Verify the steps in the proof of Theorem 3.1.

4.3.13. Verify (3.58) and (3.59). Show that

$$\| I \| \geq 1 \quad (4.3.152)$$

for any choice of norm. Hint: Apply (3.7.10) and (3.7.13) to $\| I^2 \|$. Show that if $\det(M) < 0$, no choice of λ will make the right side of (3.59) less than 1. Show that applying the method of Theorem 3.1 in the 2×2 case produces the symplectic polar decomposition given by (3.60) and (3.61).

4.3.14. Suppose the matrices M and M' belong to the same $GL(2n, \mathbb{R})/Sp(2n, \mathbb{R})$ coset. Show that they then have the same determinant. Is the converse true?

4.3.15. Verify (3.66).

4.3.16. Suppose that, for the matrix M given by (3.63), there are the conditions $\mu_1\mu_2 > 0$ and $\nu_1\nu_2 > 0$. Show that in this case a J -symmetric solution to (6.71) is given by the relation

$$Q = [N(\lambda M)]^{1/2} = \lambda \begin{pmatrix} [\mu_1\mu_2]^{1/2} & 0 & 0 & 0 \\ 0 & [\mu_1\mu_2]^{1/2} & 0 & 0 \\ 0 & 0 & [\nu_1\nu_2]^{1/2} & 0 \\ 0 & 0 & 0 & [\nu_1\nu_2]^{1/2} \end{pmatrix}. \quad (4.3.153)$$

Instead, suppose one or both of the conditions (3.70) holds. Then, from (3.118), one might surmise that (3.71) has only imaginary solutions. This is not the case. Show, for example if both conditions (3.70) hold, then (3.71) has the *real* solution

$$Q = \lambda \begin{pmatrix} [-\mu_1\mu_2]^{1/2}J_2 & 0 \\ 0 & [-\nu_1\nu_2]^{1/2}J_2 \end{pmatrix}. \quad (4.3.154)$$

However note that this solution Q is *not* J -symmetric. Consequently, there is no contradiction with the results of Section 4.3.5.

4.3.17. Graph the function $g(x)$ given by (3.103). Verify all claims made for $g(x)$. Determine the coefficients d_ℓ and the domain of convergence of this series. Verify (3.106) and (3.108). Show that Q_0 as defined by (3.106) is J -symmetric. Show that use of (3.106) gives the same result as that of Theorem 3.1.

4.3.18. We know that $N(M)$ is J -symmetric and therefore, by Lemma 3.6, there is an antisymmetric matrix A' such that

$$N(M) = JA'. \quad (4.3.155)$$

By the same lemma, if Q is J -symmetric there is an antisymmetric A such that (3.15) holds. Now suppose that there is a Q of the form (3.15) such that (3.38) is satisfied. Show, using the representations (3.15) and (3.120), that there is the relation

$$JA' = JAJA, \quad (4.3.156)$$

from which it follows that

$$A' = AJA. \quad (4.3.157)$$

Since A is antisymmetric, (3.122) can also be written in the form

$$-A' = AJA^T. \quad (4.3.158)$$

Recall the work of Section 3.12. We see that if there exists a J -symmetric Q such that (3.38) holds, then the antisymmetric matrix A associated with this Q is *also* a Darboux matrix that transforms J to $-A'$.

4.3.19. In Section 4.3.5 we studied the space of all *diagonal* matrices in $GL(4, \mathbb{R})$ to determine which of them had symplectic polar decompositions. Ideally we would like to do the same for all matrices in $GL(4, \mathbb{R})$, but this seems to be a formidable task because $GL(4, \mathbb{R})$ is 16 dimensional.³ We know that in principle it is sufficient to examine the coset space $GL(4, \mathbb{R})/Sp(4, \mathbb{R})$, which is 6 dimensional. However, the parameterization of the coset space $GL(2n, \mathbb{R})/Sp(2n, \mathbb{R})$ is complicated. See Appendix P. As a simpler task, this exercise begins to examine the subset of matrices $SO(4, \mathbb{R}) \subset GL(4, \mathbb{R})$. This set (also 6 dimensional), while incomplete in the sense of not embracing all cosets, is easier to study. The work of this exercise and the next will show that every element in $SO(4, \mathbb{R})$ has a symplectic polar decomposition. It will also provide information about $SO(4, \mathbb{R})$ that will be valuable for subsequent use.

The Lie algebra $so(4, \mathbb{R})$ consists of all real 4×4 antisymmetric matrices A . As with the case of symmetric matrices S , it is convenient to decompose A into matrices A^a and A^c that *anticommute* and *commute* with J , respectively,

$$A = A^a + A^c. \quad (4.3.159)$$

Show that

$$A^a = (1/2)(A - JAJ^{-1}) \quad (4.3.160)$$

³In fact, we would like to do the same for $GL(2n, \mathbb{R})$ for all n ; but at least $GL(4, \mathbb{R})$ is the first nontrivial case.

and

$$A^c = (1/2)(A + JAJ^{-1}). \quad (4.3.161)$$

Verify that the Lie algebra formed by the set of antisymmetric matrices has the property

$$\{A^c, (A^c)'\} = (A^c)'', \quad (4.3.162)$$

$$\{A^c, A^a\} = (A^a)', \quad (4.3.163)$$

$$\{A^a, (A^a)'\} = A^c. \quad (4.3.164)$$

If A is sufficiently small, the BCH and Zassenhaus series converge, and we may achieve the factorization

$$\exp(A) = \exp(A^a + A^c) = \exp[(A^a)'] \exp[(A^c)']. \quad (4.3.165)$$

Show that any element of the form $\exp(A^a)$ is J -symmetric, and any element of the form $\exp(A^c)$ is symplectic. The relation (3.130) shows that any $SO(4, \mathbb{R})$ element sufficiently near the identity has a symplectic polar decomposition. This is to be expected because we already know that *every* matrix sufficiently near the identity has a symplectic polar decomposition.

The most general 4×4 antisymmetric matrix A can be written in the form

$$A = \begin{pmatrix} 0 & \alpha & \beta & \gamma \\ -\alpha & 0 & \delta & \epsilon \\ -\beta & -\delta & 0 & \zeta \\ -\gamma & -\epsilon & -\zeta & 0 \end{pmatrix}. \quad (4.3.166)$$

Using the form of J given by (3.65), show that

$$A^c = (1/2) \begin{pmatrix} 0 & 2\alpha & \beta + \epsilon & \gamma - \delta \\ -2\alpha & 0 & -\gamma + \delta & \beta + \epsilon \\ -\beta - \epsilon & \gamma - \delta & 0 & 2\zeta \\ -\gamma + \delta & -\beta - \epsilon & -2\zeta & 0 \end{pmatrix} \quad (4.3.167)$$

and

$$A^a = (1/2) \begin{pmatrix} 0 & 0 & \beta - \epsilon & \gamma + \delta \\ 0 & 0 & \gamma + \delta & -\beta + \epsilon \\ -\beta + \epsilon & -\gamma - \delta & 0 & 0 \\ -\gamma - \delta & \beta - \epsilon & 0 & 0 \end{pmatrix}. \quad (4.3.168)$$

Evidently the space of matrices of the form A^c is 4 dimensional, and the space of matrices of the form A^a is 2 dimensional. Let us seek a convenient basis for each.

Begin with the A^c . Evidently matrices of the form JS^c are antisymmetric and commute with J . Verify that there is the one-to-one correspondence

$$JS^c \leftrightarrow A^c. \quad (4.3.169)$$

We already know that the matrices JS^c are associated with the $u(2)$ part of $sp(4, \mathbb{R})$. Looking ahead, a convenient basis for these matrices, in the case that J is of the form (3.1.1), will be found in Exercise 5.7.8. They are the matrices B^0 through B^3 given in (5.7.44). If we can find their counterparts for the case that J is given by (3.65), then we will have found a

convenient basis for the A^c . This is easily done. Review Section 3.2. Show that in the 4×4 case the matrix P of (3.2.5) is given by the relation

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (4.3.170)$$

Evidently P is both symmetric and orthogonal. Show that the desired basis for the A^c can be taken to be the matrices C^j defined by the rule

$$C^j = PB^jP. \quad (4.3.171)$$

Verify that the matrices C^j are given by the relations

$$C^0 = PB^0P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \quad (4.3.172)$$

$$C^1 = PB^1P = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}, \quad (4.3.173)$$

$$C^2 = PB^2P = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \quad (4.3.174)$$

$$C^3 = PB^3P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad (4.3.175)$$

and are of the form (3.132). Verify that they satisfy the commutation rules

$$\{C^0, C^j\} = 0, \quad j = 0, 1, 2, 3; \quad (4.3.176)$$

$$\{C^1, C^2\} = -2C^3, \quad (4.3.177)$$

$$\{C^2, C^3\} = -2C^1, \quad (4.3.178)$$

$$\{C^3, C^1\} = -2C^2. \quad (4.3.179)$$

Next find a basis for the A^a . By looking at (3.133), show that a convenient choice is

$$E^1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \quad (4.3.180)$$

and

$$E^2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}. \quad (4.3.181)$$

Show that they, together with the C^j , obey the commutation relations

$$\{E^1, E^2\} = 2C^0, \quad (4.3.182)$$

$$\{C^0, E^1\} = 2E^2, \quad (4.3.183)$$

$$\{C^0, E^2\} = -2E^1, \quad (4.3.184)$$

$$\{C^j, E^1\} = \{C^j, E^2\} = 0, \quad j = 1, 2, 3. \quad (4.3.185)$$

After a bit of algebraic experimentation (and in anticipation of Exercise 11.1.6), one finds that it is convenient to relabel and renormalize the basis just found by making the definitions

$$\begin{aligned} G^1 &= -(1/2)E^1 = (1/2) \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \\ G^2 &= -(1/2)E^2 = (1/2) \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \\ G^3 &= (1/2)C^0 = (1/2) \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}. \end{aligned} \quad (4.3.186)$$

$$\begin{aligned} H^1 &= (1/2)C^3 = (1/2) \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \\ H^2 &= (1/2)C^2 = (1/2) \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \\ H^3 &= (1/2)C^1 = (1/2) \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}. \end{aligned} \quad (4.3.187)$$

Show that the G^j and H^k satisfy the pleasing commutation rules

$$\{G^1, G^2\} = G^3, \quad (4.3.188)$$

$$\{G^2, G^3\} = G^1, \quad (4.3.189)$$

$$\{G^3, G^1\} = G^2, \quad (4.3.190)$$

$$\{H^1, H^2\} = H^3, \quad (4.3.191)$$

$$\{H^2, H^3\} = H^1, \quad (4.3.192)$$

$$\{H^3, H^1\} = H^2, \quad (4.3.193)$$

$$\{G^j, H^k\} = 0 \text{ for } j, k = 1, 2, 3; \quad (4.3.194)$$

and the anticommutation relations

$$\{G^j, G^k\}_+ = \{H^j, H^k\}_+ = -(1/2)\delta_{jk}I. \quad (4.3.195)$$

Show also that there are the relations

$$(G^1)^2 + (G^2)^2 + (G^3)^2 = (H^1)^2 + (H^2)^2 + (H^3)^2 = -(3/4)I. \quad (4.3.196)$$

You have verified, as advertised in the discussion associated with Table 3.7.1, that the Lie algebra $so(4, \mathbb{R})$ is the direct sum of two mutually commuting $su(2)$ Lie algebras. [Strictly speaking, based only on the commutation rules, we cannot tell at this stage whether it is $su(2)$ or $so(3, \mathbb{R})$ that we have found. In the next exercise you will verify that it is indeed $su(2)$.] Note that all the matrices G^j and H^k are real and antisymmetric, and form a basis for the 6-dimensional set of antisymmetric 4×4 matrices. Verify that G^1 through G^3 are linear combinations of pair-wise commuting generators for rotations in the (1,4 and 2,3), (1,3 and 2,4), and (1,2 and 3,4) planes, respectively. Verify that H^1 through H^3 are also linear combinations of pair-wise commuting generators for rotations in the (1,2 and 3,4), (1,3 and 2,4), and (1,4 and 2,3) planes, respectively. Verify that, given a four-element set, there are three ways of forming pairs of disjoint two-element subsets. This combinatorial fact lies behind the possible construction of the three G^j and the three H^k .

4.3.20. Review Exercise 3.19 above. It set up the machinery for a study of $SO(4, \mathbb{R})$. The purpose of this exercise is to show that *all* elements of $SO(4, \mathbb{R})$ have symplectic polar decompositions. In so doing we will also learn more about $so(4, \mathbb{R})$ and the two mutually commuting $su(2)$ Lie algebras within it.

Introduce the notation

$$\mathbf{G} = (G^1, G^2, G^3), \quad \mathbf{H} = (H^1, H^2, H^3). \quad (4.3.197)$$

Also introduce the vectors

$$\mathbf{s} = (s^1, s^2, s^3), \quad \mathbf{t} = (t^1, t^2, t^3). \quad (4.3.198)$$

Finally employ the notation

$$\mathbf{s} \cdot \mathbf{G} = s_1 G^1 + s_2 G^2 + s_3 G^3, \text{ etc.} \quad (4.3.199)$$

We know that the G^j and H^k form a basis for the Lie algebra $so(4, \mathbb{R})$. It follows from Section 3.8.1 that the most general element in $SO(4, \mathbb{R})$ can be written in the form

$$O(\mathbf{s}, \mathbf{t}) = \exp(\mathbf{s} \cdot \mathbf{G} + \mathbf{t} \cdot \mathbf{H}). \quad (4.3.200)$$

Now, since the G^j and H^k commute, we may also write

$$O(\mathbf{s}, \mathbf{t}) = \exp(\mathbf{s} \cdot \mathbf{G}) \exp(\mathbf{t} \cdot \mathbf{H}). \quad (4.3.201)$$

We observe that the factor $\exp(\mathbf{t} \cdot \mathbf{H})$ is an element in $Sp(4, \mathbb{R})$. Therefore, in order to achieve a symplectic polar decomposition for $O(\mathbf{s}, \mathbf{t})$, we only need to achieve a symplectic polar decomposition for $\exp(\mathbf{s} \cdot \mathbf{G})$.

At this point let us pause to explore more of the properties of the G^j and H^k . Review Exercise 8.7.12. Verify that the G^j and H^j satisfy the *same* multiplication rules as the K^j . But, unlike some of the K^j , they are purely real. Verify in particular that there are the relations

$$(\mathbf{s} \cdot \mathbf{G})^2 = -(1/4)(\mathbf{s} \cdot \mathbf{s})I \quad (4.3.202)$$

and

$$(\mathbf{t} \cdot \mathbf{H})^2 = -(1/4)(\mathbf{t} \cdot \mathbf{t})I. \quad (4.3.203)$$

Use these relations to show that there are the explicit results

$$\exp(\mathbf{s} \cdot \mathbf{G}) = I \cos(s/2) + (\mathbf{s} \cdot \mathbf{G})(2/s) \sin(s/2), \quad (4.3.204)$$

$$\exp(\mathbf{t} \cdot \mathbf{H}) = I \cos(t/2) + (\mathbf{t} \cdot \mathbf{H})(2/t) \sin(t/2) \quad (4.3.205)$$

where

$$s = (\mathbf{s} \cdot \mathbf{s})^{1/2}, \quad t = (\mathbf{t} \cdot \mathbf{t})^{1/2}. \quad (4.3.206)$$

It follows that the G^j and H^k generate bona fide realizations of the group $SU(2)$ rather than the group $SO(3, \mathbb{R})$. See Exercise 3.7.30 for the distinction. Note also the coefficient $(3/4)$ occurring in (3.161) is the same as that in (3.7.203) for $su(2)$, and not that in (3.7.204) for $so(3, \mathbb{R})$.

After this pleasant interruption, let us return to the main discussion. Write $\exp(\mathbf{s} \cdot \mathbf{G})$ in the Euler angle form

$$\exp(\mathbf{s} \cdot \mathbf{G}) = \exp(\phi G_3) \exp(\theta G_2) \exp(\psi G_3). \quad (4.3.207)$$

Then we may also write

$$\exp(\mathbf{s} \cdot \mathbf{G}) = \{\exp(\phi G_3) \exp(\theta G_2) \exp(-\phi G_3)\} \{\exp[(\phi + \psi) G_3]\}. \quad (4.3.208)$$

We know that $\exp(\theta G_2)$ is J -symmetric and $\exp(\phi G_3)$ is symplectic. Therefore, by Lemma 3.8, the first curly-bracketed factor in (3.173) is J -symmetric. Also, the second curly-bracketed factor in (3.173) is symplectic. Consequently, we have achieved the desired symplectic polar decomposition.

We end this exercise with a few more observations about $SO(4, \mathbb{R})$. Let us write (3.166) in the form

$$O(U, V) = UV \quad (4.3.209)$$

where

$$U(\mathbf{s}) = \exp(\mathbf{s} \cdot \mathbf{G}) \quad (4.3.210)$$

and

$$V\mathbf{t}) = \exp(\mathbf{t} \cdot \mathbf{H}). \quad (4.3.211)$$

Evidently the matrices U and V form two separate subgroups of $SO(4, \mathbb{R})$ and each of these subgroups has the same topology as $SU(2)$. From (3.169) we see that

$$U(\mathbf{s}) = -I \text{ when } s = 2\pi, \quad (4.3.212)$$

with an analogous result for V . Show that it follows that if U is a matrix of the form (3.175), then so is $-U$. Verify an analogous result for V . We also conclude from (3.174) that

$$O(-U, -V) = O(U, V). \quad (4.3.213)$$

Therefore, (3.174) provides a two-to-one homomorphism of $SU(2) \otimes SU(2)$ onto $SO(4, \mathbb{R})$.

4.3.21. The two previous exercises showed that all elements in $SO(4, \mathbb{R})$ have symplectic polar decompositions. This exercise examines a particular one-parameter subgroup of $SO(4, \mathbb{R})$. Since it is a subgroup of $SO(4, \mathbb{R})$, all its elements must have symplectic polar decompositons. We will apply the methods of Theorem 3.1 to matrices in this subgroup; and in so doing we will discover that the conditions of Theorem 3.1, while sufficient, are not necessary.

Consider a rotation by angle θ in the q_1, q_2 plane. It has the effect

$$q'_1 = q_1 c + q_2 s, \quad (4.3.214)$$

$$q'_2 = -q_1 s + q_2 c, \quad (4.3.215)$$

$$p'_1 = p_1, \quad (4.3.216)$$

$$p'_2 = p_2 \quad (4.3.217)$$

where

$$c = \cos \theta \quad (4.3.218)$$

and

$$s = \sin \theta. \quad (4.3.219)$$

Show that in the (q_1, p_1, q_2, p_2) basis this rotation is represented by the matrix $O(\theta)$ given by the relation

$$O(\theta) = \begin{pmatrix} c & 0 & s & 0 \\ 0 & 1 & 0 & 0 \\ -s & 0 & c & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (4.3.220)$$

Seek to write $O(\theta)$ in exponential form. For small θ , and through terms of degree one, show that (3.185) has the expansion

$$O(\theta) = I + \theta A \quad (4.3.221)$$

with

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (4.3.222)$$

Verify that

$$O(\theta) = \exp(\theta A). \quad (4.3.223)$$

Next, using (3.125) and (3.126), verify that A has the decomposition (3.124) with

$$A^a = (1/2)E^2 = -G^2 \quad (4.3.224)$$

and

$$A^c = (1/2)C^2 = H^2. \quad (4.3.225)$$

Observe from (3.159) that in this case A^a and A^c commute. Verify, therefore, that there is the relation

$$O(\theta) = \exp(\theta A) = \exp[\theta(A^a + A^c)] = \exp(\theta A^a) \exp(\theta A^c). \quad (4.3.226)$$

For future use show that there are the explicit matrix results

$$\exp(\theta A^a) = \exp[(\theta/2)E^2] = I \cos(\theta/2) + E^2 \sin(\theta/2) = \begin{pmatrix} c' & 0 & s' & 0 \\ 0 & c' & 0 & -s' \\ -s' & 0 & c' & 0 \\ 0 & s' & 0 & c' \end{pmatrix} \quad (4.3.227)$$

and

$$\exp(\theta A^c) = \exp[(\theta/2)C^2] = I \cos(\theta/2) + C^2 \sin(\theta/2) = \begin{pmatrix} c' & 0 & s' & 0 \\ 0 & c' & 0 & s' \\ -s' & 0 & c' & 0 \\ 0 & -s' & 0 & c' \end{pmatrix} \quad (4.3.228)$$

where

$$c' = \cos(\theta/2), \quad (4.3.229)$$

and

$$s' = \sin(\theta/2). \quad (4.3.230)$$

Verify, by explicit matrix multiplication, that (3.191) holds.

Define matrices $Q(\theta)$ and $R(\theta)$ by the rules

$$Q(\theta) = \exp(\theta A^a) = \exp[(\theta/2)E^2] \quad (4.3.231)$$

and

$$R(\theta) = \exp(\theta A^c) = \exp[(\theta/2)C^2] \quad (4.3.232)$$

so that (3.191) can be written in the form

$$O = QR. \quad (4.3.233)$$

Verify that Q is J -symmetric and R is symplectic. You have shown, as expected, that $O(\theta)$ has a symplectic polar decomposition for all θ . Verify that $Q(\theta)$ can be written in the form

$$Q(\theta) = \exp[JA'(\theta)] \quad (4.3.234)$$

Find $A'(\theta)$ explicitly and verify that it is real and antisymmetric.

Suppose now that we are just given the matrix M , with

$$M = O(\theta), \quad (4.3.235)$$

and we attempt to find a symplectic polar decomposition for M using the method of Theorem 3.1. Show that

$$N(\lambda M) = \lambda^2 \begin{pmatrix} c & 0 & s & 0 \\ 0 & c & 0 & -s \\ -s & 0 & c & 0 \\ 0 & s & 0 & c \end{pmatrix}. \quad (4.3.236)$$

As a sanity check, verify that

$$N(M) = Q^2 \quad (4.3.237)$$

using (3.196) and the explicit matrix results (3.192) and (3.201).

The next task is to compute the spectral norm of $[N(\lambda M) - I]$. Define the matrix V by the rule

$$V = N(\lambda M) - I = \begin{pmatrix} e & 0 & \lambda^2 s & 0 \\ 0 & e & 0 & -\lambda^2 s \\ -\lambda^2 s & 0 & e & 0 \\ 0 & \lambda^2 s & 0 & e \end{pmatrix} \quad (4.3.238)$$

where

$$e = \lambda^2 c - 1. \quad (4.3.239)$$

Verify that

$$\begin{aligned} V^T V &= \begin{pmatrix} e & 0 & -\lambda^2 s & 0 \\ 0 & e & 0 & \lambda^2 s \\ \lambda^2 s & 0 & e & 0 \\ 0 & -\lambda^2 s & 0 & e \end{pmatrix} \begin{pmatrix} e & 0 & \lambda^2 s & 0 \\ 0 & e & 0 & -\lambda^2 s \\ -\lambda^2 s & 0 & e & 0 \\ 0 & \lambda^2 s & 0 & e \end{pmatrix} \\ &= \begin{pmatrix} e^2 + \lambda^4 s^2 & 0 & 0 & 0 \\ 0 & e^2 + \lambda^4 s^2 & 0 & 0 \\ 0 & 0 & e^2 + \lambda^4 s^2 & 0 \\ 0 & 0 & 0 & e^2 + \lambda^4 s^2 \end{pmatrix}. \end{aligned} \quad (4.3.240)$$

Evidently $V^T V$ is diagonal and has the repeated eigenvalue $(e^2 + \lambda^4 s^2)$. Therefore, the matrix $[N(\lambda M) - I]$ has the spectral norm

$$\begin{aligned} \|N(\lambda M) - I\| &= (e^2 + \lambda^4 s^2)^{1/2} = [(\lambda^2 c - 1)^2 + \lambda^4 s^2]^{1/2} \\ &= (1 - 2\lambda^2 c + \lambda^4 c^2 + \lambda^4 s^2)^{1/2} = [1 + \lambda^4 - 2\lambda^2 c]^{1/2}. \end{aligned} \quad (4.3.241)$$

We see that when $c > 0$ there is a $\lambda > 0$ such that $\|N(\lambda M) - I\| < 1$. That is, the ray $\lambda^2 N(M)$ passes through the unit ball about I . However, this is not true when $c \leq 0$. That is,

$$\|N(\lambda M) - I\| > 1 \text{ when } \lambda > 0 \text{ and } c \leq 0. \quad (4.3.242)$$

We conclude from (3.206) that when $\theta \in (-\pi/2, \pi/2)$ there is a $\lambda > 0$ such that there is the inequality $\|N(\lambda M) - I\| < 1$. That is, the ray $\lambda^2 N(M)$ passes through the unit ball about I . However, this is not true for $\theta \notin (-\pi/2, \pi/2)$. That is,

$$\|N(\lambda M) - I\| > 1 \text{ when } \lambda > 0 \text{ and } \theta \notin (-\pi/2, \pi/2). \quad (4.3.243)$$

But we know that symplectic polar decomposition is possible for $M = O(\theta)$ for all θ . We have discovered examples where symplectic polar decomposition is possible but the ray $\lambda^2 N(M)$ does not pass through the unit ball about I .

Finally, to study uniqueness, let us compute the spectral norm of $[Q(\theta) - I]$. Similar to what was done in the case of $N(\lambda M)$, now write

$$T = Q - I = \exp(\theta A^a) - I = \begin{pmatrix} e & 0 & s' & 0 \\ 0 & e & 0 & -s' \\ -s' & 0 & e & 0 \\ 0 & s' & 0 & e \end{pmatrix} \quad (4.3.244)$$

where now

$$e = c' - 1. \quad (4.3.245)$$

Here we have used (3.196) and (3.192). Verify that in this case

$$\begin{aligned} V^T V &= \begin{pmatrix} e & 0 & -s' & 0 \\ 0 & e & 0 & s' \\ s' & 0 & e & 0 \\ 0 & -s' & 0 & e \end{pmatrix} \begin{pmatrix} e & 0 & s' & 0 \\ 0 & e & 0 & -s' \\ -s' & 0 & e & 0 \\ 0 & s' & 0 & e \end{pmatrix} \\ &= \begin{pmatrix} e^2 + (s')^2 & 0 & 0 & 0 \\ 0 & e^2 + (s')^2 & 0 & 0 \\ 0 & 0 & e^2 + (s')^2 & 0 \\ 0 & 0 & 0 & e^2 + (s')^2 \end{pmatrix}. \end{aligned} \quad (4.3.246)$$

Evidently $V^T V$ is diagonal and has the repeated eigenvalue $[e^2 + (s')^2]$. Therefore, $(Q - I)$ has the spectral norm

$$\begin{aligned} \|Q - I\| &= [e^2 + (s')^2]^{1/2} = [(c' - 1)^2 + (s')^2]^{1/2} \\ &= [1 - 2c' + (c')^2 + (s')^2]^{1/2} = (2 - 2c')^{1/2}. \end{aligned} \quad (4.3.247)$$

We see that Q lies outside the unit ball about I when $c' < 1/2$. This occurs when $|\theta/2| > 60^\circ$ and therefore $|\theta| > 120^\circ$. So, for $|\theta| < 120^\circ$, there is a symplectic polar decomposition that is unique. But the ray $\lambda^2 N(M)$ lies outside the unit circle about I when $\theta \in (90^\circ, 120^\circ)$ or $\theta \in (-120^\circ, -90^\circ)$. Thus we have found situations where symplectic polar decomposition is possible and unique, but for which the ray $\lambda^2 N(M)$ lies outside the unit circle about I . Finally, we note that $Q(\pm 2\pi) = -I$.

4.3.22. Section 4.3.6 showed that 4×4 diagonal matrices do not have symplectic polar decompositions when $\mu_1\mu_2 < 0$ and $\nu_1\nu_2 < 0$ (and $\mu_1\mu_2 \neq \nu_1\nu_2$). Exercise 3.16 showed that they do when when $\mu_1\mu_2 > 0$ and $\nu_1\nu_2 > 0$. Note that in both cases $\det(M) > 0$. One might wonder if these two cases are joined by a continuous path in $GL(4, \mathbb{R})$. You are to show that they are. Thus there must be some point along the path where symplectic polar decomposition becomes impossible.

Let D be the diagonal matrix given by (3.63), and assume $\mu_1\mu_2 > 0$ and $\nu_1\nu_2 > 0$ so that D has a symplectic polar decomposition. Show that any matrix M sufficiently close to D must also have a symplectic polar decomposition. Define a continuous family of matrices $M(\theta)$ by the rule

$$M(\theta) = O(\theta)D \quad (4.3.248)$$

where $O(\theta)$ is the matrix given by (3.185). Verify that

$$M(\theta) \in GL(4, \mathbb{R}) \text{ for all } \theta. \quad (4.3.249)$$

Show that

$$M(0) = D = \begin{pmatrix} \mu_1 & 0 & 0 & 0 \\ 0 & \mu_2 & 0 & 0 \\ 0 & 0 & \nu_1 & 0 \\ 0 & 0 & 0 & \nu_2 \end{pmatrix} \quad (4.3.250)$$

and

$$M(\pi) = \begin{pmatrix} -\mu_1 & 0 & 0 & 0 \\ 0 & \mu_2 & 0 & 0 \\ 0 & 0 & -\nu_1 & 0 \\ 0 & 0 & 0 & \nu_2 \end{pmatrix}. \quad (4.3.251)$$

Thus, the continuous path (3.213) joins the two cases. It would be interesting to know where on the path $M(\theta)$ ceases to have a symplectic polar decomposition.

4.3.23. Let H be the subgroup consisting of all elements $g \in GL(2n, R, +)$ such that

$$\{g, J\} = 0. \quad (4.3.252)$$

According Exercise 3.9.18, H is isomorphic to $GL(n, C)$. We know that symplectic polar decomposition is possible for all elements $g \in H$ that are sufficiently near the identity. Is symplectic polar decomposition possible for all elements $g \in H$?

Show that (3.217) implies the commutation relation

$$\{g^T, J\} = 0. \quad (4.3.253)$$

That is, if g commutes with J , so does g^T , and vice versa. Use this result to show that

$$N(g) = gJg^TJ^T = gg^TJJ^T = gg^T. \quad (4.3.254)$$

Does N have a J -symmetric square root? Use the correspondence relation (3.9.19) to obtain the representation

$$g = M(m). \quad (4.3.255)$$

Show, using (3.9.25) and (3.9.21), that there is the result

$$N(g) = gg^T = M(m)M^T(m) = M(m)M(m^\dagger) = M(mm^\dagger). \quad (4.3.256)$$

Verify that mm^\dagger is Hermitian and positive definite. Verify that there is a unitary matrix u such that

$$mm^\dagger = udu^\dagger \quad (4.3.257)$$

where d is diagonal with real positive entries. Define $d^{1/2}$ to be the diagonal matrix whose entries are the positive square roots of the corresponding entries in d . Use this definition to write the relation

$$mm^\dagger = udu^\dagger = ud^{1/2}d^{1/2}u^\dagger = ud^{1/2}u^\dagger ud^{1/2}u^\dagger = (ud^{1/2}u^\dagger)^2. \quad (4.3.258)$$

Show that

$$N(g) = Q^2 \quad (4.3.259)$$

where

$$Q = M(ud^{1/2}u^\dagger). \quad (4.3.260)$$

Is Q J -symmetric? Verify that

$$Q^T = M[(ud^{1/2}u^\dagger)^\dagger] = M(ud^{1/2}u^\dagger) = Q. \quad (4.3.261)$$

Also, verify that

$$\begin{aligned} JQJ^{-1} &= M(iI_n)M(ud^{1/2}u^\dagger)M(-iI_n) \\ &= M[(iI_n)(ud^{1/2}u^\dagger)(-iI_n)] = M(ud^{1/2}u^\dagger) = Q. \end{aligned} \quad (4.3.262)$$

You have shown that Q is J -symmetric, and therefore all elements $g \in H$ have a symplectic polar decomposition.

Consider the ray $\lambda^2 N(g)$. Does it intersect the unit ball around I ? Show that

$$\begin{aligned} \lambda^2 N(g) - I &= \lambda^2 M(udu^\dagger) - I = M(u)[\lambda^2 M(d) - I]M(u^\dagger) \\ &= M(u)WD(\lambda)W^{-1}M(u^\dagger) \end{aligned} \quad (4.3.263)$$

where

$$D(\lambda) = \begin{pmatrix} \lambda^2 d - I_n & 0 \\ 0 & \lambda^2 d - I_n \end{pmatrix}. \quad (4.3.264)$$

Use the properties of a matrix norm to show that

$$\|\lambda^2 N(g) - I\| \leq \|M(u)\| \|W\| \|D\| \|W^{-1}\| \|M(u^\dagger)\|. \quad (4.3.265)$$

Verify that when the spectral norm is used, there are the relations

$$\|M(u)\| = \|W\| = \|W^{-1}\| = \|M(u^\dagger)\| = 1. \quad (4.3.266)$$

Consequently, verify that for this norm

$$\|\lambda^2 N(g) - I\| \leq \|D\|. \quad (4.3.267)$$

We are ready for the final step. Since d is diagonal with all entries positive, show that there is a $\lambda_0 > 0$ such that

$$\|D(\lambda_0)\| < 1. \quad (4.3.268)$$

Conclude that

$$\|\lambda_0^2 N(g) - I\| < 1 \quad (4.3.269)$$

so that the ray $\lambda^2 N(g)$ does indeed intersect the unit ball around I .

4.3.24. Consider matrices M of the form (3.3.10) where C is an arbitrary $n \times n$ real matrix. Show that in this case

$$N(M) = \begin{pmatrix} I & 0 \\ (C - C^T) & I \end{pmatrix} \quad (4.3.270)$$

and therefore

$$[N(M)]^{1/2} = \begin{pmatrix} I & 0 \\ (C - C^T)/2 & I \end{pmatrix}. \quad (4.3.271)$$

Show that such matrices M have a symplectic polar decomposition of the form (3.10) with

$$Q = \begin{pmatrix} I & 0 \\ (C - C^T)/2 & I \end{pmatrix} \quad (4.3.272)$$

and

$$R = \begin{pmatrix} I & 0 \\ (C + C^T)/2 & I \end{pmatrix}. \quad (4.3.273)$$

Carry out an analogous demonstration for matrices of the form (3.3.9).

4.4 Finding the Closest Symplectic Matrix

4.4.1 Background

Let M be any $2n \times 2n$ matrix, and let N be the matrix associated with M by the rule (3.30). Since $N = I$ when M is symplectic, we may define a measure f of the *failure* of M to be symplectic by the rule

$$f = f(M) = \|N(M) - I\|. \quad (4.4.1)$$

Suppose f is small. Then M is nearly symplectic, and we might hope to find a matrix R that is both near M and exactly symplectic. One way to enforce this nearness condition would be to require the relation

$$\|M - R\| \sim f. \quad (4.4.2)$$

However, there is also another possibility. If R is close to M , then MR^{-1} is close to the identity I . Consequently, we could equally well require the relation

$$\|MR^{-1} - I\| \sim f. \quad (4.4.3)$$

Both (4.2) and (4.3) state the hope that if M fails to be symplectic by an amount f , then there should be a symplectic matrix R that is, so to speak, roughly within a distance f from M .

Given (4.1) with $f < 1$, we will show that there is a symplectic R that satisfies both (4.2) and (4.3). Such a matrix R is entitled to be called a *symplectification* of M . Our proof will be based on the results of the previous section. Recall the function $g(x)$ defined by (3.103). Similar to what was done before, use it to define Q by the rule

$$Q = (N)^{1/2} = [I - (I - N)]^{1/2} = I - \sum_{\ell=1}^{\infty} d_{\ell}(I - N)^{\ell}. \quad (4.4.4)$$

This series will converge if $f < 1$, and in that case we have the inequality

$$\| Q - I \| = \left\| \sum_{\ell=1}^{\infty} d_{\ell}(I - N)^{\ell} \right\| \leq \sum_{\ell=1}^{\infty} d_{\ell} \| I - N \|^{\ell} \leq \sum_{\ell=1}^{\infty} d_{\ell} f^{\ell} \leq f. \quad (4.4.5)$$

We may also define Q^{-1} by the series

$$Q^{-1} = [I - (I - Q)]^{-1} = \sum_{\ell=0}^{\infty} (I - Q)^{\ell}. \quad (4.4.6)$$

According to (4.5) this series also converges if $f < 1$. Since Q has been defined and is invertible, we may use (3.36) to define a symplectic matrix R and thereby achieve the symplectic polar decomposition (3.10).

Let us use this R and this decomposition to test the relations (4.2) and (4.3). For (4.2) we find the result

$$\| M - R \| = \| QR - R \| = \| (Q - I)R \| \leq \| Q - I \| \| R \| \leq \| R \| f. \quad (4.4.7)$$

Testing the relation (4.3) gives the result

$$\| MR^{-1} - I \| = \| QRR^{-1} - I \| = \| Q - I \| \leq f. \quad (4.4.8)$$

We have learned that if M is such that its failure f to be symplectic satisfies $f < 1$, then it has a symplectic polar decomposition and the factor R in this decomposition provides a symplectification that satisfies the nearness relations (4.7) and (4.8).

Suppose R' is a symplectic matrix that is sufficiently near R in the sense that

$$\| R' - R \| \leq f. \quad (4.4.9)$$

(Because symplectic matrices form a Lie group there are many such R' .) Then we find the result

$$\begin{aligned} \| M - R' \| &= \| (M - R) + (R - R') \| \\ &\leq \| M - R \| + \| R - R' \| \leq \| R \| f + f = (\| R \| + 1)f, \end{aligned} \quad (4.4.10)$$

and conclude that R' satisfies the nearness requirement (4.2) and hence is also an acceptable symplectification of M . Alternatively, suppose R' is a symplectic matrix that is sufficiently near R in the sense that

$$\| R(R')^{-1} - I \| \leq f. \quad (4.4.11)$$

Then we find the result

$$\begin{aligned}\| M(R')^{-1} - I \| &= \| MR^{-1}R(R')^{-1} - I \| = \| QR(R')^{-1} - I \| \\ &= \| (Q - I) + Q[R(R')^{-1} - I] \| \leq \| Q - I \| + \| Q[R(R')^{-1} - I] \| \\ &\leq f + \| Q \| \| R(R')^{-1} - I \| \leq f + \| Q \| f = (\| Q \| + 1)f,\end{aligned}\quad (4.4.12)$$

and conclude that R' satisfies the nearness requirement (4.3) and hence is also an acceptable symplectification of M . We have learned that a matrix M , whose failure f to be symplectic satisfies $f < 1$ in some norm, has many acceptable symplectifications R' that satisfy (4.2) or (4.3) and, in particular, (4.10) or (4.12). Sections 4.5 through 4.8 describe four methods for finding such symplectifications.

Since there are many symplectifications R' that meet our requirements, we may wonder which one is actually *closest* to M . The discussion so far has made only rather general assumptions about the matrix norm $\| * \|$ employed to determine nearness. It has served only as a tool to establish the convergence of various series; but, of course, the quantities defined by these series, if they are defined at all, are independent of the choice of norm. Now, however, we have a more specific question than those discussed above: Imagine we are given a matrix M and we consider all symplectic matrices R' . Is there a closest symplectic matrix R_c that *minimizes* the quantity $\| M - R' \|$? Alternatively, is there a closest symplectic matrix R_c that minimizes the quantity $\| M(R')^{-1} - I \|$? The answers to these questions do depend on the choice of matrix norm.

4.4.2 Use of Euclidean Norm

Let us explore the question of determining the closest symplectic matrix using the nearness condition (4.2) and the Euclidean matrix norm. Consider the set of all (real) $2n \times 2n$ matrices. It obviously forms a linear vector space under the operations of scalar multiplication and matrix addition. Let A and B be any two vectors (matrices) in this space. We define an inner product between them by the rule

$$(A, B) = \text{tr}(A^T B). \quad (4.4.13)$$

It is easily verified that this rule satisfies all the requirements for a positive-definite inner product. (See Exercise 4.3.) Let O' and O'' denote any $2n \times 2n$ orthogonal matrices. Then it can also be shown that the inner product (4.13) is invariant under the action of the orthogonal group in the sense that

$$(O'AO'', O'BO'') = (A, B). \quad (4.4.14)$$

Next, as in Exercise 3.7.1, we define the *Euclidean* norm $\| M \|_E$ of any matrix M by the rule

$$\| M \|_E = (M, M)^{1/2}. \quad (4.4.15)$$

This rule satisfies all the conditions (3.7.10) through (3.7.14) required for a norm. The Euclidean norm is not particularly powerful for establishing convergence in some circumstances because it gives for the $2n \times 2n$ identity matrix the result

$$\| I \|_E = (2n)^{1/2}. \quad (4.4.16)$$

By contrast there are more powerful norms, the maximum column sum norm (3.7.15) and spectral norm (3.7.17) for examples, that give the optimal result

$$\| I \| = 1. \quad (4.4.17)$$

However, as a consequence of (4.14), the Euclidean norm does have the convenient feature that

$$\| O' M O'' \|_E = \| M \|_E \quad (4.4.18)$$

where O' and O'' are any orthogonal matrices.

As stated earlier, it can be shown that if M is any matrix, then the *orthogonal* matrix O that is closest to M , in the sense of minimizing $\| M - O \|_E$, is given by the orthogonal matrix appearing in the polar decomposition (2.7). The situation with regard to the closest symplectic matrix is more complicated. One might entertain the analogous conjecture that the *symplectic* matrix R that is closest to any symplectifiable M , in the sense of minimizing $\| M - R \|_E$, is given by the symplectic matrix appearing in the symplectic polar decomposition (3.10). However, this conjecture is wrong.

As a counter example in the 2×2 case, consider the matrix M given by the relation

$$M = \mu K. \quad (4.4.19)$$

Here K is a symplectic diagonal matrix of the form

$$K = \begin{pmatrix} k & 0 \\ 0 & k^{-1} \end{pmatrix}, \quad (4.4.20)$$

and we assume that $\det(M) > 0$ so that μ has the value

$$\mu = +[\det(M)]^{1/2}. \quad (4.4.21)$$

For this M we find from (3.61) the result

$$R = K. \quad (4.4.22)$$

Next, let X be the symplectic diagonal matrix

$$X(x) = \begin{pmatrix} x & 0 \\ 0 & x^{-1} \end{pmatrix}. \quad (4.4.23)$$

Let us examine whether there is a choice of x such that $\| M - X \|_E$ has a value smaller than $\| M - K \|_E$. To make such a study, consider the function $h(x)$ defined by the relation

$$h(x) = [\| M - X \|_E]^2. \quad (4.4.24)$$

Does h have a minimum at $x = k$? From the definitions (4.13) and (4.15) we find that

$$h(x) = \text{tr} [(M - X)^T (M - X)] = (\mu k - x)^2 + (\mu k^{-1} - x^{-1})^2. \quad (4.4.25)$$

Suppose we differentiate h with respect to x and evaluate the result at $x = k$. Doing so gives the result

$$h'(k) = 2(\mu - 1)(k^{-3} - k). \quad (4.4.26)$$

We see that in general $h'(k) \neq 0$. It follows that, for this example, setting $X = K = R$ does not minimize $\|M - X\|_E$.

One can also construct counter examples for which the symplectic matrix R produced by the symplectic polar decomposition of M does not give the symplectic matrix closest to M in the sense of minimizing $\|M(R')^{-1} - I\|_E$. We also remark that if the transpose operation in (4.13) is omitted, which produces a different inner product about to be discussed, these conclusions remain unchanged. See Exercise 4.5.

4.4.3 Geometric Interpretation of Symplectic Polar Decomposition

Although the symplectic matrix R produced by the symplectic polar decomposition of M does not necessarily give the symplectic matrix closest to M as defined by either the nearness condition (4.2) or (4.3) and the use of some inner product norm, one might still wonder if it has some other *geometric* interpretation. It does, but some further concepts need to be developed to show that this is the case. We note that the nearness condition (4.2) is related to the matrix operation of *addition* (actually, in this case, subtraction) while the nearness condition (4.3) is related to the operation of matrix *multiplication*. We will explore the use of nearness conditions related to *group* properties. Since group properties are based on the operation of matrix multiplication, these nearness conditions are similar in spirit to the condition (4.3).

Consider the group $GL(2n, \mathbb{R})$. Near the identity any matrix M can be written in the form

$$M = \exp(B) \quad (4.4.27)$$

where B , an arbitrary matrix of $gl(2n, \mathbb{R})$, has the decomposition (3.1). Let B_0, B_1, \dots be a set of basis vectors (matrices) for this space. There are $(2n)^2$ such matrices. For example, for the simplest case $n = 1$, a convenient basis is given by the choice

$$B_0 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad (4.4.28)$$

$$B_1 = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \quad (4.4.29)$$

$$B_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (4.4.30)$$

$$B_3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (4.4.31)$$

We note that the first 3 basis matrices, B_0 through B_2 , are of the form JS [see (5.6.7), (5.6.13), and (5.6.14)], and the last is of the form JA with $A = -J$:

$$B_3 = J(-J) = I. \quad (4.4.32)$$

In terms of the basis provided by the B_ℓ , any matrix in $gl(2n, \mathbb{R})$ can be written in the form

$$B(b) = \sum_{\ell} b^\ell B_\ell, \quad (4.4.33)$$

where the b^ℓ are real, but otherwise arbitrary, coefficients. We may view the b^ℓ as the entries of a vector b in a $(2n)^2$ dimensional vector space. Correspondingly, we may view (4.27) and (4.33) as a mapping between points in this vector space and elements in the group $GL(2n, \mathbb{R})$ near the identity. Put another way, this mapping and the entries in b constitute a *coordinate patch* for $GL(2n, \mathbb{R})$ at the identity. Alternatively, we may regard the b^ℓ as the *components* of the vectors in the *tangent space* of $GL(2n, \mathbb{R})$ at the identity. Consequently, the vectors b are in and span the *cotangent* space of $GL(2n, \mathbb{R})$ at the identity.

Next, suppose b and b' are any two vectors. Let us introduce an inner product between them by the rule

$$(b, b') = \text{tr}[B(b)B(b')] = \text{tr}\left[\sum_{\ell\ell'} b^\ell (b')^{\ell'} B_\ell B_{\ell'}\right] = \sum_{\ell\ell'} b^\ell (b')^{\ell'} \text{tr}(B_\ell B_{\ell'}). \quad (4.4.34)$$

This inner product can be expressed in the form

$$(b, b') = \sum_{\ell\ell'} b^\ell (b')^{\ell'} g_{\ell\ell'} \quad (4.4.35)$$

where g is the *metric tensor*

$$g_{\ell\ell'} = \text{tr}(B_\ell B_{\ell'}). \quad (4.4.36)$$

In general, this metric tensor is *not* positive definite. For example, use of the basis given by (4.28) through (4.31) in the case $n = 1$ gives the result

$$g = \begin{pmatrix} -2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}. \quad (4.4.37)$$

Although the metric g is not positive definite, it does have two attractive features. The first feature is this: Suppose b is a vector such that $B(b)$ is a matrix of the form JS , and b' is a vector such that $B(b')$ is a matrix of the form JA . Then b and b' are orthogonal,

$$(b, b') = \text{tr}[B(b)B(b')] = \text{tr}(JSJA) = 0. \quad (4.4.38)$$

To verify this assertion, we make the observation

$$\text{tr}(JSJA) = \text{tr}[(JSJA)^T] = \text{tr}(A^T J^T S^T J^T) = \text{tr}(-AJSJ) = -\text{tr}(JSJA). \quad (4.4.39)$$

A description of the second attractive feature requires some background discussion. Suppose M^1 is some matrix, not necessarily near the identity, and we wish to examine matrices near M^1 . To do so, we might consider all matrices ${}^L M^1$ of the form

$${}^L M^1(b) = M^1 \exp[B(b)] \quad (4.4.40)$$

where $B(b)$ is again given by (4.33). The relation (4.40) provides a coordinate patch for $GL(2n, \mathbb{R})$ at the point M^1 . Indeed, looking at (4.40), we may say that this coordinate patch is obtained by *translating* the coordinate patch at the identity to the point M^1 . This translation may be called *left* translation since (4.40) involves multiplication by M^1 on the

left. [Hence the notation ${}^L M^1$ in (4.40).] Alternatively, we may view the entries in b as the components of the vectors in the tangent space of $GL(2n, \mathbb{R})$ at the point M^1 , and vectors in this tangent space are to be regarded as associated with those at the identity by the operation of left translation.

There is an obvious alternative to (4.40). Namely, we might equally well use *right* translation to examine matrices near M^1 by considering matrices ${}^R M^1$ of the form

$${}^R M^1(c) = \{\exp[B(c)]\}M^1. \quad (4.4.41)$$

Here the entries in c may again be viewed as the components of vectors in the tangent space of $GL(2n, \mathbb{R})$ at the point M^1 , but vectors in this tangent space are now to be regarded as associated with those at the identity by the operation of right translation.

Suppose both (4.40) and (4.41) are used to represent the same element,

$${}^L M^1(b) = {}^R M^1(c). \quad (4.4.42)$$

Then, using (4.40) and (4.41), we find the equation

$$M^1 \exp[B(b)] = \{\exp[B(c)]\}M^1, \quad (4.4.43)$$

which is essentially a relation between b and c . Indeed, this relation may be rewritten in the form

$$\exp[B(c)] = M^1 \{\exp[B(b)]\}(M^1)^{-1} = \exp\{M^1[B(b)](M^1)^{-1}\}, \quad (4.4.44)$$

from which we conclude that

$$B(c) = M^1[B(b)](M^1)^{-1}. \quad (4.4.45)$$

Also, since the matrices B_ℓ form a basis, we must have relations of the form

$$M^1 B_\ell (M^1)^{-1} = \sum_{\ell'} d_{\ell\ell'}(M^1) B_{\ell'} \quad (4.4.46)$$

where the $d_{\ell\ell'}(M^1)$ are coefficients that depend on M^1 . Correspondingly, we find the result

$$M^1[B(b)](M^1)^{-1} = \sum_{\ell} b^\ell M^1 B_\ell (M^1)^{-1} = \sum_{\ell\ell'} b^\ell d_{\ell\ell'} B_{\ell'} = \sum_{\ell'} (\sum_{\ell} b^\ell d_{\ell\ell'}) B_{\ell'}, \quad (4.4.47)$$

from which we conclude that c is given in terms of b by the equation

$$c^\ell' = \sum_{\ell} b^\ell d_{\ell\ell'}. \quad (4.4.48)$$

Now a question arises: We have seen how (4.34) can be used to define an inner product between two vectors b and b' whose entries are the components for the two vectors in the tangent space at the identity. What can be done for pairs of tangent vectors at the general point M^1 ? These tangent-vector pairs can be associated with either of the coordinate pairs b, b' or c, c' depending on whether left or right translation is used. We will define their inner products to be the same as those for their counterparts at the origin,

$$(b, b')^L = \text{tr}[B(b)B(b')], \quad (4.4.49)$$

$$(c, c')^R = \text{tr}[B(c)B(c')], \quad (4.4.50)$$

where we have used the superscripts L and R to indicate that either left or right translation has been used to make a correspondence between the tangent space at the identity and the tangent space at the general point M^1 .

However, we know that b and c are related by (4.45), and the same is true for b' and c' . Consequently, from (4.45), (4.49), and (4.50), we find the relation

$$\begin{aligned} (c, c')^R &= \text{tr}[B(c)B(c')] = \text{tr}\{M^1 B(b)(M^1)^{-1} M^1 B(b')(M^1)^{-1}\} \\ &= \text{tr}\{M^1 B(b)B(b')(M^1)^{-1}\} = \text{tr}[B(b)B(b')] = (b, b')^L. \end{aligned} \quad (4.4.51)$$

We have learned that the inner product definition (4.34) has the feature that its extension from the identity to an arbitrary point M^1 is *independent* of whether left or right translations are used.

Now that an inner product has been defined on the tangent (and cotangent) spaces at any point in $GL(2n, \mathbb{R})$, we can discuss the *lengths* of paths in matrix space. Let M^0 and M^1 be any two matrices. Suppose they are joined by a path $M(\tau)$, where τ is a parameter lying in the range $[0,1]$,

$$M(0) = M^0, \quad (4.4.52)$$

$$M(1) = M^1. \quad (4.4.53)$$

Consider the two nearby points $M(\tau)$ and $M(\tau+d\tau)$ on the path. Suppose we view $M(\tau+d\tau)$ as being related to $M(\tau)$ by right translation of elements near the identity. That is, we write a relation of the form

$$M(\tau + d\tau) = \exp[d\tau C(\tau)]M(\tau) = \{I + d\tau C(\tau) + O[(d\tau)^2]\}M(\tau). \quad (4.4.54)$$

At this juncture we note that (4.54) can be rewritten in the form

$$M(\tau + d\tau)M^{-1}(\tau) - I = d\tau C(\tau) + O[(d\tau)^2], \quad (4.4.55)$$

and we see that the left side of (4.55) is reminiscent of the nearness condition (4.3). Now make the Taylor expansion

$$M(\tau + d\tau) = M(\tau) + (dM/d\tau)d\tau + O[(d\tau)^2] = M(\tau) + \dot{M}(\tau)d\tau + O[(d\tau)^2]. \quad (4.4.56)$$

Upon comparing (4.54) and (4.56), we may solve for $C(\tau)$, which we may view as the tangent vector to the path at the point $M(\tau)$, to find the result

$$C(\tau) = \dot{M}(\tau)M^{-1}(\tau). \quad (4.4.57)$$

Let us define an *energy functional* $E[M]$ associated with any path $M(\tau)$ by the relation

$$E[M] = (1/2) \int_0^1 d\tau \text{tr}[C(\tau)C(\tau)]. \quad (4.4.58)$$

[Note that this definition employs the inner product (4.34).] Then an *affine geodesic* in matrix space is defined to be a path ${}^{ag}M(\tau)$ that extremizes the energy functional. For a discussion of geodesics and affine geodesics, see Exercise 1.6.17.

Why should these definitions interest us? Suppose M^0 and M^1 are close in the sense that $[M^1(M^0)^{-1}]$ is near the identity. Then there exists a matrix B such that

$$M^1(M^0)^{-1} = \exp(B) \quad (4.4.59)$$

or

$$M^1 = \exp(B)M^0. \quad (4.4.60)$$

Consider the particular path $M(\tau)$ given by the rule

$$M(\tau) = \exp(\tau B)M^0. \quad (4.4.61)$$

Evidently this path satisfies (4.52) and (4.53), and therefore joins M^0 and M^1 . Moreover, this path satisfies the differential equation

$$\dot{M}(\tau) = B[\exp(\tau B)]M^0 = BM(\tau), \quad (4.4.62)$$

and consequently by (4.57) has the *constant* tangent vector

$$C(\tau) = B. \quad (4.4.63)$$

Thus, in this sense, the path (4.61) in matrix space is the analog of a *straight line* in Euclidean space, which also has a constant tangent. But even more can be said about this analogy. The path (4.61) is also an affine geodesic! See Exercise 4.6.

Let us now apply these general considerations to the problem at hand. Suppose that M is a matrix that meets the condition of Theorem 3.1. Define a path $M(\tau)$ in matrix space by the rule

$$M(\tau) = \exp(\tau JA)R, \quad (4.4.64)$$

where R is the symplectic matrix in the factorization (3.10) and A is defined by (3.57). Evidently this path joins R , the symplectic factor of M , to M itself,

$$M(0) = R, \quad (4.4.65)$$

$$M(1) = M. \quad (4.4.66)$$

See Figure 4.1. Comparison of (4.61), (4.63), and (4.64) shows that this path has the constant tangent vector JA . Consider the point R at which the path $M(\tau)$ meets the group of symplectic matrices. We know that any vector in the tangent space of the group of symplectic matrices is of the form JS . We see from (4.39) that the path $M(\tau)$ is *perpendicular* to the subspace of symplectic matrices at the point R . Finally, we know that the path $M(\tau)$ is an affine geodesic. We conclude that R has the special geometric property that it is connected to M by a path that [in terms of the tangent-space metric (4.36)] is both an affine geodesic and is perpendicular to the subspace of symplectic matrices at the point R .

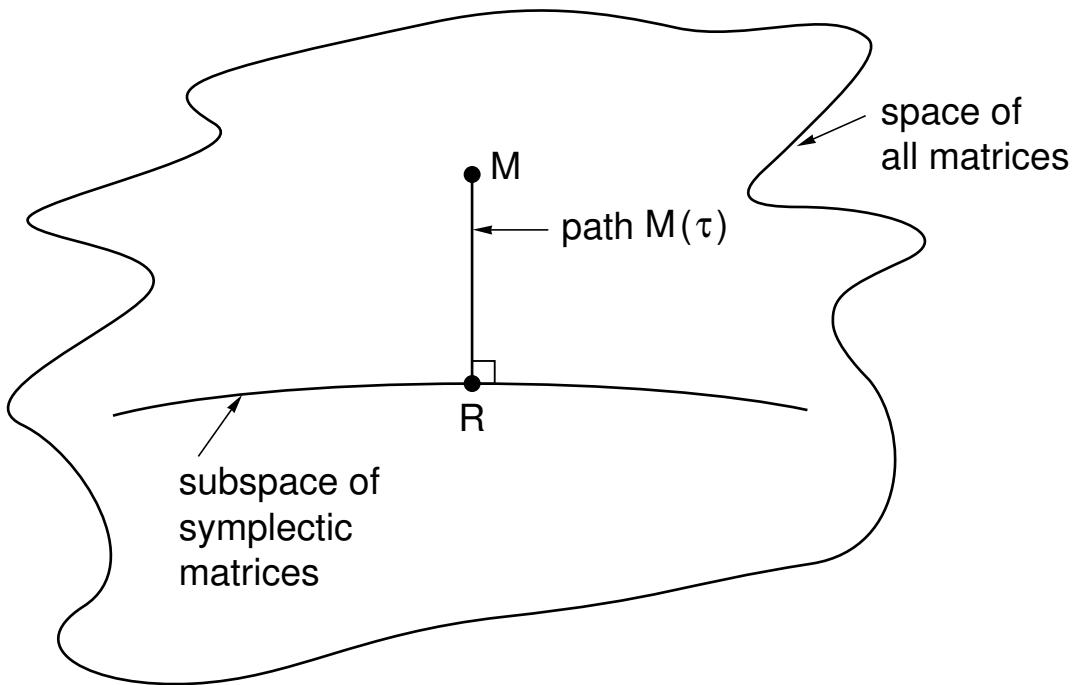


Figure 4.4.1: The matrices R and M are connected by a path that is both an affine geodesic and is perpendicular to the subspace of symplectic matrices at the point R .

Exercises

4.4.1. Verify that the rule (4.13) satisfies all the requirements for an inner product including the positive-definite conditions

$$(A, A) \geq 0, \quad (4.4.67)$$

$$(A, A) = 0 \Leftrightarrow A = 0. \quad (4.4.68)$$

4.4.2. Verify (4.14) through (4.18).

4.4.3. If you have not already worked Exercise 3.7.1, verify that (4.15) satisfies all the conditions (3.7.10) through (3.7.14) required for a matrix norm.

4.4.4. Verify (4.25) and (4.26).

4.4.5.

4.4.6. Suppose that, instead of using (4.57), which involves right translation, the tangent vector $C(\tau)$ is defined by the left-translation relation

$$C(\tau) = M^{-1}(\tau)\dot{M}(\tau). \quad (4.4.69)$$

Show that the value of the energy functional $E[M]$ given by (4.58) remains unchanged.

The remainder of this exercise is devoted to showing that $M(\tau)$ as given by (4.61) is an affine geodesic. To do so, we will need to evaluate $E[M]$ for paths near (4.61) and show

that, through first order, $E[M]$ remains unchanged when small changes are made about the path (4.61). Parameterize paths near (4.61) by writing

$$M(\epsilon, \tau) = \exp[\epsilon F(\tau)] \exp(\tau B) M^0 = \exp[\epsilon F(\tau)] M(\tau), \quad (4.4.70)$$

where $F(\tau)$ is an arbitrary matrix function save for the end-point conditions

$$F(0) = F(1) = 0. \quad (4.4.71)$$

Evaluate $E(\epsilon)$ for paths of the form (4.70) and show that

$$(dE/d\epsilon)|_{\epsilon=0} = 0. \quad (4.4.72)$$

Hints: Using (4.57) and (4.70), show that

$$C(\tau) = C_0 + \epsilon C_1 + O(\epsilon^2) \quad (4.4.73)$$

where

$$C_0 = B, \quad (4.4.74)$$

$$C_1 = \dot{F}(\tau) + \{F(\tau), B\}. \quad (4.4.75)$$

Next show that

$$E = E_0 + \epsilon E_1 + O(\epsilon^2) \quad (4.4.76)$$

where

$$E_1 = \int_0^1 d\tau \operatorname{tr}(C_0 C_1). \quad (4.4.77)$$

Finally, show that

$$E_1 = 0. \quad (4.4.78)$$

For extra credit, suppose B is of the form JA as in (4.64), $F(0)$ is of the form JS , and $F(1) = 0$. Show that (4.78) still holds in this case, and give a geometrical interpretation of this fact in terms of Figure 4.1.

4.4.7. Consider using the positive-definite inner product (4.13) instead of the indefinite inner product (4.34). See Exercise 4.1. Show that in this case matrix pairs of the form JS and JA are again orthogonal,

$$\operatorname{tr}[(JS)^T JA] = \operatorname{tr}(S^T J^T JA) = \operatorname{tr}(SA) = 0. \quad (4.4.79)$$

Use (4.13) to define an energy functional by writing

$$E[M] = (1/2) \int_0^1 d\tau \operatorname{tr}[C^T(\tau) C(\tau)]. \quad (4.4.80)$$

Following the discussion in Exercise 4.6, show that in this case

$$E_1 = \int_0^1 d\tau \operatorname{tr}(C_0^T C_1) \quad (4.4.81)$$

with C_0 and C_1 given by (4.74) and (4.75) as before. Show that in this case E_1 has the value

$$E_1 = \int_0^1 d\tau \operatorname{tr}[\{B, B^T\}F], \quad (4.4.82)$$

and that the necessary and sufficient condition for E_1 to vanish for all F satisfying (4.71) is the requirement

$$\{B, B^T\} = 0. \quad (4.4.83)$$

For the case of the path (4.64), B is the matrix given by the relation

$$B = JA. \quad (4.4.84)$$

Verify that (4.83) holds in the case of $gl(2, \mathbb{R})$, but need not be true in the cases of $gl(4, \mathbb{R})$, $gl(6, \mathbb{R})$, etc.

4.4.8. Consider orthogonal polar decompositions of the form (2.7). Suppose M is invertible. Show that there exists a real symmetric matrix S such that M can be written in the form

$$M = \exp(S)O. \quad (4.4.85)$$

See Exercise 2.3. It follows that O and M can be joined by the path

$$M(\tau) = \exp(\tau S)O, \quad (4.4.86)$$

with constant tangent vector S . We know that the orthogonal matrices form a group, and that any vector in the tangent space of this group at any point in the group is of the form A where A is an antisymmetric matrix. Show that matrix pairs of the form S and A are orthogonal for both the inner products (4.13) and (4.34). Show that $M(\tau)$ as given by (4.86) is an affine geodesic for both the energy functionals (4.58) and (4.80). Show that the length of this affine geodesic is the same independent of whether (4.57) or (4.69) is used to define the tangent vector $C(\tau)$.

4.5 Symplectification Using Symplectic Polar Decomposition

We are now prepared to discuss various symplectification processes. The first uses symplectic polar decomposition. Closely related are iterative procedures.⁴ If successful, they produce the same result as symplectic polar decomposition. We will begin with a review of the process and properties of symplectic polar decomposition, and then proceed to describe how and when iteration may be used to obtain the same results.

⁴The first iterative procedure was introduced by *Furman*. See the references at the end of this chapter.

4.5.1 Properties of Symplectification Using Symplectic Polar Decomposition

Let M denote any $2n \times 2n$ matrix. Consider the mapping \mathcal{S} of the space of such matrices into itself defined by the rule

$$\mathcal{S}(M) = (M J M^T J^T)^{-1/2} M = [N(M)]^{-1/2} M. \quad (4.5.1)$$

Here we have used (3.30). Moreover, to define $N^{-1/2}$ we will write

$$\begin{aligned} N^{-1/2} &= [I + (N - I)]^{-1/2} = I + \sum_{\ell=1}^{\infty} e_{\ell}(N - I)^{\ell} \\ &= I - (1/2)(N - I) + (3/8)(N - I)^2 - \dots \end{aligned} \quad (4.5.2)$$

and assume that in some norm (4.1) is satisfied with $f < 1$ in order to ensure convergence. Of course, this assumption places some restrictions on the domain of \mathcal{S} .

Suppose R' and R'' are any two symplectic matrices. Then we find from (3.33) the result

$$[N(R' M R'')] - I = R'[N(M) - I](R')^{-1}, \quad (4.5.3)$$

and hence

$$[N(R' M R'')] - I^{\ell} = R'[N(M) - I]^{\ell}(R')^{-1}. \quad (4.5.4)$$

Consequently, since matrix multiplication and infinite summation can be interchanged, we find from (5.2) the result

$$[N(R' M R'')]^{-1/2} = R'[N(M)]^{-1/2}(R')^{-1}. \quad (4.5.5)$$

See Exercise 5.1. It follows from the definition (5.1) that \mathcal{S} has the property

$$\begin{aligned} \mathcal{S}(R' M R'') &= R'[N(M)]^{-1/2}(R')^{-1}(R' M R'') \\ &= R'\mathcal{S}(M)R''. \end{aligned} \quad (4.5.6)$$

As a special case of (5.6) we have the result

$$\mathcal{S}(M R'') = \mathcal{S}(M)R''. \quad (4.5.7)$$

We note that this result can be proved directly without concern about the effect of interchanging the operations of matrix multiplication and infinite summation since we have as a special case of (3.33) the relation

$$N(M R'') = N(M). \quad (4.5.8)$$

Next suppose Q' is any (invertible) J -symmetric matrix. Then we find the result

$$\mathcal{S}(Q') = [Q' J(Q')^T J^T]^{-1/2} Q' = [(Q')^2]^{-1/2} Q' = I. \quad (4.5.9)$$

Finally, suppose M has the symplectic polar decomposition (3.10). Then, using (5.7) and (5.9), we find the result

$$\mathcal{S}(M) = \mathcal{S}(Q R) = \mathcal{S}(Q)R = R. \quad (4.5.10)$$

Consequently, as one might expect from (3.38) and (3.39), the map \mathcal{S} is a *symplectifying* map that sends M into the symplectic factor in its symplectic polar decomposition.

There are three properties of the symplectifying map \mathcal{S} provided by symplectic polar decomposition that are worth noting. First, we have the result

$$\mathcal{S}(R) = R \quad (4.5.11)$$

for *any* symplectic matrix R . Thus, if M is already symplectic, the map \mathcal{S} given by (5.1) leaves M in peace. The second property is that already stated in (5.6): Suppose the matrix M is given left and right symplectic translations by sending it to the matrix $(R'MR'')$, and this translated matrix is then symplectified using symplectic polar decomposition. Equation (5.6) states that the result is the same as that obtained by first symplectifying M and then giving the symplectified M the same translations. We may say that the the symplectification process provided by \mathcal{S} is *invariant* under left and right *symplectic translations*. Finally, suppose M has the symplectic polar decomposition (3.10). Then we find for M^{-1} the result

$$M^{-1} = R^{-1}Q^{-1} = (R^{-1}Q^{-1}R)R^{-1} = \dot{Q}R^{-1}. \quad (4.5.12)$$

By Lemmas 3.5 and 3.8 the matrix $\dot{Q} = (R^{-1}Q^{-1}R)$ is J -symmetric. Therefore we have the result

$$\mathcal{S}(M^{-1}) = R^{-1} = [\mathcal{S}(M)]^{-1}. \quad (4.5.13)$$

We may say that the the symplectification process provided by \mathcal{S} is also *invariant* under *inversion*.

4.5.2 Iteration

Suppose we define a map \mathcal{S}_1 , related to \mathcal{S} , by retaining only the $\ell = 1$ term in the series appearing in (5.2). This map has the definition

$$\begin{aligned} \mathcal{S}_1(M) &= [I - (1/2)(N - I)]M = (1/2)[3I - N(M)]M \\ &= (1/2)(3I - MJM^TJ^T)M. \end{aligned} \quad (4.5.14)$$

It is readily verified that \mathcal{S}_1 also satisfies a relation of the form (5.6),

$$\mathcal{S}_1(R'MR'') = R'\mathcal{S}_1(M)R'', \quad (4.5.15)$$

and now, because the series (5.2) has been truncated, there is no concern about convergence.

Now let Q' be any J -symmetric matrix. Using (3.9) and (5.14), we find that

$$\mathcal{S}_1(Q') = (3/2)Q' - (Q')^3/2. \quad (4.5.16)$$

It follows from this result and Lemmas 3.2 and 3.5 that \mathcal{S}_1 maps the space of J -symmetric matrices into itself. In addition note that $Q' = I$ is a fixed point of \mathcal{S}_1 ,

$$\mathcal{S}_1(I) = I. \quad (4.5.17)$$

Let us examine the nature of the fixed point $Q' = I$. To do so, write Q' in the form

$$Q' = I + W, \quad (4.5.18)$$

where W is “small”. By Lemmas 3.1 and 3.2, $W = Q' - I$ is also J -symmetric if Q' is J -symmetric. Upon inserting the form (5.18) into (5.16), we find the result

$$\mathcal{S}_1(Q') = S_1(I + W) = I - (3/2)W^2 - (1/2)W^3. \quad (4.5.19)$$

At this point it is convenient to introduce the map \mathcal{U}_1 defined on J -symmetric matrices by the rule

$$\mathcal{U}_1(W) = S_1(I + W) - I. \quad (4.5.20)$$

From this definition it follows, by combining (5.20) with (5.19), that

$$\mathcal{U}_1(W) = -(3/2)W^2 - (1/2)W^3. \quad (4.5.21)$$

Evidently, \mathcal{U}_1 has the fixed point $W = 0$, which corresponds precisely to the fixed point $Q' = I$ of \mathcal{S}_1 .

To exploit this correspondence, define *translation* maps \mathcal{T} and \mathcal{T}^{-1} by the rules

$$\mathcal{T}(W) = W + I, \quad (4.5.22)$$

$$\mathcal{T}^{-1}(W) = W - I. \quad (4.5.23)$$

With these definitions, we have the relations

$$\mathcal{U}_1 = \mathcal{T}^{-1}\mathcal{S}_1\mathcal{T}, \quad (4.5.24)$$

$$\mathcal{S}_1 = \mathcal{T}\mathcal{U}_1\mathcal{T}^{-1}. \quad (4.5.25)$$

From (5.25) we see that

$$\mathcal{S}_1^m = \mathcal{T}\mathcal{U}_1^m\mathcal{T}^{-1}, \quad (4.5.26)$$

and conclude that the behavior of \mathcal{S}_1^m on J -symmetric matrices of the form $Q' = I + W$ is governed by the behavior of \mathcal{U}_1^m on the matrices W . Moreover, since the right side of (5.21) is quadratic in W , we expect that $W = 0$ will be an *attractor* of \mathcal{U}_1 , and correspondingly $Q' = I$ will be an *attractor* of \mathcal{S}_1 .

An estimate of the basin of attraction of \mathcal{U}_1 can be obtained by requiring that

$$\| \mathcal{U}_1(W) \| = \| -(3/2)W^2 - (1/2)W^3 \| < \| W \| . \quad (4.5.27)$$

This condition is difficult to work with, and we will use instead a poorer estimate. Suppose we require that

$$[(3/2) \| W \|^2 + (1/2) \| W \|^3] < \| W \| . \quad (4.5.28)$$

By the properties (3.7.11) through (3.7.13) of a norm we will then have the result

$$\| -(3/2)W^2 - (1/2)W^3 \| < \| W \| . \quad (4.5.29)$$

Consequently, W that satisfy (5.28) will lie in the basin of $W = 0$. It is easily verified that (5.28) is equivalent to the condition

$$\| W \| < (-3/2 + (1/2)\sqrt{17}) \simeq (.56). \quad (4.5.30)$$

We conclude that if $\|W\| < .56$, then we have the result

$$\lim_{m \rightarrow \infty} \mathcal{U}_1^m(W) = 0. \quad (4.5.31)$$

That is, repeated application (iteration) of \mathcal{U}_1 will drive such W to 0. Moreover, in view of (5.21), once convergence gets underway it will be quadratic and therefore very rapid.

We next show that W as given by (5.18) satisfies the inequality

$$\|W\| \leq f, \quad (4.5.32)$$

where f is given by (4.1) and N is defined in terms of Q' . When $M = Q'$, we find from (3.27) the result

$$N(Q') = Q'J(Q')^TJ^T = (Q')^2. \quad (4.5.33)$$

Consequently, following (4.4) and (4.5), we may write the relations

$$Q' = (N)^{1/2} = [I - (I - N)]^{1/2} = I - \sum_{\ell=1}^{\infty} d_{\ell}(I - N)^{\ell}, \quad (4.5.34)$$

$$\|W\| = \|Q' - I\| \leq f. \quad (4.5.35)$$

We conclude that if $f < (.56)$, then we again have the result (5.31). Correspondingly, we also have the result

$$\lim_{m \rightarrow \infty} \mathcal{S}_1^m(Q') = I. \quad (4.5.36)$$

We are now ready for the master stroke. Suppose M is some matrix whose failure f to be symplectic satisfies $f < (.56)$. Then since $f < 1$, we know that such a matrix has the symplectic polar decomposition (3.10), and that [according to (4.5)] the J -symmetric factor Q of M must satisfy the relation

$$\|Q - I\| \leq (.56). \quad (4.5.37)$$

Let us compute the matrices $\mathcal{S}_1^m(M)$ for successive values of m . From (3.10) and (5.15) we find the result

$$\mathcal{S}_1^m(M) = \mathcal{S}_1^m(QR) = \mathcal{S}_1^m(Q)R. \quad (4.5.38)$$

Now take the limit $m \rightarrow \infty$. In view of (5.36), doing so gives the result

$$\lim_{m \rightarrow \infty} \mathcal{S}_1^m(M) = R. \quad (4.5.39)$$

We see that repeated application (iteration) of \mathcal{S}_1 drives M to its symplectification R . Since \mathcal{S}_1 is simple to evaluate, see (5.14), and the convergence is very rapid, we conclude that this iterative method is well suited to numerical computation.

As an example of how well the iterative method works, consider the 2×2 case. In this case W as defined by (5.18) must be a multiple of the identity matrix so that we may write

$$W = wI \quad (4.5.40)$$

and

$$\mathcal{T}_1(W) = -[(3/2)w^2 + (1/2)w^3]I. \quad (4.5.41)$$

Exhibit 5.1 below shows successive values of w given by the recursion relation

$$w_{n+1} = -(3/2)(w_n)^2 - (1/2)(w_n)^3 \quad (4.5.42)$$

for various initial conditions w_0 . Evidently the convergence is very rapid as expected.

Exhibit 4.5.1: Convergence of symplectification by iteration in the 2×2 case. Successive values of w_n for various initial conditions w_0 .

n	wn
0	0.1000000000000000
1	-1.550000000000000E-02
2	-3.5851306250000000E-04
3	-1.9277438384305627E-07
4	-5.5742941017167727E-14
5	-4.6609132098651603E-27
6	0.000000000000000E+00
0	-0.1000000000000000
1	-1.450000000000000E-02
2	-3.1385068750000000E-04
3	-1.4773792356625795E-07
4	-3.2739739477203758E-14
5	-1.6078358115527613E-27
6	0.000000000000000E+00
0	0.6000000000000000
1	-0.6480000000000000
2	-0.4938071040000000
3	-0.3055618747255026
4	-0.1257872293112257
5	-2.2738510956402836E-02
6	-7.6968146227769021E-04
7	-8.8838634673523115E-07
8	-1.1838451010276436E-12
9	-2.1022338348398979E-24
10	0.000000000000000E+00
0	-0.6000000000000000
1	-0.4320000000000000
2	-0.2396252160000000
3	-7.9250697011794843E-02
4	-9.1721356097234983E-03
5	-1.2580629042388899E-04
6	-2.3739838483241409E-08
7	-8.4536989012593641E-16
8	-1.0719753766973067E-30
9	0.000000000000000E+00

At this point at least two thoughts come to mind. First, it would be nice to have a procedure that would work whenever $f < 1$ rather than the condition $f < (.56)$, which is

more restrictive. Of course, the map \mathcal{S} does meet this requirement; but its use requires summing the infinite series (5.2), which may be only slowly convergent. It is easily verified that the series for $\mathcal{S}(M)$ and $\mathcal{S}(M^{\text{tr}})$ have the same convergence properties when M and M^{tr} are related by a condition of the form (3.33) for some symplectic matrix R . Note that (3.33) defines an *equivalence relation*. (For the definition of an equivalence relation, see Exercise 5.12.7.) It follows that the convergence of the series for $\mathcal{S}(M)$ depends only on the equivalence class to which M belongs. The same is true for the convergence of the sequence $\mathcal{S}_1^m(M)$. From (5.15) we see that its behavior also depends only on the equivalence class to which M belongs. We note that we have proved that $f < (.56)$ is sufficient to ensure convergence of the sequence $\mathcal{S}_1^m(M)$. However, there may be equivalence classes for which it is not necessary. For example, in the $2n \times 2n$ case, one equivalence class consists of matrices Q' of the form (5.18) with W given by (5.40) with I now being the $2n \times 2n$ identity matrix. It can be shown that the fixed point $w = 0$ of the sequence (5.42) has a larger basin of attraction than that given by the condition $|w| < (.56)$. See Exercise 5.3. Indeed, examination of Exhibit 5.1 shows that convergence occurs when $|w| = (.60)$.

A second thought that comes to mind concerns the properties of maps \mathcal{S}_k produced by discarding in the series (5.2) all terms beyond $\ell = k$. They have properties analogous to those of \mathcal{S}_1 , and they can also be iterated to produce R . What would be their basins of attraction and their rates of convergence? See Exercise 5.4 for a discussion of the properties of \mathcal{S}_2 .

Although these questions may be interesting, they do not seem to be of practical importance for problems encountered to date. That is, the condition $f < (.56)$ always seems to be well satisfied in practice whenever a symplectification is required. Consequently, for the present, we will not pursue these questions further.

Exercises

4.5.1. Verify the expansion (5.2) and compute the first few coefficients e_ℓ . Show that the series $\sum e_\ell x^\ell$ has a radius of convergence of 1. Verify that the series (5.2) converges in norm when $f < 1$, and therefore verify (5.6).

4.5.2. Verify (5.14) through (5.26). Verify the steps that led from (5.27) to (5.30).

4.5.3. Consider the map \mathcal{M} given by (5.42). Show that it has the four fixed points

$$w^f = -2, -1, 0, \pm\infty \quad (4.5.43)$$

where the points $\pm\infty$ are to be identified in a manner similar to the way that all points at infinity are identified by use of the Riemann sphere. Examine the stability of each. You should find that $w^f = -1$ is unstable, and the rest are stable. Show that \bar{w} defined by the equation

$$\bar{w} = -1 + \sqrt{3} \approx .732 \quad (4.5.44)$$

is the positive root of the cubic equation

$$\bar{w}^3 + 3\bar{w}^2 = 2. \quad (4.5.45)$$

Show that the open interval $w \in (-1, \bar{w})$ is in the basin of attraction of the fixed point $w^f = 0$, and points just outside the interval are not. Show that \mathcal{M} sends the two endpoints of the interval into the unstable fixed point $w^f = -1$. Make a numerical study of the w axis to see if there are any other points in the basin of attraction of $w^f = 0$. You should find, for example, that points near $w = -3$ are in the basin of $w^f = 0$. Color the w axis in three colors depending on whether a point on the axis is in the basin of $-2, 0$, or $\pm\infty$. As already illustrated in Figure 1.2.8, the basin of an attracting fixed point can have disjoint pieces.

4.5.4. Study the properties of \mathcal{S}_2 .

4.5.5. Suppose x, p_x, y, p_y, t, p_t is a set of canonical coordinates as in Exercise 1.6.1. With this order of variables the J' of (3.2.10) should be used. In this context a matrix M is called *static* if it is of the form

$$M = \begin{pmatrix} * & * & * & * & 0 & * \\ * & * & * & * & 0 & * \\ * & * & * & * & 0 & * \\ * & * & * & * & 0 & * \\ * & * & * & * & 1 & * \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (4.5.46)$$

Here the entries denoted by $*$ are arbitrary. Later, in Chapter 7, it will be evident that static symplectic matrices are related to Lie transformations generated by quadratic polynomials f_2 with the property $\partial f_2 / \partial t = 0$. Show that if M is a static (but not necessarily symplectic) matrix, the result of symplectifying M by iteration is a static symplectic matrix. That is, symplectification by iteration preserves the property of being static.

4.5.6. Show that if Q is an invertible J -symmetric matrix, then so is Q^T . Review Section 4.5.1. Show that under the transposition operation the symplectifying map \mathcal{S} also has the invariance property

$$\mathcal{S}(M^T) = [\mathcal{S}(M)]^T. \quad (4.5.47)$$

4.5.7. Section 4.5.2 studied symplectification by iteration. This exercise explores orthogonalization by iteration. Suppose M is any real matrix with nonzero determinant and we wish to make the polar decomposition (2.7). In analogy to the symplectic case, define a matrix function $N(M)$ by the rule

$$N(M) = MM^T. \quad (4.5.48)$$

Also, in analogy to (5.1), define a mapping \mathcal{S} by the rule

$$\mathcal{S}(M) = [N(M)]^{-1/2}M. \quad (4.5.49)$$

Show, using (2.8) through (2.10), that

$$\mathcal{S}(M) = O. \quad (4.5.50)$$

Again compute $N^{-1/2}$ using (5.2). Define, in analogy to (5.14), the map \mathcal{S}_1 by the rule

$$\mathcal{S}_1(M) = (1/2)(3I - MM^T)M. \quad (4.5.51)$$

Show that

$$\mathcal{S}_1(O'MO'') = O'\mathcal{S}_1(M)O'' \quad (4.5.52)$$

where O' and O'' are any two orthogonal matrices. Show that \mathcal{S}_1 maps the set of symmetric matrices into itself, and that

$$\mathcal{S}_1(I) = I \quad (4.5.53)$$

so that $M = I$ is a fixed point of \mathcal{S}_1 . Show that, on the set of symmetric matrices, this fixed point is an attractor of \mathcal{S}_1 . Show that if $N(M)$ is sufficiently near I , then

$$\lim_{m \rightarrow \infty} \mathcal{S}_1^m(M) = O \quad (4.5.54)$$

where O is the orthogonal factor appearing in the polar decomposition (2.7).

Ricecar

Let us compare the orthogonal polar decomposition of the symplectic group $Sp(2n, \mathbb{R})$ and the symplectic polar decomposition of the general linear group $GL(2n, \mathbb{R})$.

Symplectic Polar Decomposition

Let us now turn to symplectic polar decomposition. We will be working with all real $2n \times 2n$ matrices. Consider all matrices of the form JS and JA where S and A are symmetric and antisymmetric, respectively. Verify that these matrices obey the commutation rules

$$\{JS, JS'\} = JS'', \quad (4.5.55)$$

$$\{JS, JA\} = JA', \quad (4.5.56)$$

$$\{JA, JA'\} = JS, \quad (4.5.57)$$

and constitute a basis for $gl(2n, \mathbb{R})$. Also recall that the JS constitute a basis for the subalgebra $sp(2n, \mathbb{R})$. Let G be an element in $GL(2n, \mathbb{R})$ that is sufficiently *near* the identity. Then we may write G in the form.

$$G = \exp(JA) \exp(JS). \quad (4.5.58)$$

In writing (5.59) we have employed a hybrid of Lie coordinates of the first and second kinds. See Section 7.9.

Any matrix R of the form

$$R = \exp(JS) \quad (4.5.59)$$

is symplectic, and therefore satisfies the relation

$$JR^T J^{-1} = R^{-1}. \quad (4.5.60)$$

See Sections 3.1 and 3.7. What can be said about the first factor, $\exp(JA)$, in (5.59)? Define a matrix Q by writing

$$Q = \exp(JA). \quad (4.5.61)$$

It satisfies the relation

$$JQ^T J^{-1} = Q. \quad (4.5.62)$$

Any matrix Q that satisfies (5.63) is defined to be *J-symmetric*. Conversely, if Q is *J*-symmetric, it must be of the form (5.62). Finally, motivated by (5.59), we say that G has a symplectic polar decomposition if it can be written in the form

$$G = QR \quad (4.5.63)$$

where Q is *J* symmetric and R is symplectic. Here, in this definition, R may be an arbitrary symplectic matrix not necessarily expressible in the single-exponent form $\exp(JS)$.

We have shown that G has a symplectic polar decomposition if it is sufficiently near the identity. From a counter example we know that there are matrices G that do not have symplectic polar decompositions. We also know from various theorems that there are broad conditions that guarantee the existence of a symplectic polar decomposition. In particular, given G , we have found it useful to define the matrix $N(G)$ by the rule

$$N(G) = G J G^T J^T, \quad (4.5.64)$$

and have shown that its properties can be used to infer in certain important cases the existence of a symplectic polar decomposition. Our next task, given G , is to determine, if possible, explicit formulas for the factors Q and R in its symplectic polar decomposition.

Note that if G is symplectic, then

$$N(G) = G J G^T J^T = (G J G^T) J^T = J J^T = I \quad (4.5.65)$$

and

$$\|N(G) - I\| = 0. \quad (4.5.66)$$

Since $N = I$ when G is symplectic, we may define a measure $f(G)$ of the *failure* of G to be symplectic by the rule

$$f(G) = \|N(G) - I\|. \quad (4.5.67)$$

Suppose $f(G)$ is small. Then G is nearly symplectic, and we might hope to find a matrix R that is both near G and exactly symplectic. Suppose G is invertible, and suppose there exists a *J*-symmetric matrix Q such that

$$N(G) = Q^2 \quad (4.5.68)$$

with N defined by (5.65). We will see that the factorization (5.64) can be achieved if $f(G) < 1$.

To prove this result, we first observe that Q is invertible: we know from (3.32) that N is invertible if G is, and from (3.38) we see that Q is invertible if N is. Next, since Q is invertible, (5.64) can be solved for R to give

$$R = Q^{-1}G. \quad (4.5.69)$$

Finally the computation

$$\begin{aligned} R J R^T &= (Q^{-1} G) J (Q^{-1} G)^T = Q^{-1} G J G^T (Q^{-1})^T \\ &= Q^{-1} G J G^T J^T J (Q^{-1})^T J^{-1} J = Q^{-1} N Q^{-1} J \\ &= Q^{-1} Q^2 Q^{-1} J = J \end{aligned} \quad (4.5.70)$$

shows that R is symplectic, and we will say that this R is the *symplectification* of G ,

Conversely, suppose that Q and R in (3.10) are J -symmetric and symplectic, respectively. Then we find the result

$$\begin{aligned} N &= G J G^T J^T = Q R J (Q R)^T J^T = Q R J R^T Q^T J^T \\ &= Q J Q^T J^{-1} = Q^2. \end{aligned} \quad (4.5.71)$$

We have learned that establishing the factorization (5.64) is equivalent to finding a J -symmetric matrix Q that satisfies (5.69).

We now turn to the task of given G , find Q . According to (5.69) we may hope to find a relation of the form

$$Q = [N(G)]^{1/2}, \quad (4.5.72)$$

with the further requirement that Q be J -symmetric. We will see that this is possible if $f(G) < 1$.

Begin by writing the identity

$$N = I + (N - I) \quad (4.5.73)$$

and, inspired by the binomial theorem, write

$$N^{1/2} = [I + (N - I)]^{1/2} = I + (1/2)(N - I) - (1/8)(N - I)^2 + \dots = I + \sum_{\ell=1}^{\infty} c_{\ell}(N - I)^{\ell} \quad (4.5.74)$$

where the c_{ℓ} are the binomial coefficients

$$c_{\ell} = \binom{1/2}{\ell}. \quad (4.5.75)$$

Note that we may also write

$$N^{-1/2} = N^{-1} N^{1/2} = N^{-1} [I + (1/2)(N - I) + \dots] = N^{-1} + N^{-1} \sum_{\ell=1}^{\infty} c_{\ell}(N - I)^{\ell}. \quad (4.5.76)$$

Note that (by Lemmas 3.1, 3.2, 3.5, and 3.9) all the terms I , N , and $[(I - N)^{\ell}/\ell]$ are J -symmetric matrices. Consequently, if the series (5.74) converges, then (by Lemma 3.2) Q will be a J -symmetric matrix.

Assuming these relations make sense, we have the result

$$Q = N^{1/2} = I + \sum_{\ell=1}^{\infty} c_{\ell}(N - I)^{\ell}. \quad (4.5.77)$$

The infinite sum on the right side of (5.78) is defined if the sum

$$\sum_{\ell=1}^{\infty} |c_{\ell}| \|N - I\|^{\ell} \quad (4.5.78)$$

converges which, in view of (5.68), is equivalent to the requirement

$$\sum_{\ell=1}^{\infty} |c_{\ell}| f^{\ell} < \infty. \quad (4.5.79)$$

It is easily checked that this requirement is met if $f < 1$.

We now move on to the calculation of R . In view of (5.70), we see that what is also needed now is Q^{-1} . And according to (5.73), or (5.72), we may hope for a relation of the form

$$Q^{-1} = [N(G)]^{-1/2}, \quad (4.5.80)$$

and we further require that Q^{-1} be J -symmetric. We will see that this is possible if $f(G) < 1$. Again write (5.74) and, inspired by the binomial theorem, write

$$N^{-1/2} = [I + (N - I)]^{-1/2} = I - (1/2(N - I)) + \dots = I + \sum_{\ell=1}^{\infty} d_{\ell}(N - I)^{\ell} \quad (4.5.81)$$

where the d_{ℓ} are the binomial coefficients

$$d_{\ell} = \binom{-1/2}{\ell}. \quad (4.5.82)$$

This result for $N^{-1/2}$ is to be compared with the equally valid result (5.77). We also note that for $N^{1/2}$ there is the result

$$N^{1/2} = NN^{-1/2} = N[I - (1/2(N - I)) + \dots] = N + N \sum_{\ell=1}^{\infty} d_{\ell}(N - I)^{\ell}, \quad (4.5.83)$$

which is to be compared with (5.75).

Assuming these relations make sense, we have the result

$$Q^{-1} = N^{-1/2} = I + \sum_{\ell=1}^{\infty} d_{\ell}(N - I)^{\ell}. \quad (4.5.84)$$

The infinite sum on the right side of (5.85) is defined if the sum

$$\sum_{\ell=1}^{\infty} |d_{\ell}| \|N - I\|^{\ell} \quad (4.5.85)$$

converges which, in view of (5.68), is equivalent to the requirement

$$\sum_{\ell=1}^{\infty} |d_{\ell}| f^{\ell} < \infty. \quad (4.5.86)$$

It is easily checked that this requirement is met if $f < 1$. Also, again by Lemmas 3.1, 3.2, 3.5, and 3.9, all the terms I , N , and $[(I - N)^\ell/\ell]$ are J -symmetric matrices. Consequently, if the series (5.81) converges, then (by Lemma 3.2) Q^{-1} will be a J -symmetric matrix.

With Q^{-1} in hand we are able to compute R using (5.70) to find

$$R = Q^{-1}G = G + \left[\sum_{\ell=1}^{\infty} d_\ell (N - I)^\ell \right] G. \quad (4.5.87)$$

In view of (5.65), (5.78), and (5.88) we have now found, in terms of G , both the symplectic polar decomposition factors Q and R .

This series converges for $|a - 1| < 1$ in which case a must be in the interval

$$a \in (0, 2). \quad (4.5.88)$$

This domain can be extended by using the group property of ordinary multiplication. Let us write

$$a = \lambda(\lambda^{-1}a) \quad (4.5.89)$$

where λ is any positive number. Then there is the result

$$a^{1/2} = \lambda^{1/2}(\lambda^{-1}a)^{1/2}. \quad (4.5.90)$$

Suppose λ is of the form

$$\lambda = 2^{2m} \quad (4.5.91)$$

where m is a positive integer. Then

$$\lambda^{1/2} = 2^m. \quad (4.5.92)$$

and

$$a^{1/2} = \lambda^{1/2}(\lambda^{-1}a)^{1/2} = 2^m[a/(2^{2m})]^{1/2} = 2^m[a/(4^m)]^{1/2}. \quad (4.5.93)$$

Combining this result with (5.89) gives the final result

$$a^{1/2} = 2^m \left\{ 1 + \sum_{\ell=1}^{\infty} c_\ell [a/(4^m) - 1]^\ell \right\}. \quad (4.5.94)$$

According to (5.90) this relation will be valid for

$$[a/(4^m)] \in (0, 2) \Leftrightarrow a \in (0, 8^m). \quad (4.5.95)$$

Note that, given a , by a suitable choice of m there is a Taylor representation for $a^{1/2}$ whose interval of validity contains a .

Application of Taylor Method to Numbers

Let us explore the Taylor method (5.75) when it is applied to numbers a . In the case of numbers the Taylor method reads

$$a^{1/2} = [1 + (a - 1)]^{1/2} = 1 + (1/2)(a - 1) - (1/8)(a - 1)^2 + \dots = 1 + \sum_{\ell=1}^{\infty} c_{\ell}(a - 1)^{\ell}. \quad (4.5.96)$$

This series converges for $|a - 1| < 1$ in which case a must be in the interval

$$a \in (0, 2). \quad (4.5.97)$$

This domain can be extended by using the group property of ordinary multiplication. Let us write

$$a = \lambda(\lambda^{-1}a) \quad (4.5.98)$$

where λ is any positive number. Then there is the result

$$a^{1/2} = \lambda^{1/2}(\lambda^{-1}a)^{1/2}. \quad (4.5.99)$$

Suppose λ is of the form

$$\lambda = 2^{2m} \quad (4.5.100)$$

where m is a positive integer. Then

$$\lambda^{1/2} = 2^m. \quad (4.5.101)$$

and

$$a^{1/2} = \lambda^{1/2}(\lambda^{-1}a)^{1/2} = 2^m[a/(2^{2m})]^{1/2} = 2^m[a/(4^m)]^{1/2}. \quad (4.5.102)$$

Combining this result with (5.89) gives the final result

$$a^{1/2} = 2^m \left\{ 1 + \sum_{\ell=1}^{\infty} c_{\ell}[a/(4^m) - 1]^{\ell} \right\}. \quad (4.5.103)$$

According to (5.90) this relation will be valid for

$$[a/(4^m)] \in (0, 2) \Leftrightarrow a \in (0, 8m). \quad (4.5.104)$$

Note that, given a , by a suitable choice of m there is a Taylor representation for $a^{1/2}$ whose interval of validity contains a .

Solution by Iteration

Given the matrix G , we have defined the matrix $N(G)$ by the rule (5.65). Subsequently we sought to define and then find the matrices $N(G)^{1/2}$ and $N(G)^{-1/2}$. We were able to do so in terms of the power series (5.75) and (5.82), which were shown to be convergent providing $f(G) < 1$.

Surprisingly, the matrices $N(G)^{1/2}$ and $N(G)^{-1/2}$ can also be found using an iterative process. For a *number*, (Robert E.) Goldschmidt discovered an iterative process for finding its square root and the inverse of its square root. The iterative process we will ultimately

describe is the matrix version of Goldschmidt's method for numbers. It takes the matrix N as an input and delivers, when convergent, the matrices $N^{1/2}$ and $N^{-1/2}$ as outputs. We remark that this process makes no particular assumption about how N depends on G . In retrospect, the power series method does not either. It only assumes that $\|N - I\| < 1$.

Shortly we will describe Goldschmidt's algorithm for numbers. We begin with Newton's method for functions of numbers. Suppose a is a given number and we wish to find $a^{1/2}$, its square root. Define a function $f(x)$ by the rule

$$f(x) = x^2 - a. \quad (4.5.105)$$

The number we seek, $a^{1/2}$, is evidently a zero of f ,

$$f(a^{1/2}) = 0. \quad (4.5.106)$$

According to Newton, see Section *, it can be found by setting up the recursion relation

$$x_{n+1} = x_n - f(x_n)/f'(x_n), \quad (4.5.107)$$

which will converge to

$$x_\infty = a^{1/2} \quad (4.5.108)$$

provided the recursion relation is begun with an *estimated* initial value x_0 sufficiently near $a^{1/2}$,

$$x_0 = x_{est} \simeq a^{1/2}. \quad (4.5.109)$$

For the problem at hand,

$$f'(x) = 2x \quad (4.5.110)$$

so that Newton's method for this case becomes

$$x_{n+1} = x_n - (1/2)(x_n^2 - a)/x_n, \quad (4.5.111)$$

which can be rewritten as

$$x_{n+1} = (1/2)x_n + (1/2)(a)(1/x_n). \quad (4.5.112)$$

[Newton's method, when applied to computing a square root, is sometimes referred to as the Babylonian method or Hero's method since it was known to the Babylonians (1500 BC) and Greeks (100 AD).] Note that each iteration requires performing the division $1/x_n$, which is a relatively slow process. More about this point later.

In summary, given a , Newton's method for computing $a^{1/2}$ is as follows:

- Initialize

$$x_0 = a_{est}^{1/2}, \quad (4.5.113)$$

- Iterate

$$x_{n+1} = (1/2)x_n + (1/2)(a)(1/x_n), \quad (4.5.114)$$

- Final Result

$$x_\infty = a^{1/2}. \quad (4.5.115)$$

Assuming convergence occurs, (5.98) becomes

$$x_\infty = (1/2)x_\infty + (1/2)(a)(1/x_\infty), \quad (4.5.116)$$

from which we conclude

$$(x_\infty)^2 = a, \quad (4.5.117)$$

in accord with (5.99).

As a sanity check, let us compare this method with the Taylor method (5.75) when it is applied to numbers. In the case of numbers the Taylor method reads

$$a^{1/2} = [1 + (a - 1)]^{1/2} = 1 + (1/2)(a - 1) - (1/8)(a - 1)^2 + \dots = 1 + \sum_{\ell=1}^{\infty} c_\ell(a - 1)^\ell. \quad (4.5.118)$$

Now we are ready for an inspired trick: Set up a related iterative process that will compute $a^{-1/2}$. This is easily done. Define a function $g(y)$ by the rule

$$g(y) = y^2 - (1/a). \quad (4.5.119)$$

Evidently $a^{-1/2}$ a zero of g ,

$$g(a^{-1/2}) = [a^{-1/2}]^2 - (1/a) = a^{-1} - (1/a) = 0. \quad (4.5.120)$$

And, applying Newton's method to g yields the iterative process

$$y_{n+1} = (1/2)y_n + 1/2)(a^{-1})(1/y_n). \quad (4.5.121)$$

It will converge to

$$y_\infty = a^{-1/2} \quad (4.5.122)$$

provided the recursion relation is begun with an initial value y_0 sufficiently near $a^{-1/2}$,

$$y_0 = y_{est} \simeq a^{-1/2}. \quad (4.5.123)$$

Now consider the two relations (5.95) and (5.98). Initiate (5.95) with x_0 as described, and initiate (5.98) with

$$y_0 = 1/x_0 \Leftrightarrow x_0 y_0 = 1. \quad (4.5.124)$$

Then y_0 will be near $a^{-1/2}$ since x_0 is near $a^{1/2}$. Choose x_0 sufficiently near $a^{1/2}$ so both the iterative processes (5.95) and (5.98) are convergent. Since both are convergent we have

$$x_n \rightarrow x_\infty = a^{1/2} \text{ and } y_n \rightarrow y_\infty = a^{-1/2}. \quad (4.5.125)$$

Observe that

$$x_\infty y_\infty = 1. \quad (4.5.126)$$

Since the recursion relations are initiated with (5.100) and end with (5.102), it is reasonable to hope that, as iteration proceeds, there will be the relation

$$x_n y_n \simeq 1 \Leftrightarrow (1/x_n) \simeq y_n \text{ and } (1/y_n) \simeq x_n. \quad (4.5.127)$$

Let us compute x_1y_1 . We find

$$\begin{aligned} x_1y_1 &= [(1/2)x_0 + (1/2)(a)(1/x_0)][(1/2)y_0 + (1/2)(a^{-1})(1/y_0)] \\ &= [(1/4)x_0y_0 + (1/4)(x_0y_0)^{-1}] \\ &+ [(1/4)a^{-1}(x_0/y_0) + (1/4)(a)(y_0/x_0)] \\ &= [1/2] + (1/4)[a^{-1}(x_0)^2 + a(y_0)^2]. \end{aligned} \quad (4.5.128)$$

From (5.93) and (5.101) we see that

$$a^{-1}(x_0)^2 \simeq 1 \text{ and } a(y_0)^2 \simeq 1 \quad (4.5.129)$$

and therefore

$$(1/4)[a^{-1}(x_0)^2 + a(y_0)^2] \simeq (1/2). \quad (4.5.130)$$

Combining (5.106) and (5.108) gives the result

$$x_1y_1 \simeq 1, \quad (4.5.131)$$

which is consistent with (5.105). Note that (5.105) can be checked numerically. In view of (5.104), it should become evermore exact as $n \rightarrow \infty$.

Assuming (5.105) to be true, make in (5.95) and (5.98) the substitutions

$$1/x_n \rightarrow y_n \text{ and } 1/y_n \rightarrow x_n \quad (4.5.132)$$

to propose the grand Ansatz

$$x_{n+1} = (1/2)x_n + (1/2)(a)(y_n), \quad (4.5.133)$$

$$y_{n+1} = (1/2)y_n + (1/2)(a^{-1})(x_n) \quad (4.5.134)$$

with the iteration to be begun with the condition (5.100) and x_0 sufficiently near $a^{1/2}$ and y_0 sufficiently near $a^{-1/2}$. This is a version of the algorithm of Goldschmidt for numbers. Note that, unlike its Newton parents, it has the feature that it is never necessary to invert x_n or y_n during the iteration process. Finally, if convergence occurs and all goes well, there should be the results

$$x_\infty = (1/2)x_\infty(a) + (1/2)(a)(y_\infty) \Leftrightarrow x_\infty = (a)(y_\infty), \quad (4.5.135)$$

$$y_\infty = (1/2)y_\infty + (1/2)(a^{-1})(x_\infty) \Leftrightarrow y_\infty = (a^{-1})(x_\infty). \quad (4.5.136)$$

Combining the relations on the far right sides of (5.113) and (5.114) with (5.104) gives the relations

$$x_\infty = a^{1/2} \text{ and } y_\infty = a^{-1/2}, \quad (4.5.137)$$

as expected.

Presumably because of (5.105) and its increasing validity as $n \rightarrow \infty$, the substitutions (5.110) should not disturb the final convergence properties of (5.111) and (5.112) from those of its underlying Newton methods, but they may affect the size of the convergence basin. This question could also be studied numerically.

There is another version of Goldschmidt's algorithm that is also convenient. It is achieved by change of some variables.

- Initialize

$$x_0 = (a^{1/2})_{est}, \quad (4.5.138)$$

$$h_0 = (1/2)[(a^{1/2})_{est}]^{-1}, \quad (4.5.139)$$

$$r_0 = (1/2)(1 - 2x_0 h_0) \quad (4.5.140)$$

- Iterate

$$x_{n+1} = x_n + r_n x_n \quad (4.5.141)$$

$$h_{n+1} = h_n + r_n h_n \quad (4.5.142)$$

$$r_{n+1} = (1/2)(1 - 2x_{n+1} h_{n+1}) \quad (4.5.143)$$

- Final Result

$$x_\infty = a^{1/2} \quad (4.5.144)$$

$$2h_\infty = a^{-1/2} \quad (4.5.145)$$

$$r_\infty = 0 \quad (4.5.146)$$

- Initialize. Begin with a input matrix N . Guess a matrix y_0 , an approximation to $N^{-1/2}$,

$$y_0 \simeq N^{-1/2}, \quad (4.5.147)$$

in the sense that $(y_0)^2 \simeq N^{-1}$. For example, use the first few terms in the series (5.82) as an estimate for y_0 . Next define x_0 by the rule

$$x_0 = Ny_0. \quad (4.5.148)$$

It is an approximation to $N^{1/2}$. See (5.89). Finally, define h_0 and r_0 by the rules

$$h_0 = (1/2)y_0 \quad (4.5.149)$$

and

$$r_0 = x_0 h_0 - (1/2)I. \quad (4.5.150)$$

Note that, depending on the accuracy of the guess y_0 , r_0 has the property

$$r_0 = (1/2)N(y_0)^2 - (1/2)I \simeq (1/2)NN^{-1} - (1/2)I \simeq 0. \quad (4.5.151)$$

- Iterate using the rule

$$x_{n+1} = x_n + x_n r_n = x_n(I + r_n), \quad (4.5.152)$$

$$h_{n+1} = h_n + h_n r_n = h_n(I + r_n), \quad (4.5.153)$$

$$r_{n+1} = h_n r_n - (1/2)I. \quad (4.5.154)$$

The relations (5.94) and (5.95) provide a map of x, h , space into itself. We hope that this map will have a fixed point x_∞, h_∞ and that this fixed point will be an attractor with sizable basin so that x_0, h_0 is within the basin. The fixed point will then provide the results

$$N^{-1}x_\infty = N^{-1/2}, \quad (4.5.155)$$

$$2Nh_\infty = N^{1/2} \quad (4.5.156)$$

so that, as desired, $N^{1/2}$ and $N^{-1/2}$ have been found. And, as iterations proceed, there should be the limiting behavior $\lim_{n \rightarrow \infty} r_n = 0$, which demonstrates convergence and can conveniently be monitored by observing the $\|r_n\|$. We also remark that, unlike some other possible algorithms, the relations (5.94) though (5.96) involve only matrix multiplication and not the slower operation of matrix inversion. Finally, once underway, the convergence to x_∞, h_∞ is quadratic.

As a sanity check, let us apply the iterative method starting with the lowest order approximation for y_0 given by (5.82), namely

$$y_0 = I. \quad (4.5.157)$$

Then use of (5.87) through (5.89) gives the initiation results

$$x_0 = Ny_0 = NI = N, \quad (4.5.158)$$

$$h_0 = (1/2)y_0 = (1/2)I, \quad (4.5.159)$$

$$r_0 = x_0h_0 - (1/2)I = (1/2)N - (1/2)I = (1/2)(N - I). \quad (4.5.160)$$

And use of (5.94) through (5.96) gives, for $n = 0$, the first iteration results

$$x_1 = x_0 + x_0r_0 = N + N(1/2)(N - I), \quad (4.5.161)$$

$$h_1 = h_0 + h_0r_0 = (1/2)I + (1/2)I(1/2)(N - I) = (1/2)I + (1/4)(N - I), \quad (4.5.162)$$

$$r_1 = h_0r_0 - (1/2)I = (1/2)(N - I). \quad (4.5.163)$$

From (5.103) and (5.73) we see that

$$N^{-1}x_1 = I + (1/2)(N - I) \simeq N^{1/2}; \quad (4.5.164)$$

and from (5.104) and * we see that

$$2Nh_1 = N + (1/2)N(N - I) \simeq N^{-1/2}. \quad (4.5.165)$$

These results are consistent with the asserted limiting behaviors (5.97) and (5.98), respectively. Finally, let us compute r_2 as given by (5.93) with $n = 1$. So doing gives the result

$$\begin{aligned} r_2 &= h_1r_1 - (1/2)I = [(1/2)I + (1/4)(N - I)][(1/2)(N - I)] - (1/2)I \\ &= [(1/4)(N - I) + (1/8)(N - I)^2] = \end{aligned} \quad (4.5.166)$$

A standard general method for finding roots of a general matrix N is to first find, if possible, its logarithm.⁵ Therefore, let us first try to compute $\log(N)$. From (3.7.2) we find the result

$$\log(N) = - \sum_{\ell=1}^{\infty} (I - N)^{\ell}/\ell. \quad (4.5.167)$$

Note that (by Lemmas 3.1, 3.2, 3.5, and 3.9) all the terms $[(I - N)^{\ell}/\ell]$ are J -symmetric matrices. Consequently, if the series (5.74) converges, then (by Lemma 3.2) $\log(N)$ will be a J -symmetric matrix. If the series does converge, let us define a matrix Q by the rule

$$Q = \exp[(1/2)\log(N)]. \quad (4.5.168)$$

The matrix Q will also be J -symmetric. [Apply to the series (3.7.1) arguments similar to those just made for $\log(N)$.] Moreover, Q will satisfy (3.38),

$$Q^2 = \{\exp[(1/2)\log(N)]\}^2 = \exp[\log(N)] = N. \quad (4.5.169)$$

Therefore we can achieve the factorization (3.10) if the series (3.43) converges.

The series (3.43) will converge if N is sufficiently near I . Specifically, the series will converge if $\|N - I\| < 1$ for some choice of matrix norm, which is the case if $f(G) < 1$. [Moreover, according to the remark made in Lemma 3.9, $N = I$ if G is symplectic. Consequently, the series will converge if G is sufficiently near a symplectic matrix.] Taken together, (5.65), (5.74), and (5.75) provide an explicit formula for Q . And (5.70) provides an explicit formula for R .

To see how this strategy works in practice, suppose we retain only the first few terms in the various series involved. Retaining the first two terms in (5.74) yields

$$\log(N) = (N - I) - (1/2)(N - I)^2 + O[(N - I)^3]. \quad (4.5.170)$$

Next expand (5.75) to find

$$\begin{aligned} Q &= \exp[(1/2)\log(N)] = I + [(1/2)\log(N)] + \cdots \\ &= I + (1/2)(N - I) + \cdots. \end{aligned} \quad (4.5.171)$$

Note that there is the binomial theorem result

$$N^{1/2} = [I + (N - I)]^{1/2} = I + (1/2)(N - I) + \cdots \quad (4.5.172)$$

so that

$$Q = N^{1/2} = I + (1/2)(N - I) + \cdots, \quad (4.5.173)$$

in agreement with (5.78).

Let us continue on to compute R using (5.70). From (5.80) we find

$$Q^{-1} = I - (1/2)(N - I) + \cdots. \quad (4.5.174)$$

Combining (5.70) and (5.81) yields

$$R = [I - (1/2)(N - I)]G + \cdots = G - (1/2)(N - I)G + \cdots. \quad (4.5.175)$$

⁵Recall from algorithms Exercise 2.3 that there are special methods for finding matrix roots if the matrix is positive symmetric.

Other Symplectification Methods

We have described at length matrix symplectification using symplectic polar decomposition. We next consider some other possible matrix symplectification methods.

4.6 Modified Darboux Symplectification

Suppose one is given a matrix M whose determinant is nonzero. The columns of M may be regarded as vectors m^1, m^2, m^3, \dots , and the condition $\det(M) \neq 0$ is equivalent to the statement that the vectors m^j are linearly independent. Given a set of linearly independent vectors m^j , there is the Darboux process for constructing an associated set of symplectic vectors r^j . Finally, the vectors r^j may be viewed as the columns of a matrix R , and this matrix will be symplectic. Thus, given any nonsingular matrix M , there is a procedure for constructing a corresponding symplectic matrix R . Moreover, if M itself is nearly symplectic, then R will be near M . Indeed, if M happens to be symplectic, then R will coincide with M . See Sections 3.6.3 and 3.6.5. In this section we will describe what we will call *modified* Darboux symplectification, and will examine how close R is to M if M is nearly symplectic.

Let M be a $2n \times 2n$ matrix. Rather than using (4.1), we will describe the *failure* of M to be symplectic in terms of an antisymmetric matrix F defined by the relation

$$F = M^T JM - J. \quad (4.6.1)$$

From (4.1) and (6.1) we have the result

$$\begin{aligned} \| F \| &= \| (M^T JM J^T - I)J \| \leq \| M^T JM J^T - I \| \| J \| \\ &\leq \| M^T JM J^T - I \| = \| M^T J(M^T)^T J^T - I \| = f(M^T). \end{aligned} \quad (4.6.2)$$

Here we have assumed that the matrix norm employed has the property

$$\| J \| = 1, \quad (4.6.3)$$

which is true for the maximum column sum norm (3.7.15) and the spectral norm (3.7.17). If the norm also has the property (3.7.97), which we shall also assume, then the matrix elements of F are bounded by the relation

$$|F_{jk}| \leq f(M^T). \quad (4.6.4)$$

Suppose we view M as a collection of column vectors m^1, m^2, \dots, m^{2n} . Let m_i^j denote the i th component of the j th such vector. Then, following the usual matrix element labelling scheme, we have the relation

$$m_i^j = M_{ij}. \quad (4.6.5)$$

In terms of the vectors m^j , the relation (6.1) can be rewritten in the form

$$(M^T JM)_{jk} = (m^j, Jm^k) = J_{jk} + F_{jk}. \quad (4.6.6)$$

Correspondingly if R is a symplectic matrix and we view it as a collection of column vectors r^j , then the symplectic condition (3.1.2) can be written in the form

$$(r^j, Jr^k) = J_{jk}. \quad (4.6.7)$$

Assume we are given an M for which $f(M^T)$ is sufficiently small. From M we extract the vectors m^j using (6.5). Our task is to use these m^j , which obey (6.6), to construct a set of vectors r^j that obey (6.7). Moreover, this construction is to be made in such a way that the corresponding symplectic matrix R is near M in the sense that

$$\| M - R \| \sim f(M^T). \quad (4.6.8)$$

We will construct the vectors r^j two at a time, beginning with r^1 and r^2 . To simplify our presentation, we will use a J matrix of the form (3.2.10). For this choice we have the relation

$$J_{12} = 1, \quad (4.6.9)$$

and (6.6) gives the result

$$(m^1, Jm^2) = 1 + F_{12}. \quad (4.6.10)$$

According to (6.4) and (6.10), if $f(M^T)$ is sufficiently small, the quantity (m^1, Jm^2) will be positive and hence will have a positive square root γ_{12} ,

$$\gamma_{12} = +[(m^1, Jm^2)]^{1/2}. \quad (4.6.11)$$

We can therefore define “normalized” vectors r^1 and r^2 by the rules

$$r^1 = m^1 / \gamma_{12}, \quad (4.6.12)$$

$$r^2 = m^2 / \gamma_{12}. \quad (4.6.13)$$

Note that by (6.4) and (6.10), γ_{12} will be near 1 if $f(M^T)$ is sufficiently small. Correspondingly r^1 and r^2 will be near m^1 and m^2 , respectively. By construction, these vectors satisfy the relation

$$(r^1, Jr^2) = 1 = J_{12}, \quad (4.6.14)$$

as required by (6.7). Also, because J is antisymmetric, we automatically get from (6.14) the relations

$$(r^j, Jr^k) = J_{jk} \text{ when } j = 1, 2 \text{ and } k = 1, 2, \quad (4.6.15)$$

as is also required by (6.7).

Next we construct the vectors r^3 and r^4 . We begin by defining intermediate vectors s^3 and s^4 according to the rule

$$s^3 = m^3 + \alpha_{31}r^1 + \alpha_{32}r^2, \quad (4.6.16)$$

$$s^4 = m^4 + \alpha_{41}r^1 + \alpha_{42}r^2, \quad (4.6.17)$$

where the α 's are coefficients still to be determined. According to (6.7) we must have the relations

$$(r^j, Jr^k) = 0 \text{ when } j = 1, 2 \text{ and } k = 3, 4. \quad (4.6.18)$$

Let us therefore require the relations

$$(r^j, Js^k) = 0 \text{ when } j = 1, 2 \text{ and } k = 3, 4. \quad (4.6.19)$$

Doing so determines the values of the coefficients α :

$$\alpha_{31} = (r^2, Jm^3), \quad (4.6.20)$$

$$\alpha_{32} = -(r^1, Jm^3), \quad (4.6.21)$$

$$\alpha_{41} = (r^2, Jm^4), \quad (4.6.22)$$

$$\alpha_{42} = -(r^1, Jm^4). \quad (4.6.23)$$

If $f(M^T)$ is sufficiently small then, according to (6.4), (6.6), (6.11) through (6.13), and (6.20) through (6.23), all the α 's are of order $f(M^T)$. It also follows that the quantity (s^3, Js^4) will be positive and consequently will have the positive square root

$$\gamma_{34} = +[(s^3, Js^4)]^{1/2}. \quad (4.6.24)$$

Finally, we define the normalized vectors r^3 and r^4 by the rules

$$r^3 = s^3 / \gamma_{34}, \quad (4.6.25)$$

$$r^4 = s^4 / \gamma_{34}. \quad (4.6.26)$$

Upon reflection we see that we have now constructed four vectors r^1 through r^4 that are, respectively, near m^1 through m^4 if $f(M^T)$ is small; and these vectors satisfy the relations

$$(r^j, Jr^k) = J_{jk} \text{ when } j, k = 1, 2, 3, 4. \quad (4.6.27)$$

Moreover the general pattern is now clear. We see that the construction can be continued to include r^5 and r^6 (and still more r 's if we are dealing with more than a 6-dimensional phase space). We simply write the analogs of (6.16) and (6.17), for example

$$s^5 = m^5 + \alpha_{51}r^1 + \alpha_{52}r^2 + \alpha_{53}r^3 + \alpha_{54}r^4, \quad (4.6.28)$$

$$s^6 = m^6 + \alpha_{61}r^1 + \alpha_{62}r^2 + \alpha_{63}r^3 + \alpha_{64}r^4, \quad (4.6.29)$$

determine the α 's, and then normalize the results. Finally, we may view all the r^j we have constructed in this manner as the columns of a matrix R . This matrix will be symplectic and will be close to M in the sense of satisfying (6.8).

There is one last nuisance to be resolved. All our estimates have involved the quantity $f(M^T)$ whereas it would be more pleasant to work with $f(M)$. This defect can be overcome by using the modified Darboux procedure just described to symplectify the matrix M^T instead of M . Call the resulting symplectic matrix R' . Using (3.7.51) and (6.8) we will then have the result

$$|(M^T)_{jk} - R'_{jk}| \sim f(M). \quad (4.6.30)$$

Finally we define R , which is to be the symplectification of M , by writing

$$R = (R')^T. \quad (4.6.31)$$

Combining (6.30) and (6.31) then gives the desired result

$$|M_{jk} - R_{jk}| \sim f(M). \quad (4.6.32)$$

We close this discussion with the comment that, like the Darboux symplectification procedure described in Section 3.6.5, modified Darboux Symplectification also does not treat the vectors m^j democratically.

Exercises

4.6.1. Show that F as defined by (6.1) is antisymmetric.

4.6.2. Refer to Exercise 5.4. Show that modified Darboux symplectification also preserves the property of being static.

4.7 Exponential and Cayley Symplectifications

Both the exponential and Cayley representations of a matrix provide additional methods for matrix symplectification. We will first describe the use of the exponential representation. Subsequently we will consider the use of the Cayley representation, which is based on the exponential representation.

4.7.1 Exponential Symplectification

As before, let M be a (real) $2n \times 2n$ matrix. Consider, in matrix space, the ray λM where λ lies in the range $0 < \lambda < \infty$. Suppose that for some value λ_0 the matrix $\lambda_0 M$ lies *within* the unit ball about I . [The geometric picture for this situation is similar to that of Figure 4.1 except that the ray $N(\lambda M)$ is replaced by the ray λM .] Then M can be written in the exponential form

$$M = \exp(B) \quad (4.7.1)$$

where B is a real matrix. The proof for this assertion is straightforward: Since by hypothesis $\lambda_0 M$ lies within the unit ball about I , the series of the form (3.43) for $\log(\lambda_0 M)$ converges, and we may write

$$\lambda_0 M = \exp[\log(\lambda_0 M)]. \quad (4.7.2)$$

It follows that M can be written in the form

$$M = [(\lambda_0)^{-1} I][\lambda_0 M] = \exp[-I \log(\lambda_0)] \exp[\log(\lambda_0 M)] = \exp(B) \quad (4.7.3)$$

where B is defined by the equation

$$B = \log(\lambda_0 M) - I \log(\lambda_0). \quad (4.7.4)$$

It is now a simple matter to find a symplectification R for M . Without loss of generality, the matrix B can be written in the form (3.1) where S and A are uniquely defined. We simply take R to be the symplectic matrix given by the relation

$$R = \exp(JS). \quad (4.7.5)$$

4.7.2 Cayley Symplectification

The symplectification provided by (7.5) has the defect that it requires the summation of the infinite exponential series. Although this problem can be overcome by the method of

Section 4.1, it is worthwhile to explore other possibilities. Suppose M can be written in the exponential form (7.1). Then we may write the relations

$$\begin{aligned} M &= \exp(B) = [\exp(B/2)]/[\exp(-B/2)] \\ &= [\cosh(B/2) + \sinh(B/2)]/[\cosh(B/2) - \sinh(B/2)] \\ &= [I + \tanh(B/2)]/[I - \tanh(B/2)]. \end{aligned} \quad (4.7.6)$$

Define a matrix T by the equation

$$T = \tanh(B/2). \quad (4.7.7)$$

With the aid of T , M as given by (7.6) has the Cayley representation

$$M = (I + T)(I - T)^{-1} = (I - T)^{-1}(I + T). \quad (4.7.8)$$

The relation (7.8) can be solved for T to give the result

$$T = (M + I)^{-1}(M - I) = (M - I)(M + I)^{-1}. \quad (4.7.9)$$

Now view (7.9) as the *definition* of T in terms of M . That is, this definition can be made without any reference to B . Define the matrix V by the equation

$$V = J^{-1}T. \quad (4.7.10)$$

We know that V will be symmetric if M is symplectic, and vice versa. See Section 3.11. Consequently, V will be nearly symmetric if M is nearly symplectic. Let us define a symmetric matrix W by taking the symmetric part of V ,

$$W = (V + V^T)/2. \quad (4.7.11)$$

Then we may define a symplectic matrix R by writing

$$R = (I + JW)(I - JW)^{-1} = (I - JW)^{-1}(I + JW), \quad (4.7.12)$$

and R will be a symplectification of M that we will call the *Cayley* symplectification. Note that while the evaluation of (7.5) requires the summation of an infinite series, the evaluation of (7.9) and (7.12) requires only matrix inversion.

Let us view R , the result of this Cayley symplectification process applied to M , as the outcome of a Cayley symplectifying map \mathcal{S}_C applied to M ,

$$R = \mathcal{S}_C(M). \quad (4.7.13)$$

Then it is easily verified that Cayley symplectification has the feature

$$\mathcal{S}_C(M^{-1}) = [\mathcal{S}_C(M)]^{-1}. \quad (4.7.14)$$

That is, Cayley symplectification, like symplectic polar decomposition symplectification, is invariant under inversion. See (5.13). Moreover, suppose \tilde{R} is any symplectic matrix. Then it can be shown that Cayley symplectification has the feature

$$\mathcal{S}_C(\tilde{R}M\tilde{R}^{-1}) = \tilde{R}[\mathcal{S}_C(M)]\tilde{R}^{-1}. \quad (4.7.15)$$

We may say that Cayley symplectification is invariant under symplectic similarity transformation. This property, although weaker than and a special case of the symplectic translational invariance described by (5.6), is still significant.

4.7.3 Cayley Symplectification Near the Identity

Cayley symplectification is particularly useful near the identity. Consider the problem of evaluating $\exp(\epsilon JS)$ where ϵ is small and S is symmetric and may itself have the form of a power series in ϵ beginning with constant terms. As discussed at the beginning of this chapter, such is the problem in evaluating linear transformations of the form $\exp(: k_2 :)$ where k_2 arises solely from nonlinear feed-down effects. See Chapter 9. According to (7.6) we have the result

$$R = \exp(\epsilon JS) = [I + \tanh(\epsilon JS/2)][I - \tanh(\epsilon JS/2)]^{-1}. \quad (4.7.16)$$

The hyperbolic tangent function has the Taylor expansion

$$\begin{aligned} \tanh(\epsilon JS/2) &= \sum_{\ell=1}^{\infty} a_{\ell} (\epsilon JS/2)^{\ell} = (\epsilon JS/2) - (1/3)(\epsilon JS/2)^3 + (2/15)(\epsilon JS/2)^5 \\ &\quad - (17/315)(\epsilon JS/2)^7 + (62/2835)(\epsilon JS/2)^9 - \dots . \end{aligned} \quad (4.7.17)$$

Note that the coefficients a_{ℓ} vanish for even ℓ .⁶ Suppose we *truncate* the series (7.17) by omitting terms beyond $\ell = k$, and use this truncated series to define a matrix W_t by the relation

$$W_t = J^{-1} \sum_{\ell=1}^k a_{\ell} (\epsilon JS/2)^{\ell}. \quad (4.7.18)$$

Let us use W_t to define the matrix R_a , which will be an *approximation* to the matrix R , by the equation

$$R_a = (I + JW_t)(I - JW_t)^{-1} = (I - JW_t)^{-1}(I + JW_t). \quad (4.7.19)$$

It is easily verified that W_t is a symmetric matrix, and hence R_a will be symplectic. Moreover, R_a will be near to R in the sense of satisfying relations of the form

$$\| R - R_a \| \sim \epsilon^{k+2}, \quad (4.7.20)$$

$$\| R(R_a)^{-1} - I \| \sim \epsilon^{k+2}. \quad (4.7.21)$$

We conclude that the use of (7.18) and (7.19) is well suited to the calculation of $\exp(\epsilon JS)$ where S , although symmetric, is only known through some power in some smallness parameter ϵ . Correspondingly, in the language of and as will be needed for Chapter 9, this method is well suited to the calculation of $\exp(: k_2 :)$ when k_2 itself is only known through some power in some smallness parameter ϵ .

Exercises

4.7.1. Show that the two factors in (7.8) commute as indicated. Show the same for the two factors in (7.9).

⁶It is tempting to regard (7.16) through (7.18) as a diagonal Padé approximate to the exponential function. However, it is not. For example, the 3,3 diagonal Padé approximate (approximation through cubic terms in the numerator and denominator) for the exponential function has different coefficients. In particular, it contains both even and odd powers: $\exp(z) \simeq (1 + z/2 + z^2/10 + z^3/120)/(1 - z/2 + z^2/10 - z^3/120)$.

4.7.2. Verify the invariance properties (7.14) and (7.15).

4.7.3. Show that W_t as given by (7.18) is symmetric.

4.7.4. Verify the estimates (7.20) and (7.21).

4.8 Generating Function Symplectification

It is well known that canonical transformations (symplectic maps as defined in Section 6.1) can be produced by the method of mixed-variable generating functions, often referred to as F_1 through F_4 . The generating functions are called *mixed* because they involve both “old” and “new” variables. In this section we will outline how quadratic mixed-variable generating functions can be used to symplectify matrices. See section 6.5 for a more extensive discussion of the mixed-variable generating functions F_1 through F_4 .⁷

Since the method of generating functions does not treat coordinate and momentum variables on a common footing, it is convenient to introduce the notation

$$z = (q_1 \cdots q_n, p_1 \cdots p_n), \quad (4.8.1)$$

$$Z = (Q_1 \cdots Q_n, P_1 \cdots P_n). \quad (4.8.2)$$

Let R be a symplectic matrix that maps z to Z according to the rule

$$Z = Rz. \quad (4.8.3)$$

Then, under certain conditions, the transformation (8.3) can be produced by a mixed-variable generating function.

For example, let us attempt to use a generating function of the second kind, $F_2(q, P)$. Its use gives the implicit equations

$$p_\ell = \partial F_2 / \partial q_\ell, \quad (4.8.4)$$

$$Q_\ell = \partial F_2 / \partial P_\ell. \quad (4.8.5)$$

In view of (8.4) and (8.5), and since the relation (8.3) is linear, we will consider a quadratic generating function. The most general such function (of the second kind) can be written in the form

$$F_2(q, P) = (1/2) \sum_{i,j} \alpha_{ij} q_i q_j + \sum_{i,j} \beta_{ij} q_i P_j + (1/2) \sum_{i,j} \delta_{ij} P_i P_j, \quad (4.8.6)$$

where the matrices α and δ are symmetric,

$$\alpha^T = \alpha, \quad (4.8.7)$$

⁷We remark that the adjective *generating* often occurs in an “infinitesimal” context” in the sense that one says that Lie algebras generate Lie groups or Hamiltonians generate symplectic maps. That is, generation involves some sort of “exponentiation/integration” process. By contrast, in the case of mixed-variable generating functions, results are immediate with no need to pass from the infinitesimal to the finite. Still, there is no free lunch. The complexity of exponentiation/integration is replaced by the complexity of making initially implicit relations explicit.

$$\delta^T = \delta, \quad (4.8.8)$$

and the matrix β is arbitrary. (Soon, however, we will require that β be invertible. Also, here the matrix δ is not to be confused with the Kronecker delta.)

Applying the rules (8.4) and (8.5) to this F_2 gives the set of implicit equations

$$p = \alpha q + \beta P, \quad (4.8.9)$$

$$Q = \beta^T q + \delta P. \quad (4.8.10)$$

These equations may be made explicit to give the relations

$$P = -\beta^{-1} \alpha q + \beta^{-1} p, \quad (4.8.11)$$

$$Q = (\beta^T - \delta \beta^{-1} \alpha) q + \delta \beta^{-1} p. \quad (4.8.12)$$

(Here we have assumed that β is invertible.) Suppose R is written in the $n \times n$ block form

$$R = \begin{pmatrix} A & B \\ C & D \end{pmatrix}. \quad (4.8.13)$$

Then comparison of (8.3) with (8.11) and (8.12) gives the relations

$$A = \beta^T - \delta \beta^{-1} \alpha, \quad (4.8.14)$$

$$B = \delta \beta^{-1}, \quad (4.8.15)$$

$$C = -\beta^{-1} \alpha, \quad (4.8.16)$$

$$D = \beta^{-1}. \quad (4.8.17)$$

These relations may be solved for the matrices α , β , and δ to give the results

$$\alpha = -D^{-1} C, \quad (4.8.18)$$

$$\beta = D^{-1}, \quad (4.8.19)$$

$$\delta = B D^{-1}. \quad (4.8.20)$$

We conclude that a necessary condition for (8.3) to be produced by an $F_2(q, P)$ is that the matrix D be invertible. Moreover, it is easily checked that the matrices A through D given by (8.14) through (8.17) satisfy the symplectic conditions (3.3.3) through (3.3.5). Consequently, both the necessary and sufficient condition for the linear symplectic transformation (8.3) to be produced by the F_2 defined in (8.6) is that the D matrix associated with R be invertible.

We momentarily interrupt our discussion to observe for future use that the relations (8.13) through (8.17), which relate R to the matrices α , β , and δ , can be written in a more compact form. Let W be the *symmetric* matrix defined by the equation

$$W = \begin{pmatrix} \alpha & \beta \\ \beta^T & \delta \end{pmatrix}, \quad (4.8.21)$$

and define matrices E through H by the rules

$$E = \begin{pmatrix} 0 & 0 \\ I & 0 \end{pmatrix}, \quad (4.8.22)$$

$$F = \begin{pmatrix} 0 & I \\ 0 & 0 \end{pmatrix}, \quad (4.8.23)$$

$$G = \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}, \quad (4.8.24)$$

$$H = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}. \quad (4.8.25)$$

Here, as with R , all blocks in W and in the matrices E through H are $n \times n$. With these definitions, it can be verified that R can be written in terms of W in the compact form

$$R = (FW + G)(EW + H)^{-1}. \quad (4.8.26)$$

Equation (8.26) is an example of symplectic and symmetric matrices being related by a Möbius transformation. See Section 5.13 for further discussion of this topic.

To continue our discussion of symplectification, suppose M is an arbitrary $2n \times 2n$ matrix written in the $n \times n$ block form

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \quad (4.8.27)$$

Let us seek to symplectify M . First we make the restriction that d is invertible, and define a matrix β by the rule

$$\beta = d^{-1}. \quad (4.8.28)$$

Next, following (8.18), we form the matrix $(-d^{-1}c)$ and define a matrix α by taking its symmetric part,

$$\alpha = -[(d^{-1}c) + (d^{-1}c)^T]/2. \quad (4.8.29)$$

Also, following (8.20), we form the matrix (bd^{-1}) and define a matrix δ by taking its symmetric part,

$$\delta = [(bd^{-1}) + (bd^{-1})^T]/2. \quad (4.8.30)$$

Finally, from the α , β , and δ matrices just defined, we construct the matrices A through D given by (8.14) through (8.17). In so doing we have constructed a *symplectic* matrix R of the form (8.13), and this matrix may be taken to be a symplectification of M .

There are also the generating functions $F_1(q, Q)$, $F_3(p, Q)$, and $F_4(p, P)$. They too can be used for symplectification in ways analogous to that described for F_2 . We close this section by noting that there are nearly symplectic matrices that cannot be symplectified by using any of the generating functions F_1 through F_4 . For example the matrix R given by

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \quad (4.8.31)$$

is symplectic, but cannot be produced by any of the generating functions $F_1(q, Q)$ through $F_4(p, P)$. Correspondingly, there are nonsymplectic matrices M near R that cannot be symplectified by use of the generating functions $F_1(q, Q)$ through $F_4(p, P)$. However, there are other mixed-variable generating functions that can be used. See Section 6.7.4.

Exercises

4.8.1. Verify the relations (8.9) through (8.20). Show that the symplectic conditions (3.3.3) through (3.3.5) are satisfied.

4.8.2. Verify (8.26). Hint: Along the way you will have to verify the relation

$$(EW + H)^{-1} = \begin{pmatrix} I & 0 \\ -\beta^{-1}\alpha & \beta^{-1} \end{pmatrix}. \quad (4.8.32)$$

4.8.3. Referring to (8.27) through (8.30), work out explicitly the relations giving the matrices A through D in terms of the matrices a through d . Show that R coincides with M if M is symplectic, and is near M if M is nearly symplectic.

4.8.4. Verify that the matrix R given by (8.31) is symplectic. Verify that the matrices A through D that compose R as in (8.13) have the properties

$$\det(A) = \det(B) = \det(C) = \det(D) = 0, \quad (4.8.33)$$

and therefore fail to have inverses.

4.8.5. Show that orthogonal matrices, matrices satisfying (6.1), have the property (6.2).

Bibliography

Taylor Series with Remainder

- [1] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, Dover (1972).
Also available on the Web by Googling “abramowitz and stegun 1972”.
- [2] F. Olver, D. Lozier, R. Boisvert, and C. Clark, Editors, *NIST Handbook of Mathematical Functions*, Cambridge (2010). See also the Web site <http://dlmf.nist.gov/>.

Matrix Exponentiation

- [3] C. Moler and C. Van Loan, “Nineteen Dubious Ways to Compute the Exponential of a Matrix”, *SIAM Review* **20**, 801 (1978); “Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later”, *SIAM Review* **45**, 3 (2003).
- [4] C. Kenney and A. Laub, “A Schur-Fréchet Algorithm for Computing the Logarithm and Exponential of a Matrix”, to appear in *SIAM Journal of Matrix Analysis and Applications* (1998).

- [5] N. Higham, *Functions of Matrices: Theory and Computation*, SIAM (2008).

- [6] E. Celledoni and A. Iserles, “Methods for the approximation of the matrix exponential in a Lie-algebraic setting”, arXiv:math/9904122v1 (2008).

(Orthogonal) Polar Decomposition

- [7] F.R. Gantmacher, *The Theory of Matrices*, Vols. One and Two, Chelsea (1959).
- [8] N. Jacobson, *Lectures in Abstract Algebra, Vol. II – Linear Algebra*, p. 188, D. Van Nostrand (Princeton, 1953).
- [9] J.B. Keller, “Closest Unitary, Orthogonal and Hermitian Operators to a Given Operator”, *Mathematics Magazine* **48**, p. 192 (1975).
- [10] N. J. Higham, D. S. Mackey, N. Mackey, and F. Tisseur, “Computing the Polar Decomposition and the Matrix Sign Decomposition in Matrix Groups”, *SIAM J. of Matrix Analysis and Appl.*, v.25, No. 4, pp. 1178 - 1192 (2004).
- [11] N. Higham, *Functions of Matrices: Theory and Computation*, SIAM (2008).

Analyticity

- [12] The treatment of analyticity in part *d* of Exercise 2.2 was suggested by R. Goodman.
Symplectification by Iteration
- [13] M. Furman, “Simple Method to Symplectify Matrices”, SSC Central Design Group Report SSC-TM-4001 (1985).

Chapter 5

Preliminary Lie Concepts for Classical Mechanics and Related Delights

In this chapter we will begin a study of the Lie algebraic structure of Classical Mechanics. We will learn about Lie operators and Lie transformations, and how they can be used to represent the symplectic group. We will see that the symplectic group is related to the quaternion field just as the orthogonal group and the unitary group are related to the real and complex fields. We will also find that there is a close connection between symplectic and symmetric matrices. Along the way we will learn something about Cartan's method for understanding the nature of simple Lie algebras and their representations, Clebsch-Gordan series, the topology of $Sp(2n)$, Siegel and homogeneous spaces, Möbius transformations, and Lagrangian planes.

5.1 Properties of the Poisson Bracket

The Poisson bracket has already been defined in Section 1.7. The purpose of this section is to review its properties. Suppose that f and g are any two functions of the variables q, p, t . We recall that the *Poisson bracket* of f and g , denoted by the symbol $[f, g]$, is defined by the equation

$$[f, g] = \sum_i [(\partial f / \partial q_i)(\partial g / \partial p_i) - (\partial f / \partial p_i)(\partial g / \partial q_i)]. \quad (5.1.1)$$

We also recall from Section 1.7 that it is convenient to introduce the $2n$ variables $(z_1 \dots z_{2n})$ by the rule

$$z = (z_1 \dots z_n, z_{n+1} \dots z_{2n}) = (q_1 \dots q_n, p_1 \dots p_n). \quad (5.1.2)$$

When this is done, the Poisson bracket in terms of the variables z, t can be written more compactly in the forms

$$[f, g] = \sum_{a,b} (\partial f / \partial z_a) J_{ab} (\partial g / \partial z_b), \quad (5.1.3)$$

$$[f, g] = (\partial_z f, J \partial_z g). \quad (5.1.4)$$

Here J is the fundamental $2n \times 2n$ matrix given by (1.7.11) or (3.1.1), and used in defining symplectic matrices. Note that the Poisson bracket symbol $[,]$ is the same as that used

earlier for a commutator. This is somewhat awkward, but unfortunately there are not always enough convenient symbols to go around.

We also saw in Section 1.7 that the Poisson bracket has several obvious properties. These are again listed below along with one less obvious property, the Jacobi identity. You are instructed to verify it in Exercise 1.3.

1. Distributive property

$$[(af + bg), h] = a[f, h] + b[g, h] \quad (5.1.5)$$

for arbitrary constants a, b .

2. Antisymmetry condition,

$$[f, g] = -[g, f]. \quad (5.1.6)$$

3. Derivation with respect to ordinary multiplication,

$$[f, gh] = [f, g]h + g[f, h]. \quad (5.1.7)$$

4. Jacobi identity,

$$[f, [g, h]] + [g, [h, f]] + [h, [f, g]] = 0. \quad (5.1.8)$$

Now the stage is set for a subtle conclusion. Observe that the set of all functions of the variables q, p, t or z, t forms a *linear vector space*. That is, any linear combination of two such functions is again such a function. Thus, we have the first ingredient for a Lie algebraic structure. Now define the Lie product of any two functions to be the Poisson bracket (1.1). Equations (1.5) and (1.6) show that conditions 1 through 4 for a Lie algebra are satisfied. See Section 3.7. And (1.8) shows that condition 5 is satisfied. Consequently, the set of functions of the variables q, p, t or z, t forms a Lie algebra! This Lie algebra will be called the Poisson bracket Lie algebra of dynamical variables. It is evidently infinite dimensional since the set of all functions on phase space is infinite dimensional.

Exercises

5.1.1. If you have not already done so, work out Exercises 1.7.1 through 1.7.4.

5.1.2. Determine the dimensionality of the Poisson bracket Lie algebra of dynamical variables.

Answer: The set of functions of q, p, t or z, t is an infinite dimensional vector space.

5.1.3. Verify the Jacobi identity (1.8). Hint: Use the relation (1.3).

5.1.4. Verify the relation

$$[f, g] = - \sum_{a,b} [f, z_a][z_a, z_b][z_b, g]. \quad (5.1.9)$$

5.2 Equations, Constants, and Integrals of Motion

It has already been shown in Section 1.7 that any dynamical variable $f(z, t)$ of a dynamical system governed by a Hamiltonian H obeys the equation of motion

$$df/dt = \partial f/\partial t + [f, H]. \quad (5.2.1)$$

A special case of this relation is the fact that the dynamical variables z_a obey the equations of motion

$$\dot{z}_a = (J\partial_z H)_a, \quad (5.2.2)$$

or, in more compact vector notation,

$$\dot{z} = J\partial_z H. \quad (5.2.3)$$

A dynamical variable f is called a *constant of motion* if its total time derivative vanishes. In view of (2.1), a constant of motion satisfies the equation

$$\partial f/\partial t + [f, H] = 0. \quad (5.2.4)$$

It can be shown in general that any Hamiltonian dynamical system with n degrees of freedom has $2n$ functionally independent constants of motion. See Exercise 2.4.

Suppose that a constant of motion f does not explicitly depend on the time t ,

$$\partial f/\partial t = 0. \quad (5.2.5)$$

A constant of motion that does not explicitly depend on the time will be called an *integral of motion*. By definition, an integral of motion is a constant of motion, but a constant of motion is not an integral of motion if it has explicit time dependence. Evidently, an integral of motion obeys the equation

$$[f, H] = 0. \quad (5.2.6)$$

The question of the existence of integrals of motion is quite complicated. Observe that if $f(z)$ is an integral of motion, then any given trajectory must remain for all time on a general hypersurface in phase space defined by an equation of the form

$$f(z) = \text{constant}. \quad (5.2.7)$$

If there are several functionally independent integrals of motion, then the general trajectory is further restricted to lie in the intersection of several hypersurfaces for all time. Thus, the greater the number of integrals, the more that can be said about the behavior of a dynamical system.

Consider a time-independent Hamiltonian $H(z)$. A point z^c in phase space for which the vector $\partial_z H$ is zero is called a *critical point*. Evidently, according to (2.2) or (2.3), a critical point is some kind of equilibrium point. Now suppose some small region R of phase space contains *no* critical points. Then it can be shown that, provided R is small enough, the dynamical system described by the Hamiltonian $H(z)$ has $2n - 1$ functionally independent integrals of motion in the region R . Furthermore, n of these integrals can be arranged to be

in *involution*. (Two functions f and g are said to be in involution if their Poisson bracket $[f, g]$ is zero.)¹ See Exercises 2.5 and 2.6.

The result just stated is of limited use unless all trajectories starting in R happen to remain in R . In general, and contrary to the impression given by most textbooks, most dynamical Hamiltonian systems do not have global integrals of motion. If a time-independent Hamiltonian dynamical system with n degrees of freedom has n functionally independent global integrals of motion in involution, the system is said to be *completely integrable*. In general, only the soluble problems found in textbooks fall into this category. Most Hamiltonian dynamical systems, including the majority encountered in real life (e.g. Celestial Mechanics, Accelerator Physics, Penning Traps, Mirror Machines, etc.), are not completely integrable and are therefore sufficiently complicated to be in some sense insoluble. In particular, the behavior of most Hamiltonian systems is sufficiently complicated that the trajectories are not generally confined to lie on hypersurfaces in phase space. See the references on *Non-Integrability* listed at the end of this chapter.

Exercises

5.2.1. Verify (2.2).

5.2.2. Suppose that the Hamiltonian H for a dynamical system does not depend explicitly on the time t . Show that then H is an integral of motion.

5.2.3. Suppose that the dynamical variables f and g are constants of motion. Verify *Poisson's theorem*, which states that the quantity $[f, g]$ is then also a constant of motion. Suppose that f and g are integrals of motion. Show that $[f, g]$ is then also an integral of motion.
Hint: Use the Jacobi identity.

5.2.4. Suppose that f_1, f_2, \dots, f_n are n constants of motion. Let c be any function of the f_j ,

$$c = c(f_1, f_2, \dots, f_n). \quad (5.2.8)$$

Show that c is then also a constant of motion. Suppose that f_1, f_2, \dots, f_n are n integrals of motion, and that c is again defined as above. Show that c is then also an integral of motion.

5.2.5. Let t^i denote some *initial* time. Given t and $z(t)$, we can always integrate the equations of motion backward (or forward) in time to the time t^i to find the initial conditions z^i . The result of this process will generally depend on t and $z(t)$. Thus, we obtain $2n$ functions $z_a^i(z, t)$. Show that these functions are functionally independent and are constants of motion. Carry out this construction explicitly for the case of the one-dimensional simple harmonic oscillator.

5.2.6. Problem on constructing local integrals of motion (Hamiltonian flow-box or straightening-out theorem). ▀

¹It is a confusing fact that the term *involution* has multiple meanings. It can also refer to a map or operator whose square is the identity. See, for example, Exercise 3.12.5.

5.3 Lie Operators

Let $f(z, t)$ be any function of the phase-space variables z and perhaps the time t . Associated with each f is a *Lie operator* that we denote by the symbol $: f :$. The Lie operator $: f :$ is a *differential* operator defined by the rule

$$: f : \stackrel{\text{def}}{=} \sum_i (\partial f / \partial q_i) (\partial / \partial p_i) - (\partial f / \partial p_i) (\partial / \partial q_i). \quad (5.3.1)$$

In particular, if $: f :$ acts on any phase-space function g , one finds the result

$$: f : g = \sum_i (\partial f / \partial q_i) (\partial g / \partial p_i) - (\partial f / \partial p_i) (\partial g / \partial q_i) = [f, g]. \quad (5.3.2)$$

Thus, one may heuristically view a Lie operator as a Poisson bracket waiting to happen. Note that in view of (1.3), the defining relation (3.1) can also be written in the form

$$: f := \sum_{a,b} (\partial f / \partial z_a) J_{ab} (\partial / \partial z_b). \quad (5.3.3)$$

We also remark that in the Mathematics literature the Lie operator $: f :$ is sometimes referred to as *ad*(f) where *ad* is shorthand for *adjoint*. Note the similarity of the relations (3.7.71) and (3.2). See also the discussion in Section 8.1. We use the notation $: f :$ instead of *ad*(f) because it facilitates the writing of complicated expressions.

Powers of $: f :$ can be defined by repeated application, which amounts to taking repeated Poisson brackets. For example, $: f :^2$ is defined by the relation

$$: f :^2 g =: f :: f : g =: f : [f, g] = [f, [f, g]]. \quad (5.3.4)$$

Finally, $: f :$ to the zero power is defined to be the identity operator,

$$: f :^0 = \mathcal{I} \Leftrightarrow : f :^0 g = g. \quad (5.3.5)$$

We note that Lie operators, as well as their powers, are linear operators because of (1.5) and (1.6)

As result of (1.5), the sum of two Lie operators is again a Lie operator. Specifically, one finds the relation

$$a : f : + b : g :=: (af + bg) : \quad (5.3.6)$$

for any two scalars a, b and any two functions f, g . Therefore, the set of Lie operators forms a linear vector space.

A Lie operator is also a *derivation* with respect to the operation of ordinary multiplication. That is, a Lie operator satisfies the product rule analogous to that for differentiation: Let g and h be any two functions. Then, according to (1.8), $: f :$ obeys the rule

$$: f : (gh) = (: f : g)h + g(: f : h). \quad (5.3.7)$$

In addition to being a derivation with respect to ordinary multiplication, a Lie operator is also a derivation with respect to Poisson bracket multiplication. Suppose g and h are any two functions. Then the Jacobi identity (1.8) can be written in the form

$$[f, [g, h]] = [[f, g], h] + [g, [f, h]]. \quad (5.3.8)$$

or equivalently, using Lie operator notation,

$$:f:[g,h] = [:f:g,h] + [g,:f:h]. \quad (5.3.9)$$

Since the set of Lie operators forms a linear vector space, it is of interest to inquire whether the vector space can be given a multiplication rule that will convert it into a Lie algebra. The answer is yes, as is nearly obvious, since Lie operators are linear operators and linear operators are quite similar to matrices. The Lie product of two Lie operators $:f:$ and $:g:$ is simply taken to be their commutator. Denoting the Lie product of two Lie operators by the symbol $\{ :f: , :g: \}$, the Lie product is defined by the rule

$$\{ :f: , :g: \} = :f::g: - :g::f:. \quad (5.3.10)$$

See Exercise (3.5). Note that there are now two Lie algebras that have to be kept in mind. First, there is the Lie algebra of functions of z, t with the Lie product defined to be the Poisson bracket. Second, there is the Lie algebra of Lie operators with the Lie product defined to be the commutator.

One point, however, has been overlooked. Namely, is the right side of (3.10) a Lie operator? To answer this question, it is useful to view the Jacobi identity (1.8) for Poisson brackets from yet another perspective. For any function h , the Jacobi identity can be written in the form

$$[f, [g, h]] - [g, [f, h]] = [[f, g], h]. \quad (5.3.11)$$

However, using Lie operator notation, this same equation can be written in the form

$$:f::g:h - :g::f:h =: [f, g] : h, \quad (5.3.12)$$

or more compactly, using (3.10),

$$\{ :f: , :g: \} h =: [f, g] : h. \quad (5.3.13)$$

But, since h is an arbitrary function, (3.13) can also be viewed as the operator identity

$$\{ :f: , :g: \} =: [f, g] :. \quad (5.3.14)$$

Evidently, the commutator of two Lie operators $:f:$ and $:g:$ is again a Lie operator, and is in fact the Lie operator associated with the function $[f, g]$.

Put another way, (3.14) shows that there is a close connection between the Lie algebra of functions and the Lie algebra of Lie operators. Specifically, the Lie product (commutator) of two Lie operators is the Lie operator of the Lie product (Poisson bracket) of the two associated functions. Mathematicians have a word for such a situation. They would say that the two Lie algebras are *homomorphic*. To see that this relation between the two Lie algebras is a homomorphism and not an isomorphism, suppose two Lie operators $:f:$ and $:g:$ are equal,

$$:f: =: g: . \quad (5.3.15)$$

Then from (3.15) we can only deduce the relation

$$f = g + c, \quad (5.3.16)$$

where c is an arbitrary constant. That is, as is obvious from the definition (3.1), the Lie operator associated with any constant is identically zero.

We close this subsection by noting that what we have called a Lie operator is actually a special case of a more general object. Let x denote a collection of N variables x_1, x_2, \dots, x_N . Also, let $\mathbf{g} = (g_1, g_2, \dots, g_N)$ be a collection of N functions of x and perhaps the time t . The Lie operator $\mathcal{L}_{\mathbf{g}}$ associated with the collection of functions $g_b(x, t)$ is defined to be the differential operator given by the rule

$$\mathcal{L}_{\mathbf{g}} = \sum_{b=1}^N g_b(x, t) (\partial/\partial x_b). \quad (5.3.17)$$

The relation (3.17) is the general definition of a Lie operator. It is also sometimes called a *vector field*. With the introduction of the notation $\boldsymbol{\partial} = (\partial/\partial x_1, \partial/\partial x_2, \dots, \partial/\partial x_N)$, it is often convenient to write $\mathcal{L}_{\mathbf{g}}$ in the suggestive form

$$\mathcal{L}_{\mathbf{g}} = \mathbf{g} \cdot \boldsymbol{\partial}. \quad (5.3.18)$$

There is an intimate connection between vector fields and ordinary differential equations. Consider the set of first-order differential equations

$$\dot{x}_a = g_a(x, t). \quad (5.3.19)$$

Then, using (3.17), this set can also be written in the form

$$\dot{x}_a = \mathcal{L}_{\mathbf{g}} x_a. \quad (5.3.20)$$

Also, let h be any function of x and perhaps the time t . Then, by the chain rule, the time derivative of h along a trajectory is given by the relation

$$dh/dt = \partial h/\partial t + \sum_b (\partial h/\partial x_b) \dot{x}_b = \partial h/\partial t + \sum_b g_b (\partial h/\partial x_b) = \partial h/\partial t + \mathcal{L}_{\mathbf{g}} h. \quad (5.3.21)$$

Upon comparison of (3.17) with (3.3), we see that we have assumed $N = 2n$ and

$$g_b(z, t) = \sum_a (\partial f/\partial z_a) J_{ab}. \quad (5.3.22)$$

For future reference we note that (3.22) can also be written in the form

$$g_b(z, t) = [f, z_b] = [z_b, (-f)]. \quad (5.3.23)$$

We conclude that, in the case of interest for Hamiltonian systems, the collection of functions g_b arises from a *single* function f according to the relation (3.22). Thus, to be more precise, what we have called and will continue to call a Lie operator could better be called a *Hamiltonian* Lie operator or a *Hamiltonian* vector field. Non-Hamiltonian vector fields are of use for describing dissipative effects including, in the field of accelerator physics, synchrotron radiation effects and electron and ionization cooling. Our primary attention will be focused on Hamiltonian Lie operators. However, where applicable, we will also present

results for general Lie operators. General polynomial vector fields, both Hamiltonian and non-Hamiltonian, are treated and classified in Chapter 27.

Finally, in the case $N = 2n$, define quantities η_c by the rule

$$\eta_c = \sum_b J_{bc} g_b. \quad (5.3.24)$$

If the g_b arise from a single function f as in (3.22), we find the result

$$\begin{aligned} \eta_c &= \sum_{ab} J_{bc} J_{ab} (\partial f / \partial z_a) = \sum_{ab} J_{ab} J_{bc} (\partial f / \partial z_a) \\ &= \sum_a (J^2)_{ac} (\partial f / \partial z_a) = -\partial f / \partial z_c. \end{aligned} \quad (5.3.25)$$

It follows from (3.25) that the collection of functions η_c then has the property

$$\partial \eta_c / \partial z_d - \partial \eta_d / \partial z_c = -\partial^2 f / \partial z_d \partial z_c + \partial^2 f / \partial z_c \partial z_d = 0. \quad (5.3.26)$$

Evidently (3.26) is a necessary condition for a vector field to be Hamiltonian. In Section 6.4 we will see that it is also sufficient.

It is easily verified that the set of all vector fields in N variables forms a Lie algebra with the commutator taken as the Lie product (see Exercise 3.8), and (in the even dimensional case) the set of Hamiltonian vector fields forms a Lie subalgebra of the Lie algebra of all vector fields. In subsequent chapters we will learn that the set of all vector fields is the Lie algebra of the group of all diffeomorphisms, and the set of all Hamiltonian vector fields is the Lie algebra of the subgroup of all symplectic maps.

Exercises

5.3.1. Starting from (3.7), show that $:f:n$ obeys the *Leibniz* rule

$$:f:n(gh) = \sum_{m=0}^n \binom{n}{m} (:f:m g)(:f:n-m h), \quad (5.3.27)$$

where $\binom{n}{m}$ is the binomial coefficient defined by

$$\binom{n}{m} = \frac{n!}{(m!)(n-m)!}. \quad (5.3.28)$$

Suggestion: Use induction and the relations $\binom{n}{n} = 1$, $\binom{n}{0} = 1$, and $\binom{n}{m-1} + \binom{n}{m} = \binom{n+1}{m}$.

According to some, perhaps apocryphal, lore it took Leibniz seven years to discover this rule (in the context of how to differentiate a product of two functions). He and others first assumed that the derivative of a product would be the product of the derivatives (which, in fact, is the case for the chain rule that applies to functions of functions).

5.3.2. Verify (3.8) and (3.9).

5.3.3. State and verify the analog of the Leibniz rule of Exercise (3.1) for the case of $f : f :^n [g, h]$.

5.3.4. Verify that the Lie product defined by (3.10) satisfies the properties 1 through 5 required to make the set of Lie operators into a Lie algebra. See Section 3.7.

5.3.5. Verify (3.11), (3.12), and (3.13).

5.3.6. Let h_0 be any constant function. Verify that

$$: h_0 := 0. \quad (5.3.29)$$

5.3.7. Let G be any function of the variables z . A set of differential equations of the form

$$\dot{z}_a = -\partial G / \partial z_a \quad (5.3.30)$$

is called a *gradient* system, and the corresponding vector field

$$\mathcal{L}_G = - \sum_a (\partial G / \partial z_a) (\partial / \partial z_a) \quad (5.3.31)$$

is called a gradient vector field. At this point it is interesting to contrast Hamiltonian systems, see (2.2), with gradient systems. Both are derived from master functions, H and G , respectively. But their behavior can be very different. Verify the relation

$$dG/dt = \mathcal{L}_G G = - \sum_a (\partial G / \partial z_a)^2 \leq 0. \quad (5.3.32)$$

It follows that, for a gradient system, points on a trajectory move away from maxima of G and toward minima of G . Compare the behavior of Hamiltonian and gradient systems near and at local extrema of H and G , respectively. What happens at and near saddle points?

5.3.8. Suppose \mathcal{L}_f and \mathcal{L}_g are any two vector fields. Show that their commutator is also a vector field. That is, given f and g , show that there is a relation of the form

$$\{\mathcal{L}_f, \mathcal{L}_g\} = \mathcal{L}_h \quad (5.3.33)$$

and find a formula for h in terms of f and g . Show that, for vector fields, double commutators composed of three vector fields obey the Jacobi identity,

$$\{\mathcal{L}_f, \{\mathcal{L}_g, \mathcal{L}_h\}\} + \{\mathcal{L}_g, \{\mathcal{L}_h, \mathcal{L}_f\}\} + \{\mathcal{L}_h, \{\mathcal{L}_f, \mathcal{L}_g\}\} = 0. \quad (5.3.34)$$

Show that the set of all vector fields forms an infinite-dimensional Lie algebra with the commutator playing the role of a Lie product.

5.3.9. Suppose some N -dimensional Lie algebra L has structure constants $c_{\alpha\beta}^\gamma$. Consider an N -dimensional Euclidean space with coordinates x_1, x_2, \dots, x_N . Define N vector fields \mathcal{L}_α by the rule

$$\mathcal{L}_\alpha = - \sum_{\beta\gamma} c_{\alpha\beta}^\gamma x_\beta \partial/\partial x_\gamma. \quad (5.3.35)$$

Show that these vector fields satisfy the commutation relations

$$\{\mathcal{L}_\alpha, \mathcal{L}_\beta\} = \sum_\gamma c_{\alpha\beta}^\gamma \mathcal{L}_\gamma, \quad (5.3.36)$$

and therefore provide a vector-field realization of L . Since the vector fields are manufactured from the structure constants, might this realization be related to the adjoint representation of L ? Using (3.7.54), show that

$$\mathcal{L}_\alpha x_\beta = - \sum_\gamma (\hat{B}_\alpha)_{\beta\gamma} x_\gamma \quad (5.3.37)$$

or, more compactly,

$$\mathcal{L}_\alpha x = - \hat{B}_\alpha x. \quad (5.3.38)$$

Suggestion: First verify (3.38) and then (3.36).

5.3.10. Let \mathcal{L}_f be a vector field and suppose g and h are any two functions. In analogy to (3.7), prove the derivation property

$$\mathcal{L}_f(gh) = (\mathcal{L}_f g)h + g(\mathcal{L}_f h). \quad (5.3.39)$$

Find the Leibniz rule for $(\mathcal{L}_f)^n$ analogous to (3.27).

5.4 Lie Transformations

5.4.1 Definition and Some Properties

Since powers of $:f:$ have been defined, it is also possible to deal with power series in $:f:$. Of particular importance is the power series $\exp(:f:)$. This particular object is called the *Lie transformation* associated with $:f:$ or f .² The Lie transformation is also a linear operator, and is formally defined as expected by the exponential series

$$e^{:f:} = \exp(:f:) = \sum_{n=0}^{\infty} :f:^n / n!. \quad (5.4.1)$$

In particular, the action of $\exp(:f:)$ on any function g is given by the rule

$$\exp(:f:)g = g + [f, g] + [f, [f, g]]/2! + \dots \quad (5.4.2)$$

²Some authors use the terms *Lie transformation* and *Lie series* interchangeably. We prefer to refer to any power series in $:f:$ as a Lie series, and to refer to the particular power series $\exp(:f:)$ as a Lie transformation.

The fact that $:f:$ is a derivation with respect to ordinary multiplication, see (3.7), implies that the Lie transformation $\exp(:f:)$ is an *isomorphism* with respect to ordinary multiplication. (This is another remarkable property of the exponential function!) That is, suppose g and h are any two functions. Then the Lie transformation $\exp(:f:)$ has the property

$$\exp(:f:)(gh) = [\exp(:f:)g][\exp(:f:)h]. \quad (5.4.3)$$

In words, (4.3) says that one can either let a Lie transformation act on the product of two functions, or act on each function separately and then take the product of the results. Both operations give the same net result.

The relation (4.3) may be proved as follows. First, use the definition (4.1) to get the result

$$\exp(:f:)(gh) = \sum_{n=0}^{\infty} (:f:^n / n!)(gh). \quad (5.4.4)$$

Next, use the Leibniz rule (3.27), which is a consequence of the derivation property (3.7), to get the result

$$\exp(:f:)(gh) = \sum_{n=0}^{\infty} (1/n!) \sum_{m=0}^n \binom{n}{m} (:f:^m g)(:f:^{n-m} h). \quad (5.4.5)$$

The binomial coefficients obey the relation

$$(1/n!) \binom{n}{m} = 1/[(m!)(n-m)!]. \quad (5.4.6)$$

Consequently, (4.5) can also be written in the form

$$\exp(:f:)(gh) = \sum_{n=0}^{\infty} \sum_{m=0}^n \{[:f:^m / m!]g\} \{[:f:^{n-m} / (n-m)!]h\}. \quad (5.4.7)$$

Observe that the double sum on the right side of (4.7) can be rearranged to give the result

$$\exp(:f:)(gh) = \sum_{m=0}^{\infty} \sum_{n=m}^{\infty} \{[:f:^m / m!]g\} \{[:f:^{n-m} / (n-m)!]h\}. \quad (5.4.8)$$

See Figure 4.1 and Exercise 4.8. Finally, let $\ell = n - m$ be a new summation index. Then (4.8) takes the final form

$$\begin{aligned} \exp(:f:)(gh) &= \sum_{m=0}^{\infty} [:f:^m / m!]g \sum_{\ell=0}^{\infty} [:f:^\ell / \ell!]h \\ &= [\exp(:f:)g][\exp(:f:)h]. \end{aligned} \quad (5.4.9)$$

The relation (4.3) may be extended to products of Lie transformations acting on products of functions. Let a and b be any two functions, and let $\exp(:f:)$ and $\exp(:g:)$ be any two Lie transformations. Then we have, by using (4.3) repeatedly, the result

$$\begin{aligned} \exp(:f:)\exp(:g:)(ab) &= \exp(:f:)\{[\exp(:g:)a][\exp(:g:)b]\} \\ &= [\exp(:f:)\exp(:g:)a][\exp(:f:)\exp(:g:)b]. \end{aligned} \quad (5.4.10)$$

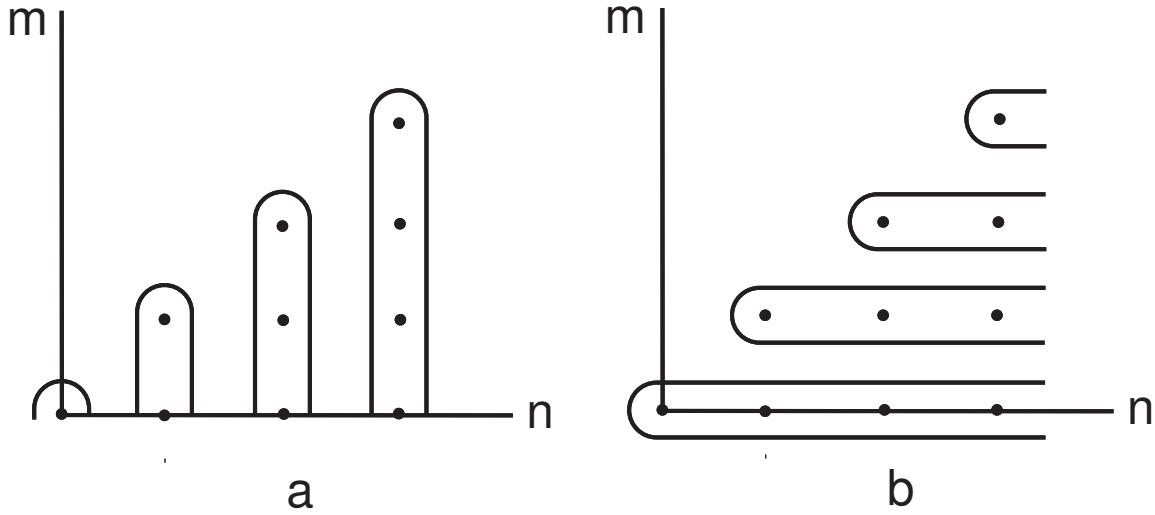


Figure 5.4.1: a) The summation points in m, n space for the sum (4.7) indicating that the inner sum is over m followed by a sum over n . b) The summation points for the sum (4.8) illustrating that the points are the same, but the inner sum is now over n followed by a sum over m .

Analogous results evidently hold for any number of Lie transformations and any number of functions.

The isomorphism property of $\exp(: f :)$ described by (4.3) often facilitates computations involving Lie transformations. Let the symbol z stand, as usual, for the collection of quantities $z_1 \dots z_{2n}$. Similarly, let the symbol $\exp(: f :)z$ stand for the collection of quantities $\exp(: f :)z_1, \dots, \exp(: f :)z_{2n}$. Now let $g(z)$ be any function. Then it follows from (4.3) that

$$\exp(: f :)g(z) = g[\exp(: f :)z]. \quad (5.4.11)$$

That is, the action of a Lie transformation on a function is to perform a Lie transformation on its arguments.

To see the truth of (4.11), suppose first that g were a polynomial in the quantities $z_1 \dots z_{2n}$. But a polynomial is just a sum of monomials of the form

$$z_1^{m_1} z_2^{m_2} \dots z_{2n}^{m_{2n}}.$$

It follows from (4.9) that

$$\exp(: f :)z_1^{m_1} z_2^{m_2} \dots z_{2n}^{m_{2n}} = [\exp(: f :)z_1]^{m_1} \dots [\exp(: f :)z_{2n}]^{m_{2n}}. \quad (5.4.12)$$

Also, as mentioned earlier, $\exp(: f :)$ is a linear operator. Therefore a Lie transformation has the advertised property (4.11) when acting on polynomials. But, according to the Weierstrass approximation theorem, the set of monomials is dense in the complete set of functions on any bounded domain. Consequently, (4.11) holds in general by continuity.

As a consequence of (4.10), there is a result analogous to (4.11) for any product of Lie transformations acting on a function. For example, in the case of two Lie transformations $\exp(: f :)$ and $\exp(: g :)$ and a function h , we have the result

$$\exp(: f :) \exp(: g :) h(z) = h[\exp(: f :) \exp(: g :) z]. \quad (5.4.13)$$

Similar results hold for any number of Lie transformations. The proof of these results is similar to that just given for (4.11).

The last observation to be made is that since $: f :$ is also a derivation with respect to Poisson bracket multiplication, the Lie transformation $\exp(: f :)$ must also be an isomorphism with respect to Poisson bracket multiplication. That is, suppose g and h are any two functions. Then the Lie transformation $\exp(: f :)$ has the property

$$\exp(: f :)[g, h] = [\exp(: f :)g, \exp(: f :)h]. \quad (5.4.14)$$

This property will be essential for subsequent discussions of symplectic maps and charged-particle beam transport. Its proof is exactly analogous to that just given for the case of ordinary multiplication. Also, there are results analogous to (4.10) for products of Lie transformations acting on Poisson brackets. Suppose, for example, that a and b are any two functions. Then we have the result

$$\exp(: f :) \exp(: g :)[a, b] = [\exp(: f :) \exp(: g :)a, \exp(: f :) \exp(: g :)b], \quad (5.4.15)$$

and similar results hold for any number of Lie transformations.

5.4.2 Applications

Subsequent chapters and sections will be devoted to the use of Lie transformations for representing, manipulating, and analyzing symplectic maps. They can also be used to transform Hamiltonians to normal form. Indeed, this was their original use as envisioned by Hori and Deprit. Similarly, they can be used to transform vector and tensor fields. See the references listed at the end of this chapter.

Exercises

5.4.1. Let q and p be the phase-space coordinates for a system having one degree of freedom. Let f be the function

$$f = -\lambda p^2/2. \quad (5.4.16)$$

Show that

$$\begin{aligned} \exp(: f :)p &= p, \\ \exp(: f :)q &= q + \lambda p. \end{aligned} \quad (5.4.17)$$

Here λ is an arbitrary parameter.

Hint: Observe that the series (4.2) terminates in this case.

5.4.2. Repeat Exercise 4.1 for the case $f = \lambda q^2/2$.

5.4.3. Repeat Exercise 4.1 for the case $f = \lambda q^3/3$.

5.4.4. Repeat Exercise 4.1 for the case $f = -\lambda pq$. Now you must sum an infinite series.

Answer:

$$\begin{aligned}\exp(:f:)q &= (e^\lambda)q, \\ \exp(:f:)p &= (e^{-\lambda})p.\end{aligned}\tag{5.4.18}$$

5.4.5. Repeat Exercise 4.1 for the case $f = -\lambda(p^2 + q^2)/2$.

Answer:

$$\begin{aligned}\exp(:f:)q &= q \cos \lambda + p \sin \lambda, \\ \exp(:f:)p &= -q \sin \lambda + p \cos \lambda.\end{aligned}\tag{5.4.19}$$

5.4.6. Repeat Exercise 4.1 for the case $f = -\lambda(p^2 - q^2)/2$.

Answer:

$$\begin{aligned}\exp(:f:)q &= q \cosh \lambda + p \sinh \lambda, \\ \exp(:f:)p &= q \sinh \lambda + p \cosh \lambda.\end{aligned}\tag{5.4.20}$$

5.4.7. Repeat Exercise 4.1 for the case $f = \lambda qp^2$.

Answer:

$$\begin{aligned}\exp(:f:)q &= q(1 - \lambda p)^2, \\ \exp(:f:)p &= p/(1 - \lambda p).\end{aligned}\tag{5.4.21}$$

See the end of Section 1.4.

5.4.8. Verify (4.3) for the case $f = \lambda q^2$ and $g = h = p$.

5.4.9. Verify the rearrangement required to go from (4.7) to (4.8). Hint: Mark out, in m, n space, the lattice of points that are summed over in (4.7). Show that the same points are summed over in (4.8). See Figure 4.1.

5.4.10. Prove (4.13).

5.4.11. Derive (4.14) from the definition (4.1) and the results of Exercise 3.3.

5.4.12. Prove (4.15).

5.4.13. Let c be any constant. Verify the result

$$\exp(:f:)c = c.\tag{5.4.22}$$

5.4.14. Let \mathcal{L}_f be a general vector field. In analogy to (4.1) define an associated Lie transformation by the rule

$$\exp(\mathcal{L}_f) = \sum_{n=0}^{\infty} (\mathcal{L}_f)^n / n!.\tag{5.4.23}$$

Show, in analogy to (4.3), that this Lie transformation is also an isomorphism with respect to function multiplication,

$$\exp(\mathcal{L}_f)(gh) = [\exp(\mathcal{L}_f)g][\exp(\mathcal{L}_f)h].\tag{5.4.24}$$

5.5 Realization of the $sp(2n, \mathbb{R})$ Lie Algebra

According to Exercise (1.2), the Poisson bracket Lie algebra of dynamical variables is infinite dimensional. The purpose of this section is to show that, for a $2n$ -dimensional phase space, the Poisson bracket Lie algebra of dynamical variables contains $sp(2n, \mathbb{R})$ as a subalgebra.

Suppose f and g are homogeneous polynomials of degree 2 in the variables z . Then, inspection of (1.3) indicates that their Poisson bracket $[f, g]$ is also a homogeneous polynomial of degree two. We conclude that second-degree polynomials form a subalgebra of the Poisson bracket Lie algebra of all functions. In fact, calculation shows that this subalgebra is a realization of $sp(2n, \mathbb{R})$.

To verify this assertion, suppose that f and g are any two homogeneous second-degree polynomials in the variables z . They can be written in the form

$$f = (1/2) \sum_{a,b} S_{ab}^f z_a z_b = (1/2)(z, S^f z), \quad (5.5.1)$$

$$g = (1/2) \sum_{c,d} S_{cd}^g z_c z_d = (1/2)(z, S^g z), \quad (5.5.2)$$

where S^f and S^g are real *symmetric* matrices. Evidently, there is a one-to-one correspondence between homogeneous second-degree polynomials and symmetric matrices. We will indicate a one-to-one correspondence by the symbol \Leftrightarrow . Since J is invertible, there is also an associated one-to-one correspondence between homogeneous second degree polynomials and matrices of the form JS . Indeed, the relations (5.1) and (5.2) can be written also in the form

$$f \Leftrightarrow JS^f \Leftrightarrow f = (1/2)(Jz, JS^f z), \quad (5.5.3)$$

$$g \Leftrightarrow JS^g \Leftrightarrow g = (1/2)(Jz, JS^g z). \quad (5.5.4)$$

Recall (3.1.6). Here the symbol \Leftrightarrow denotes logical implication in both directions.

Now use the representations (5.1) and (5.2) to compute the Poisson bracket $[f, g]$. This calculation is facilitated by the relation

$$\begin{aligned} [z_a z_b, z_c z_d] &= z_a z_c J_{bd} + z_a z_d J_{bc} + z_b z_c J_{ad} + z_b z_d J_{ac} \\ &= \sum_{e,f} (\delta_{ae} \delta_{cf} J_{bd} + \delta_{ae} \delta_{df} J_{bc} + \delta_{be} \delta_{cf} J_{ad} + \delta_{be} \delta_{df} J_{ac}) z_e z_f. \end{aligned} \quad (5.5.5)$$

[Note that in this realization the structure constants are related to the entries of J and the Kronecker delta. This result is not completely surprising because J also enters the definition of the Poisson bracket. Recall (1.3).] We find from (5.1), (5.2), and (5.5) the result

$$[f, g] = (z, S^f JS^g z) = (Jz, JS^f JS^g z). \quad (5.5.6)$$

Similarly, the Poisson bracket $[g, f]$ can be evaluated to give the result

$$[g, f] = (Jz, JS^g JS^f z). \quad (5.5.7)$$

Subtract (5.7) from (5.6) and use the antisymmetry condition (1.6). Doing so gives the result

$$[f, g] = (1/2)(Jz, \{JS^f, JS^g\}z). \quad (5.5.8)$$

Here the notation $\{, \}$ indicates the matrix commutator,

$$\{JS^f, JS^g\} = JS^f JS^g - JS^g JS^f. \quad (5.5.9)$$

Suppose the second-degree polynomial h is defined by the relation

$$h = [f, g]. \quad (5.5.10)$$

Then, comparison of (5.8) and (5.10) shows that h can be written also in the form

$$h = (1/2)(Jz, JS^h z), \quad (5.5.11)$$

where the matrix JS^h is defined by the relation

$$JS^h = \{JS^f, JS^g\}. \quad (5.5.12)$$

Observe that (5.10) is a Lie-algebraic relation in the Poisson bracket Lie algebra of second-degree polynomials, and (5.12) is a Lie-algebraic relation in $sp(2n)$. Thus we have the logical implication

$$h = [f, g] \Leftrightarrow JS^h = \{JS^f, JS^g\}. \quad (5.5.13)$$

What we have just shown is that these two Lie algebras are isomorphic under the one-to-one correspondence given by (5.3), (5.4), and (5.11).

In the next three sections we will study the problem of finding suitable bases for the Lie algebras $sp(2)$, $sp(4)$, and $sp(6)$ when special attention is given to their $u(1)$, $u(2)$, and $u(3)$ subalgebras, respectively. We close this section by finding a basis for $sp(2n)$ when special attention is given to the subgroups described in Section (3.10).

We have already studied the basis for $sp(2n)$ consisting of the monomials $z_a z_b$ and found that they satisfy the Poisson bracket rules (5.5). Another possible basis can be found by decomposing these monomials into those associated with the subgroups constituted by matrices of the form (3.3.9), (3.3.10), and (3.3.11), respectively. Consider first the subgroup associated with matrices of the form (3.3.9). In this case, S is of the form (3.10.2). Correspondingly, the polynomials f given by (5.1) are linear combinations of the monomials $p_j p_k$. They satisfy the Poisson bracket relations

$$[p_j p_k, p_\ell p_m] = 0. \quad (5.5.14)$$

The vanishing of all Lie products for elements of this subalgebra is expected since the associated subgroup is Abelian.

Consider next the subgroup associated with matrices of the form (3.3.10). In this case S is of the form (3.10.7), and the polynomials f given by (5.1) are linear combinations of the monomials $q_j q_k$. They satisfy the Poisson bracket relations

$$[q_j q_k, q_\ell q_m] = 0. \quad (5.5.15)$$

Again all Poisson brackets for this Lie subalgebra vanish since the associated subgroup is also Abelian.

Finally consider the subgroup associated with matrices of the form (3.3.11). In this case S is of the form (3.10.13). Correspondingly, the polynomials f given by (5.1) are linear combinations of the monomials $q_j p_k$. They satisfy the Poisson bracket relations

$$[q_j p_k, q_\ell p_m] = \delta_{jm} q_\ell p_k - \delta_{k\ell} q_j p_m. \quad (5.5.16)$$

Since the right side of (5.16) is again of the form $q_j p_k$, these monomials constitute a Lie subalgebra as expected. This subalgebra is the Lie algebra $g\ell(n, \mathbb{R})$, the Lie algebra of the group $GL(n, \mathbb{R})$.

It remains to compute the Poisson brackets of the monomials $p_j p_k$, $q_j q_k$, and $q_j p_k$ with each other. We find the results

$$[q_j p_k, p_\ell p_m] = \delta_{j\ell} p_k p_m + \delta_{jm} p_k p_\ell, \quad (5.5.17)$$

$$[q_j p_k, q_\ell q_m] = -\delta_{k\ell} q_j q_m - \delta_{km} q_j q_\ell, \quad (5.5.18)$$

$$[q_j q_k, p_\ell p_m] = \delta_{j\ell} q_k p_m + \delta_{jm} q_k p_\ell + \delta_{k\ell} q_j p_m + \delta_{km} q_j p_\ell. \quad (5.5.19)$$

Note that (5.17) indicates that the Lie algebra formed by the monomials $p_\ell p_m$ is transformed under the action of the Lie algebra formed by the monomials $q_j p_k$. Also, (5.16) and (5.17) together indicate that the set of monomials $q_j p_k$ and $p_\ell p_m$, when combined in linear combinations, still form a Lie subalgebra. This is the subalgebra associated with the subgroup of matrices of the form (3.10.16). The fact that the monomials $p_\ell p_m$ transform under the action of the monomials $q_j p_k$ is a consequence of the fact that the subgroup of matrices (3.10.16) is a *semidirect* product of the subgroups of matrices (3.3.11) and (3.3.9). Similarly, the relations (5.16) and (5.18) indicate that the monomials $q_j p_k$ and $q_\ell q_m$ span a Lie subalgebra associated with the subgroup of matrices of the form (3.10.19), and this subgroup is a semidirect product of the subgroups of matrices (3.3.11) and (3.3.10).

Exercises

5.5.1. Verify (5.5).

5.5.2. Verify (5.6), (5.7), and (5.8).

5.5.3. Verify the following Poisson bracket relation:

$$[z_a z_b, z_c] = z_a J_{bc} + z_b J_{ac}. \quad (5.5.20)$$

Suppose f is given by (5.1). Show that the matrix JS^f can be computed from f by the relation

$$: f : z_c = [f, z_c] = -(JS^f z)_c = - \sum_d (JS^f)_{cd} z_d. \quad (5.5.21)$$

5.5.4. Find the dimensions of the three Lie subalgebras spanned by the monomials of the form $p_j p_k$, monomials of the form $q_j q_k$, and monomials of the form $q_j p_k$, respectively.

5.5.5. Find the dimension of the Lie subalgebra spanned by the monomials of the form $q_j p_k$ plus monomials of the form $p_\ell p_m$. Find the dimension of the Lie subalgebra spanned by the monomials of the form $q_j p_k$ plus monomials of the form $q_\ell q_m$.

5.5.6. Show that the monomials $q_j q_k$ and $p_\ell p_m$ generate the full Lie algebra of $sp(2n)$ in the sense that taking suitable Poisson brackets of them produces all possible monomials $z_a z_b$.

5.6 Basis for $sp(2, \mathbb{R})$

The symplectic Lie algebras of primary interest for accelerator applications are $sp(2, \mathbb{R})$, $sp(4, \mathbb{R})$, and $sp(6, \mathbb{R})$.³ The purpose of this and the next two sections is to discuss suitable bases for these Lie algebras when special attention is given to their unitary subalgebras $u(1)$, $u(2)$, and $u(3)$. What we will be finding are the defining or fundamental representations of $sp(2, \mathbb{R})$, $sp(4, \mathbb{R})$, and $sp(6, \mathbb{R})$. See Section 3.7.6. We remark that these are the lowest dimensional representations that are faithful, in the sense of being isomorphic, to the underlying abstract Lie algebra.

One way to specify a basis is to select suitable matrices of the form JS . In Section (5.5) we learned that there is an isomorphism between the Poisson bracket Lie algebra of quadratic polynomials and the commutator Lie algebra of the matrices JS . Therefore, another way to specify a basis for $sp(2n, \mathbb{R})$ is to select suitable second-degree polynomials. We will mostly choose this second approach because of its convenience for later use. However, some of the calculations employed in selecting suitable polynomials will involve the associated matrices. Moreover, as indicated by (5.1) through (5.4), the associated matrices can easily be constructed from a knowledge of the associated second degree polynomials, and vice versa.

Because (as discussed in Section 3.9 and Exercise 3.9.1) matrices of the form JS^c form a Lie algebra in their own right, it is convenient to find the polynomials associated with these matrices first. By making use of the results of Sections 3.9 and 5.5, these polynomials can be arranged to give a realization of the Lie algebra $u(n)$. Then, when this is done, the polynomials associated with matrices of the form JS^a can be selected in a suitable manner. In particular, these polynomials can be selected in such a way that they have convenient transformation properties under the action of $u(n)$.

We begin with the case of $sp(2, \mathbb{R})$. In the 2×2 realization of $sp(2, \mathbb{R})$, the most general symmetric matrix S is of the form

$$S = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}, \quad (5.6.1)$$

and J is simply the matrix

$$J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad (5.6.2)$$

Requiring that J commute with S gives the restrictions

$$\beta = 0, \quad \gamma = \alpha. \quad (5.6.3)$$

Consequently, the most general S^c in the 2×2 case is just a multiple of the identity,

$$S^c = \alpha I, \quad (5.6.4)$$

and JS^c is simply a multiple of J ,

$$JS^c = \alpha J. \quad (5.6.5)$$

³In addition, the Lie algebra $sp(8, \mathbb{R})$ is useful for the treatment of errors. See Section 9.4.

Let b^0 be the polynomial associated with S^c by a relation of the form (5.1). Set $\alpha = 1$ so that $S^c = I$, in which case b^0 is given by the relation

$$b^0 = (1/2)(z_1^2 + z_2^2) = (1/2)(q^2 + p^2). \quad (5.6.6)$$

Let B^0 denote the associated matrix of the form JS^c . Then, according to (6.5), B^0 is given by the relation

$$B^0 = JS^c = JI = J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = i\sigma^2. \quad (5.6.7)$$

[Here, and in (6.13) and (6.14), we have also referenced a Pauli matrix σ^α . See Exercise 3.7.31. This referencing will be useful later.] We observe that Exercise 3.7.23 shows that $u(1)$ is one dimensional. The fact that in the 2×2 case we have found only one linearly independent matrix of the form JS^c is consistent with this observation.

Next study matrices S^a that anticommute with J . Requiring that J anticommute with the S of (6.1) gives only the restriction

$$\alpha = -\gamma. \quad (5.6.8)$$

Consequently, S^a is of the general form

$$S^a = \gamma \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} + \beta \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (5.6.9)$$

and JS^a is of the general form

$$JS^a = \gamma \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \beta \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (5.6.10)$$

Suppose we set $\gamma = 1$ and $\beta = 0$ in (6.9). Let f be the polynomial corresponding to this choice for S^a . It is given by the relation

$$f = (1/2)(-z_1^2 + z_2^2) = (1/2)(-q^2 + p^2). \quad (5.6.11)$$

Alternatively, suppose we set $\gamma = 0$ and $\beta = 1$ in (6.9). Let g be the polynomial corresponding to this choice for S^a . It is given by the relation

$$g = z_1 z_2 = qp. \quad (5.6.12)$$

[Again see (5.1).] Let F and G be the matrices associated with f and g . According to (6.10), F and G are given by the relations

$$F = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \sigma^1, \quad (5.6.13)$$

$$G = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \sigma^3. \quad (5.6.14)$$

It is readily verified that the polynomials b^0, f , and g obey the Poisson bracket rules

$$[b^0, f] = 2g, \quad (5.6.15)$$

$$[b^0, g] = -2f, \quad (5.6.16)$$

$$[f, g] = -2b^0. \quad (5.6.17)$$

Correspondingly, the matrices B^0, F , and G obey the analogous commutation rules,

$$\{B^0, F\} = 2G, \quad (5.6.18)$$

$$\{B^0, G\} = -2F, \quad (5.6.19)$$

$$\{F, G\} = -2B^0. \quad (5.6.20)$$

This is one version of the commutation rules for $sp(2, \mathbb{R})$. All others can be obtained by making the transformations (3.7.56) with a real invertible matrix T . Note that all the matrices B^0, F , and G are of the form JS with S real and symmetric. They therefore belong to $sp(2, \mathbb{R})$. Also, B^0 is real anti-Hermitian/antisymmetric, and therefore generates elements in $SO(2, \mathbb{R})$ upon exponentiation. By contrast, F and G are real Hermitian/symmetric and generate noncompact subgroups upon exponentiation.

Exercises

5.6.1. Verify that the requirement that J commute with S does indeed give the restrictions (6.3).

5.6.2. Verify that the requirement that J anticommute with S gives the restriction (6.8).

5.6.3. Verify that b^0, f , and g are associated with B^0, F , and G by relations of the form (5.3).

5.6.4. Verify the Lie algebraic relations (6.15) through (6.20).

5.6.5. Let g be the 2×2 matrix

$$g = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (5.6.21)$$

Let $U(1, 1)$ be the set of all complex 2×2 matrices that satisfy the relation

$$U^\dagger g U = g. \quad (5.6.22)$$

Show that $U(1, 1)$ is a group. Let $SU(1, 1)$ be the subset of matrices in $U(1, 1)$ that have unit determinant. Show that $SU(1, 1)$ is a group. Find the corresponding Lie algebras $u(1, 1)$ and $su(1, 1)$. Show that $su(1, 1)$ and $sp(2)$ are equivalent over the complex field.

5.6.6. Let g be the 3×3 matrix

$$g = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \quad (5.6.23)$$

Let $O(2, 1)$ be the set of all real 3×3 matrices that satisfy the relation

$$O^T g O = g. \quad (5.6.24)$$

Show that $O(2, 1)$ is a group. Let $SO(2, 1)$ be the subset of matrices in $O(2, 1)$ that have unit determinant. Show that $SO(2, 1)$ is a group. Find the corresponding Lie algebra $so(2, 1)$. Show that $so(2, 1)$ and $sp(2)$ are equivalent over the complex field.

5.6.7. This exercise studies polar decomposition for two interesting symplectic matrices, call them L and M , defined by the equations

$$L = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad (5.6.25)$$

$$M = -L = \begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix}. \quad (5.6.26)$$

Verify that both L and M are indeed symplectic. The matrix M is interesting because we know from Exercise 3.7.12 that it cannot be written in single exponential form. By contrast, verify that L can be written in the form

$$L = \exp(JS) \quad (5.6.27)$$

with

$$JS = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad (5.6.28)$$

where S is the symmetric matrix

$$S = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \quad (5.6.29)$$

Let us first work on finding the polar decomposition for L . That is, according to Subsection 3.8.2, we wish to write L in the form

$$L = PO. \quad (5.6.30)$$

Verify that from the properties of P and O it follows that

$$LL^T = P O O^T P^T = P^2. \quad (5.6.31)$$

Show that the matrix (LL^T) is real positive-definite symmetric since L is real symplectic. Next show that (LL^T) has a unique real positive-definite symmetric square root. Thus, P is determined by the equation

$$P = (LL^T)^{1/2}. \quad (5.6.32)$$

Show for the problem at hand that

$$LL^T = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \quad (5.6.33)$$

and P is given by the relation

$$P = \frac{1}{\sqrt{5}} \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}. \quad (5.6.34)$$

Observe that P is symplectic and symmetric as desired. Moreover, let us check that P as given by (6.34) is indeed positive definite. Let v be a two-component real vector given by

$$v = \{v_1, v_2\}. \quad (5.6.35)$$

Verify that

$$(v, Pv) = (1/5)(3v_1^2 + 2v_1v_2 + 2v_2^2). \quad (5.6.36)$$

The *discriminant* D of a binary quadratic form

$$av_1^2 + bv_1v_2 + cv_2^2 \quad (5.6.37)$$

is defined by the relation

$$D = 4ac - b^2. \quad (5.6.38)$$

From the theory of binary quadratic forms it is known that such a form is positive definite if $D > 0$. Verify that for the form (6.36)

$$D = (1/25)(24 - 4) > 0, \quad (5.6.39)$$

and therefore P is positive definite.

Now that P is known, O is given by (4.2.10). Verify that

$$O = P^{-1}L = -JP^TJL = -JPJL = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}. \quad (5.6.40)$$

Here we have used the fact that P is symplectic to compute its inverse. Verify that O is symplectic and orthogonal as desired. Finally, (6.30) holds because of the construction (6.40).

Let us now turn our attention to finding the polar decomposition for M , which we seek to write as

$$M = P'O'. \quad (5.6.41)$$

Show that

$$P' = P \quad (5.6.42)$$

and

$$O' = -O. \quad (5.6.43)$$

The last items we might wonder about for L and M are the matrices S^a and S^c and the associated polynomials f_2^a and f_2^c . The computation of these items is the task of Exercise 7.6.14.

5.7 Basis for $sp(4, \mathbb{R})$

The case of $sp(4, \mathbb{R})$ is somewhat more complicated. We again begin with the $u(n)$ Lie algebra, in this case $u(2)$. Any matrix v in $U(2)$ can be written in the form

$$v = e^{i\tau} \quad (5.7.1)$$

where τ is some linear combination (with real coefficients) of the Hermitian matrices $\sigma^0, \sigma^1, \sigma^2$, and σ^3 , which will be specified shortly. Comparison of (7.1) with (3.9.15) gives the relation

$$\tau = A + iB. \quad (5.7.2)$$

It is convenient to select σ^0 to be the 2×2 identity matrix, and to require that the remaining σ^j be the *Pauli* matrices,

$$\begin{aligned} \sigma^0 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \\ \sigma^2 &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned} \quad (5.7.3)$$

(See Exercise 3.7.31.) Note that the Pauli matrices are all Hermitian and, save for σ^0 , are traceless. Then successively setting $\tau = \sigma^j$, with $j = 0, 1, 2, 3$, and using (7.2) specifies four pairs of A, B matrices. Correspondingly, according to (3.9.10), these four pairs of A, B matrices specify four matrices of the form S^c . Finally, using the correspondence (5.1), these four S^c matrices specify four second-degree polynomials. Call these polynomials b^0, b^1, b^2 , and b^3 , respectively. Carrying out the required calculations gives the results

$$\begin{aligned} b^0 &= (1/2)(z_1^2 + z_2^2 + z_3^2 + z_4^2) = (1/2)(q_1^2 + p_1^2 + q_2^2 + p_2^2), \\ b^1 &= z_1 z_2 + z_3 z_4 = q_1 q_2 + p_1 p_2, \\ b^2 &= -z_1 z_4 + z_2 z_3 = -q_1 p_2 + q_2 p_1, \\ b^3 &= (1/2)(z_1^2 - z_2^2 + z_3^2 - z_4^2) = (1/2)(q_1^2 + p_1^2 - q_2^2 - p_2^2). \end{aligned} \quad (5.7.4)$$

It is readily verified that the polynomials b^0 through b^3 obey the Poisson bracket rules

$$[b^0, b^j] = 0, \quad j = 0, 1, 2, 3; \quad (5.7.5)$$

$$\begin{aligned} [b^1, b^2] &= -2b^3, \\ [b^2, b^3] &= -2b^1, \\ [b^3, b^1] &= -2b^2. \end{aligned} \quad (5.7.6)$$

These are the rules for the Lie algebra $u(2)$. Observe also that the relations (7.6) are a variant of the rules for the Lie algebra $su(2)$. That is, the rules (7.6) can be written in the form

$$[b^j, b^k] = -2 \sum_{\ell} \epsilon_{jkl} b^{\ell}, \quad (5.7.7)$$

where ϵ_{jkl} is the *Levi-Civita* tensor.

The reader is probably aware that the treatment of angular momentum in quantum mechanics, which essentially amounts to a study of the representations of $su(2)$, is facilitated by the introduction of *raising* and *lowering ladder* operators J_{\pm} as well as the *diagonal* operator J_z . For our purposes it is convenient to employ the analogous polynomials $r(\pm)$ and c defined by the relations

$$r(\pm) = (i/2)(b^1 \pm ib^2) = (i/2)(q_1 \pm ip_1)(q_2 \mp ip_2), \quad (5.7.8)$$

$$c = (-i/\sqrt{2})b^3 = (-i/\sqrt{8})(q_1^2 + p_1^2 - q_2^2 - p_2^2). \quad (5.7.9)$$

They obey the Poisson bracket rules

$$[c, r(\pm)] = \pm(\sqrt{2})r(\pm), \quad (5.7.10)$$

$$[r(+), r(-)] = (\sqrt{2})c. \quad (5.7.11)$$

We note that these rules can also be written in the form

$$:c:r(\pm) = \pm(\sqrt{2})r(\pm), \quad (5.7.12)$$

$$:r(\pm):c = \mp(\sqrt{2})r(\pm). \quad (5.7.13)$$

$$:r(+):r(-) = (\sqrt{2})c, \quad (5.7.14)$$

We now turn to the problem of determining the matrices S^a that anticommute with J . As described earlier, the most general real symmetric S can be written in the form (3.9.1) subject to the conditions (3.9.2). Requiring that S^a anticommute with J gives the further restrictions

$$B^T = B, \quad (5.7.15)$$

$$C = -A. \quad (5.7.16)$$

Consequently, the most general S^a is of the form

$$S^a = \begin{pmatrix} A & B \\ B & -A \end{pmatrix}, \quad (5.7.17)$$

with both A and B real and symmetric,

$$A^T = A, \quad B^T = B. \quad (5.7.18)$$

In the 4×4 case of $sp(4)$, both A and B are 2×2 . Thus, since they are symmetric, they can be written in the form

$$A = a \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + b \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + c \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (5.7.19)$$

$$B = d \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + e \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + f \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (5.7.20)$$

where the coefficients a through f are arbitrary. It follows that in the 4×4 case the vector space spanned by matrices of the form JS^a is six dimensional. Since the Lie algebra generated by matrices of the form JS^c is four dimensional as has already been seen, the dimension of the complete Lie algebra generated by both the matrices JS^c and JS^a is $4+6=10$ dimensional, in accord with (3.7.35) for $n = 4$.

The last step is to make a suitable choice of six second-degree polynomials corresponding to six different choices for the matrices S^a . We have found it convenient to first introduce 3 complex polynomials h^\pm , h^0 by the rules

$$h^+ = -(1/2)(q_1 + ip_1)^2, \quad (5.7.21)$$

$$h^0 = -(1/\sqrt{2})(q_1 + ip_1)(q_2 + ip_2), \quad (5.7.22)$$

$$h^- = (1/2)(q_2 + ip_2)^2. \quad (5.7.23)$$

Under the action of $r(\pm)$ and c they are transformed according to the rules

$$: c : h^\pm = \pm(\sqrt{2})h^\pm, \quad (5.7.24)$$

$$: c : h^0 = 0, \quad (5.7.25)$$

$$: r(+) : h^- = (\sqrt{2})h^0, \quad (5.7.26)$$

$$: r(+) : h^0 = -(\sqrt{2})h^+. \quad (5.7.27)$$

Note that these rules are analogous to the relations (7.12) through (7.14). Consequently, we may view the 3 objects h^\pm and h^0 as the components of a “spin” 1 (vector) object in a spherical basis. [Strictly speaking, we should invent a special terminology for this and related situations. Perhaps we should talk about *unitary* spin 1 to emphasize the fact that the spin we are referring to is with respect to an $SU(2)$ group, and is not an angular momentum spin related to some rotation group.] Finally, under the action of b^0 , they are transformed according to the rules

$$: b^0 : h^\pm = 2ih^\pm, \quad (5.7.28)$$

$$: b^0 : h^0 = 2ih^0. \quad (5.7.29)$$

We now form 6 real polynomials f^j and g^j by taking suitable linear combinations of real and imaginary parts of h^\pm and h^0 ,

$$f^3 = -\sqrt{2} \operatorname{Re}(h^0) = q_1q_2 - p_1p_2, \quad (5.7.30)$$

$$\begin{aligned} f^1 &= -(1/2)[b^2, f^3] = (1/2)(p_1^2 - q_1^2 - p_2^2 + q_2^2), \\ f^2 &= -(1/2)[b^3, f^1] = -q_1p_1 - q_2p_2, \\ g^3 &= -\sqrt{2} \operatorname{Im}(h^0) = q_1p_2 + q_2p_1, \\ g^1 &= -(1/2)[b^2, g^3] = -q_1p_1 + q_2p_2, \\ g^2 &= -(1/2)[b^3, g^1] = (1/2)(-p_1^2 + q_1^2 - p_2^2 + q_2^2). \end{aligned} \quad (5.7.31)$$

The f 's and g 's have been selected in such a way that they obey the Lie algebraic rules

$$[b^j, f^k] = -2 \sum_{\ell} \epsilon_{jk\ell} f^{\ell}, \quad (5.7.32)$$

$$[b^j, g^k] = -2 \sum_{\ell} \epsilon_{jk\ell} g^{\ell}. \quad (5.7.33)$$

That is, they behave like the Cartesian components of spin 1 objects under the action of $su(2)$. Under the action of b^0 , the f 's and g 's are transformed into each other,

$$[b^0, f^j] = -2g^j, \quad (5.7.34)$$

$$[b^0, g^j] = 2f^j.$$

Finally, the Poisson brackets of the f 's and g 's with each other are given by the relations

$$[f^j, f^k] = 2 \sum_{\ell} \epsilon_{jk\ell} b^{\ell}, \quad (5.7.35)$$

$$[g^j, g^k] = 2 \sum_{\ell} \epsilon_{jk\ell} b^{\ell}, \quad (5.7.36)$$

$$[f^j, g^k] = 2\delta_{jk} b^0. \quad (5.7.37)$$

Note that according to the relations (7.32) through (7.34), the f 's and g 's are transformed among each other under the action of $u(2)$; and the right sides of (7.35) through (7.37) are elements of $u(2)$. This result is in accord with Exercise (3.9.1).

When taken all together, the rules (7.5) through (7.7) and (7.32) through (7.37) specify the Lie algebra $sp(4)$.

Exercises

5.7.1. Carry out the calculations that produce the results (7.4).

5.7.2. Verify the Poisson bracket relations (7.5), (7.6), and (7.7). Show that the Pauli matrices satisfy the analogous commutation rules

$$\{i\sigma^0, i\sigma^j\} = 0, \quad (5.7.38)$$

$$\{i\sigma^j, i\sigma^k\} = -2 \sum_{\ell} \epsilon_{jk\ell} (i\sigma^{\ell}) \text{ or } \{\sigma^j, \sigma^k\} = 2i \sum_{\ell} \epsilon_{jk\ell} \sigma^{\ell} \Leftrightarrow \{\sigma^1, \sigma^2\} = 2i\sigma^3, \text{ etc.} \quad (5.7.39)$$

5.7.3. Verify that S^a is of the form (7.10) subject to the conditions (7.11).

5.7.4. Show that the f 's and g 's given by (7.30) and (7.31) do indeed correspond to matrices of the form S^a , and find these matrices.

5.7.5. Verify the Poisson bracket rules (7.32) through (7.37).

5.7.6. Consider the complex conjugates of the polynomials h^{\pm} and h^0 . Show that they are also transformed among each other as a spin 1 object under the action of $u(2)$. Thus, as already evidenced by the existence of the f^j and g^j , there are *two* spin 1 objects in $sp(4)$ corresponding to the 6 independent matrices of the form JS^a .

5.7.7. Review Exercise 7.2 above. The purpose of this exercise is to further explore properties of the Pauli matrices. Show that they obey the multiplication rules

$$\sigma^j \sigma^k = \delta_{jk} \sigma^0 + i \sum_{\ell} \epsilon_{jkl} \sigma^{\ell} \text{ for } j, k, \ell = 1, 2, 3 \Leftrightarrow (\sigma^j)^2 = I \text{ and } \sigma^1 \sigma^2 = i \sigma^3, \text{ etc.} \quad (5.7.40)$$

Show, as a special case of (7.40), that they obey the *anticommutation* rules

$$\{\sigma^j, \sigma^k\}_+ = \sigma^j \sigma^k + \sigma^k \sigma^j = 2\delta_{jk} \sigma^0 ; \quad j, k = 1, 2, 3. \quad (5.7.41)$$

In particular, they anticommute ($\sigma^j \sigma^k = -\sigma^k \sigma^j$) when $j \neq k$.

Show that the Pauli matrices σ^j for $j = 1, 2, 3$ span the vector space of 2×2 *traceless* Hermitian matrices, and obey the relations

$$\text{tr}(\sigma^j \sigma^k) = 2\delta_{jk} ; \quad j, k = 1, 2, 3. \quad (5.7.42)$$

Show that there are the additional trace relations

$$\text{tr}(\sigma^j \{\sigma^k, \sigma^{\ell}\}_+) = 0, \quad (5.7.43)$$

$$\text{tr}(\sigma^j \{\sigma^k, \sigma^{\ell}\}) = 4i\epsilon_{jkl} = -4i(L^j)_{kl}, \quad (5.7.44)$$

$$\text{tr}(\sigma^j \sigma^k \sigma^{\ell}) = 2i\epsilon_{jkl} = -2i(L^j)_{kl}. \quad (5.7.45)$$

Recall (3.7.182).

Let \mathbf{a} be a three-component vector with entries (a_1, a_2, a_3) . Introduce the notation

$$\mathbf{a} \cdot \boldsymbol{\sigma} = \sum_{j=1}^3 a_j \sigma^j. \quad (5.7.46)$$

Verify that

$$\mathbf{a} \cdot \boldsymbol{\sigma} = \begin{pmatrix} a_3 & a_1 - ia_2 \\ a_1 + ia_2 & -a_3 \end{pmatrix}. \quad (5.7.47)$$

Verify that

$$\det(\mathbf{a} \cdot \boldsymbol{\sigma}) = -\mathbf{a} \cdot \mathbf{a}. \quad (5.7.48)$$

Show that there are the multiplication, commutation, and anticommutation relations

$$(\mathbf{a} \cdot \boldsymbol{\sigma})(\mathbf{b} \cdot \boldsymbol{\sigma}) = (\mathbf{a} \cdot \mathbf{b})\sigma^0 + i(\mathbf{a} \times \mathbf{b}) \cdot \boldsymbol{\sigma}, \quad (5.7.49)$$

$$\{(\mathbf{a} \cdot \boldsymbol{\sigma}), (\mathbf{b} \cdot \boldsymbol{\sigma})\} = 2i(\mathbf{a} \times \mathbf{b}) \cdot \boldsymbol{\sigma}, \quad (5.7.50)$$

$$\{(\mathbf{a} \cdot \boldsymbol{\sigma}), (\mathbf{b} \cdot \boldsymbol{\sigma})\}_+ = 2(\mathbf{a} \cdot \mathbf{b})\sigma^0. \quad (5.7.51)$$

Show that the Pauli matrices and σ^0 span the vector space of 2×2 Hermitian matrices, and obey the relations

$$\text{tr}(\sigma^j \sigma^k) = 2\delta_{jk} ; \quad j, k = 0, 1, 2, 3. \quad (5.7.52)$$

5.7.8. The relation (5.3) associates a matrix JS with every quadratic polynomial. Find the matrices B^i (for $i = 0, 1, 2, 3$) associated with the polynomials b^i . Find the matrices F^j and G^j (for $j = 1, 2, 3$) associated with the polynomials f^j and g^j . Use (5.21) if you wish. The B^i , F^j , and G^j provide a basis for the 4×4 matrix representation of $sp(4)$. Find their commutation rules.

Answer:

$$\begin{aligned}
 B^0 = J &= \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \quad B^1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}, \\
 B^2 &= \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \quad B^3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \\
 F^1 &= \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \quad F^2 = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \\
 F^3 &= \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}, \quad G^1 = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \\
 G^2 &= \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \quad G^3 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix}.
 \end{aligned} \tag{5.7.53}$$

Note that these generators/matrices are given for the case that J has the form (3.1.1). In the case that J' as given by (3.2.10) is employed, one may find the related generators by means of the permutation matrix P given by (3.2.18).

5.7.9. Consider the matrices $i\sigma^j$ that obey the $su(2)$ commutation rules (7.39). Also, review Exercise 3.7.36. Form the associated hatted representation given by (3.7.218). Verify that this representation is equivalent to the original representation using the matrix

$$E = i\sigma^2 = J. \tag{5.7.54}$$

Form the associated checked representation given by (3.7.219). Show that in this case the result is the same as using the hatting operation (3.7.218). Consider the matrices B^1, B^2, B^3 given by (7.47). Verify that, as expected, they also provide a representation of $su(2)$. Show that these matrices are unaffected by either of the hatting or checking operations (3.7.218) and (3.7.219).

5.7.10. Review Exercise 3.7.36. Consider the representation of $sp(4, \mathbb{R})$ provided by the matrices (7.47). Since they are real, they are unaffected by the ‘check’ operation (3.7.219). Find the hatted representation given by (3.7.218). Verify that, as expected, this representation is equivalent to the original representation using

$$E = J. \quad (5.7.55)$$

5.7.11. Review Exercise 4.3.19. Let z' denote the collection of phase-space variables with the ordering

$$z' = (q_1, p_1, q_2, p_2). \quad (5.7.56)$$

The purpose of this exercise is to find a relation between the polynomials b^j and the matrices C^j . Show that there is the relation

$$: b^j : z'_c = - \sum_d C_{cd}^j z'_d. \quad (5.7.57)$$

5.7.12. Consider the linear transformation on 4-dimensional phase space given by the rules

$$\bar{q}_1 = q_2, \quad (5.7.58)$$

$$\bar{q}_2 = q_1, \quad (5.7.59)$$

$$\bar{p}_1 = p_2, \quad (5.7.60)$$

$$\bar{p}_2 = p_1. \quad (5.7.61)$$

Verify that it is described by the matrix

$$R = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (5.7.62)$$

Evidently R interchanges the q_1, p_1 and q_2, p_2 planes. Verify that this transformation is symplectic and also that R is orthogonal. Thus, R is in the $U(2)$ subgroup of $Sp(4, \mathbb{R})$ and must be expressible in the form

$$R = \exp(JS^c). \quad (5.7.63)$$

Your task is to find S^c . Verify that R can be written in the form

$$R = M(v) \quad (5.7.64)$$

as described in Section 3.9 and show that

$$v = \sigma^1. \quad (5.7.65)$$

Using (3.7.159), show that σ^1 satisfies the relation

$$\sigma^1 = \exp[(i\pi/2)(\sigma^1 - \sigma^0)]. \quad (5.7.66)$$

Use this result to find S^c .

5.8 Basis for $sp(6, \mathbb{R})$

The case of $sp(6, \mathbb{R})$ is even more complicated, yet the procedure will still be the same. Again we will begin with the unitary Lie algebra, in this case $u(3)$, corresponding to matrices of the form JS^c . Then we will select a basis for matrices of the form JS^a (or, equivalently, a basis for the corresponding second degree polynomials) in such a way that these matrices (polynomials) have convenient transformation properties under the action of $u(3)$ or $su(3)$. This second step will require some discussion of representations of $su(3)$. Fortunately for us $su(3)$ has been well studied, initially by mathematicians, and subsequently by physicists because of its applications to Elementary Particle and Nuclear Physics and the Three-Body problem.

5.8.1 $U(3)$ Preliminaries

Any matrix in $U(3)$ can be written in the form (7.1) where τ is some linear combination (with real coefficients) of $3^2 = 9$ Hermitian matrices $\lambda^0, \lambda^1, \dots, \lambda^8$ that will be listed below. Once these matrices are specified, use of (7.2) in turn specifies 9 pairs of A, B matrices, which in turn according to (3.9.10) specifies 9 matrices of the form S^c . Finally, using the correspondence (5.1), these 9 S^c matrices specify 9 second-degree polynomials.

We select λ^0 to be the 3×3 identity matrix, and require that the remaining λ^j be the *Gell-Mann* (1929-2019) matrices,

$$\begin{aligned} \lambda^0 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \lambda^1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda^2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\ \lambda^3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda^4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \lambda^5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \\ \lambda^6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \lambda^7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \lambda^8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \end{aligned} \quad (5.8.1)$$

Note that all the λ matrices are Hermitian and the λ^j (for $j > 0$) are traceless. They satisfy the commutation rules

$$\{i\lambda^0, i\lambda^j\} = 0; \quad (5.8.2)$$

$$\{i\lambda^j, i\lambda^k\} = -2 \sum_{\ell} f_{jkl}(i\lambda^{\ell}) \text{ for } j, k, \ell \neq 0. \quad (5.8.3)$$

Taken together, the rules (8.2) and (8.3) are the commutation rules for $u(3)$. The rules (8.3), for which $j, k, \ell = 1, 2, \dots, 8$, are the commutation rules for $su(3)$.⁴ The coefficients f_{jkl} (up to a multiplicative constant) are the *structure constants* of $su(3)$. They are real and

⁴We remark that quantum physicists prefer, when possible, to work with Hermitian matrices because in Quantum Mechanics observables are associated with Hermitian operators. (See Exercise 3.7.43.) Therefore judicious factors of i are employed in definitions like (7.3) and (8.1) to achieve this end. In so doing, mathematically extraneous factors of i appear elsewhere. However, in writing (7.38), (7.39), (8.2), and (8.3),

antisymmetric under the interchange of any two (adjacent) indices. Thus most of them are zero. See Exercise 8.2. The table below lists some of them. All the rest are zero, or can be obtained from those listed by permutation of indices and use of the antisymmetry property.

Table 5.8.1: Structure Constants of $su(3)$.

$\underline{jk\ell}$	$f_{jk\ell}$	$\underline{jk\ell}$	$f_{jk\ell}$	$\underline{jk\ell}$	$f_{jk\ell}$
123	1	246	$1/2$	367	$-1/2$
147	$1/2$	257	$1/2$	458	$\sqrt{3}/2$
156	$-1/2$	345	$1/2$	678	$\sqrt{3}/2$

Before going on, we remark that the Gell-Mann matrices also satisfy the *anticommutation* rules

$$\{\lambda^j, \lambda^k\}_+ = \lambda^j \lambda^k + \lambda^k \lambda^j = (4/3)\delta_{jk}\lambda^0 + 2 \sum_\ell d_{jk\ell} \lambda^\ell. \quad (5.8.4)$$

Here the coefficients $d_{jk\ell}$, called the *symmetric coupling coefficients*, are *symmetric* under the interchange of any two indices. See Exercise 8.4. Table 8.2 below lists some of them. All the rest are zero, or can be gotten from those listed by permutation of indices and use of the symmetry property.

Table 5.8.2: Symmetric Coupling Coefficients of $su(3)$.

$\underline{jk\ell}$	$d_{jk\ell}$	$\underline{jk\ell}$	$d_{jk\ell}$	$\underline{jk\ell}$	$d_{jk\ell}$	$\underline{jk\ell}$	$d_{jk\ell}$
118	$1/\sqrt{3}$	247	$-1/2$	355	$1/2$	558	$-\sqrt{3}/6$
146	$1/2$	256	$1/2$	366	$-1/2$	668	$-\sqrt{3}/6$
157	$1/2$	338	$1/\sqrt{3}$	377	$-1/2$	778	$-\sqrt{3}/6$
228	$1/\sqrt{3}$	344	$1/2$	448	$-\sqrt{3}/6$	888	$-1/\sqrt{3}$

5.8.2 Polynomials for $u(3)$

Now successively set $\tau = \lambda^j$ with $j = 0, 1, \dots, 8$, and compute the corresponding second-degree polynomials b^0, b^1, \dots, b^8 . Doing so gives the results

$$b^0 = (1/2)(q_1^2 + p_1^2 + q_2^2 + p_2^2 + q_3^2 + p_3^2), \quad (5.8.5)$$

$$b^1 = q_1 q_2 + p_1 p_2,$$

$$b^2 = -q_1 p_2 + q_2 p_1,$$

we have compensated for this mischief by explicitly displaying i factors in the commutation rules. From a Lie algebraic perspective, the natural basis for any $su(n)$ Lie algebra consists of anti-Hermitian matrices. Thus, for a mathematician, the natural basis for $su(3)$ consists of the matrices $i\lambda^j$ with $j = 1, 2 \dots, 8$.

$$\begin{aligned}
b^3 &= (1/2)(q_1^2 + p_1^2 - q_2^2 - p_2^2), \\
b^4 &= q_1 q_3 + p_1 p_3, \\
b^5 &= -q_1 p_3 + q_3 p_1, \\
b^6 &= q_2 q_3 + p_2 p_3, \\
b^7 &= -q_2 p_3 + q_3 p_2, \\
b^8 &= (1/\sqrt{12})(q_1^2 + p_1^2 + q_2^2 + p_2^2 - 2q_3^2 - 2p_3^2).
\end{aligned}$$

It is readily verified that the polynomials b^0 through b^8 obey the Poisson bracket rules

$$[b^0, b^j] = 0 \text{ for } j = 0, 1, \dots, 8; \quad (5.8.6)$$

$$[b^j, b^k] = -2 \sum_{\ell} f_{jkl} b^{\ell} \text{ for } j, k, \ell = 1, \dots, 8. \quad (5.8.7)$$

Taken together, these are the rules for the Lie algebra $u(3)$. By themselves, the relations (8.7) are the rules for the Lie algebra $su(3)$.

5.8.3 Plan for the Remaining Polynomials

We next turn to the problem of finding the second-degree polynomials corresponding to the matrices JS^a . As was the case for $sp(4)$, the most general S^a is of the form (7.17) with the matrices A, B subject to the symmetry conditions (7.18). In the 6×6 case of $sp(6)$, both A and B are 3×3 . Since there are 6 linearly independent 3×3 symmetric matrices, the space spanned by matrices of the form JS^a is $2 \times 6 = 12$ dimensional. This is as it should be since $9 + 12 = 21$, the dimension of $sp(6)$. What we wish to do is select 12 second-degree polynomials corresponding to the matrices JS^a in such a way that these polynomials have convenient transformation properties under $su(3)$. To do so will require some discussion of what is called the *Cartan basis* for $su(3)$ and of the representations of $su(3)$.

5.8.4 Cartan Basis for $su(3)$

As already mentioned in the previous section, a study of the representations of $su(2)$ is facilitated by the introduction of *raising* and *lowering ladder* operators J_{\pm} as well as the *diagonal* operator J_z . As discovered by *Killing* and *Cartan*, the same is true for $su(3)$ and all *simple* Lie algebras.⁵ In the case of $su(3)$ there are 6 ladder elements that play roles analogous to J_{\pm} ; and there are 2 commuting elements that play roles analogous to J_z [for this reason $su(3)$ is said to be of *rank* 2].

Abstractly speaking, a Lie algebra is any set of elements with the properties (3.7.43) through (3.7.45) and (3.7.48) and (3.7.49). For many purposes (including illustrative purposes) it is convenient to work with concrete matrix or differential operator *representations* of Lie algebras. In the matrix case, the Lie algebra consists of linear operators acting on a (usually finite-dimensional) vector space, and the Lie product is matrix commutation. In

⁵Recall that a Lie algebra is called simple if it has no ideals. See Section 8.9. For the use of ladder operators in the case of the symplectic Lie algebras, see Chapter 27.

the differential operator case, the Lie algebra consists of linear differential operators acting on a function space, and the Lie product is differential operator commutation. [As indicated, both these kinds of realizations (*representations*) of Lie algebras and their associated Lie groups are *linear*. There are also *nonlinear* realizations of groups as illustrated, for example, in Section 5.11.]

In the case of $su(3)$, as might be imagined, the smallest matrix representation is realized in terms of 3×3 matrices. For our purposes it is again convenient to employ the Gell-Mann matrices. [However, just as in the case of $su(2)$ for which the smallest matrix representation is realized in terms of the 2×2 Pauli matrices but there are also representations in terms of larger $(2j+1) \times (2j+1)$ matrices, so too there are also representations of $su(3)$ in terms of larger matrices.] We will therefore begin by illustrating for $su(3)$ how the 2 commuting elements and 6 ladder elements are set up in the 3×3 case.

Call the commuting elements C^1 and C^2 . In the 3×3 case they are defined by the relations

$$C^1 = (1/\sqrt{2})\lambda^3, \quad C^2 = (1/\sqrt{2})\lambda^8. \quad (5.8.8)$$

It is easily checked that they do indeed commute,

$$\{C^1, C^2\} = 0. \quad (5.8.9)$$

The 6 ladder elements are conveniently labelled by 3 two-component vectors and their negatives, collectively called *root vectors*. Let e^1 and e^2 be orthogonal unit vectors. Define three vectors α, β, γ by the relations

$$\begin{aligned} \alpha &= (\sqrt{2})e^1, \\ \beta &= (1/\sqrt{2})e^1 + (\sqrt{6}/2)e^2, \\ \gamma &= -(1/\sqrt{2})e^1 + (\sqrt{6}/2)e^2. \end{aligned} \quad (5.8.10)$$

Figure 8.1 shows these vectors and their negatives in what is called a *root diagram*. (Note that all the root vectors have length $\sqrt{2}$, and the angle between any two successive root vectors as one goes around the root diagram is 60 degrees.) We denote the ladder elements by $R(\mu)$ where μ is one of the root vectors, i.e. one of the vectors (8.10) or their negatives. The ladder elements are defined by the relations

$$R(\pm\alpha) = (-1/2)(\lambda^1 \pm i\lambda^2), \quad (5.8.11)$$

$$R(\pm\beta) = (-1/2)(\lambda^4 \pm i\lambda^5),$$

$$R(\pm\gamma) = (-1/2)(\lambda^6 \pm i\lambda^7).$$

The choice of elements given by the relations (8.8) and (8.11) is called the *Cartan basis* for $su(3)$. The commuting elements C^j are referred to as the *Cartan subalgebra*, and the *rank* of a simple Lie algebra is the dimension of its Cartan subalgebra.⁶ Inspection of (8.8), (8.11), and the Gell-Mann matrices (8.1) reveals that all the C^j and $R(\mu)$ are *real* matrices. This is a general feature of the Cartan basis for simple Lie algebras. See, for example, Chapter 27

⁶To be true to history, the Cartan subalgebra could better be called the *Killing* subalgebra since it was he who first recognized and employed it. We also remark that Killing discovered Lie algebras independently of Lie.

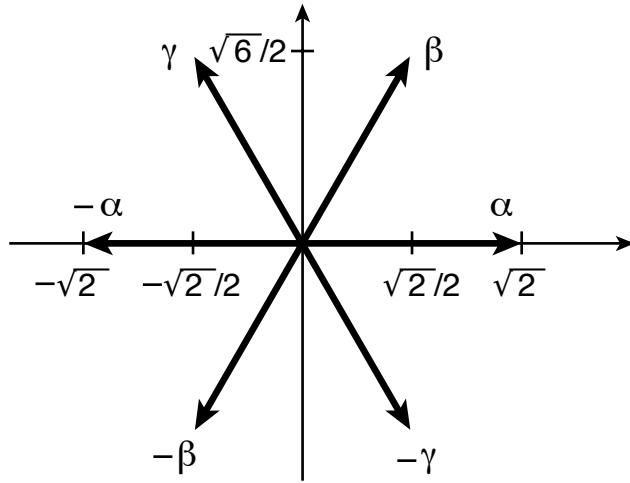


Figure 5.8.1: Root diagram showing the root vectors for $su(3)$.

for the case of $sp(2n, \mathbb{R})$. We also note that matrices of the form $\exp(i\theta_1 C^1 + i\theta_2 C^2)$ produce a *torus*, indeed a 2-torus, in $SU(3)$. See (8.1), (8.5), and Section 3.9. Moreover, this torus has largest dimension for any torus in $SU(3)$. Thus, exponentiating the Cartan subalgebra produces a maximal torus.

We remark that in the Lie algebraic mathematics literature it is customary to denote the elements of the Cartan subalgebra by the symbols H^j rather than our C^j , and the ladder elements by $E(\boldsymbol{\mu})$ rather than our $R(\boldsymbol{\mu})$. We have departed from this common notation because of our desire to generally reserve the symbol H for Hamiltonians and to employ the symbol E for other purposes.

The virtue of the Cartan basis is that the commutation rules take a particularly illuminating form. The commutator of C^j with $R(\boldsymbol{\mu})$ is

$$\{C^j, R(\boldsymbol{\mu})\} = (\mathbf{e}^j \cdot \boldsymbol{\mu})R(\boldsymbol{\mu}). \quad (5.8.12)$$

The C^j thus serve to establish the coordinate system for the root vectors.⁷ The commutators between pairs of R 's, $R(\boldsymbol{\mu})$ and $R(\boldsymbol{\nu})$, are of two types. If the root vectors $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are equal and opposite, the commutator is given by the relation

$$\{R(\boldsymbol{\mu}), R(-\boldsymbol{\mu})\} = \sum_j (\mathbf{e}^j \cdot \boldsymbol{\mu})C^j. \quad (5.8.13)$$

If the sum of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ is again a root vector, the commutator takes the form

$$\{R(\boldsymbol{\mu}), R(\boldsymbol{\nu})\} = N(\boldsymbol{\mu}, \boldsymbol{\nu})R(\boldsymbol{\mu} + \boldsymbol{\nu}). \quad (5.8.14)$$

All other commutators vanish. Here $N(\boldsymbol{\mu}, \boldsymbol{\nu})$ is a numerical factor equal to ± 1 . The positive N 's are $N(\boldsymbol{\alpha}, -\boldsymbol{\beta})$, $N(\boldsymbol{\gamma}, \boldsymbol{\alpha})$, $N(-\boldsymbol{\beta}, \boldsymbol{\gamma})$, $N(\boldsymbol{\beta}, -\boldsymbol{\alpha})$, $N(-\boldsymbol{\alpha}, -\boldsymbol{\gamma})$, and $N(-\boldsymbol{\gamma}, \boldsymbol{\beta})$.

⁷Note that 8.12 can also be written in the form $(\text{ad } C^j)R(\boldsymbol{\mu}) = (\mathbf{e}^j \cdot \boldsymbol{\mu})R(\boldsymbol{\mu})$. Thus, the components of the root vectors are the eigenvalues of the linear operators $(\text{ad } C^j)$.

5.8.5 Representations of $su(3)$: Cartan's Approach

Suppose the C^j and $R(\boldsymbol{\mu})$ are *any* set of matrices obeying the commutation rules (8.9) and (8.12) through (8.14). Suppose further that a scalar product can be set up in the underlying vector space in such a way that the C^j are Hermitian. See Section 7.3. Let $|\mathbf{w}\rangle = |w_1 w_2\rangle$ denote an eigenvector of the C^j with the property

$$C^j |w_1 w_2\rangle = w_j |w_1 w_2\rangle \quad \text{or} \quad C^j |\mathbf{w}\rangle = (\mathbf{e}^j \cdot \mathbf{w}) |\mathbf{w}\rangle. \quad (5.8.15)$$

Since the C^j are Hermitian, the w_j are real. It is convenient, as shown, to treat them together as the components of a single labeling vector denoted as \mathbf{w} and called a *weight*. Consider the vector $R(\boldsymbol{\mu})|\mathbf{w}\rangle$. From the commutation rules (8.12) we have the relation

$$\begin{aligned} C^j R(\boldsymbol{\mu}) |\mathbf{w}\rangle &= R(\boldsymbol{\mu}) C^j |\mathbf{w}\rangle + (\mathbf{e}^j \cdot \boldsymbol{\mu}) R(\boldsymbol{\mu}) |\mathbf{w}\rangle \\ &= R(\boldsymbol{\mu}) w_j |\mathbf{w}\rangle + (\mathbf{e}^j \cdot \boldsymbol{\mu}) R(\boldsymbol{\mu}) |\mathbf{w}\rangle \\ &= [\mathbf{e}^j \cdot (\mathbf{w} + \boldsymbol{\mu})] R(\boldsymbol{\mu}) |\mathbf{w}\rangle. \end{aligned} \quad (5.8.16)$$

It follows that if $R(\boldsymbol{\mu})|\mathbf{w}\rangle$ is different from zero, then it is an eigenvector of the C^j with weight $\mathbf{w} + \boldsymbol{\mu}$. Consequently, from a single weight we can produce a whole set of weights. The set of weight vectors can be *ordered* by means of the following definitions:

1. A vector is *positive* if its first nonvanishing component is positive.
2. A vector \mathbf{w} is *higher* than the vector \mathbf{w}' if $\mathbf{w} - \mathbf{w}'$ is positive.

We can now state the fundamental theorems of Cartan concerning representations:

1. In any irreducible representation, there is an eigenvector with highest weight, and this eigenvector is unique, i.e., non-degenerate.
2. Two irreducible representations are equivalent if they have the same highest weight.
3. Every highest weight \mathbf{w}^h is a linear combination, with non-negative integer coefficients, of what are called *fundamental* weights. For a rank ℓ Lie algebra there are ℓ such fundamental weights. Thus, for a rank ℓ Lie algebra, each irreducible representation is (uniquely) specified by an ℓ -tuple of non-negative integers.

For example in the case of $su(3)$, which is of rank 2, the two fundamental weights ϕ^1 and ϕ^2 are given by the relations

$$\phi^1 = (1/\sqrt{2})\mathbf{e}^1 + (1/\sqrt{6})\mathbf{e}^2, \quad (5.8.17)$$

$$\phi^2 = (1/\sqrt{2})\mathbf{e}^1 - (1/\sqrt{6})\mathbf{e}^2. \quad (5.8.18)$$

These fundamental weights are shown in Figure 8.2 along with the $su(3)$ root vectors. Consequently, for $su(3)$, every highest weight \mathbf{w}^h is of the form

$$\mathbf{w}^h = m\phi^1 + n\phi^2 = m[(1/\sqrt{2})\mathbf{e}^1 + (1/\sqrt{6})\mathbf{e}^2] + n[(1/\sqrt{2})\mathbf{e}^1 - (1/\sqrt{6})\mathbf{e}^2], \quad (5.8.19)$$

where m and n are arbitrary non-negative integers.

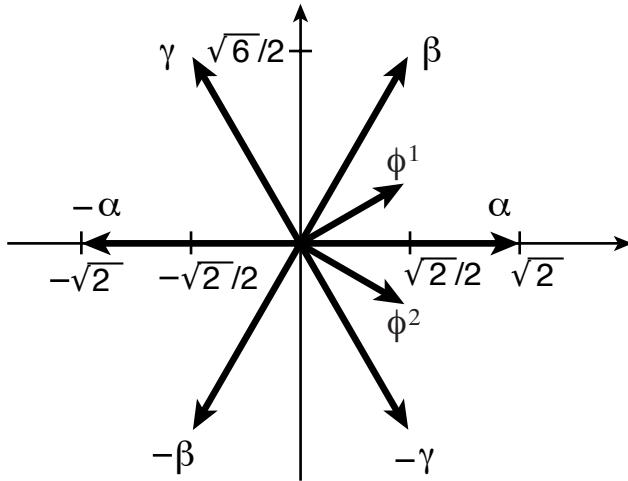


Figure 5.8.2: Fundamental weights ϕ^1 and ϕ^2 for $su(3)$. The root vectors are also shown.

Taken together, Cartan's theorems show that an irreducible representation of $su(3)$ is completely characterized by the two non-negative integers m and n . We denote this representation by $\Gamma(m, n)$. It can be shown that the conjugate representation is given by $\Gamma(n, m)$. That is,

$$\bar{\Gamma}(m, n) = \Gamma(n, m). \quad (5.8.20)$$

For discussion and examples see Exercises 3.7.36, 8.29, and 8.30. We also note for future use that the dimension of the representation $\Gamma(m, n)$ is given by the relation

$$\dim \Gamma(m, n) = (m + 1)(n + 1)(m + n + 2)/2. \quad (5.8.21)$$

For quick reference, the dimensions of the first few representations are listed in Table 8.3 below. Note that, as expected, $\Gamma(m, n)$ and $\Gamma(n, m)$ have the same dimension. Finally, for simplicity and where no ambiguity is involved, we sometimes refer to a representation by its dimension. That is, in view of (8.20) and (8.21), we use the shorthand notation $1 = \Gamma(0, 0)$, $3 = \Gamma(1, 0)$, $\bar{3} = \Gamma(0, 1)$, $6 = \Gamma(2, 0)$, $\bar{6} = \Gamma(0, 2)$, $8 = \Gamma(1, 1)$, etc. Note however that $\Gamma(2, 1)$ and $\Gamma(4, 0)$ as well as their conjugates all have dimension 15.

Table 5.8.3: Dimensions of Representations of $su(3)$.

m	n	$\dim \Gamma(m, n)$	m	n	$\dim \Gamma(m, n)$
0	0	1	4	0	15
1	0	3	0	4	15
0	1	3	3	1	24
2	0	6	1	3	24
0	2	6	2	2	27
1	1	8	5	0	21
3	0	10	0	5	21
0	3	10	4	1	35
2	1	15	1	4	35
1	2	15	3	2	42
			2	3	42

5.8.6 Weight Diagrams for the First Few $su(3)$ Representations

We begin this subsection with the preparatory remark that the overall normalization of the root vectors μ (and, correspondingly, that of the related fundamental weights) is arbitrary. We have chosen a normalization that facilitates comparison of the root vectors for $su(3)$, $sp(2)$, $sp(4)$, and $sp(6)$. See Chapter 27. For the normalization we have adopted, the $su(3)$ root vectors obey the relation

$$\sum_{\mu} (\mathbf{e}^i \cdot \mu)(\mu \cdot \mathbf{e}^j) = 6\delta_{ij}. \quad (5.8.22)$$

To continue our discussion, consider the representation $\Gamma(0, 0)$. According to (8.19) its highest weight is the vector zero, and according to (8.21) this representation is one dimensional. Thus, $\Gamma(0, 0)$ has only one weight vector. Figure 8.3 displays this vector in what is called a *weight diagram*. Since $\Gamma(0, 0)$ is one dimensional, and as described earlier, it is often referred to by its dimension, 1.

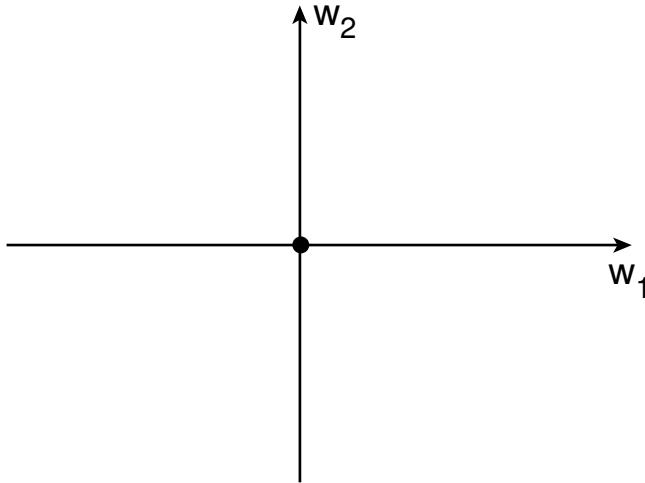


Figure 5.8.3: Weight diagram for the representation $1 = \Gamma(0,0)$.

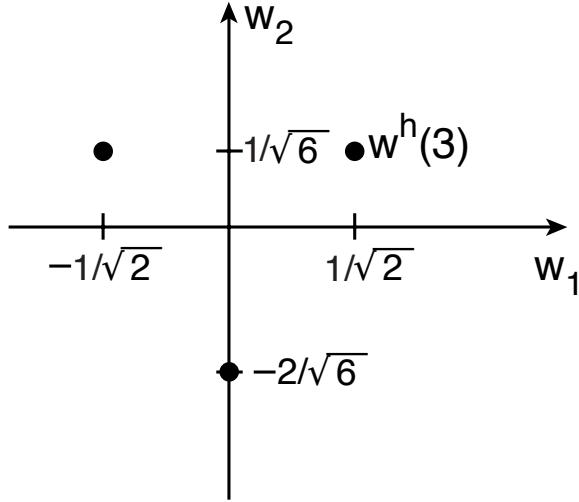
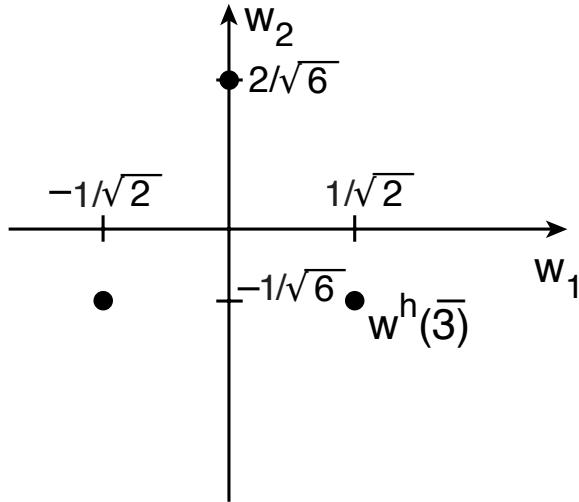
Consider the representation $\Gamma(1,0)$. The highest weight \mathbf{w}^h for this representation is shown in Figure 8.4. Also shown are all other weights obtained from \mathbf{w}^h by adding and subtracting integer multiples of α , β , and γ . We observe that there are 3 different weights. Correspondingly, in accord with (8.21), $\Gamma(1,0)$ is a 3-dimensional representation. It is often referred to by its dimension, 3.

Next consider the representation $\Gamma(0,1)$. Its weights are shown in Figure 8.5. Evidently this representation is also 3 dimensional. In view of (8.20) and (8.21), it is often referred to as $\bar{3}$. From (8.19) we find that the highest weights for the representations 3 and $\bar{3}$ are given by the relations

$$\mathbf{w}^h(3) = (1/\sqrt{2})\mathbf{e}^1 + (1/\sqrt{6})\mathbf{e}^2, \quad (5.8.23)$$

$$\mathbf{w}^h(\bar{3}) = (1/\sqrt{2})\mathbf{e}^1 - (1/\sqrt{6})\mathbf{e}^2. \quad (5.8.24)$$

Note also that all the weights for $\bar{3}$ are related to those for 3 by the operation of reflection across the w_1 axis. This is a general result. The weights for $\bar{\Gamma}(m,n)$ are related to those of $\Gamma(m,n)$ by reflection across the w_1 axis. It is a consequence of (8.20) and the fact that the fundamental weights ϕ^1 and ϕ^2 are interchanged by reflection across the \mathbf{e}^1 axis. See (8.17), (8.18), and Figure 8.2.

Figure 5.8.4: Weight diagram for the representation $3 = \Gamma(1, 0)$.Figure 5.8.5: Weight diagram for the representation $\bar{3} = \Gamma(0, 1)$.

Figures 8.6 through 8.8 show the weight diagrams for the representations $\Gamma(2, 0)$, $\Gamma(0, 2)$, and $\Gamma(1, 1)$. The highest weights in these cases are

$$\mathbf{w}^h(6) = (\sqrt{2})\mathbf{e}^1 + (2/\sqrt{6})\mathbf{e}^2, \quad (5.8.25)$$

$$\mathbf{w}^h(\bar{6}) = (\sqrt{2})\mathbf{e}^1 - (2/\sqrt{6})\mathbf{e}^2, \quad (5.8.26)$$

$$\mathbf{w}^h(8) = (\sqrt{2})\mathbf{e}^1 = \boldsymbol{\alpha}. \quad (5.8.27)$$

According to (8.21) the dimensionality of these representations are 6, 6, and 8, respectively. Observe that Figures 8.6 and 8.7 for $\Gamma(2, 0)$ and $\Gamma(0, 2)$ each contain 6 weights. Correspondingly, since 6 is the dimension of each of these representations, we conclude that the corresponding eigenvector for each weight \mathbf{w} is unique. By contrast, Figure 8.8 for $\Gamma(1, 1)$

only contains 7 weights while we know that the dimension of $\Gamma(1, 1)$ is 8. It can be shown that the eigenvectors corresponding to the 6 weight vectors \mathbf{w} in the diagram at the hexagonal vertices are nondegenerate. However, there are *two* linearly independent eigenvectors corresponding to the weight at the origin. That is, an additional label, beyond the weight itself, is necessary to completely specify these vectors. Note that $6 + 2 = 8$, the dimension of $\Gamma(1, 1)$.

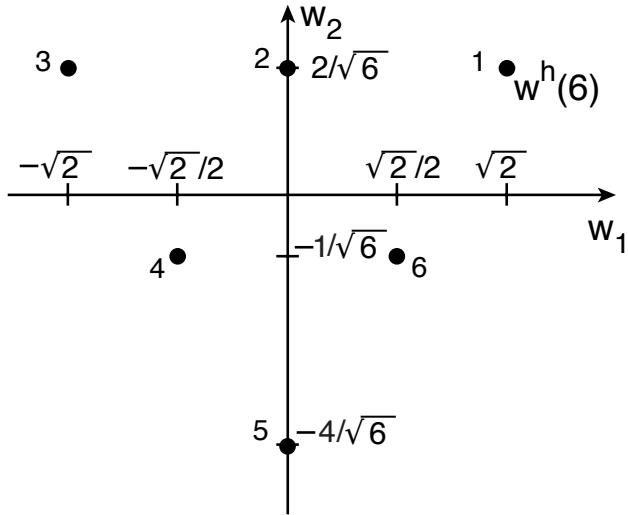


Figure 5.8.6: Weight diagram for the representation $6 = \Gamma(2, 0)$.

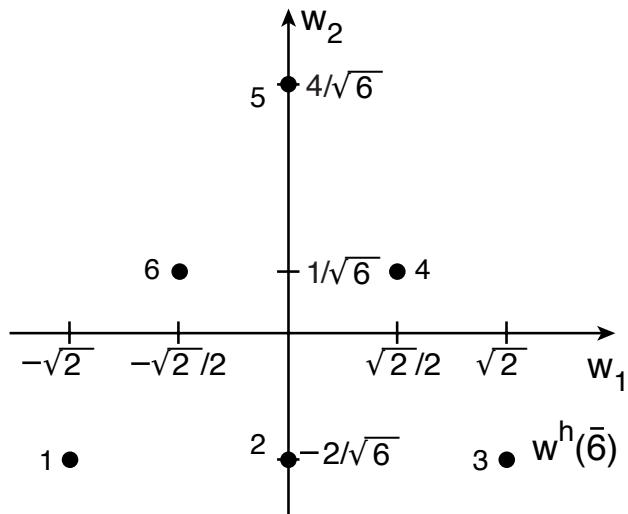


Figure 5.8.7: Weight diagram for the representation $\bar{6} = \Gamma(0, 2)$.

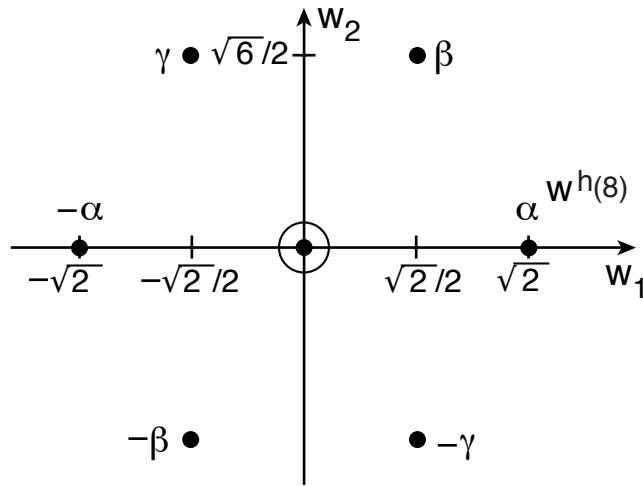


Figure 5.8.8: Weight diagram for the adjoint representation $8 = \Gamma(1, 1)$. The 6 weights at the hexagonal vertices lie at the tips of the root vectors $\pm\alpha, \pm\beta, \pm\gamma$ shown in Figure 8.1. The highest weight lies at the tip of the vector α . There are two eigenvectors corresponding to the weight at the origin.

5.8.7 Weight Diagram for the General $su(3)$ Representation

Consider the general representation $\Gamma(m, n)$. Figure 8.9 shows the general form of the weight diagram for this representation. It consists of concentric layers that may be constructed as follows:

1. Find and plot the highest weight \mathbf{w}^h using (8.19).
2. Plot the points $\mathbf{w}^h + \gamma, \mathbf{w}^h + 2\gamma, \dots, \mathbf{w}^h + n\gamma$ and the points $\mathbf{w}^h - \beta, \mathbf{w}^h - 2\beta, \dots, \mathbf{w}^h - m\beta$.
3. Reflect the points obtained in step 2 above across the w_2 axis and plot them.
4. Taken together, steps 2 and 3 produce the weights that lie on the left and right boundaries of the weight diagram. Now we need to fill in the top and bottom boundaries. They are the points $\mathbf{w}^h + n\gamma - \alpha, \mathbf{w}^h + n\gamma - 2\alpha, \dots, \mathbf{w}^h + n\gamma - m\alpha$ and $\mathbf{w}^h - m\beta - \alpha, \mathbf{w}^h - m\beta - 2\alpha, \dots, \mathbf{w}^h - m\beta - n\alpha$.
5. The weights on the boundary have now been found, and only the weights in the interior remain to be determined. Next, if the boundary is not triangular, find the point $\mathbf{w}^h - \alpha$. Starting from this point, form the next outermost layer by repeating steps 2 through 4 with (m, n) replaced by $(m - 1, n - 1)$.
6. Form successive concentric layers following step 5 starting from the points $\mathbf{w}^h - 2\alpha, \mathbf{w}^h - 3\alpha$, etc., until a triangular layer (or the origin) is reached. If a triangular layer is reached, all successive layers will also be triangles, and the innermost layer will be either the point at the origin or a triangle that is the same as one of those shown in Figures 8.4 through 8.7.

The result of this process is a set of weights that are related under translation in the directions $\pm\alpha, \pm\beta, \pm\gamma$. It can be shown that all eigenvectors $|\mathbf{w}\rangle$ corresponding to weights \mathbf{w} on a given layer have the same multiplicity. Those on the boundary have multiplicity 1 (are nondegenerate). If the boundary is not triangular, the eigenvectors $|\mathbf{w}\rangle$ corresponding to weights on the next outermost layer will have multiplicity 2. The multiplicity will continue to increase by 1 for each consecutive layer until a triangular layer (which may be the origin) is reached. The vectors $|\mathbf{w}\rangle$ corresponding to this layer will also have a multiplicity one unit larger than those of the previous layer. However, all vectors $|\mathbf{w}\rangle$ corresponding to consecutive layers *inside* the triangle will have the same multiplicity as those of the outermost triangle. That is, the multiplicity remains constant after a triangular layer is reached. For example, referring to Figure 8.9, the multiplicity of the eigenvectors $|\mathbf{w}\rangle$ corresponding to the boundary layer is 1, and the multiplicities for the next two layers in are 2 and 3 respectively. The multiplicity of the eigenvectors for the first triangular layer is 4, and the multiplicity for the triangular layer inside it is also 4.

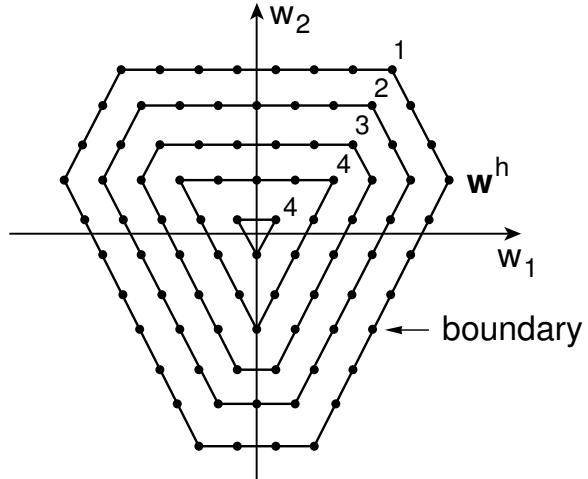


Figure 5.8.9: General form of the weight diagram for the representation $\Gamma(m, n)$. Shown here is the case $(m, n) = (7, 3)$. All eigenvectors $|\mathbf{w}\rangle$ corresponding to weights \mathbf{w} on a given layer have the same multiplicity. Those corresponding to sites on the boundary have multiplicity 1. Those corresponding to sites on the next two layers have multiplicities 2 and 3, respectively. Those corresponding to sites on the two triangular layers have multiplicity 4.

5.8.8 The Clebsch-Gordan Series for $su(3)$

Review of $su(2)$ Results

In the quantum-mechanical treatment of angular momentum, which is essentially an exercise in the properties of $su(2)$, there is the result that two spin $1/2$ entities can be combined to form entities with spin 0 and spin 1. If we denote the spin j representation of $su(2)$ by the symbols $\Gamma(j)$, then we may summarize this result by writing

$$\Gamma(1/2) \otimes \Gamma'(1/2) = \Gamma(0) \oplus \Gamma(1). \quad (5.8.28)$$

Here we have denoted the second spin $1/2$ entity on the left side of (8.28) by a prime to acknowledge that the two spin $1/2$ entities may be different. Indeed, the spin 0 combination [appearing on the right side of (8.28) and called the *singlet* state] is odd under the interchange/permuation of the spin $1/2$ entities, and the spin 1 combination (called the *triplet* state) is even under the interchange of the spin $1/2$ entities. Consequently, if the two spin $1/2$ entities are the same, the $\Gamma(0)$ entry on the right side of (8.28) is empty.

Similarly, two spin 1 entities can be combined to form entities with spins 0, 1, and 2; and we may summarize this result by writing

$$\Gamma(1) \otimes \Gamma'(1) = \Gamma(0) \oplus \Gamma(1) \oplus \Gamma(2). \quad (5.8.29)$$

If we view the two spin 1 entities on the left side of (8.29) as being the three-component vectors \mathbf{u} and \mathbf{v} then, for the entries on the right side of (8.29), there are the correspondences

$$\Gamma(0) \leftrightarrow \mathbf{u} \cdot \mathbf{v}, \quad (5.8.30)$$

$$\Gamma(1) \leftrightarrow \mathbf{u} \times \mathbf{v}, \quad (5.8.31)$$

$$\Gamma(2) \leftrightarrow (1/2)(u_a v_b + u_b v_a) - (1/3)\delta_{ab}(\mathbf{u} \cdot \mathbf{v}). \quad (5.8.32)$$

Note that $\Gamma(1)$ as given by (8.31) is odd under the interchange of \mathbf{u} and \mathbf{v} . Consequently the $\Gamma(1)$ entry is empty in the case that $\mathbf{u} = \mathbf{v}$. By contrast $\Gamma(2)$ (which is a symmetric traceless tensor) and $\Gamma(0)$ are even under the interchange of \mathbf{u} and \mathbf{v} .

We have given specific instances of the Clebsch-Gordan series for $su(2)$. It can be shown that for any two spins there is the general Clebsch-Gordan relation

$$\Gamma(j) \otimes \Gamma'(j') = \Gamma(j + j') \oplus \Gamma(j + j' - 1) \oplus \Gamma(j + j' - 2) \oplus \cdots \oplus \Gamma(|j - j'|). \quad (5.8.33)$$

Here all the representations on the right side occur once and only once unless some happen to be empty in the case of some possible interchange symmetry.

The Case of $su(3)$

We have briefly reviewed the Clebsch-Gordan series for representations of $su(2)$. There are similar combination rules for representations of $su(3)$ and, indeed, for all the representations of all the simple groups.⁸ The remaining task of this subsection is to review these rules for

⁸See Chapter 27 for the case of the symplectic group.

the case of $su(3)$. We begin with some specific cases. For some of the first few representations of $su(3)$ there are the results

$$3 \otimes 3' = \bar{3} \oplus 6, \quad (5.8.34)$$

$$\bar{3} \otimes \bar{3}' = 3 \oplus \bar{6}, \quad (5.8.35)$$

$$3 \otimes \bar{3} = 1 \oplus 8, \quad (5.8.36)$$

$$6 \otimes 6' = \Gamma(2, 0) \otimes \Gamma'(2, 0) = \Gamma(0, 2) \oplus \Gamma(2, 1) \oplus \Gamma(4, 0), \quad (5.8.37)$$

$$6 \otimes \bar{6} = 1 \oplus 8 \oplus 27, \quad (5.8.38)$$

$$8 \otimes 6 = \Gamma(1, 1) \otimes \Gamma(2, 0) = \Gamma(0, 1) \oplus \Gamma(2, 0) \oplus \Gamma(1, 2) \oplus \Gamma(3, 1), \quad (5.8.39)$$

$$8 \otimes 8' = 1 \oplus 8 \oplus 8 \oplus 10 \oplus \bar{10} \oplus 27. \quad (5.8.40)$$

In (8.37) and (8.39) the $\Gamma(m, n)$ notation is used to specify a representation since not all the representations appearing in (8.37) and (8.39) are uniquely specified by their dimensions. See Table 8.3. We also remark that if the two factors appearing on the left side of a Clebsch-Gordan relation are potentially the same, as for example in (8.34), (8.35), (8.37), and (8.40), then some of the terms appearing on the right side are potentially empty. See, for example, Exercises 8.21 and 8.23.

Just as is the case for $su(2)$ where (8.33) provides an explicit result for combining any two spins, there is also an explicit formula for the general case of $su(3)$. Use the shorthand notation (j_1, j_2) to denote the $su(3)$ representation $\Gamma(j_1, j_2)$. The Clebsch-Gordan series for $su(3)$ in the general case is given by the relation

$$(j_1, j_2) \otimes (j'_1, j'_2)' = \sum_{i=0}^{\min(j_1, j'_2)} \sum_{k=0}^{\min(j_2, j'_1)} (j_1 - i, j'_1 - k; j_2 - k, j'_2 - i), \quad (5.8.41)$$

where the quantity $(n, n'; m, m')$ is defined by the relation

$$\begin{aligned} (n, n'; m, m') &= (n + n', m + m') \oplus \sum_{i=1}^{\min(n, n')} (n + n' - 2i, m + m' + i) \\ &\oplus \sum_{k=1}^{\min(m, m')} (n + n' + k, m + m' - 2k). \end{aligned} \quad (5.8.42)$$

All the sums in the expressions above are direct sums.

5.8.9 Representations of $su(3)$: the Approach of Schur and Weyl

Subsections 8.4 through 8.7 have illustrated how, following Cartan, the representations of $su(3)$ can be described in terms of ladder operators and weight diagrams. We will employ the same approach in Chapter 27 for the case of $sp(2n)$. However, we take here the opportunity to mention an alternate approach due to Schur (1875-1941) and Weyl.

The method of Cartan has the feature that the properties of any given representation are described without reference to the properties of any other representation. Each representation is treated in isolation. By contrast, the approach of Schur and Weyl capitalizes

on the fact (as illustrated by the Clebsch-Gordan series) that suitable tensor products of low-dimensional representations contain higher-dimensional representations. In particular it can be shown for the case of $su(3)$ that by forming suitable tensor products of multiple copies of $3 = \Gamma(1, 0)$ and $\bar{3} = \Gamma(0, 1)$ one obtains a multi-index tensor representation of $su(3)$ that contains any desired irreducible representation. With this result in hand, the remaining task is to extract and label from such a representation the desired irreducible representation. This is done by possibly tracing over some index pairs and by forming linear combinations over various permutations of other indices with these linear combinations being described by *Young (1873-1940) tableaux*. Thus, in the approach of Schur and Weyl, each representation is labeled by a Young tableau.

5.8.10 Remaining Polynomials

With this brief background on representation theory for $su(3)$, we are prepared to construct 12 second-degree polynomials corresponding to the matrices JS^a in such a way that these polynomials have convenient transformation properties under $su(3)$.

$su(3)$ Decomposition of Homogenous Polynomials

First we state a general result: Let f_ℓ be a homogeneous polynomial of degree ℓ in the six phase-space variables $z_1 \cdots z_6$. Then it is easily verified that the quantities : b^1 : f_ℓ through : b^8 : f_ℓ are also homogeneous polynomials of degree ℓ . Consequently, the subspace of homogeneous polynomials of degree ℓ is sent into itself under the action of $su(3)$. Next, it can be shown that each f_ℓ subspace can itself be decomposed into smaller subspaces that are each sent into themselves separately under the action of $su(3)$. Indeed, this can be done in such a way that each smaller subspace forms an irreducible representation of $su(3)$. See Section 34.2.4. When this is done, the following results are found:

1. Suppose ℓ is even. Then f_ℓ has the direct sum decomposition

$$f_\ell = \sum_{m+n=\ell} \Gamma(m, n) \oplus \sum_{m+n=\ell-2} \Gamma(m, n) \oplus \sum_{m+n=\ell-4} \Gamma(m, n) \oplus \cdots \oplus \Gamma(0, 0). \quad (5.8.43)$$

Each representation listed in (8.43) occurs once and only once. For example, f_0 , f_2 , and f_4 have the decompositions

$$f_0 = \Gamma(0, 0), \quad (5.8.44)$$

$$f_2 = \Gamma(2, 0) \oplus \Gamma(1, 1) \oplus \Gamma(0, 2) \oplus \Gamma(0, 0), \quad (5.8.45)$$

$$f_4 = \Gamma(4, 0) \oplus \Gamma(3, 1) \oplus \Gamma(2, 2) \oplus \Gamma(1, 3) \oplus \Gamma(0, 4) \oplus \Gamma(2, 0) \oplus \Gamma(1, 1) \oplus \Gamma(0, 2) \oplus \Gamma(0, 0). \quad (5.8.46)$$

2. Suppose ℓ is odd. Then f_ℓ has the direct sum decomposition

$$\begin{aligned} f_\ell = & \sum_{m+n=\ell} \Gamma(m, n) \oplus \sum_{m+n=\ell-2} \Gamma(m, n) \oplus \sum_{m+n=\ell-4} \Gamma(m, n) \oplus \cdots \\ & \oplus \Gamma(1, 0) \oplus \Gamma(0, 1). \end{aligned} \quad (5.8.47)$$

Each representation listed in (8.47) occurs once and only once. For example, f_1 and f_3 have the decompositions

$$f_1 = \Gamma(1, 0) \oplus \Gamma(0, 1), \quad (5.8.48)$$

$$f_3 = \Gamma(3, 0) \oplus \Gamma(2, 1) \oplus \Gamma(1, 2) \oplus \Gamma(0, 3) \oplus \Gamma(1, 0) \oplus \Gamma(0, 1). \quad (5.8.49)$$

We will use these results below for the special case of quadratic polynomials. But we remark that these results are also useful for the construction of Cremona maps and determining the long-term behavior of particles in storage rings. Again see Section 34.2.4.

Explicit Results for Remaining Quadratic Polynomials

For our present discussion we are interested in the case of quadratic polynomials, the generators of $sp(6)$. According to the previous paragraph, they have the decomposition (8.45). It can be shown that the $\Gamma(0, 0)$ part in (8.45) corresponds to a b^0 part as given in (8.5), and the $\Gamma(1, 1)$ part corresponds to the b^1 through b^8 parts given in (8.5). See Exercise 8.19. What remains is the $\Gamma(2, 0) \oplus \Gamma(0, 2)$ part. It has dimension $6 + 6 = 12$, which is the dimension of the set of matrices JS^a . This circumstance suggests that the second-degree polynomials corresponding to the matrices JS^a might be arranged to transform under the action of $su(3)$ according to the representation $\Gamma(2, 0) \oplus \Gamma(0, 2) = 6 \oplus \bar{6}$. This is indeed the case. As a sanity check on our hypothesis, let us do a dimension count. We know that $sp(6)$ has dimension 21. Together b^0 and the b^1 through b^8 span a space of dimension $1 + 8 = 9$. Observe that $9 + 6 + 6 = 21$, as desired.

Let the symbols c^j and $r(\boldsymbol{\mu})$ denote the second-degree polynomials corresponding to the C^j and $R(\boldsymbol{\mu})$. They are selected and normalized in such a way that their Lie algebra (with the Poisson bracket as the Lie product) is the same as the Lie algebra of the C^j and $R(\boldsymbol{\mu})$ (with the commutator as the Lie product). Calculation shows that they are given by the relations

$$c^1 = -(i/\sqrt{2})b^3 = (-i/\sqrt{8})(q_1^2 + p_1^2 - q_2^2 - p_2^2), \quad (5.8.50)$$

$$c^2 = -(i/\sqrt{2})b^8 = (-i/\sqrt{24})(q_1^2 + p_1^2 + q_2^2 + p_2^2 - 2q_3^2 - 2p_3^2);$$

$$r(\pm\boldsymbol{\alpha}) = (i/2)(b^1 \pm ib^2) = (i/2)(q_1 \pm ip_1)(q_2 \mp ip_2), \quad (5.8.51)$$

$$r(\pm\boldsymbol{\beta}) = (i/2)(b^4 \pm ib^5) = (i/2)(q_1 \pm ip_1)(q_3 \mp ip_3),$$

$$r(\pm\boldsymbol{\gamma}) = (i/2)(b^6 \pm ib^7) = (i/2)(q_2 \pm ip_2)(q_3 \mp ip_3).$$

Define six weight vectors $\mathbf{w}^1 \cdots \mathbf{w}^6$ for $\Gamma(2, 0)$ by the rules

$$\begin{aligned} \mathbf{w}^1 &= \mathbf{w}^h(6) , \quad \mathbf{w}^2 = \mathbf{w}^1 - \boldsymbol{\alpha}, \\ \mathbf{w}^3 &= \mathbf{w}^2 - \boldsymbol{\alpha} , \quad \mathbf{w}^4 = \mathbf{w}^3 - \boldsymbol{\gamma}, \\ \mathbf{w}^5 &= \mathbf{w}^4 - \boldsymbol{\gamma} , \quad \mathbf{w}^6 = \mathbf{w}^5 + \boldsymbol{\beta}. \end{aligned} \quad (5.8.52)$$

See Figures 8.1 and 8.6, and note that the weights shown in Figure 8.6 are numbered in accord with (8.52). Define six corresponding polynomials $h^1 \cdots h^6$ by the relations

$$h^1 = (1/2)(q_1 + ip_1)^2,$$

$$\begin{aligned}
h^2 &= (q_1 + ip_1)(q_2 + ip_2), \\
h^3 &= (1/2)(q_2 + ip_2)^2, \\
h^4 &= (q_2 + ip_2)(q_3 + ip_3), \\
h^5 &= (1/2)(q_3 + ip_3)^2, \\
h^6 &= (q_3 + ip_3)(q_1 + ip_1).
\end{aligned} \tag{5.8.53}$$

It is easy to check that the h^k are all simultaneous eigenvectors of the $:c^j:$ with eigenvalues corresponding to the weights \mathbf{w}^k ,

$$:c^j:h^k = (\mathbf{e}^j \cdot \mathbf{w}^k)h^k. \tag{5.8.54}$$

Also, there are ladder relations, corresponding to the relations (8.52), of the form

$$\begin{aligned}
h^2 &\propto r(-\boldsymbol{\alpha}):h^1, \\
h^3 &\propto r(-\boldsymbol{\alpha}):h^2, \quad h^4 \propto r(-\boldsymbol{\gamma}):h^3, \\
h^5 &\propto r(-\boldsymbol{\gamma}):h^4, \quad h^6 \propto r(+\boldsymbol{\beta}):h^5.
\end{aligned} \tag{5.8.55}$$

Finally, calculation shows that the action of b^0 is given by the relation

$$:b^0:h^k = 2ih^k. \tag{5.8.56}$$

We conclude that the six polynomials $h^1 \dots h^6$ transform according to the representation 6 under the action of $su(3)$, and also are transformed among each other under the action of the full $u(3)$. See Exercise 8.14.

With the h^k determined, the construction of a second set of six polynomials corresponding to the representation $\bar{6}$ is easy. Take the complex conjugate of both sides of the relations (8.54) and (8.56). Doing so gives the results

$$:\bar{c}^j:\bar{h}^k = (\mathbf{e}^j \cdot \mathbf{w}^k)\bar{h}^k, \tag{5.8.57}$$

$$:\bar{b}^0:\bar{h}^k = -2i\bar{h}^k. \tag{5.8.58}$$

However, inspection of (8.5) and (8.50) gives the relations

$$\bar{b}^0 = b^0, \quad \bar{c}^j = -c^j. \tag{5.8.59}$$

Consequently, we also have the results

$$:c^j:\bar{h}^k = -(\mathbf{e}^j \cdot \mathbf{w}^k)\bar{h}^k, \tag{5.8.60}$$

$$:b^0:\bar{h}^k = -2i\bar{h}^k. \tag{5.8.61}$$

Upon comparing the weight diagrams in Figures 8.6 and 8.7 for the representations 6 and $\bar{6}$, we conclude that the polynomials \bar{h}^k transform according to the representation $\bar{6}$.

Our task of finding a suitable set of polynomials corresponding to the matrices JS^a is almost finished. For physical applications, we will want to work with real polynomials

instead of the complex polynomials h^k given by (8.53). This task is easily accomplished. Define real polynomials f^k and g^k by writing the relations

$$h^k = f^k + ig^k. \quad (5.8.62)$$

Doing so gives the results

$$\begin{aligned} f^1 &= (1/2)(q_1^2 - p_1^2), \\ f^2 &= q_1q_2 - p_1p_2, \\ f^3 &= (1/2)(q_2^2 - p_2^2), \\ f^4 &= q_2q_3 - p_2p_3, \\ f^5 &= (1/2)(q_3^2 - p_3^2), \\ f^6 &= q_3q_1 - p_3p_1; \\ g^1 &= q_1p_1, \\ g^2 &= q_1p_2 + q_2p_1, \\ g^3 &= q_2p_2, \\ g^4 &= q_2p_3 + q_3p_2, \\ g^5 &= q_3p_3, \\ g^6 &= q_3p_1 + q_1p_3. \end{aligned} \quad (5.8.63)$$

Since the h^k are transformed among themselves under the action of the full $u(3)$, the Poisson brackets $[b^j, h^k]$ can be written in the form

$$[b^j, h^k] = \sum_{\ell} \zeta_{jk\ell} h^{\ell}. \quad (5.8.64)$$

The results for the cases $j = 0, 3, 8$ follow from (8.56), (8.54), and (8.50):

$$\begin{aligned} [b^0, h^k] &= 2ih^k, \\ [b^3, h^k] &= i(\sqrt{2})(\mathbf{e}^1 \cdot \mathbf{w}^k)h^k, \\ [b^8, h^k] &= i(\sqrt{2})(\mathbf{e}^2 \cdot \mathbf{w}^k)h^k. \end{aligned} \quad (5.8.65)$$

Calculation of the other Poisson brackets requires somewhat more work, the results of which will be presented shortly in tabular form. Suppose the coefficients $\zeta_{jk\ell}$ are decomposed into real and imaginary parts by writing the relations

$$\zeta_{jk\ell} = \xi_{jk\ell} + i\eta_{jk\ell}. \quad (5.8.66)$$

Then equating real and imaginary parts of (8.64), observing that the b^j are real, and using the decomposition (8.62) give the results

$$[b^j, f^k] = \sum_{\ell} \xi_{jk\ell} f^{\ell} - \eta_{jk\ell} g^{\ell}, \quad (5.8.67)$$

$$[b^j, g^k] = \sum_{\ell} \eta_{jk\ell} f^{\ell} + \xi_{jk\ell} g^{\ell}.$$

The nonzero values of the $\xi_{jk\ell}$ and $\eta_{jk\ell}$ are tabulated below.

Table 5.8.4: Some Structure Constants of $sp(6)$.

$jk\ell$	$\xi_{jk\ell}$	$\eta_{jk\ell}$	$jk\ell$	$\xi_{jk\ell}$	$\eta_{jk\ell}$	$jk\ell$	$\xi_{jk\ell}$	$\eta_{jk\ell}$
011	0	2	311	0	2	634	0	1
022	0	2	333	0	-2	643	0	2
033	0	2	344	0	-1	645	0	2
044	0	2	366	0	1	654	0	1
055	0	2	416	0	1	662	0	1
066	0	2	424	0	1	726	-1	0
112	0	1	442	0	1	734	-1	0
121	0	2	456	0	1	743	2	0
123	0	2	461	0	2	745	-2	0
132	0	1	465	0	2	754	1	0
146	0	1	516	-1	0	762	1	0
164	0	1	524	-1	0	811	0	$2/\sqrt{3}$
212	-1	0	542	1	0	822	0	$2/\sqrt{3}$
221	2	0	556	1	0	833	0	$2/\sqrt{3}$
223	-2	0	561	2	0	844	0	$-1/\sqrt{3}$
232	1	0	565	-2	0	855	0	$-4/\sqrt{3}$
246	1	0	626	0	1	866	0	$-1/\sqrt{3}$
264	-1	0						

It remains to compute the Poisson brackets of the f 's and g 's with themselves. First we observe, as can be easily verified, that the h^k are all in involution,

$$[h^j, h^k] = 0. \quad (5.8.68)$$

Next, since the commutator of two matrices of the form JS^a is a matrix of the form JS^c (see Exercise 3.9.1), we must have a relation of the form

$$[h^j, \bar{h}^k] = \sum_{\ell} \tau_{jk\ell} b^{\ell}. \quad (5.8.69)$$

Using the decomposition (8.62) and taking real and imaginary parts of (8.68) give the results

$$[f^j, f^k] = [g^j, g^k], \quad (5.8.70)$$

$$[f^j, g^k] = [f^k, g^j].$$

To complete the calculation, decompose $\tau_{jk\ell}$ into real and imaginary parts by writing the relations

$$\tau_{jk\ell} = \rho_{jk\ell} + i\sigma_{jk\ell}. \quad (5.8.71)$$

Then taking real and imaginary parts of (8.69) and using (8.70) give the results

$$[f^j, f^k] = [g^j, g^k] = +(1/2) \sum_{\ell} \rho_{jk\ell} b^{\ell}, \quad (5.8.72)$$

$$[f^j, g^k] = -(1/2) \sum_{\ell} \sigma_{jk\ell} b^\ell.$$

Note that by (8.70) and (8.72), $\rho_{jk\ell}$ is antisymmetric in its first two indices, and $\sigma_{jk\ell}$ is symmetric,

$$\rho_{jk\ell} = -\rho_{kj\ell}, \quad (5.8.73)$$

$$\sigma_{jk\ell} = \sigma_{kj\ell}.$$

The table below lists the needed values of $\rho_{jk\ell}$ and $\sigma_{jk\ell}$. All the rest are zero, or can be obtained from the symmetry conditions (8.73). Taken together, (8.6), (8.7), (8.67), and (8.72) specify the Lie algebra $sp(6)$ in all its beauty.

Table 5.8.5: Remaining Structure Constants of $sp(6)$.

$jk\ell$	$\rho_{jk\ell}$	$\sigma_{jk\ell}$	$jk\ell$	$\rho_{jk\ell}$	$\sigma_{jk\ell}$	$jk\ell$	$\rho_{jk\ell}$	$\sigma_{jk\ell}$
110	0	-4/3	245	2	0	456	0	-2
113	0	-2	266	0	-2	457	2	0
118	0	-2/ $\sqrt{3}$	267	2	0	461	0	-2
121	0	-2	330	0	-4/3	462	-2	0
122	2	0	333	0	2	550	0	-4/3
164	0	-2	338	0	-2/ $\sqrt{3}$	558	0	4/ $\sqrt{3}$
165	2	0	346	0	-2	564	0	-2
220	0	-8/3	347	2	0	565	-2	0
228	0	-4/ $\sqrt{3}$	440	0	-8/3	660	0	-8/3
231	0	-2	443	0	2	663	0	2
232	2	0	448	0	2/ $\sqrt{3}$	668	0	2/ $\sqrt{3}$
244	0	-2						

Closing Remarks

We close this section with two remarks. First, we note that the $sp(6)$ polynomials c^1 , $r(\pm\alpha)$, h^1 , h^2 , and h^3 are the same (up to normalizations) as the $sp(4)$ polynomials c , $r(\pm)$, h^\pm , and h^0 . This correspondence indicates, as expected, that $sp(6)$ contains $sp(4)$ as a subgroup.

The second remark concerns $su(3)$. As mentioned earlier, it can be shown that the quadratic polynomials b^0 through b^8 , which correspond to the Gell-Mann matrices λ^0 through λ^8 , transform under the action of $su(3)$ according to the representations $1 = \Gamma(0,0)$ and $8 = \Gamma(1,1)$. See Exercise 8.19.

Now look at the relation (8.3) and compare it with (8.40). The left side of (8.3) may be viewed as the antisymmetric part of a second-order tensor consisting of ingredients which each transform according to the representation 8, and the right side of (8.3) is a sum of entities which also transform according to the representation 8. This result is consistent with the relation (8.40), which states that the tensor product of an 8 and an 8 is expected to contain an 8. Moreover the coefficients $f_{jk\ell}$, which are the structure constants for $su(3)$,

specify which $su(3)$ elements occur in each entry in the antisymmetric part of the second-order tensor product. Therefore these coefficients may also be viewed as particular instances of the Clebsch-Gordan coefficients for $su(3)$, namely those associated with the tensor product of the adjoint representation with itself! Indeed, the structure constants of any Lie algebra may be viewed as particular instances of the Clebsch-Gordan coefficients for that algebra, specifically those associated with the tensor product of the adjoint representation with itself.

Next look at (8.4). The left side of (8.4) may be viewed as the symmetric part of a second-order tensor consisting of ingredients which each transform according to the representation 8, and the right side of (8.4) is a sum of entities which transform according to the representations 1 and 8. This result is also consistent with the relation (8.40), which states that the tensor product of an 8 and an 8 is also expected to contain a 1 and a second 8. Thus the coefficients δ_{jk} and d_{jkl} are also particular instances of the Clebsch-Gordan coefficients for $su(3)$.

What about the entries 10, $\overline{10}$, and 27 which occur in (8.40) but do not occur in (8.3) and (8.4) and their composite (8.78)? They do not occur because both factors on the left sides of (8.3), (8.4), and (8.78) involve the same 8 rather than an 8 and an $8'$.

Note that in general the kind of Clebsch-Gordan analysis we have been making only predicts what representations can possibly occur in a product when each factor in the product has known transformation properties. For example, consider all entities of the form $b^i b^j$ where i and j range from 1 to 8. Since each factor belongs to an 8, the product can possibly contain the representations appearing on the right side of (8.40). But each entity is also a homogeneous polynomial of degree 4, and therefore can potentially have the $su(3)$ content given by (8.46). Observe that $10 = \Gamma(3, 0)$ and $\overline{10} = \Gamma(0, 3)$ do not occur in (8.46), but $27 = \Gamma(2, 2)$ does.

Exercises

5.8.1. Suppose, for the purposes of this exercise, that λ^0 is redefined by the relation

$$\lambda^0 = (2/3)^{1/2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (5.8.74)$$

Show that the matrices $\lambda^0 \cdots \lambda^8$ span the vector space of all 3×3 Hermitian matrices. Show that λ^0 , together with the Gell-Mann matrices, obey the relations

$$tr(\lambda^j \lambda^k) = 2\delta_{jk}; \quad j, k = 0, 1 \cdots 8. \quad (5.8.75)$$

5.8.2. Show that the commutator of two Gell-Mann matrices must be of the general form (8.3) with the f_{jkl} real. Use (8.3) and (8.75) to derive the relation

$$f_{jkl} = -(i/4)tr(\{\lambda^j, \lambda^k\}\lambda^l); \quad j, k, l = 1, 2, \cdots 8. \quad (5.8.76)$$

Prove from this relation that f_{jkl} is antisymmetric under the interchange of any two (adjacent) indices. Verify Table 8.1.

5.8.3. Show that the anticommutator of two Gell-Mann matrices must be of the general form (8.4) with the $d_{jk\ell}$ real. Use (8.4) and (8.75) to derive the relation

$$d_{jk\ell} = (1/4)\text{tr}(\{\lambda^j, \lambda^k\}_+ \lambda^\ell); \quad j, k, \ell = 1, 2, \dots, 8. \quad (5.8.77)$$

Prove from this relation that $d_{jk\ell}$ is symmetric under the interchange of any two indices. Verify Table 8.2.

5.8.4. Show that the Gell-Mann matrices obey the multiplication rules

$$\lambda^j \lambda^k = (2/3)\delta_{jk}\lambda^0 + \sum_\ell (d_{jk\ell} + i f_{jk\ell})\lambda^\ell; \quad j, k, \ell = 1, 2, \dots, 8. \quad (5.8.78)$$

5.8.5. Verify the results (8.5).

5.8.6. Verify the Poisson bracket rules (8.6) and (8.7).

5.8.7. Show that the Cartan-basis matrices given by (8.8) and (8.11) are *real* and satisfy the relations

$$(C^j)^\dagger = C^j, \quad (5.8.79)$$

$$R(\boldsymbol{\mu})^\dagger = R(-\boldsymbol{\mu}). \quad (5.8.80)$$

5.8.8. Verify the Cartan-basis commutation rules (8.9) and (8.12) through (8.14).

5.8.9. Verify that the root vectors given by (8.10) and their negatives satisfy the relation (8.22).

5.8.10. Verify that the Clebsch-Gordan relations (8.34) through (8.40) are specific cases of the general Clebsch-Gordan relation given by (8.41) and (8.42).

5.8.11. Look at the Clebsch-Gordan relations (8.34) through (8.40). As a sanity check, the dimensions of the left and right sides should agree. For example, the dimension (the number of entities) on the left side of (8.34) is $3 \times 3 = 9$. And the dimension of the right side of (8.34) is $3 + 6 = 9$. Verify that analogous results hold for (8.35) through (8.40). If you are algebraically ambitious, verify that analogous results hold in the general case described by (8.41) and (8.42) using (8.21).

5.8.12. Look at the relations (8.43) through (8.49) for the case of a six-dimensional phase space. As a sanity check, verify that the dimensions of the left and right sides agree using (7.3.40) and (8.21).

5.8.13. Review the relations (8.50) and (8.51). Verify the relation

$$r(-\boldsymbol{\mu}) = -\bar{r}(\boldsymbol{\mu}). \quad (5.8.81)$$

Next consider the Lie operators associated with the c^j and $r(\boldsymbol{\mu})$. It can be shown that a suitable scalar product can be defined so that, in analogy to (8.79) and (8.80), they satisfy the relations (7.3.22) and (7.3.23). See Section 7.3. Review Exercise 8.8. Show that the Lie operators associated with the c^j and $r(\boldsymbol{\mu})$ obey the same commutation rules as the C^j and $R(\boldsymbol{\mu})$.

5.8.14. The purpose of this exercise is to construct the polynomials (8.53) corresponding to matrices of the form JS^a . Consider some second-degree polynomial corresponding to some matrix of the form JS^a . For example, we may set $B = 0$ and $A = I_3$ in (7.10). Show that use of (7.10) and (5.1) then produces a polynomial, call it a^1 , given by the relation

$$a^1 = (1/2)(q_1^2 - p_1^2 + q_2^2 - p_2^2 + q_3^2 - p_3^2). \quad (5.8.82)$$

If our suspicions about polynomials associated with the JS^a belonging to the representation $\Gamma(2, 0) \oplus \Gamma(0, 2)$ are correct, the polynomial a^1 should be some linear combination of polynomials corresponding to the weights of Figures 8.6 and 8.7. With luck, it may be possible to produce a polynomial corresponding to the highest weights $\mathbf{w}^h(6)$ and $\mathbf{w}^h(\bar{6})$ by repeatedly applying $:r(\alpha):$ to a^1 . See Figure 8.1. Verify the results

$$a^2 =: r(\alpha) : a^1 = [r(\alpha), a^1] = (-i)(q_1 p_2 + q_2 p_1), \quad (5.8.83)$$

$$a^3 =: r(\alpha) : a^2 = (1/2)[(q_1 + ip_1)^2 + (q_2 - ip_2)^2], \quad (5.8.84)$$

$$a^4 =: r(\alpha) : a^3 = 0. \quad (5.8.85)$$

The relation (8.85) shows that a^3 cannot be raised any further in the α direction, and suggests that a^3 is, as desired, a polynomial corresponding to the highest weights $\mathbf{w}^h(6)$ and $\mathbf{w}^h(\bar{6})$. Indeed, show that

$$:c^1 : a^3 = [c^1, a^3] = (\sqrt{2})a^3 = [\mathbf{e}^1 \cdot \mathbf{w}^h(6)]a^3 = [\mathbf{e}^1 \cdot \mathbf{w}^h(\bar{6})]a^3. \quad (5.8.86)$$

Finally, the components of a^3 corresponding to $\mathbf{w}^h(6)$ and $\mathbf{w}^h(\bar{6})$ separately can be removed from a^3 by using the operators $[:c^2 : -\mathbf{e}^2 \cdot \mathbf{w}^h(6)]$ and $[:c^2 : -\mathbf{e}^2 \cdot \mathbf{w}^h(\bar{6})]$, respectively. Do so by defining further polynomials a^5 and a^6 by the rules

$$a^5 = [:c^2 : -\mathbf{e}^2 \cdot \mathbf{w}^h(6)]a^3 = -(2/\sqrt{6})(q_2 - ip_2)^2, \quad (5.8.87)$$

$$a^6 = [:c^2 : -\mathbf{e}^2 \cdot \mathbf{w}^h(\bar{6})]a^3 = (2/\sqrt{6})(q_1 + ip_1)^2. \quad (5.8.88)$$

Verify (8.87) and (8.88), and show that these polynomials are simultaneous eigenvectors of both the $:c^j:$,

$$\begin{aligned} :c^j : a^5 &= [\mathbf{e}^j \cdot \mathbf{w}^h(\bar{6})]a^5, \\ :c^j : a^6 &= [\mathbf{e}^j \cdot \mathbf{w}^h(6)]a^6. \end{aligned} \quad (5.8.89)$$

Now verify that

$$h^1 \propto a^6, \quad (5.8.90)$$

and verify the results (8.55). The particular normalizations used in defining the h^k as given in (8.53) have been chosen for convenience.

5.8.15. Verify the relations (8.54) and (8.56).

5.8.16. Verify the relations (8.57) through (8.61).

5.8.17. Verify Table 8.4.

5.8.18. Verify Table 8.5.

5.8.19. Thanks to the work of Subsection 8.10, we know that the quadratic polynomials h^k transform under $su(3)$ according to the representation 6, and the \bar{h}^k transform according to $\bar{6}$. The purpose of this exercise is to study how the remaining polynomials b^j transform. Recall that the c^j and $r(\mu)$ defined by (8.50) and (8.51) satisfy the same Lie algebra as the C^j and $R(\mu)$. See Exercise 8.13.

Begin by considering the polynomial b^0 . The relations (8.6) can be rewritten in the form

$$: b^j : b^0 = [b^j, b^0] = 0. \quad (5.8.91)$$

The relations (8.91) may be understood to say that b^0 transforms according to the representation $\Gamma(0, 0)$. Show from (8.21) that $\dim \Gamma(0, 0) = 1$, and from (8.19) that $w^h(1) = 0$.

What about the remaining 8 polynomials b^1, \dots, b^8 ? Show that $\dim \Gamma(1, 1) = 8$, and that $w^h(8) = \alpha$. Figure 8.8 displays the weight diagram for the representation $8 = \Gamma(1, 1)$. There are 6 points on the vertices of a hexagon at the ends of the vectors $\pm \alpha, \pm \beta, \pm \gamma$. In addition, there are 2 eigenvectors corresponding to the weight at the origin (indicated by a dot and a concentric circle) to make a total of $6+2=8$ states. Verify the relations

$$: c^j : r(\nu) = (\mathbf{e}^j \cdot \nu) r(\nu), \quad (5.8.92)$$

$$: r(\mu) : r(\nu) = N(\mu, \nu) r(\mu + \nu), \quad (\text{when } \mu + \nu \text{ is a root vector}). \quad (5.8.93)$$

The relations (8.92) indicate that each $r(\nu)$ has a weight ν corresponding to a particular vertex of the hexagon, and the relations (8.93) indicate that the $: r(\mu) :$ act on the $r(\nu)$ to produce polynomials with raised and lowered weights. Also, verify the relations

$$: c^j : c^k = 0, \quad (5.8.94)$$

$$: r(\mu) : c^k = -(\mathbf{e}^k \cdot \mu) r(\mu). \quad (5.8.95)$$

The relations (8.94) indicate that c^1 and c^2 correspond to the two eigenvectors for the weight at the origin of the weight diagram, and the relations (8.95) indicate that the $: r(\mu) :$ raise and lower these eigen vectors. Finally, show that the polynomials b^1, \dots, b^8 are related to the c^j and $r(\mu)$ by a nonsingular matrix.

In summary, we conclude that the polynomial b^0 transforms under $su(3)$ according to the representation 1, and the 8 polynomials b^1, \dots, b^8 transform according to the representation 8. The representation $8 = \Gamma(1, 1)$ is called the *adjoint* or *regular* representation because it arises from the action of the Lie algebra on itself. See the discussion at the end of Section 3.7.

5.8.20. Consider the first-degree polynomials t^1, t^2, t^3 defined by the relations

$$t^j = q_j + i p_j. \quad (5.8.96)$$

Consider also the representation $\Gamma(1, 0)$. Show that $\dim \Gamma(1, 0) = 3$ and compute $w^h(3)$. Figure 8.4 shows the weight vectors for $\Gamma(1, 0)$. Show that the t^j transform under $su(3)$ according to the representation 3. Also compute $: b^0 : t^j$. Figure 8.5 shows the weight vectors for $\Gamma(0, 1)$. Show that the \bar{t}^j transform under $su(3)$ according to the representation $\bar{3}$. Also compute $: b^0 : \bar{t}^j$. It follows that the 6 monomials $q_1, q_2, q_3, p_1, p_2, p_3$ transform according to the representation $3 \oplus \bar{3}$. This is in accord with (8.48).

5.8.21. Consider the vector space spanned by the quadratic polynomials of the form $t^j \times t^k$. See (8.96). Show that this vector space is 6 dimensional, and is spanned by the polynomials h^ℓ of (8.53). Exercise 8.20 showed that the t^j transform under $su(3)$ according to the representation 3, and the \bar{t}^j transform under $su(3)$ according to the representation $\bar{3}$. It follows from group theory and the derivation property (3.7) that the products $t^j \times t^k$ must transform as some portion of the direct product representation $3 \otimes 3$. In the case of $su(3)$, the general direct product representation $3 \otimes 3'$ has the Clebsch-Gordan series decomposition (8.34). For the present application, both “3” factors on the left of (8.34) are the same, and correspondingly the $\bar{3}$ portion of the direct product representation is absent. (The $\bar{3}$ portion is antisymmetric under the interchange of the two 3 factors, and the 6 portion is symmetric.) It follows that the h^ℓ should transform according to the representation 6, which is indeed the case. Similarly, the general direct product representation $\bar{3} \otimes \bar{3}'$ has the Clebsch-Gordan series decomposition (8.35). It follows that the \bar{h}^ℓ should transform according to the representation $\bar{6}$, which is also the case.

5.8.22. Consider the vector space spanned by the quadratic polynomials of the form $t^j \times \bar{t}^k$. See (8.96). Show that this vector space is 9 dimensional, and is spanned by the polynomials b^ℓ of (8.5). Exercise 8.20 showed that the t^j and \bar{t}^k transform under $su(3)$ according to the representations 3 and $\bar{3}$, respectively. It follows from group theory and the derivation property (3.7) that the products $t^j \times \bar{t}^k$ must transform according to the direct product representation $3 \otimes \bar{3}$. In the case of $su(3)$, the general direct product representation $3 \otimes \bar{3}$ has the Clebsch-Gordan series decomposition (8.36). It follows that the b^ℓ should transform according to the representations 1 and 8. This surmise is indeed the case since Exercise 8.19 showed that b^0 transforms according to the representation 1, and the remaining b 's transform according to the representation 8.

5.8.23. Verify the relations (8.68). The polynomials h^j transform according to the representation 6. Also, the Poisson bracket operation may be viewed as a kind of multiplication. It follows from group theory and the derivation property (3.9) that the Lie products $[h^j, h^k]$ must transform as some portion of the direct product representation $6 \otimes 6'$. In the case of $su(3)$, the general direct product representation $6 \otimes 6'$ has the Clebsch-Gordan series decomposition (8.37). On the other hand, from the structure of $sp(6)$, the Poisson brackets $[h^j, h^k]$ can only yield terms of the form b^ℓ , which transform according to $1 = \Gamma(0, 0)$ and $8 = \Gamma(1, 1)$. See (3.9.3) and Exercise 8.19. We seem to have arrived at an apparent contradiction because $\Gamma(0, 0)$ and $\Gamma(1, 1)$ do not appear on the right side of (8.37). The only resolution to this apparent dilemma is for the Poisson brackets (8.68) to vanish, which they indeed do.

By contrast the general direct product representation $6 \otimes \bar{6}$ has the Clebsch-Gordan series decomposition (8.38). It follows that the Lie products $[h^j, \bar{h}^k]$ must transform as some portion of the representations $1 \oplus 8 \oplus 27$. This surmise is indeed the case since the Poisson brackets $[h^j, \bar{h}^k]$ yield terms of the form b^ℓ [see (8.53)], and these terms transform according to 1 and 8. Show that similar considerations apply to the Poisson bracket relation (8.64). The relevant Clebsch-Gordan series decomposition in this cases is (8.39).

5.8.24. Verify that the polynomials b^1, b^2 , and b^3 form a Lie subalgebra under the Poisson bracket operation. This subalgebra is the Lie algebra for an $su(2)$ subalgebra of $su(3)$.

5.8.25. Verify that the polynomials b^2, b^5 , and b^7 form a Lie subalgebra under the Poisson bracket operation. This subalgebra is the Lie algebra for an $so(3)$ subalgebra of $su(3)$. See Exercises 7.2.5 and 7.2.6.

5.8.26. For the case of a 6-dimensional phase space, consider the quadratic polynomials defined by the relations

$$T_{jk} = q_j q_k + p_j p_k, \quad (5.8.97)$$

$$L_j = \sum_{k\ell} \epsilon_{jkl} q_k p_\ell. \quad (5.8.98)$$

Show that there are 9 such elements, and that they can be written as linear combinations of the quantities b^0 through b^8 , and vice versa. The quantities T and L therefore provide an alternate basis for $u(3)$. Note that T is symmetric. Relate b^0 to the trace of T . Show that the quantities L_j form a basis for $so(3)$. Show that the quantities T transform as a tensor under $so(3)$. That is, evaluate the Poisson brackets $[L_j, L_k]$ and $[L_j, T_{k\ell}]$. Finally, evaluate the Poisson brackets $[T_{k\ell}, T_{mn}]$.

5.8.27. The relation (5.3) associates a matrix JS with every quadratic polynomial. Find the matrices B^i (for $i = 0, 1, \dots, 8$) associated with the polynomials b^i . Find the matrices F^j and G^j (for $j = 1, 2, \dots, 6$) associated with the polynomials f^j and g^j . Use (5.21) if you wish. The B^i, F^j , and G^j provide a basis for the 6×6 matrix representation of $sp(6)$. Find their commutation rules.

Partial Answer:

$$B^0 = J = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \end{pmatrix}, \quad B^1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad (5.8.99)$$

$$B^2 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad B^3 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$B^4 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad B^5 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \end{pmatrix},$$

$$G^4 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & 0 \end{pmatrix}, \quad G^5 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix},$$

$$G^6 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \end{pmatrix}.$$

Note that these generators/matrices are given for the case that J has the form (3.1.1). In the case that J' as given by (3.2.10) is employed, one may find the related generators by means of the permutation matrix P given by (3.2.19).

5.8.28. The purpose of this exercise is to explore the *dynamic* role of the basis polynomials b^0 through b^8 , f^1 through f^6 , and g^1 through g^6 .

- a) Consider first the $u(3)$ basis polynomials b^0 through b^8 . In place of b^0 , b^3 , and b^8 it is convenient to use the polynomials $(p_1^2 + q_1^2)/2$, $(p_2^2 + q_2^2)/2$, $(p_3^2 + q_3^2)/2$. Show that $(p_1^2 + q_1^2)/2$ generates rotations in the q_1, p_1 plane, etc. See Exercise 5.4.5. Next consider b^2 , b^5 , and b^7 . Show that b^2 generates rotations in the p_1, p_2 and q_1, q_2 planes simultaneously, etc. See Section 7.2. Finally consider b^1 , b^4 , and b^6 . Show that b^1 generates rotations in the q_1, p_2 and q_2, p_1 planes simultaneously, etc.
- b) Next consider the remaining polynomials f^1 through f^6 and g^1 through g^6 . The polynomials f^1 , f^3 , and f^5 are analogous. Show that f^1 generates motions on hyperbolas in the q_1, p_1 plane, etc. The polynomials g^1 , g^3 , and g^5 are also analogous. Show that g^1 also generates motions on hyperbolas in the q_1, p_1 plane, etc. See Exercise 5.4.4. The polynomials f^2 , f^4 , and f^6 are analogous. Show that f^2 generates motions on hyperbolas in the q_1, p_2 and q_2, p_1 planes simultaneously, etc. Finally, the polynomials g^2 , g^4 , and g^6 are analogous. Show that g^2 generates motions on hyperbolas in the p_1, p_2 and q_1, q_2 planes simultaneously, etc.

5.8.29. The purpose of this exercise is to explore conjugacy relations for the case of $su(3)$. Review Exercise 3.7.36. Suppose, for some representation, there is a basis for which the elements in the Lie algebra $su(3)$ are anti-Hermitian matrices (and the structure constants are real). Recall that for such matrices the hattening and checking operations given by (3.7.219) and (3.7.222) have the same effect.

Specifically, consider the matrices $i\lambda^j$ that obey the $su(3)$ commutation rules (8.3). Form the associated hatted representation given by (3.7.219). Next form the associated checked representation given by (3.7.222). Show that, as expected, in this case both operations produce the same result. Verify that in this case the conjugate representation is *not* equivalent

to the original representation. Hint: Show that C^1 and C^2 as given by (8.8) have the eigenvalues displayed in the weight diagram of Figure 8.4. Hence the matrices iC^1 and iC^2 will have these eigenvalues multiplied by i . Consequently, the matrix pairs iC^1 , iC^2 , and $-iC^1$, $-iC^2$ cannot have the same eigenvalues, and therefore cannot be related by a similarity transformation. Indeed, the eigenvalues of the pair $-iC^1$, $-iC^2$ are those shown in Figure 8.5 multiplied by i .

Repeat the above analysis for the basis matrices given by (8.8) and (8.11). Hint: Review Exercise 8.7.

5.8.30. Use the quantities f_{jkl} to form the adjoint representation of $su(3)$. See (8.3). Show that this representation has dimension 8 and is therefore $\Gamma(1, 1)$. See Table 8.3 and Figure 8.8. Review Exercise 3.7.36. Show that the adjoint representation of $su(3)$ is unaffected by either the hatting operation (3.7.218) or the checking operation (3.7.219). We say that the adjoint representation is *self conjugate* in accord with the $su(3)$ representation conjugacy relation (8.20).

5.8.31. Construct the weight diagrams for the representations 10, $\overline{10}$, and 27. Indicate the multiplicity of each weight.

5.8.32. A *Chevalley* basis for a Lie algebra is one for which the structure constants are all integers. Sometimes one also requires that the entries in matrices used to represent the algebra have all integer entries. Show that the basis found for $sp(2)$ and $sp(4)$ in Sections 5.6 and 5.7 are Chevalley bases. Find Chevalley bases for $su(3)$ and $sp(6)$.

5.9 Some Topological Questions

In this section we will learn something about the topology of $Sp(2n, \mathbb{R})$ and how the stable elements of $Sp(2n, \mathbb{R})$ reside within it.

5.9.1 Nature and Connectivity of $Sp(2n, \mathbb{R})$

$Sp(2n, \mathbb{R})$ Is Connected

We begin by showing that the symplectic group $Sp(2n, \mathbb{R})$ is connected, and indeed infinitely connected. Let us start with the connected claim, which is easy to demonstrate. Suppose M and N are any two matrices in $Sp(2n, \mathbb{R})$. Define the symplectic matrix R by the rule

$$R = MN^{-1} \tag{5.9.1}$$

so that there is the relation

$$M = RN. \tag{5.9.2}$$

Since R is symplectic, from (3.8.24) we know there are symmetric matrices S^a and S^c such that R can be written in the form

$$R = \exp(JS^a) \exp(JS^c), \tag{5.9.3}$$

and the matrices $R(\lambda)$ defined by

$$R(\lambda) = \exp(\lambda JS^a) \exp(\lambda JS^c) \quad (5.9.4)$$

will form a one-parameter family of symplectic matrices with

$$R(0) = I \quad (5.9.5)$$

and

$$R(1) = R. \quad (5.9.6)$$

Now consider the one-parameter family of symplectic matrices $M(\lambda)$ defined by the rule

$$M(\lambda) = R(\lambda)N. \quad (5.9.7)$$

From this and the previous definitions we have the results

$$M(0) = N \quad (5.9.8)$$

and

$$M(1) = M. \quad (5.9.9)$$

Thus, the matrices $M(\lambda)$ provide a path, in the space of symplectic matrices, that connects N to M .

Sp(2n, \mathbb{R}) Is Infinitely Connected

We next turn to the harder task of examining the topology of $Sp(2n, \mathbb{R})$ in more detail and showing that the space of symplectic matrices is infinitely connected. We will begin with the case of $Sp(2, \mathbb{R})$.

Suppose the basis elements B^0 , F , and G given by (6.7), (6.13), and (6.14) are used to evaluate (3.8.24). Doing so shows that the most general real 2×2 symplectic matrix can be written in the form

$$M = \exp(\phi F + \gamma G) \exp(\beta_0 B^0), \quad (5.9.10)$$

where β_0 , ϕ , and γ are arbitrary real coefficients. [Note that there are indeed three coefficients as predicted by (3.7.35) evaluated for $n=1$.] Thus, (9.10) gives a complete parameterization of the 2×2 symplectic group. The quantities $\exp(\phi F + \gamma G)$ and $\exp(\beta_0 B^0)$ can be evaluated using (3.7.1) to give the results

$$\begin{aligned} \exp(\phi F + \gamma G) &= I \cosh[(\phi^2 + \gamma^2)^{1/2}] \\ &\quad + [(\phi F + \gamma G)/(\phi^2 + \gamma^2)^{1/2}] \sinh[(\phi^2 + \gamma^2)^{1/2}], \end{aligned} \quad (5.9.11)$$

$$\exp(\beta_0 B^0) = I \cos \beta_0 + B^0 \sin \beta_0 = \begin{pmatrix} \cos \beta_0 & \sin \beta_0 \\ -\sin \beta_0 & \cos \beta_0 \end{pmatrix}. \quad (5.9.12)$$

Observe that, according to (9.11), the factor $\exp(\phi F + \gamma G)$ has the topology of two-dimensional Euclidean space E^2 since ϕ and γ can each range over $\pm\infty$ without any duplication of results. By contrast, the factor $\exp(\beta_0 B^0)$, according to (9.12), has the topology of a circle T^1 since it is periodic in β_0 with period 2π . Indeed, the matrix on the far right

of (9.12) represents $SO(2, \mathbb{R})$ which has the topology of T^1 , and also of $U(1)$. (Here, as in Section 3.9, we use the notation T^n to denote an n -torus, the topological product of n circles. Thus, T^1 denotes a 1-torus, which is just a circle.) It follows that $Sp(2, \mathbb{R})$ has the product topology $E^2 \times T^1$.⁹ Since T^1 is infinitely connected, $Sp(2, \mathbb{R})$ is infinitely connected. Finally, in view of (3.9.88), we note that there is the relation

$$W \exp(\beta_0 B^0) W^{-1} = \begin{pmatrix} \exp(i\beta_0) & 0 \\ 0 & \exp(-i\beta_0) \end{pmatrix}, \quad (5.9.13)$$

which shows explicitly that $\exp(\beta_0 B^0)$ is isomorphic to the representation $U(1) \oplus \overline{U(1)}$ of $U(1)$, as expected.

In an analogous way, with $m = n(n + 1)$, it can be seen that $Sp(2n, \mathbb{R})$ has the product topology $E^m \times [U(n) \oplus \overline{U(n)}]$. First, again by (3.8.24), any M in $Sp(2n, \mathbb{R})$ can be written in the product form

$$M = \exp(JS^a) \exp(JS^c). \quad (5.9.14)$$

Now, according to Exercise 3.9.10, there are $m = n(n + 1)$ linearly independent matrices of the form JS^a . Note that from its form, m is an *even* integer, and hence $k = m/2$ is an integer. In analogy to the cases of $sp(2, \mathbb{R})$, $sp(4, \mathbb{R})$, and $sp(6, \mathbb{R})$, let F^1, \dots, F^k and G^1, \dots, G^k be a basis for the set of matrices of the form JS^a . Then the $\exp(JS^a)$ factor can be written in the form

$$\exp(JS^a) = \exp\left[\sum_{j=1}^k (\phi_j F^j + \gamma_j G^j)\right]. \quad (5.9.15)$$

Since the real symmetric logarithm of a real symmetric positive definite matrix is unique, the m parameters ϕ_j and γ_j can all range from $\pm\infty$ without any duplication of results. Consequently, the factor $\exp(JS^a)$ has the topology of E^m . Finally, according to Section 3.9, matrices of the form $\exp(JS^c)$ are isomorphic to $U(n)$. Thus, as stated at the beginning of this paragraph, $Sp(2n, \mathbb{R})$ has the product topology $E^m \times [U(n) \oplus \overline{U(n)}]$.

We pause at this point to observe that some of the matrix entries on the right side of (9.11) grow in magnitude without bound as ϕ and γ range over $\pm\infty$. Thus, the group $Sp(2, \mathbb{R})$ is not compact. Moreover, since $Sp(2, \mathbb{R})$ is a subgroup of $Sp(2n, \mathbb{R})$ and $Sp(2n, \mathbb{C})$ for any n , it follows that these groups are also not compact.¹⁰

In this vein, what can be said about what we have called the $U(n)$ subgroup, the $[U(n) \oplus \overline{U(n)}]$ factor of $Sp(2n, \mathbb{R})$? From the discussion of Section 3.9 we know that all matrices of the form $\exp(JS^c)$ are in the orthogonal group $SO(2n, \mathbb{R})$. From the work of the first part of Section 3.6.3 we know that the rows (and columns) of an orthogonal matrix are orthonormal. In particular, the rows (and columns) are unit vectors. It follows that all entries in an orthogonal matrix are bounded in magnitude by 1. Consequently, the $U(n)$

⁹ Because ϕ and γ are unrestricted, this set is sometimes referred to as a *solid* torus.

¹⁰ *Compactness* is a topological property of sets that may be defined in a variety of ways. For our purposes, since we are generally dealing with matrices which may be viewed as being imbedded in some high dimensional Euclidean space, we will say that a set of matrices is compact if all matrix elements of these matrices are confined to lie within some closed and bounded set within this Euclidean space. (A set is *closed* if it contains all its limit points.) Conversely, if any matrix elements for some sequence of matrices in the set are unbounded (grow in magnitude without bound), we will say that the set of matrices is *noncompact*.

subgroup of $Sp(2n, \mathbb{R})$ is compact. Indeed, it can be shown to be the largest compact subgroup of $Sp(2n, \mathbb{R})$.

We have already seen that $Sp(2n, \mathbb{R})$ has the product topology $E^m \times [U(n) \oplus \overline{U(n)}]$. What remains is to study the topology of the $U(n)$ subgroup of $Sp(2n, \mathbb{R})$. In the $n = 1$ case we have already found that $Sp(2, \mathbb{R})$ has the product topology $E^2 \times U(1)$ and that $U(1)$ has the topology of T^1 . We might hope to proceed in a similar fashion for the case $n > 1$. Suppose, for specificity, we consider the case $Sp(4, \mathbb{R})$ for which $n = 2$ and we are therefore interested in the topology of $U(2)$. In the 4×4 case a basis for the Lie algebra of the matrices of the form JS^c can be taken to be the matrices B^0 through B^3 given displayed in (7.45). We also note that B^0 commutes with the B^1 through B^3 . See (7.5). Therefore in the 4×4 case the most general $\exp(JS^c)$ can be written in the form

$$\exp(JS^c) = [\exp\left(\sum_1^3 \beta_j B^j\right)] \exp(\beta_0 B^0). \quad (5.9.16)$$

Matrices of the form $\exp(\sum_1^3 \beta_j B^j)$ carry the $SU(2) \oplus \overline{SU(2)}$ representation of $SU(2)$, and all the groups $SU(n)$ for $n > 1$ are known to be simply connected. For example, $SU(2)$ has the topology of the 3-sphere S^3 . See Exercise 10.13. And S^3 is simply connected. What remains is to examine the factor $\exp(\beta_0 B^0)$.

A remark is in order before doing so. It is common in the physics literature to see the assertion

$$U(n) = SU(n) \otimes U(1) \quad (5.9.17)$$

where the symbol \otimes denotes a direct product. [A particular case of (9.17) is the assertion that $U(2) = SU(2) \otimes U(1)$.] If this were true, since $SU(n)$ is simply connected, $U(n)$ would have the *connectivity* of $U(1)$, which is T^1 . And, in particular, $U(2)$ would have the connectivity T^1 . Correspondingly, $Sp(2n, \mathbb{R})$ would have the product topology $E^m \times SU(n) \times T^1$, and consequently all the $Sp(2n, \mathbb{R})$ would be *infinitely* connected. It turns out that these topological statements are correct, but the argument is wrong. The assertion (9.17) is not *globally* true. What is true is the weaker result that (9.17) holds only in some vicinity of the identity.

Let us continue. In view of the result given for B^0 in (7.45) and the relation (3.8.30) it follows that

$$\exp(\beta_0 B^0) = I \cos \beta_0 + J \sin \beta_0. \quad (5.9.18)$$

And, again in view of (3.9.88), we see that in the 4×4 case there is the result

$$W \exp(\beta_0 B^0) W^{-1} = \begin{pmatrix} \exp(i\beta_0) & 0 & 0 & 0 \\ 0 & \exp(i\beta_0) & 0 & 0 \\ 0 & 0 & \exp(-i\beta_0) & 0 \\ 0 & 0 & 0 & \exp(-i\beta_0) \end{pmatrix}. \quad (5.9.19)$$

We conclude, because they contain only the diagonal entries $\exp(i\beta_0)$ and $\exp(-i\beta_0)$, that for small β_0 the matrices on the right side of (9.19) behave like $U(1) \oplus \overline{U(1)}$. But observe that taking the determinants of the 2×2 matrices in the upper left and lower right blocks of (9.19) yields the results $\exp(2i\beta_0)$ and $\exp(-2i\beta_0)$, respectively. These results, when

evaluated for $\beta_0 = \pi$, both have the value +1. Thus, when for $\beta_0 = \pi$, the 2×2 matrices are in $SU(2)$ and can be absorbed into the first factor, the $SU(2) \oplus \overline{SU(2)}$ factor, on the right side of (9.16). For this value of β_0 the direct product hypothesis $U(2) = SU(2) \otimes U(1)$ has abruptly changed. Consequently the 2×2 matrices in (9.19) are *not* a global representation of $U(1) \oplus \overline{U(1)}$. By an analogous analysis it is evident that the hypothesis (9.17) is not true globally for any $n \geq 2$. See, for example, Exercise 9.4.

Where does our exploration now stand? The approach we have been following has not been adequate for determining the global topology of $U(2)$, and evidently it will also fail for all $U(n)$ with $n > 1$. However, by more powerful methods beyond the scope of this book, it can be shown that all the $U(n)$ have the connectivity of T^1 . Consequently $Sp(2n, \mathbb{R})$ has the product topology $E^m \times SU(n) \times T^1$. It follows, because of the presence of T^1 , that the groups $Sp(2n, \mathbb{R})$ are infinitely connected for all n .

Since $Sp(2n, \mathbb{R})$ is infinitely connected, it must have a multiplicity of covering groups.¹¹ In particular, it has a two-fold covering group. This group is called the *metaplectic* group, and is of interest for paraxial wave optics (*Fourier* optics) and quantum mechanics.

Finally we remark that, contrary to the case of $Sp(2n, \mathbb{R})$, $Sp(2n, \mathbb{C})$ is *simply* connected.

5.9.2 Where Are the Stable Elements?

With the topology of $Sp(2n, \mathbb{R})$ in view, it would be useful to know where the stable elements (those with distinct eigenvalues on the unit circle) reside. In general this is a difficult question because $Sp(2n, \mathbb{R})$ is $n(2n + 1)$ dimensional. However, $Sp(2, \mathbb{R})$ is only 3 dimensional, and we will see that this case is tractable.

In the case of $Sp(2, \mathbb{R})$, combining (9.10) through (9.12) gives the result

$$\begin{aligned} M = & \{I \cosh[(\phi^2 + \gamma^2)^{1/2}] + [(\phi F + \gamma G)/(\phi^2 + \gamma^2)^{1/2}] \sinh[(\phi^2 + \gamma^2)^{1/2}]\} \times \\ & \{I \cos \beta_0 + B^0 \sin \beta_0\}. \end{aligned} \quad (5.9.20)$$

From the work of Section 3.4.4 we know that in the 2×2 case the spectrum of M is governed by the quantity

$$A = \text{tr}(M). \quad (5.9.21)$$

This quantity can be readily evaluated using (9.20) to yield the result

$$A = 2 \cosh[(\phi^2 + \gamma^2)^{1/2}] \cos \beta_0. \quad (5.9.22)$$

See Exercise 9.5. Introduce a radius r in ϕ, γ space by writing

$$r^2 = \phi^2 + \gamma^2. \quad (5.9.23)$$

With this definition, (9.22) can be rewritten in the form

$$A = 2(\cosh r)(\cos \beta_0). \quad (5.9.24)$$

From (3.4.21) and Figure 3.4.3 we know that there is stability (eigenvalues of M are on the unit circle) when

$$-2 < 2(\cosh r)(\cos \beta_0) < +2. \quad (5.9.25)$$

¹¹For a brief discussion of the concept of a covering group, see Exercise 8.2.11.

It follows that, when $(\cos \beta_0) > 0$, we can move away from the origin in ϕ, γ space while maintaining stability until $r = r_{\max}$ with

$$(\cosh r_{\max})(\cos \beta_0) = 1, \quad (5.9.26)$$

which is equivalent to the statement

$$r_{\max} = \cosh^{-1}[1/\cos(\beta_0)]. \quad (5.9.27)$$

On the other hand, when $(\cos \beta_0) < 0$, we can move away from the origin in ϕ, γ space while maintaining stability until

$$(\cosh r_{\max})(\cos \beta_0) = -1, \quad (5.9.28)$$

which is equivalent to the statement

$$r_{\max} = \cosh^{-1}[-1/\cos(\beta_0)]. \quad (5.9.29)$$

The two conditions (9.27) and (9.29) can be combined to give the net result

$$r_{\max} = \cosh^{-1}[1/|\cos(\beta_0)|]. \quad (5.9.30)$$

Figure 9.1 displays the relation (9.30) in the β_0, r plane. We observe that $r_{\max}(\beta_0)$ is periodic in β_0 with period π (and therefore also 2π) and that

$$r_{\max} = \infty \text{ when } \beta_0 = \pm\pi/2 \quad (5.9.31)$$

and

$$r_{\max} = 0 \text{ when } \beta_0 = 0, \pm\pi. \quad (5.9.32)$$

Note that $r = 0$ and $\beta_0 = \pm\pi/2$ correspond to tunes of $\pm 1/4$, and $r = 0$ and $\beta_0 = 0, \pm\pi$ correspond to tunes of $0, \pm 1/2$.

Suppose Γ is any closed path in $Sp(2, \mathbb{R})$ that goes once around the torus $SO(2, \mathbb{R})$. For example, it could begin at the identity I , that is $\beta_0 = \gamma = \phi = 0$, and end again at the identity with a 2π increase in β_0 so that at the end point $\beta_0 = 2\pi$ and again $\gamma = \phi = 0$. Then, somewhere along the path, the variable β_0 must take on the values $\beta_0 = \pi/2$ and $\beta_0 = 3\pi/2$. (Note that, by periodicity, the points $3\pi/2$ and $-\pi/2$ are equivalent.) At these points r_{\max} is infinite. Thus, at least two stable group elements must lie on any closed path Γ that goes once around the torus. Moreover, since the eigenvalues for these elements are not ± 1 (they are $\pm i$ because $A = 0$ at these points), these elements must lie in open sets comprised of stable elements. Indeed, at these points the tunes are $\pm 1/4$.

It would be pleasant to have an analogous understanding of $Sp(2n, \mathbb{R})$ for general n , or at least for $n = 2$ and $n = 3$. Perhaps this is possible for $Sp(4, \mathbb{R})$ using the parameterization of Section 5.7 and the results associated with Figure 3.4.4. And perhaps, in a National Emergency, the case of $Sp(6, \mathbb{R})$ could also be understood. But we have not attempted to do so. However, what we already do know, thanks to the discussion of Sections 3.4 and 3.5, is that when the eigenvalues of an element lie on the unit circle and are distinct, then this element is surrounded by an open set of stable elements. We reiterate that this fact should be of comfort to accelerator designers and builders because it means that, at least in the

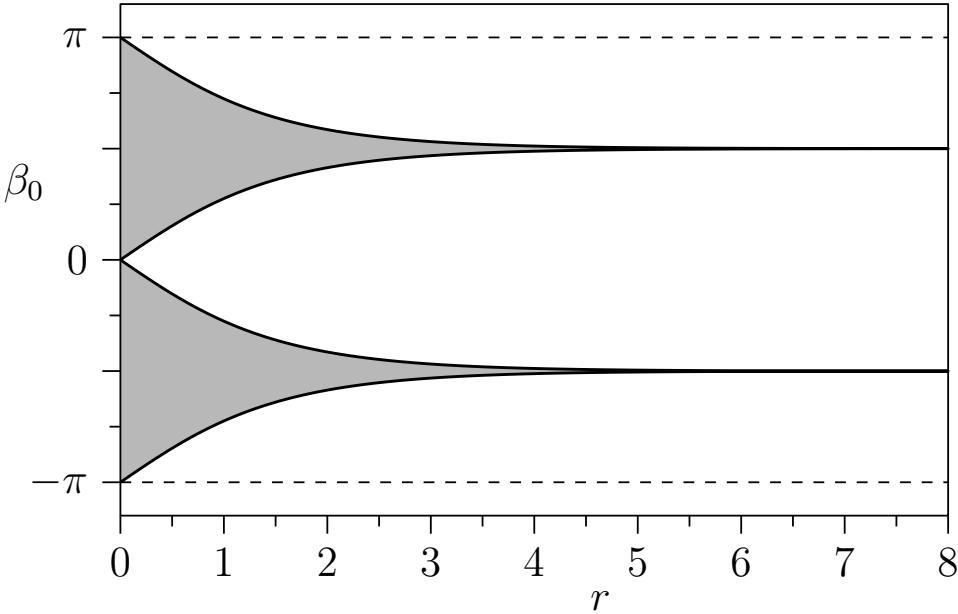


Figure 5.9.1: Stability diagram for $Sp(2, \mathbb{R})$ showing the quantity r_{\max} as a function of β_0 . All elements with $r < r_{\max}$ are stable, and all elements with $r > r_{\max}$ are unstable. That is, the shaded regions are stable, and the unshaded regions are unstable. In accord with toroidal topology, corresponding points on the dashed lines at the top and bottom of the figure ($\beta_0 = \pm\pi$) are to be identified.

linear approximation, the stability of orbits will not be damaged by small fabrication and control parameter errors.

We close this subsection with a remark that, perhaps, should have been made at the beginning of this subsection. We know that every symplectic matrix R has the unique factorization (9.3). Also, if S^a vanishes in this factorization, then R is diagonalizable and all its eigenvalues lie on the unit circle. Hence, all such R are stable elements. By contrast, if S^c vanishes in this factorization, then R has all its eigenvalues on the positive real axis, and some must exceed 1. Hence, all such R are unstable elements. See Exercise 3.8.12. What we have learned in this subsection is that there are cases where both S^a and S^c are non vanishing and R is stable, and other cases where both S^a and S^c are non vanishing and R is unstable.

5.9.3 Covering/Circumnavigating $U(n)$

We know that there is a $U(n)$ subgroup of $Sp(2n, \mathbb{R})$ and that any R in the $U(n)$ subgroup of $Sp(2n, \mathbb{R})$ can be written in the form

$$R(S^c) = \exp(JS^c). \quad (5.9.33)$$

Since $U(n)$ is compact, the matrices S^c cannot be arbitrarily large without some repetition occurring among the elements $R(S^c)$. Here we will find a result for how large S^c needs to be for all of the $U(n)$ subgroup to be covered.

According to the work of Section 3.9, given any R in the $U(n)$ subgroup of $Sp(2n, \mathbb{R})$, there is a $u \in U(n)$ such that

$$R = M(u). \quad (5.9.34)$$

Also, given any $u \in U(n)$ there is a $t \in U(n)$ such that

$$u = tvt^{-1} \quad (5.9.35)$$

where v is a diagonal matrix of the form (3.9.50). Since the mapping $M(u)$ is an isomorphism, we have the result

$$R = M(u) = M(tvt^{-1}) = M(t)M(v)[M(t)]^{-1} = M(t)V[M(t)]^{-1}. \quad (5.9.36)$$

Here we have used (3.9.51). But V is in the $U(n)$ subgroup and therefore there is a matrix \hat{S}^c such that

$$V = \exp(J\hat{S}^c). \quad (5.9.37)$$

See (3.9.55), (3.9.63), and (3.9.64). It follows from (9.36) and (9.37) that

$$R = M(t)\exp(J\hat{S}^c)[M(t)]^{-1} = \exp\{M(t)J\hat{S}^c\}[M(t)]^{-1}. \quad (5.9.38)$$

Upon comparing (9.33) and (9.37) we see that a suitable JS^c is given by the relation

$$JS^c = M(t)J\hat{S}^c[M(t)]^{-1}, \quad (5.9.39)$$

from which it follows that

$$S^c = J^{-1}M(t)J\hat{S}^c[M(t)]^{-1}. \quad (5.9.40)$$

From the symplectic condition $MJM^T = J$ and (3.9.30) we see that

$$J^{-1}MJ = (M^T)^{-1} = M. \quad (5.9.41)$$

It follows that

$$S^c = M\hat{S}^cM^{-1}, \quad (5.9.42)$$

and therefore

$$(S^c)^2 = M(\hat{S}^c)^2M^{-1}. \quad (5.9.43)$$

Now take the trace of both sides of (9.43) to find the result

$$\text{tr}[(S^c)^2] = \text{tr}[M(\hat{S}^c)^2M^{-1}] = \text{tr}[(\hat{S}^c)^2]. \quad (5.9.44)$$

The right side of (9.44) can be easily evaluated using (3.9.63) and (3.9.64). We find that

$$\text{tr}[(\hat{S}^c)^2] = 2 \sum_{\ell=1}^n \phi_\ell^2. \quad (5.9.45)$$

Since each $\phi_\ell \in [-\pi, \pi]$, we see that all of the $U(n)$ subgroup is covered when

$$\text{tr}[(S^c)^2] \leq 2n\pi^2. \quad (5.9.46)$$

When (9.46) holds some elements in the $U(n)$ subgroup are covered multiple times and some are covered only once. But each is covered at least once.

Exercises

5.9.1. Verify the results (9.11) and (9.12).

5.9.2. Verify that the first part of (9.12), namely

$$\exp(\beta_0 B^0) = I \cos \beta_0 + B^0 \sin \beta_0, \quad (5.9.47)$$

holds in general (the $2n \times 2n$ case) for $B^0 = J$.

5.9.3. Rob, Salman, and Ivan's work on exponentiating $sp(4)$.

5.9.4. The purpose of this exercise is to make an analysis of $Sp(6, \mathbb{R})$ analogous to that provided for $Sp(4, \mathbb{R})$ in Subsection 9.1. Begin by observing that (9.14) and (9.15) hold for general n . Verify that, for $n = 3$, (9.16) takes the form

$$\exp(JS^c) = [\exp\left(\sum_1^8 \beta_j B^j\right)] \exp(\beta_0 B^0). \quad (5.9.48)$$

Matrices of the form $\exp(\sum_1^8 \beta_j B^j)$ carry the $SU(3) \oplus \overline{SU(3)}$ representation of $SU(3)$. What remains is to examine the factor $\exp(\beta_0 B^0)$ with B^0 given by the first entry in (8.99).

Verify, using (3.9.88), that for $n = 3$ there is the result

$$W \exp(\beta_0 B^0) W^{-1} = \begin{pmatrix} \exp(i\beta_0) & 0 & 0 & 0 & 0 & 0 \\ 0 & \exp(i\beta_0) & 0 & 0 & 0 & 0 \\ 0 & 0 & \exp(i\beta_0) & 0 & 0 & 0 \\ 0 & 0 & 0 & \exp(-i\beta_0) & 0 & 0 \\ 0 & 0 & 0 & 0 & \exp(-i\beta_0) & 0 \\ 0 & 0 & 0 & 0 & 0 & \exp(-i\beta_0) \end{pmatrix}. \quad (5.9.49)$$

Observe that taking the determinants of the 3×3 matrices in the upper left and lower right blocks of (9.49) yields the results $\exp(3i\beta_0)$ and $\exp(-3i\beta_0)$, respectively. These results, when evaluated for $\beta_0 = 2\pi/3$, both have the value $+1$. Thus, when $\beta_0 = 2\pi/3$, the 3×3 matrices are in $SU(3)$ and can be absorbed into the first factor, the $SU(3) \oplus \overline{SU(3)}$ factor, on the right side of (9.48). For this value of β_0 the direct product hypothesis $U(3) = SU(3) \otimes U(1)$ has abruptly changed. (Verify that the same is true when $\beta_0 = 4\pi/3$.) Consequently the 3×3 matrices in (9.49) are *not* a global representation of $U(1) \oplus \overline{U(1)}$.

5.9.5. Show that carrying out the multiplication indicated in (9.20) yields a linear combination of the matrices I, F, G, B_0, FB_0 , and GB_0 . Show, with the exception of I , that all these matrices are traceless. Use this result to prove (9.22). Show that all matrices M of the form (9.20) satisfy

$$M^2 = -I \text{ when } \beta_0 = \pm\pi/2. \quad (5.9.50)$$

Suggestion: Use the normal form technology of Section 3.3.7.

5.9.6. Review Subsection 9.2 and Figure 9.1. Pick a value for r , say $r = 0.20$. Plot, in the complex plane, the eigenvalues of M as β_0 varies over the interval $\beta_0 \in [-\pi, \pi]$. You should find that they move about the unit circle, collide at the points ± 1 , and leave and re-enter the unit circle through the collision points ± 1 . Finally, when they leave the unit circle, they both lie on the positive or negative real axis. Recall Figures 3.4.1 and 3.4.3.

Consider the cases where $A = \pm 2$, but do not otherwise constrain β_0 and r , and let M_{\pm} be the *Jordan* normal form for M in these cases. Show that, generically,

$$M_+ = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad (5.9.51)$$

and

$$M_- = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}. \quad (5.9.52)$$

Show that M as given by (3.5.76) with $\alpha \neq 0$ has the Jordan normal form M_+ and $-M$ [with M again given by (3.5.76) with $\alpha \neq 0$], which is also symplectic, has Jordan normal form M_- . For M given by (9.20), under what conditions can M , by a similarity transformation, be brought to the *diagonal* forms

$$M_{d\pm} = \begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix} = \pm I? \quad (5.9.53)$$

5.10 Notational Pitfalls and Quaternions

5.10.1 The Lie Algebras $sp(2n, \mathbb{R})$ and $usp(2n)$

In the discussion of the Lie algebra $sp(6)$ we found it useful to work over the complex field even though we eventually wrote our final results in real form. The use of the complex field is both a powerful tool and a possible source of confusion. We have repeatedly made use of the particular Lie algebraic properties that the commutator of two matrices of the form JS^c is again of the same form, the commutator of a JS^c and a JS^a is a matrix of the form JS^a , and the commutator of two matrices of the form JS^a is a matrix of the form JS^c . We write these relations symbolically in the form

$$\{JS^c, JS^{c'}\} \propto JS^{c''}, \quad (5.10.1)$$

$$\{JS^c, JS^a\} \propto JS^{a'}, \quad (5.10.2)$$

$$\{JS^a, JS^{a'}\} \propto JS^c. \quad (5.10.3)$$

Here all matrices are taken to be real. That is, we are working with the Lie algebra $sp(2n, \mathbb{R})$. Now suppose that all matrices of the form JS^a are replaced by matrices of the form iJS^a , and the matrices of the form JS^c are left unchanged. Doing so converts the relations (10.1) through (10.3) into the relations

$$\{JS^c, JS^{c'}\} \propto JS^{c''}, \quad (5.10.4)$$

$$\{JS^c, (iJS^a)\} \propto (iJS^{a'}), \quad (5.10.5)$$

$$\{(iJS^a), (iJS^{a'})\} \propto JS^c. \quad (5.10.6)$$

Examination of the relations (10.4) though (10.6) shows that this replacement produces a related Lie algebra of the same dimension as before. Evidently this algebra is a subalgebra of $sp(2n, C)$. We next observe that matrices of the form JS^c and iJS^a (with S^c and S^a real) are anti-Hermitian,

$$(JS^c)^\dagger = (S^c)^\dagger J^\dagger = S^c(-J) = -JS^c, \quad (5.10.7)$$

$$\begin{aligned} (iJS^a)^\dagger &= (-i)(S^a)^\dagger (J^\dagger) = -iS^a(-J) \\ &= -iJS^a. \end{aligned} \quad (5.10.8)$$

Consequently, the Lie algebra they generate is also a subalgebra of $u(2n)$. Let us use the notation $usp(2n)$ to denote the Lie algebra generated by matrices of the form JS^c and iJS^a . Then we have the relation

$$usp(2n) = u(2n) \cap sp(2n, \mathbb{C}). \quad (5.10.9)$$

Note that although $usp(2n)$ has a complex basis if a real basis is used for $sp(2n, \mathbb{R})$, it still has *real* structure constants in terms of this complex basis. In the language of Section 3.7, we have found that the Lie algebras $sp(2n, \mathbb{R})$ and $usp(2n)$ are equivalent over the complex field. However, they are not equivalent over the real field.

Also, let $USp(2n)$ denote the group obtained by exponentiating matrices of the form JS^c and iJS^a . These matrices belong to both $U(2n)$ and $Sp(2n, \mathbb{C})$, and we have the relation

$$USp(2n) = U(2n) \cap Sp(2n, \mathbb{C}). \quad (5.10.10)$$

This group $USp(2n)$ is called the *unitary symplectic* group.

Unfortunately for Physicists, Mathematicians often refer to this group simply as $Sp(2n)$ while we have been using the same notation as shorthand for $Sp(2n, \mathbb{R})$. This dual notation can be a source of serious confusion because $USp(2n)$ and $Sp(2n, \mathbb{R})$ have very different properties. For example, $USp(2n)$ is compact [all the entries in $USp(2n)$ matrices are bounded in absolute value by 1 since these matrices are unitary] while, as can be seen from the results of Section 5.9, the matrix elements of $Sp(2n, \mathbb{R})$ matrices can be arbitrarily large. Moreover, it can be shown that $USp(2n)$ is simply connected, and we have seen that $Sp(2n, \mathbb{R})$ is infinitely connected. Finally, for completeness, we remark that $Sp(2n, \mathbb{C})$ is noncompact and simply connected.

5.10.2 $USp(2n)$ and the Quaternion Field

The group $USp(2n)$ is of mathematical interest for at least two reasons. First, because it is compact, it is much easier to analyze than is $Sp(2n, \mathbb{R})$. And, because $sp(2n, \mathbb{R})$ and $usp(2n)$ are complex equivalent, many results obtained for $usp(2n)$ are readily transferable to $sp(2n, \mathbb{R})$. Second, $USp(2n)$ is closely related to quaternions and can be viewed as the quaternion field analog of the groups $O(n, \mathbb{R})$ and $U(n)$ for the real and complex fields.¹² We will now describe briefly how this comes about.

¹²Quaternions as an algebra were discovered by *Hamilton* in 1843, and often the quaternion field is referred to as \mathbb{H} . Some aspects of them were also known in some form earlier and independently to *Euler* in 1748, *Gauss* in 1819, and *Rodrigues* in 1840.

Consider an n -dimensional real vector space with the usual real inner product. To emphasize the use of the *real* field, we denote this inner product by the symbols $(,)_{\mathbb{R}}$. Then the set of real linear transformations that preserves this inner product forms the orthogonal group, $O(n, \mathbb{R})$. Specifically, if x and y are any two vectors, we require the relation

$$(Ox, Oy)_{\mathbb{R}} = (x, y)_{\mathbb{R}} \quad (5.10.11)$$

But we also have the relation

$$(Ox, Oy)_{\mathbb{R}} = (x, O^T Oy)_{\mathbb{R}} \quad (5.10.12)$$

from which it follows that O must satisfy the condition

$$O^T O = I. \quad (5.10.13)$$

Next consider an n -dimensional *complex* vector space with the usual complex inner product. We denote this inner product by the symbols $(,)_{\mathbb{C}}$ to emphasize the use of the complex field. Then the set of complex linear transformations that preserves this inner product forms the unitary group, $U(n)$. Specifically, if x and y are any two vectors, we require the relation

$$(Ux, Uy)_{\mathbb{C}} = (x, y)_{\mathbb{C}}. \quad (5.10.14)$$

But we also have the relation

$$(Ux, Uy)_{\mathbb{C}} = (x, U^\dagger Uy)_{\mathbb{C}} \quad (5.10.15)$$

from which it follows that U must satisfy the condition

$$U^\dagger U = I. \quad (5.10.16)$$

Finally, suppose we consider an n -dimensional vector space over the *quaternion* field \mathbb{H} with a suitable inner product yet to be defined. Then the set of linear transformations with quaternion entries that preserves this inner product can be shown to be isomorphic to $U\text{Sp}(2n)$. Thus, the groups $O(n)$, $U(n)$, and $U\text{Sp}(2n)$ all arise from analogous constructions over the real field, the complex field, and the quaternion field, respectively.

We will work up to this result in stages. First we will study the structure of $usp(2n)$ and $U\text{Sp}(2n)$. Next we will represent quaternions using Pauli matrices, and define a suitable inner product. Finally, we will show that $U\text{Sp}(2n)$ preserves this inner product.

5.10.3 Quaternion Matrices

Let S^c be any real $2n \times 2n$ symmetric matrix that commutes with J . For J we shall take the form (3.2.10). That is, J is an $n \times n$ collection of 2×2 blocks. Suppose that S^c is also written as an $n \times n$ collection of 2×2 blocks,

$$S^c = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix}, \quad (5.10.17)$$

where each entry c_{jk} is a 2×2 block. Then it is easily verified that requiring S^c to commute with J is equivalent to requiring that each entry c_{jk} commute with the 2×2 matrix J_2 of (3.2.11),

$$c_{jk}J_2 - J_2c_{jk} = 0. \quad (5.10.18)$$

The condition (10.18) in turn requires that each c_{jk} be a linear combination (with arbitrary real coefficients) of σ^0 and J_2 .

Similarly, let S^a be any real $2n \times 2n$ matrix that *anticommutes* with J . Suppose that S^a is written as an $n \times n$ collection of 2×2 blocks in the form

$$S^a = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}. \quad (5.10.19)$$

Then requiring that S^a anticommute with J is equivalent to requiring that each entry a_{jk} anticommute with J_2 ,

$$a_{jk}J_2 + J_2a_{jk} = 0. \quad (5.10.20)$$

It is easily checked that the condition (10.20) implies in turn that each a_{jk} must be a linear combination (with real coefficients) of σ^3 and σ^1 .

Now consider matrices of the form iS^a where the entries in S^a itself are real. Then, every 2×2 block in S^a must be a linear combination with real coefficients of the matrices $i\sigma^3$ and $i\sigma^1$. Also, we note that J_2 is given by the relation

$$J_2 = i\sigma^2. \quad (5.10.21)$$

We conclude that all matrices of the form S^c and iS^a , and their linear combinations with real coefficients, must have in their 2×2 blocks only matrices of the form

$$Q = w_0\sigma^0 + iw_1\sigma^1 + iw_2\sigma^2 + iw_3\sigma^3, \quad (5.10.22)$$

where the coefficients w_0 through w_3 are *real*. It is convenient to regard the quantities w_1, w_2 , and w_3 as the three components of a vector \mathbf{w} , and to then write (10.22) in the more compact form

$$Q = w_0\sigma^0 + i\mathbf{w} \cdot \boldsymbol{\sigma}. \quad (5.10.23)$$

The reader will eventually have the pleasure of showing, in Exercise 10.15, that the set of all 2×2 matrices of the form (10.23) is isomorphic to the quaternion field \mathbb{H} .¹³ For this reason, these matrices will be called *quaternion matrices*.

5.10.4 Properties of Quaternion Matrices

Suppose two quaternion matrices Q and Q' are multiplied together. From the relation (7.40) we find the result

$$QQ' = (w_0w'_0 - \mathbf{w} \cdot \mathbf{w}')\sigma^0 + i(w_0\mathbf{w}' + w'_0\mathbf{w} - \mathbf{w} \times \mathbf{w}') \cdot \boldsymbol{\sigma} = Q'' \quad (5.10.24)$$

¹³This discovery was first made by Cayley.

with

$$Q'' = w_0''\sigma^0 + i\mathbf{w}'' \cdot \boldsymbol{\sigma}. \quad (5.10.25)$$

and

$$w_0'' = w_0 w_0' - \mathbf{w} \cdot \mathbf{w}', \quad (5.10.26)$$

$$\mathbf{w}'' = w_0 \mathbf{w}' + w_0' \mathbf{w} - \mathbf{w} \times \mathbf{w}'. \quad (5.10.27)$$

We conclude that the product of two quaternion matrices is again a quaternion matrix. Also, any linear combination with real coefficients of quaternion matrices is again a quaternion matrix,

$$\alpha Q + \beta Q' = (\alpha w_0 + \beta w_0')\sigma^0 + i(\alpha \mathbf{w} + \beta \mathbf{w}') \cdot \boldsymbol{\sigma} \quad (5.10.28)$$

We summarize the results of (10.24) and (10.28) by saying that the set of quaternion matrices is *closed* under the operators of multiplication and addition with real coefficients.

From the specific form of the Pauli matrices we find that Q can be written in the explicit form

$$Q = \begin{pmatrix} w_0 + iw_3 & iw_1 + w_2 \\ iw_1 - w_2 & w_0 - iw_3 \end{pmatrix}. \quad (5.10.29)$$

From this form we easily compute that the determinant of Q is given by the relation

$$\det(Q) = w_0^2 + w_1^2 + w_2^2 + w_3^2. \quad (5.10.30)$$

We conclude that all quaternion matrices are invertible save for the zero quaternion matrix. We will soon see that the inverse of a quaternion matrix is again a quaternion matrix. It follows that quaternion matrices form a *division algebra*.

Given any quaternion matrix Q specified by (10.23), we define a *conjugate* quaternion matrix Q^* by the relation

$$Q^* = w_0\sigma^0 - i\mathbf{w} \cdot \boldsymbol{\sigma}. \quad (5.10.31)$$

It is easily verified from (10.24) that the product Q^*Q is given by the relation

$$Q^*Q = QQ^* = (w_0^2 + w_1^2 + w_2^2 + w_3^2)\sigma^0 = [\det(Q)]\sigma^0. \quad (5.10.32)$$

Consequently, the inverse of a quaternion matrix is also a quaternion matrix given by the relation

$$Q^{-1} = Q^*/[\det(Q)]. \quad (5.10.33)$$

Since the Pauli matrices are Hermitian, the definition (10.31) is equivalent to the relation

$$Q^* = Q^\dagger. \quad (5.10.34)$$

We also find by explicit calculation the relation

$$Q^* = -J_2 Q^T J_2 = (J_2)^{-1} Q^T J_2. \quad (5.10.35)$$

From either (10.33), (10.34), or (10.35) we find the relation

$$(Q'Q)^* = Q^*(Q')^*. \quad (5.10.36)$$

We note that if Q is a quaternion matrix, so are the matrices Q^* , Q^{-1} , and Q^T . Finally, we observe from (10.21) that the matrix J_2 is also a quaternion matrix. It follows that the set of all nonzero quaternion matrices forms a group.

5.10.5 Quaternion Matrices and $USp(2n)$

We have learned that all matrices of the form S^c and iS^a , and their linear combinations with real coefficients, must have quaternion matrices in their 2×2 blocks when the form (3.2.10) is used for J . Moreover, since in this form J also has quaternion matrices in its 2×2 blocks, all matrices of the form JS^c and iJS^a , and their linear combinations with real coefficients, must also have quaternion matrices in their 2×2 blocks. This result follows from the fact that the set of quaternion matrices is closed under multiplication and addition with real coefficients. Finally, suppose matrices of the form JS^c and iJS^a , and their linear combinations with real coefficients, are exponentiated and multiplied together. Since these operations all reduce to the multiplication and addition (again with real coefficients) of quaternion matrices, all $n \times n$ arrays resulting from these operations must also have quaternion matrices in their 2×2 blocks. We conclude that when the form (3.2.10) is used for J , all matrices in the group $USp(2n)$ must have quaternion matrices in their 2×2 blocks.

Let M be a matrix in $USp(2n)$. Suppose M is written in an $n \times n$ array,

$$M = \begin{pmatrix} m_{11} & \cdots & m_{1n} \\ \vdots & & \vdots \\ m_{n1} & \cdots & m_{nn} \end{pmatrix}, \quad (5.10.37)$$

where, according to the preceding discussion, each block m_{jk} is a quaternion matrix. Then the matrices M^\dagger and M^T are given by the relations

$$M^\dagger = \begin{pmatrix} m_{11}^\dagger & \cdots & m_{n1}^\dagger \\ \vdots & & \vdots \\ m_{1n}^\dagger & \cdots & m_{nn}^\dagger \end{pmatrix}, \quad (5.10.38)$$

$$M^T = \begin{pmatrix} m_{11}^T & \cdots & m_{n1}^T \\ \vdots & & \vdots \\ m_{1n}^T & \cdots & m_{nn}^T \end{pmatrix}. \quad (5.10.39)$$

Because M is unitary, it must satisfy the relation

$$M^\dagger M = I. \quad (5.10.40)$$

However, because the entries in M are quaternion matrices, use of the relations (3.2.10), (10.34), and (10.35) gives the result

$$M^\dagger = J^{-1} M^T J. \quad (5.10.41)$$

Consequently (10.40) can be rewritten in the form

$$J^{-1} M^T J M = I \text{ or } M^T J M = J. \quad (5.10.42)$$

We conclude that a unitary matrix whose 2×2 blocks are quaternion matrices must also be a symplectic matrix. Conversely, if M is symplectic and is made of 2×2 quaternion matrix blocks, then M must also be unitary.

5.10.6 Quaternion Inner Product and Its Preservation

Let x and y be any two n -component vectors whose entries are quaternion matrices. We will call such vectors *quaternion vectors*. For quaternion vectors we define a quaternion inner product, which we denote by the symbols $(,)_{\mathbb{H}}$, by the relation

$$(x, y)_{\mathbb{H}} = \sum_{j=1}^n x_j^* y_j. \quad (5.10.43)$$

(Note that the result of forming a quaternion inner product is a quaternion matrix.)

Next, suppose M is any matrix in $USp(2n)$. Using the representation (10.37), we define transformed vectors x', y' by the relations

$$x'_j = \sum_k m_{jk} x_k, \quad (5.10.44)$$

$$y'_j = \sum_\ell m_{j\ell} y_\ell. \quad (5.10.45)$$

Note that the operations on the right sides of (10.44) and (10.45) involve only quaternion matrix multiplication and addition. We write (10.44) and (10.45) more compactly in the form

$$x' = Mx, \quad y' = My. \quad (5.10.46)$$

From (10.34), (10.36), and (10.44) we find the relations

$$(x'_j)^* = \sum_k (m_{jk} x_k)^* = \sum_k x_k^* m_{jk}^* = \sum_k x_k^* (m_{jk})^\dagger. \quad (5.10.47)$$

Let us compute $(Mx, My)_{\mathbb{H}}$. We find from (10.45) and (10.47) the result

$$\begin{aligned} (Mx, My)_{\mathbb{H}} &= (x', y')_Q = \sum_{j=1}^n (x'_j)^* y'_j \\ &= \sum_{j,k,\ell} x_k^* (m_{jk})^\dagger m_{j\ell} y_\ell = \sum_{k,\ell} x_k^* \delta_{k\ell} \sigma^0 y_\ell \\ &= \sum_k x_k^* y_k = (x, y)_{\mathbb{H}}. \end{aligned} \quad (5.10.48)$$

Here we have used the fact that the relation (10.40) can be written in the 2×2 block form

$$\sum_j (m_{jk})^\dagger m_{j\ell} = \delta_{k\ell} \sigma^0. \quad (5.10.49)$$

See (10.37) and (10.38). Note that (10.49) can also be written in the form

$$\sum_j (m_{jk})^* m_{j\ell} = \delta_{k\ell} \sigma^0, \quad (5.10.50)$$

and in this form only quaternion operations are involved. We conclude from (10.48) that M preserves the quaternion inner product,

$$(Mx, My)_{\mathbb{H}} = (x, y)_{\mathbb{H}}. \quad (5.10.51)$$

Further, it can be checked that if M preserves (10.51) for arbitrary quaternion vectors x and y , then M must belong to $USp(2n)$.

At this point we might wonder if there is a connection between the the quaternion inner product (10.43) and the fundamental symplectic 2-form (3.2.3). By working Exercise 10.18 you will have the pleasure of seeing that they are closely related.

5.10.7 Discussion

Comparison of (10.11), (10.14), and (10.51) shows that $O(n)$, $U(n)$, and $USp(2n)$ all arise from analogous constructions over the real field, the complex field, and the quaternion field, respectively. We remark that the only finite-dimensional associative normed division algebras over the real number field are the real number field itself, the complex field, and the quaternion field. (This proposition is known as Frobenius' theorem.) Thus, $O(n)$, $U(n)$, and $USp(2n)$ are not only analogous, they are also exhaustive.

Reference to Table 3.7.1 shows that we have accounted for all the classical Lie algebras. What can be said about the exceptional algebras? After the reals, the complex numbers, and the quaternions come the *octonions* (also called *Cayley numbers*). As their name suggests, they form an eight-dimensional vector space for which multiplication can also be defined. Like the reals, complexes, and quaternions, octonions form a normed division algebra. (In fact, these four are the *only* normed division algebras.) However, octonion multiplication is *not* associative. It can be shown that, in one way or another, all the exceptional Lie algebras are related to various properties of the octonions. Moreover, the failure of the exceptional Lie algebras to form regular infinite families (like the classical Lie algebras do) is related to the nonassociativity of octonion multiplication.

Exercises

5.10.1. Verify the commutation rules (10.4) through (10.6).

5.10.2. Look at the relations (10.9) and (10.10). Strictly speaking, our discussion has only shown that $usp(2n)$ is contained in $u(2n) \cap sp(2n, \mathbb{C})$, etc. Prove that they are in fact equal. That is, prove that (10.9) and (10.10) are correct.

5.10.3. Show that W as given by (3.9.8) belongs to $USp(2n)$.

5.10.4. Show that the matrix elements of (real) orthogonal matrices and unitary matrices are less than or equal to 1 in absolute value. Show that the matrix elements of matrices in $GL(n, \mathbb{R})$ [which, as we have seen in Section (3.10), is a subgroup of $Sp(2n, \mathbb{R})$] are unbounded. That is, there are matrices in $GL(n, \mathbb{R})$ whose matrix elements are arbitrarily large.

5.10.5. Verify that requiring S^c to commute with J is equivalent to (10.18). Verify the claim that each c_{jk} must be a linear combination of σ^0 and J_2 with real coefficients. Verify that requiring S^a to anticommute with J is equivalent to (10.20). Verify the claim that each a_{jk} must be a linear combination with real coefficients of σ^3 and σ^1 .

5.10.6. Verify the multiplication rule (10.24) through (10.27). Show that the multiplication of quaternion matrices is generally not commutative.

5.10.7. Verify the relations (10.29) through (10.34).

5.10.8. Verify (10.35) by explicit calculation. Find the same result using (3.1.7) and (10.33).

5.10.9. Verify (10.36) directly from (10.24) and the definition (10.31).

5.10.10. Verify (10.41). Is it true that any unitary matrix that is also symplectic with respect to (3.2.10) must have quaternion matrices in its 2×2 blocks? Prove your answer.

5.10.11. Verify (10.49).

5.10.12. Show that the quaternion inner product (10.43) has the property

$$(y, x)_{\mathbb{H}} = [(x, y)_{\mathbb{H}}]^*. \quad (5.10.52)$$

Suppose that the vector x has quaternion entries x_j . Let λ be any quaternion. Define $x\lambda$ to be the vector with quaternion entries $x_j\lambda$. Show that the quaternion inner product has the properties

$$(x, y\lambda)_{\mathbb{H}} = [(x, y)_{\mathbb{H}}]\lambda, \quad (5.10.53)$$

$$(x\lambda, y)_{\mathbb{H}} = \lambda^*(x, y)_{\mathbb{H}}. \quad (5.10.54)$$

We see that in the case of quaternion vectors with the quaternion inner product (10.43), what is the analog of scalar multiplication must take place by multiplication on the right.

5.10.13. Show that the set of nonzero quaternion matrices forms a group. Show that any nonzero quaternion matrix Q can be written in the form

$$Q = \exp(v_0\sigma^0 + i\mathbf{v} \cdot \boldsymbol{\sigma}), \quad (5.10.55)$$

where v_0 and \mathbf{v} are real. Thus, these quaternion matrices form a Lie group. Find the associated Lie algebra. Consider quaternions with determinant +1. In view of (10.30) and Exercise 10.14 below, we will refer to such quaternions as *unit* quaternions. Show that the set of all unit quaternion matrices forms a subgroup that is identical to $SU(2)$. Show, in view of (10.30), that $SU(2)$ may be viewed as the manifold S^3 , the 3-dimensional surface of a sphere in 4-dimensional Euclidean space, also known as the *3-sphere*. (It can be shown that among all the n -spheres, only S^1 and S^3 also have the structure of a group.) Suppose that Q is a unit quaternion matrix. Using the parameterization (10.29) and the result (3.9.20), find the matrix $M(Q)$.

5.10.14. Define a quaternion matrix norm by the relation

$$\| Q \| = \sqrt{\det(Q)}. \quad (5.10.56)$$

Show that this norm satisfies (3.7.10) through (3.7.13).

5.10.15. The purpose of this exercise is to define quaternions and to show that, as discovered by Cayley, they are faithfully represented by quaternion matrices.¹⁴ The quaternion field \mathbb{H} , often called Hamilton's quaternion algebra, is a four-dimensional linear vector space over the *real* number field. Let the basis for this vector space be denoted by the symbols e, j, k, ℓ . Impose the following laws of multiplication among the basis vectors:

$$\begin{aligned} e^2 &= e, \quad ej = je = j, \quad ek = ke = k, \quad e\ell = \ell e = \ell; \\ j^2 &= k^2 = \ell^2 = -e; \\ jk &= -kj = \ell, \quad k\ell = -\ell k = j, \quad \ell j = -j\ell = k. \end{aligned} \tag{5.10.57}$$

Note that the quantities e, j, k, ℓ all anticommute. Since the vectors e, j, k, ℓ form a basis, the most general quaternion is a vector, which we will denote by the symbol q , of the form

$$q = ae + bj + ck + d\ell, \tag{5.10.58}$$

where the quantities a, b, c, d are real numbers.¹⁵ Suppose q' is a second quaternion,

$$q' = a'e + b'j + c'k + d'\ell. \tag{5.10.59}$$

We then have the addition rule

$$q + q' = q'' = a''e + b''j + c''k + d''\ell \tag{5.10.60}$$

with

$$a'' = a + a', \quad b'' = b + b', \quad c'' = c + c', \quad d'' = d + d'. \tag{5.10.61}$$

Show, using the multiplication rules (10.57), that

$$qq' = q'' = a''e + b''j + c''k + d''\ell \tag{5.10.62}$$

with

$$\begin{aligned} a'' &= aa' - bb' - cc' - dd', \\ b'' &= ab' + ba' + cd' - dc', \\ c'' &= ac' + ca' + db' - bd', \\ d'' &= ad' + da' + bc' - cb'. \end{aligned} \tag{5.10.63}$$

Now make the following correspondence \leftrightarrow between the quaternion matrices $\sigma^0, -i\sigma^1, -i\sigma^2, -i\sigma^3$ and the quaternion basis vectors e, j, k, ℓ :

$$\sigma^0 \leftrightarrow e,$$

¹⁴For faithful representations of quaternions by real 4×4 matrices, see Exercise 11.1.7.

¹⁵Other authors, including Hamilton, commonly use the symbols $1, i, j, k$ for our e, j, k, ℓ . Our notation is designed to avoid confusion between quaternion basis vectors and the quantities 1 and $\sqrt{-1}$. Finally we remark that if the coefficients a, b, c, d are permitted to be *complex*, the resulting object is called a *biquaternion*.

$$\begin{aligned} -i\sigma^1 &\leftrightarrow j, \\ -i\sigma^2 &\leftrightarrow k, \\ -i\sigma^3 &\leftrightarrow \ell. \end{aligned} \tag{5.10.64}$$

Make the correspondence \leftrightarrow into a linear mapping by extending it from basis elements to arbitrary elements in a linear fashion. Suppose q is the quaternion (10.58). Define a corresponding quaternion matrix Q by the rule

$$Q = a\sigma^0 + b(-i\sigma^1) + c(-i\sigma^2) + d(-i\sigma^3) = \begin{pmatrix} a - id & -c - ib \\ c - ib & a + id \end{pmatrix}, \tag{5.10.65}$$

and make the correspondence

$$Q \leftrightarrow q. \tag{5.10.66}$$

Using (7.3), verify the arithmetic in (10.65). By linearity the correspondence (10.66) and the correspondence

$$Q' \leftrightarrow q' \tag{5.10.67}$$

imply the correspondence

$$Q' + Q \leftrightarrow q' + q. \tag{5.10.68}$$

Verify this assertion. Using (10.57) and the rules for matrix multiplication, show that the correspondences (10.66) and (10.67) imply the correspondence

$$Q'Q \leftrightarrow q'q. \tag{5.10.69}$$

See (7.40). Prove that quaternion multiplication is associative.

Given any quaternion q of the form (10.58), the conjugate quaternion q^* is defined by the relation

$$q^* = ae - bj - ck - d\ell. \tag{5.10.70}$$

Show that the correspondence given by (10.65) and (10.66) implies the correspondence

$$Q^\dagger \leftrightarrow q^*. \tag{5.10.71}$$

Review Exercise 10.14. Compare q^*q and qq^* with $\|Q\|^2$.

5.10.16. Suppose M is a (possibly complex) symplectic matrix. Then according to (3.1.8) and (3.1.9), M is nonsingular. Consequently, M must have a unique polar decomposition of the form

$$M = PU, \tag{5.10.72}$$

where P is positive definite Hermitian and U is unitary. Show, in analogy to (3.8.6) and (3.8.7), that P and U are also symplectic. Next show, in analogy to the derivation of (3.9.33), that P must have determinant +1. Now consider the matrix U . Since U is both unitary and symplectic, it must belong to $USp(2n)$. Show that if U is sufficiently near the identity, then it must have determinant +1. But since $USp(2n)$ is connected (indeed, simply connected), every matrix in $USp(2n)$ can be continuously deformed to the identity while remaining within $USp(2n)$. Show, by continuity arguments, that these circumstances require that all U in $USp(2n)$ must have determinant +1. Finally, use (10.72) to show that M must have determinant +1.

5.10.17. Review Exercises 3.1.2 and 3.1.3. Show that the groups $USp(2)$ and $SU(2)$, and correspondingly the Lie algebras $usp(2)$ and $su(2)$, are the same.

5.10.18. The quantities x_j and y_j appearing in (10.46) are quaternion matrices, and therefore can be written in the form

$$y_j = \begin{pmatrix} q_j & s_j \\ p_j & r_j \end{pmatrix}, \quad (5.10.73)$$

$$x_j = \begin{pmatrix} \tilde{q}_j & \tilde{s}_j \\ \tilde{p}_j & \tilde{r}_j \end{pmatrix}, \quad (5.10.74)$$

where the various entries are (possibly complex) numbers and \sim is simply a mark that distinguishes quaternion matrix entries associated with an x_j from those associated with a y_j . Let z , w , \tilde{z} , and \tilde{w} be *column* vectors with $2n$ entries of the form

$$z = (q_1, p_1, q_2, p_2, \dots, q_n, p_n)^T, \quad (5.10.75)$$

$$w = (s_1, r_1, s_2, r_2, \dots, s_n, r_n)^T, \quad (5.10.76)$$

$$\tilde{z} = (\tilde{q}_1, \tilde{p}_1, \tilde{q}_2, \tilde{p}_2, \dots, \tilde{q}_n, \tilde{p}_n)^T, \quad (5.10.77)$$

$$\tilde{w} = (\tilde{s}_1, \tilde{r}_1, \tilde{s}_2, \tilde{r}_2, \dots, \tilde{s}_n, \tilde{r}_n)^T. \quad (5.10.78)$$

The vector z is made from the entries in the first columns of the y_j and the vector w is made from the entries in the second columns of the y_j , etc. We know that the quantity $(x, y)_{\mathbb{H}}$ is a quaternion matrix. Show, using (10.38) and (10.46), that it is the quaternion matrix given by the relation

$$(x, y)_{\mathbb{H}} = \begin{pmatrix} -(\tilde{w}, J'z) & -(\tilde{w}, J'w) \\ (\tilde{z}, J'z) & (\tilde{z}, J'w) \end{pmatrix} \quad (5.10.79)$$

where J' is the matrix (3.2.10), the matrix we have been calling J in this section. Evidently the entries of $(x, y)_{\mathbb{H}}$ consist of fundamental symplectic 2-forms involving the vectors z , w , \tilde{z} , and \tilde{w} . Show that (10.79) can also be written in the more symmetric form

$$(x, y)_{\mathbb{H}} = -J_2 \begin{pmatrix} (\tilde{z}, J'z) & (\tilde{z}, J'w) \\ (\tilde{w}, J'z) & (\tilde{w}, J'w) \end{pmatrix}. \quad (5.10.80)$$

Verify that the matrix appearing in the second factor in (10.80) is a quaternion matrix. Hint: Use the fact that $(x, y)_{\mathbb{H}}$ is a quaternion matrix and that J_2 is an invertible quaternion matrix. Finally, verify that (10.45) is equivalent to the two ordinary vector and matrix relations

$$z' = Mz, \quad (5.10.81)$$

$$w' = Mw, \quad (5.10.82)$$

and that there are analogous results for (10.44). Note that, in writing relations of the form (10.73), we have not forced the y_j , etc., to be quaternion matrices. Show that to do so one should require the relations

$$s_j = -\bar{p}_j, \quad (5.10.83)$$

$$r_j = \bar{q}_j, \quad (5.10.84)$$

where the overbar denotes complex conjugation. Thus, y_j takes the form

$$y_j = \begin{pmatrix} q_j & -\bar{p}_j \\ p_j & \bar{q}_j \end{pmatrix}, \quad (5.10.85)$$

and similarly for x_j . Hint: Use (10.29), but realize the that w 's appearing in it are different from those in (10.76). Finally, show that

$$(y, y)_{\mathbb{H}} = \sigma^0 \sum_j (|q_j|^2 + |p_j|^2) = \sigma^0 \sum_j \det y_j. \quad (5.10.86)$$

5.10.19. The work of Section 3.8.2 showed that any element of $Sp(2n, \mathbb{R})$ can be written uniquely in the form

$$M = \exp(JS^a) \exp(JS^c). \quad (5.10.87)$$

Also, according to Section 3.8.1, any unitary matrix U can be written in the form

$$U = \exp(A) \quad (5.10.88)$$

where A is anti-Hermitean. What can be said about matrices in $USp(2n)$?

5.11 Möbius Transformations

Möbius transformations occur in many branches of pure mathematics, and also in some areas of applied mathematics. This section defines and lays out some general properties of Möbius transformations. Two subsequent sections use Möbius transformations to provide a relation between $Sp(2n, \mathbb{R})$ and the theory of several complex variables, and to provide a relation between symplectic and symmetric matrices. This second relation generalizes the Cayley representation of Section 3.12. Later, in Section 6.7 of Chapter 6, Möbius transformations will be used to provide a fundamental connection between symplectic maps and gradient maps, thereby producing a plethora of generating functions.

5.11.1 Definition in the Context of Complex Variables

In the theory of a single complex variable z , one set of transformations of particular interest is the set of *Möbius* or *homographic* or *fractional linear* transformations given by relations of the kind¹⁶

$$z' = (az + b)/(cz + d). \quad (5.11.1)$$

(Some authors refer to these transformations as *linear* even though they manifestly are not.) Let M be a 2×2 matrix of the form

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (5.11.2)$$

¹⁶In 1778, years before Möbius and Cayley were born, Euler employed this relation and, apart from interchanging the letters a and b in the numerator and the letters c and d in the denominator, wrote the right side of (11.1) in exactly the same form as it appears here. He also considered the possibility that z was the tangent of some other quantity as in (3.11.2).

where the coefficients a through d are those appearing in (11.1). We view (11.1) as a transformation T_M associated with the matrix M , and write (11.1) in the compact form

$$z' = T_M(z). \quad (5.11.3)$$

Suppose T_N is a second Möbius transformation sending z' to z'' ,

$$z'' = T_N(z'). \quad (5.11.4)$$

Upon combining (11.3) and (11.4), we get the composite transformation

$$z'' = T_N(z') = T_N(T_M(z)). \quad (5.11.5)$$

Direct evaluation of (11.5) shows that the composite transformation is given by the relation

$$T_N T_M = T_{NM}. \quad (5.11.6)$$

Note also that T_I , where I is the identity matrix, is the identity transformation,

$$T_I(z) = z. \quad (5.11.7)$$

Suppose we agree to work with Möbius transformations for which the matrix (11.2) has nonzero determinant. Then (11.6) and (11.7) show that such Möbius transformations form a group. Moreover, suppose we scale all the entries in (11.2) by a common factor. In particular, suppose we select the scaling factor in such a way that the matrix (11.2) has determinant +1. [This will require scaling by a complex number if $\det(M)$ is not positive.] Examination of (11.1) shows that such scaling leaves the Möbius transformation unchanged. That is, there is the relation

$$T_{\lambda M}(z) = T_M(z) \quad (5.11.8)$$

where λ is any non vanishing scalar. We may therefore restrict our attention to matrices (11.2) that are symplectic. Thus, the Möbius transformations associated with these matrices provide a realization of $Sp(2, \mathbb{R})$ or $Sp(2, \mathbb{C})$ as a set of *nonlinear* transformations of the complex plane into itself. We remark that this realization is of mathematical interest for the construction of unitary representations of $Sp(2, \mathbb{R})$. It is of physical interest for the construction of unitary representations of the Lorentz group (needed for elementary particle physics) and for laser optics. In the case of laser optics, the Möbius transformation is essentially the so-called *ABCD law* for the propagation of axially symmetric Gaussian beams.

5.11.2 Matrix Extension

The Möbius transformation can be extended/generalized to higher dimensions in the following way. Let U be a $n \times n$ matrix and let M be a $2n \times 2n$ matrix written in terms of $n \times n$ matrices A^M through D^M in the block form

$$M = \begin{pmatrix} A^M & B^M \\ C^M & D^M \end{pmatrix}. \quad (5.11.9)$$

(Here we use the superscript M in connection with A^M through D^M to indicate that these matrices depend on M .) Define a transformation T_M associated with M that sends U to U' by the rule

$$U' = T_M(U) = (A^M U + B^M)(C^M U + D^M)^{-1}. \quad (5.11.10)$$

Then it can be verified by direct but slightly tedious matrix algebra that successive transformations again obey the composition law (11.6). Indeed, the composition law holds for any set of $n \times n$ matrices U and $2n \times 2n$ matrices M and N . Thus, if we require that the matrices M be invertible, the transformations T_M provide a representation of the group $GL(2n, \mathbb{C})$. Note that according to (11.10) these transformations are again nonlinear. Moreover, in analogy to (11.8), there is the scaling relation

$$T_{\lambda M}(U) = T_M(U) \quad (5.11.11)$$

where λ is any non vanishing scalar. We may therefore restrict our attention to matrices (11.9) that have unit determinant. Thus, if we require that the matrices M have unit determinant, the associated Möbius transformations provide a realization of $SL(2n, \mathbb{C})$ as a set of *nonlinear* transformations of the set of $n \times n$ matrices into itself.

5.11.3 Invertibility Conditions

Of course, for (11.10) to make sense, we must require that the matrix $(C^M U + D^M)$ be invertible,

$$\det(C^M U + D^M) \neq 0. \quad (5.11.12)$$

We will learn that this invertibility condition entails, and is entailed by, three others. That is, we will learn that there are *four equivalent* invertibility conditions. First we must see what these equivalent invertibility conditions might be.

The relation (11.10) can be solved for U by matrix manipulation. First multiply both sides of (11.10) on the right by $(C^M U + D^M)$. So doing gives the result

$$U'(C^M U + D^M) = A^M U + B^M \quad (5.11.13)$$

which, when multiplied out, becomes

$$U'C^M U + U'D^M = A^M U + B^M. \quad (5.11.14)$$

Now rearrange terms in (11.14) so that it becomes

$$(U'C^M - A^M)U = -U'D^M + B^M. \quad (5.11.15)$$

This relation can be rewritten in the form

$$U = (U'C^M - A^M)^{-1}(-U'D^M + B^M). \quad (5.11.16)$$

This assertion makes sense if $(U'C^M - A^M)$ is invertible,

$$\det(U'C^M - A^M) \neq 0. \quad (5.11.17)$$

On the other hand, from (11.10) and the group property (11.6), we deduce that

$$T_{M^{-1}}(U') = T_{M^{-1}}(T_M(U)) = T_{M^{-1}M}(U) = T_I(U) = U. \quad (5.11.18)$$

But, from the general definition (11.10), there is also the relation

$$T_{M^{-1}}(U') = (A^{M^{-1}}U' + B^{M^{-1}})(C^{M^{-1}}U' + D^{M^{-1}})^{-1}. \quad (5.11.19)$$

Comparison of (11.18) and (11.19) gives the result

$$U = (A^{M^{-1}}U' + B^{M^{-1}})(C^{M^{-1}}U' + D^{M^{-1}})^{-1}. \quad (5.11.20)$$

For this result to make sense, the matrix $(C^{M^{-1}}U' + D^{M^{-1}})$ must be invertible,

$$\det(C^{M^{-1}}U' + D^{M^{-1}}) \neq 0. \quad (5.11.21)$$

Also, comparison of (11.20) and (11.16) gives the identity

$$(A^{M^{-1}}U' + B^{M^{-1}})(C^{M^{-1}}U' + D^{M^{-1}})^{-1} = (U'C^M - A^M)^{-1}(-U'D^M + B^M) \quad (5.11.22)$$

which must hold for any matrices U' and M as long as (11.17) and (11.21) are satisfied. Observe that the identity (11.22) can also be written in the form

$$(A^{M^{-1}}U' + B^{M^{-1}})(C^{M^{-1}}U' + D^{M^{-1}})^{-1} = (-U'C^M + A^M)^{-1}(U'D^M - B^M). \quad (5.11.23)$$

Finally, (11.20) can be solved for U' by using the same kind of matrix manipulation that was employed to solve (11.10) for U . So doing gives the result

$$U' = (UC^{M^{-1}} - A^{M^{-1}})^{-1}(-UD^{M^{-1}} + B^{M^{-1}}). \quad (5.11.24)$$

This assertion makes sense if $(UC^{M^{-1}} - A^{M^{-1}})$ is invertible,

$$\det(UC^{M^{-1}} - A^{M^{-1}}) \neq 0. \quad (5.11.25)$$

Also, comparison of (11.10) and (11.24) gives the identity

$$(A^M U + B^M)(C^M U + D^M)^{-1} = (UC^{M^{-1}} - A^{M^{-1}})^{-1}(-UD^{M^{-1}} + B^{M^{-1}}), \quad (5.11.26)$$

which must hold for any matrices U and M as long as (11.12) and (11.25) are satisfied. Note that the identities (11.22) and (11.26) are equivalent: M is simply replaced by M^{-1} .

We will now prove that the four invertibility conditions (11.12), (11.17), (11.21), and (11.25) are equivalent. Let I^n and I^{2n} be the $n \times n$ and $2n \times 2n$ identity matrices, respectively. Then, using the representation (11.9), the equations

$$MM^{-1} = M^{-1}M = I^{2n} \quad (5.11.27)$$

are equivalent to the set of equations

$$A^M A^{M^{-1}} + B^M C^{M^{-1}} = A^{M^{-1}} A^M + B^{M^{-1}} C^M = I^n, \quad (5.11.28)$$

$$C^M A^{M^{-1}} + D^M C^{M^{-1}} = C^{M^{-1}} A^M + D^{M^{-1}} C^M = 0, \quad (5.11.29)$$

$$A^M B^{M^{-1}} + B^M D^{M^{-1}} = A^{M^{-1}} B^M + B^{M^{-1}} D^M = 0, \quad (5.11.30)$$

$$C^M B^{M^{-1}} + D^M D^{M^{-1}} = C^{M^{-1}} B^M + D^{M^{-1}} D^M = I^n. \quad (5.11.31)$$

Now examine the pair of relations (11.20) and (11.10). We claim that there is the equality

$$(C^{M^{-1}} U' + D^{M^{-1}})(C^M U + D^M) = I^n. \quad (5.11.32)$$

The proof is by direct calculation. Use of (11.10) gives the result

$$(C^{M^{-1}} U' + D^{M^{-1}}) = (C^{M^{-1}})(A^M U + B^M)(C^M U + D^M)^{-1} + D^{M^{-1}}. \quad (5.11.33)$$

Here we have assumed that (11.12) holds. From (11.33) we conclude that

$$(C^{M^{-1}} U' + D^{M^{-1}})(C^M U + D^M) = C^{M^{-1}}(A^M U + B^M) + D^{M^{-1}}(C^M U + D^M). \quad (5.11.34)$$

The terms on the right side of (11.34) can be regrouped to give the result

$$C^{M^{-1}}(A^M U + B^M) + D^{M^{-1}}(C^M U + D^M) = (C^{M^{-1}} A^M + D^{M^{-1}} C^M)U + (C^{M^{-1}} B^M + D^{M^{-1}} D^M). \quad (5.11.35)$$

By (11.29), the first factor on the right side of (11.35) vanishes, and by (11.31) the second factor equals I^n . Therefore (11.32) is correct. Now take determinants of both sides of (11.32) to find the result

$$[\det(C^{M^{-1}} U' + D^{M^{-1}})][\det(C^M U + D^M)] = 1. \quad (5.11.36)$$

We conclude that (11.12) and (11.21) are logically equivalent,

$$\det(C^{M^{-1}} U' + D^{M^{-1}}) \neq 0 \Leftrightarrow \det(C^M U + D^M) \neq 0. \quad (5.11.37)$$

In a similar way, using (11.24), it can be verified that

$$(U C^{M^{-1}} - A^{M^{-1}})(U' C^M - A^M) = I^n, \quad (5.11.38)$$

providing (11.25) holds. Hence (11.17) and (11.25) are logically equivalent,

$$\det(U' C^M - A^M) \neq 0 \Leftrightarrow \det(U C^{M^{-1}} - A^{M^{-1}}) \neq 0. \quad (5.11.39)$$

It remains to be shown that the invertibility conditions (11.12) and (11.25) are logically equivalent,

$$\det(C^M U + D^M) \neq 0 \Leftrightarrow \det(U C^{M^{-1}} - A^{M^{-1}}) \neq 0. \quad (5.11.40)$$

Once this is done we will have the complete chain of inferences

$$(11.21) \Leftrightarrow (11.12) \Leftrightarrow (11.25) \Leftrightarrow (11.17). \quad (5.11.41)$$

Specifically, in terms of matrices, (11.41) states that there is the complete chain of inferences

$$\begin{aligned} \det(C^{M^{-1}} U' + D^{M^{-1}}) \neq 0 &\Leftrightarrow \det(C^M U + D^M) \neq 0 \Leftrightarrow \\ \det(U C^{M^{-1}} - A^{M^{-1}}) \neq 0 &\Leftrightarrow \det(U' C^M - A^M) \neq 0. \end{aligned} \quad (5.11.42)$$

We now check the logical equivalence (11.40). It can be verified from (11.28) through (11.31) by matrix multiplication that there is the identity

$$\begin{pmatrix} I & -U \\ C^M & D^M \end{pmatrix} \begin{pmatrix} A^{M^{-1}} & B^{M^{-1}} \\ C^{M^{-1}} & D^{M^{-1}} \end{pmatrix} = \begin{pmatrix} A^{M^{-1}} - UC^{M^{-1}} & B^{M^{-1}} - UD^{M^{-1}} \\ 0 & I^n \end{pmatrix}. \quad (5.11.43)$$

Also, there is the identity

$$\begin{pmatrix} I^n & -U \\ C^M & D^M \end{pmatrix} = \begin{pmatrix} I^n & 0 \\ C^M & I^n \end{pmatrix} \begin{pmatrix} I^n & -U \\ 0 & C^M U + D^M \end{pmatrix}. \quad (5.11.44)$$

Taking the determinant of both sides of (11.44) gives the result

$$\det \begin{pmatrix} I^n & -U \\ C^M & D^M \end{pmatrix} = \det(C^M U + D^M). \quad (5.11.45)$$

Finally, take the determinant of both sides of (11.43) and use (11.45) to get the relation

$$[\det(C^M U + D^M)][\det(M^{-1})] = \det(A^{M^{-1}} - UC^{M^{-1}}) = (-1)^n \det(UC^{M^{-1}} - A^{M^{-1}}). \quad (5.11.46)$$

Since we have assumed that M is invertible, we have $[\det(M^{-1})] \neq 0$ and therefore the relation (11.46) implies the relation (11.40).

5.11.4 Transitivity

We close this section with a simple, but useful observation. Suppose U and V are any two nonsingular matrices. Then there is a nonsingular matrix M such that

$$V = T_M(U). \quad (5.11.47)$$

That is, any nonsingular matrix can be sent into any other nonsingular matrix by a suitable Möbius transformation. To verify this assertion, simply define M by the equation

$$M = \begin{pmatrix} VU^{-1} & 0 \\ 0 & I \end{pmatrix}, \quad (5.11.48)$$

and see that (11.47) is satisfied.

Exercises

5.11.1. Verify that

$$T_I(U) = U \quad (5.11.49)$$

and that (11.6) holds for the generalized Möbius transformation (11.10).

5.11.2. Verify (11.24).

5.11.3. Verify (11.28) through (11.31).

5.11.4. The critical reader might object that the proof of the logical equivalence (11.37) is incomplete. Why? Using (11.20) under the assumption (11.21), show that

$$(C^M U + D^M)(C^{M^{-1}} U' + D^{M^{-1}}) = I^n. \quad (5.11.50)$$

Similarly, complete the proof of the logical equivalence(11.39). Verify (11.38) and show that

$$(U' C^M - A^M)(U C^{M^{-1}} - A^{M^{-1}}) = I^n \quad (5.11.51)$$

using (11.16) under the assumption (11.17).

5.11.5. Verify (11.43) through (11.46).

5.11.6. Suppose two functions $f(z)$ and $g(z)$ are connected by the relation

$$g(z) = [af(z) + b]/[cf(z) + d] \quad (5.11.52)$$

which, employing the notation of (11.2) and (11.3), we also write in the form

$$g = T_M(f). \quad (5.11.53)$$

Assume that $\det M \neq 0$. Write $g \sim f$ if (11.53) holds for some M . Show that \sim is an equivalence relation among functions. (For the definition of an equivalence relation, see Exercise 5.12.7.) Show that

$$f \sim g \quad (5.11.54)$$

if, and only if,

$$\mathcal{S}f = \mathcal{S}g \quad (5.11.55)$$

where \mathcal{S} denotes the Schwarzian derivative (1.2.16). Show that the differential equation

$$\mathcal{S}f = 0 \quad (5.11.56)$$

has, as its most general solution, the relation

$$f(z) = (az + b)/(cz + d). \quad (5.11.57)$$

5.11.7. Exercise 3.12.5 introduced the Cayley function cay defined by

$$\text{cay}(X) = (I - X)/(I + X). \quad (5.11.58)$$

Show that

$$\text{cay}(X) = T_M(X) \quad (5.11.59)$$

with

$$M = (1/\sqrt{2}) \begin{pmatrix} -I & I \\ I & I \end{pmatrix}. \quad (5.11.60)$$

Verify that

$$M^2 = I, \quad (5.11.61)$$

in agreement with (3.12.47).

5.12 Symplectic Transformations and Siegel Space

5.12.1 Action of $Sp(2n, \mathbb{C})$ on the Space of Complex Symmetric Matrices

Suppose X and Y are $n \times n$ real matrices. Define the most general *complex* $n \times n$ matrix Z by writing the relation

$$Z = X + iY. \quad (5.12.1)$$

Now, with U replaced by Z , define a generalized Möbius transformation T_M associated with any $2n \times 2n$ matrix M and sending Z to Z' by the rule

$$Z' = T_M(Z) = (AZ + B)(CZ + D)^{-1}. \quad (5.12.2)$$

Here, for the moment, we have omitted the M superscript on the matrices A through D .

Suppose M is symplectic. Then it is a remarkable fact that Z' is symmetric if Z is symmetric. (We say that Z is symmetric if X and Y are symmetric.) Thus, if we regard complex symmetric matrices as generalizations of a single complex variable, then the transformations T_M with M symplectic can be viewed as transformations of a generalized complex variable.¹⁷

Brute force verification of the assertion that Z' is symmetric if Z is symmetric (provided that M is symplectic) is difficult. However, the proof is easy if we make use of the fact that $Sp(2n, \mathbb{C})$ is generated by matrices of the form (3.3.9) through (3.3.11). The assertion can easily be verified for the transformations associated with these matrices, and use of the group property (11.6) then assures that the assertion is true for all symplectic matrices.

We now check each of the cases (3.3.9) through (3.3.11). Suppose M is of the form (3.3.9). Then we have the transformation

$$Z' = Z + B. \quad (5.12.3)$$

Because of (3.3.12), Z' is symmetric if Z is symmetric. Next suppose M is of the form (3.3.11). Then we have the transformation

$$Z' = AZA^T \quad (5.12.4)$$

where use has been made of (3.3.13). Again Z' is symmetric if Z is. Finally, suppose that M is of the form (3.3.10). Recall the conjugacy relation (3.10.8). Evidently, in view of this relation, of what has already been checked, and by the group property, verification of the case (3.3.10) is equivalent to verification of the case $M = J$. When $M = J$, we have the transformation

$$Z' = -Z^{-1}. \quad (5.12.5)$$

Again it is evident that Z' is symmetric if Z is.

¹⁷Here is an opportunity for a tangential historical comment: If $n = 1$ then X and Y become the real numbers x and y and Z becomes the complex variable $z = x + iy$. And since any complex number z may be viewed as a point in the complex plane, any Z with Z symmetric may be viewed as a point in a generalized complex plane when $n > 1$. But who invented and first published the concept of a complex plane? The answer is *Caspar Wessel* (1745-1818) in 1797. This was long after Euler (1707-1783) who one might think knew all there was to know about complex numbers. Thus, in all fairness, one should call the complex plane the *Wessel* plane.

5.12.2 Siegel Space and $Sp(2n, \mathbb{R})$

One of the properties of the Möbius transformation (11.1), when the coefficients a through d are real and $\det M = ad - bc > 0$, is that it maps the upper half plane $y > 0$ into itself, and is in fact the most general analytic mapping of the upper half plane into itself. Consider all symmetric matrices of the form (12.1) with Y positive definite. Such matrices are sometimes called a *Siegel space*, and may be viewed as a *generalized upper half plane* (guhp). Remarkably, this guhp is mapped into itself by the generalized Möbius transformation (12.2) providing M is real symplectic. (See Exercise 12.3.) Indeed, it can be shown that the most general analytic mapping of the guhp into itself must be of the form (12.2) with M an element of $Sp(2n, \mathbb{R})$.

With regard to physical applications, we remark that the generalized Möbius realization of $Sp(4, \mathbb{R})$ is of interest for the propagation of generalized Gaussian laser beams when axial symmetry is not assumed.

5.12.3 Group Actions on Homogeneous Spaces

5.12.3.1 Definition of a Homogeneous Space

What is going on here? Speaking abstractly, we have a group G [$Sp(2n, \mathbb{R})$ in our case] acting on some space \mathcal{Z} (the guhp in our case) by mapping it into itself,

$$G : \mathcal{Z} \rightarrow \mathcal{Z}. \quad (5.12.6)$$

Suppose the action of G on \mathcal{Z} is such that given any two “points” Z and Z' in \mathcal{Z} , there is some group element g in G whose action sends Z to Z' . Then the action of G on \mathcal{Z} is said to be *transitive*, and the space \mathcal{Z} is said to be *homogeneous* with respect to the group G . The remarkable fact about a homogeneous space, as we will eventually show, is that there is a natural identification between it and the coset space of the group G with respect to some subgroup H . Moreover, the action of G on the homogeneous space is equivalent, under this identification, to the action of G (under group multiplication) on its own coset space. Thus, in a sense, homogeneous spaces are really aspects of various groups masquerading as if they were independent spaces in their own right.

5.12.3.2 Siegel Space Is a Homogeneous Space

Before continuing our general abstract discussion, we will show that the guhp is a homogeneous space with respect to $Sp(2n, \mathbb{R})$. First consider the point Z^0 given by (12.1) with $X = 0$ and $Y = I$,

$$Z^0 = iI. \quad (5.12.7)$$

Note that Z^0 is symmetric and I is positive definite. Thus, Z^0 is in the guhp, and may be viewed as the analog of the point $+i$ in the ordinary upper half plane. Next consider all matrices L in $Sp(2n, \mathbb{R})$ that leave Z^0 fixed under the action (12.2). The relation

$$T_L(Z^0) = Z^0, \quad (5.12.8)$$

with Z^0 given by (12.7), is equivalent to the relation

$$(AiI + B)(CiI + D)^{-1} = iI, \quad (5.12.9)$$

which gives the relation

$$AiI + B = iI(CiI + D). \quad (5.12.10)$$

Upon equating real and imaginary parts in (12.10), we find the results

$$B = -C, \quad D = A. \quad (5.12.11)$$

Now look at (3.9.28). We see that L must be orthogonal as well as (real) symplectic. We already know from Section 3.9 that such matrices form a $U(n)$ subgroup in $Sp(2n, \mathbb{R})$. Evidently, a necessary and sufficient condition for L to satisfy (12.8) is that L belong to the $U(n)$ subgroup.

Next suppose that X is any real symmetric $n \times n$ matrix and Y is any real symmetric positive definite $n \times n$ matrix. Since Y is real symmetric positive definite, it has a square root $Y^{1/2}$ that is also real symmetric positive definite, and this matrix has an inverse $Y^{-1/2}$ that is also real symmetric positive definite. (See Exercise 12.4.) Consider the matrix $M(Z)$ defined by (12.1) and the relation

$$M(Z) = \begin{pmatrix} Y^{1/2} & 0 \\ 0 & Y^{-1/2} \end{pmatrix} \begin{pmatrix} I & Y^{-1/2}XY^{-1/2} \\ 0 & I \end{pmatrix} = \begin{pmatrix} Y^{1/2} & XY^{-1/2} \\ 0 & Y^{-1/2} \end{pmatrix}. \quad (5.12.12)$$

Look at the two factors in (12.12). The first factor is of the form (3.3.11) and satisfies (3.3.13). Therefore it is symplectic. The second factor is of the form (3.3.9) and satisfies the first of the relations (3.3.12). Therefore it is also symplectic. It follows that M is symplectic. Now let M act on Z^0 . We find the result

$$\begin{aligned} T_M(Z^0) &= (Y^{1/2}iI + XY^{-1/2})(0iI + Y^{-1/2})^{-1} \\ &= (iY^{1/2} + XY^{-1/2})Y^{1/2} = X + iY = Z, \end{aligned} \quad (5.12.13)$$

where, according to (12.1), Z is an arbitrary point in the guhp. Moreover we have the inverse relation

$$T_{M^{-1}}(Z) = T_{M^{-1}}(T_M(Z^0)) = T_{M^{-1}M}(Z^0) = T_I(Z^0) = Z^0. \quad (5.12.14)$$

This result follows from (11.37), which also appears in the logical equivalence chain (11.42), and from (11.6). We note that (11.37) insures the mutual invertibility of any relation of the form (11.10) and its inverse. Finally, let Z' be any other point in the guhp. Define a symplectic matrix M' associated with Z' by the analog of (12.2). Then we have the result

$$Z' = T_{M'}(Z^0) = T_{M'}(T_{M^{-1}}(Z)) = T_{M'M^{-1}}(Z). \quad (5.12.15)$$

We conclude that the symplectic matrix $M'M^{-1}$ sends the arbitrary point Z in the guhp to the arbitrary point Z' in the guhp. Therefore the guhp is indeed a homogeneous space with respect to $Sp(2n, \mathbb{R})$.

5.12.4 Homogeneous Spaces and Cosets

We now resume our general abstract discussion of homogeneous spaces. As before, \mathcal{Z} will denote some space on which some group G acts according to the relation

$$Z' = T_g(Z). \quad (5.12.16)$$

Here g is any element in G and T_g is some transformation rule that depends on g . In analogy with (11.6), we require that the transformation rule satisfy the group representation property

$$T_{g_1}(T_{g_2}(Z)) = T_{g_1 g_2}(Z) \quad (5.12.17)$$

for all “points” Z in \mathcal{Z} and all elements g_1 and g_2 in G . We also require the relation

$$T_e(Z) = Z \text{ for all } Z \text{ in } \mathcal{Z}, \quad (5.12.18)$$

where e is the identity element in G .

5.12.4.1 Definition of Stability Group

Now pick some point in \mathcal{Z} and call it Z^0 .¹⁸ Consider all elements h in G that keep Z^0 fixed. That is, consider all elements h such that

$$T_h(Z^0) = Z^0. \quad (5.12.19)$$

If h_1 is such an element, it follows from (12.17) through (12.19) that h_1^{-1} is also such an element,

$$T_{h_1^{-1}}(Z^0) = Z^0. \quad (5.12.20)$$

Also, if h_1 and h_2 are two such elements, it follows from (12.17) that the product $h_1 h_2$ is also such an element. We conclude that the elements h from a subgroup of G , which we will call H . This subgroup is often referred to as the *stability* (stationary, isotropy, little) *group* of Z^0 .

Suppose we had selected some other point Z^1 instead of Z^0 . Then, since we assume that the action of G is transitive, there is some g_1 in G such that

$$Z^1 = T_{g_1}(Z^0). \quad (5.12.21)$$

Consider the subgroup of elements in G that keep Z^1 fixed. From (12.17), (12.19), and (12.21) we have the relations,

$$\begin{aligned} T_{g_1 h g_1^{-1}}(Z^1) &= T_{g_1 h}(T_{g_1^{-1}}(Z^1)) = T_{g_1 h}(Z^0) \\ &= T_{g_1}(T_h(Z^0)) = T_{g_1}(Z^0) = Z^1. \end{aligned} \quad (5.12.22)$$

We conclude that the subgroup that keeps Z^1 fixed is conjugate (under g_1) to the subgroup that keeps Z^0 fixed. See Exercise 12.6. Therefore, it does not really matter what point in \mathcal{Z} we choose to be a fixed point.

¹⁸We reiterate that in this subsection and the next we are again working in the abstract. That is, we are dealing with some general point Z^0 in some abstract space \mathcal{Z} and not necessarily the specific point Z^0 given by (12.7) in the concrete case $\mathcal{Z} = \text{guhp}$.

5.12.4.2 Use of Stability Group to Define Cosets

We will now use the subgroup H to define an equivalence relation among the elements of G . Suppose g_1 and g_2 are any two elements in G . We say that g_2 is *equivalent* to g_1 (and write $g_2 \sim g_1$) if there exists an h in H such that

$$g_1^{-1}g_2 = h \text{ or, put another way, } g_2 = g_1h. \quad (5.12.23)$$

This equivalence relation can be used to partition the elements of G into disjoint equivalence classes. These equivalence classes are called the *left cosets* of G with respect to H .¹⁹ The collection of all of these cosets is called the *left coset space*, and is customarily denoted by the symbols G/H . See Exercises 12.7 and 12.15.

5.12.4.3 Identification of a Homogeneous Space with Cosets

Suppose two group elements g_1 and g_2 both send Z^0 to the same point Z' ,

$$T_{g_1}(Z^0) = Z', \quad T_{g_2}(Z^0) = Z'. \quad (5.12.24)$$

Then from (12.24) we have the result

$$T_{g_1^{-1}g_2}(Z^0) = T_{g_1^{-1}}(T_{g_2}(Z^0)) = T_{g_1^{-1}}(Z') = Z^0. \quad (5.12.25)$$

It follows from (12.25) and the definition of H that there is an h in H such that (12.23) is satisfied, and we include that g_1 and g_2 are in the same equivalence class (coset). Conversely, if g_1 and g_2 are equivalent (in the same coset), then it follows from (12.23) and (12.19) that

$$T_{g_2}(Z^0) = T_{g_1h}(Z^0) = T_{g_1}(T_h(Z^0)) = T_{g_1}(Z^0). \quad (5.12.26)$$

Thus, they then both send Z^0 to the same point Z' . We conclude that the points Z' may be used to label the cosets, and that the correspondence between cosets in G/H and points Z' in \mathcal{Z} is one-to-one. Put another way, to see what coset a particular group element g belongs to, simply compute $T_g(Z^0)$. We conclude that there is a natural identification between points in \mathcal{Z} and cosets in G/H :

$$\mathcal{Z} \leftrightarrow G/H. \quad (5.12.27)$$

5.12.5 Group Action on Cosets Equals Group Action on a Homogeneous Space

Next, suppose that g_1 is some element in G . All elements in the same coset as g_1 are of the form g_1h with h being an arbitrary element in H . For any element in this coset we have the result

$$T_{g_1h}(Z^0) = T_{g_1}(T_h(Z^0)) = T_{g_1}(Z^0) = Z^1. \quad (5.12.28)$$

¹⁹Note that while, for the second relation in (12.23), the elements of H act by multiplication on the right, the elements of G that are being acted on appear on the *left*.

Let g be any element in G . Consider the element gg_1 . It must belong to some coset. Suppose g_2 also belongs to this coset. Then we must have the relation

$$gg_1 = g_2 h', \quad (5.12.29)$$

where h' is some element in H . We note that (12.29) can also be written in the form

$$g(g_1 h) = g_2 h'', \quad (5.12.30)$$

where h and h'' are elements in H . In this form we see that the effect of g in (12.30) is to send the coset containing g_1 to the coset containing g_2 . That is, elements g in G act on “points” (cosets) in G/H by left multiplication.

Finally, suppose the coset g_1 is labelled by Z^1 as in (12.28), and that the coset containing g_2 is labelled by Z^2 ,

$$T_{g_2}(Z^0) = Z^2. \quad (5.12.31)$$

Let us compute the action of g on Z^1 . We find the result

$$\begin{aligned} T_g(Z^1) &= T_g(T_{g_1 h}(Z^0)) = T_{g g_1 h}(Z^0) \\ &= T_{g_2 h''}(Z^0) = T_{g_2}(T_{h''}(Z^0)) = T_{g_2}(Z^0) = Z^2. \end{aligned} \quad (5.12.32)$$

Upon comparing (12.32) and (12.30), we see that the action of G on \mathcal{Z} is equivalent to the left multiplicative action of G on G/H .

5.12.6 Application of Results to Action of $Sp(2n, \mathbb{R})$ on Siegel Space

How do these results work out in the case of $Sp(2n, \mathbb{R})$ and its action on the guhp? From the discussion surrounding (12.8) we learned that the subgroup H of $Sp(2n, \mathbb{R})$ that keeps Z^0 fixed [with Z^0 given by (12.7)] is $U(n)$. It follows that points Z in the guhp \mathcal{Z} are in one-to-one correspondence with the cosets in $Sp(2n, \mathbb{R})/U(n)$. Indeed, suppose we are given a real symplectic matrix N . In accord with the previous discussion, to find out what coset it belongs to we simply compute $T_N(Z^0)$. From (12.2) and (12.7) we find the result

$$Z = T_N(Z^0) = (iA + B)(iC + D)^{-1}. \quad (5.12.33)$$

Here N is assumed to be written in the block form (3.3.1). Thanks to Exercise 12.3, we know that Z is in the guhp. Finally, we note that the generalized Möbius transformation (12.2) is equivalent to the left multiplicative action of $Sp(2n, \mathbb{R})$ on $Sp(2n, \mathbb{R})/U(n)$.

Let Z be an arbitrary point in the guhp. We know that it labels a coset of $Sp(2n, \mathbb{R})/U(n)$ and that, according to (12.13), the matrix $M(Z)$ given by (12.12) belongs to this coset. The most general matrix belonging to this coset, call it N , is of the form $M(Z)L$ where L is in $U(n)$. We also know that the general element in $U(n)$ can be written in the form $\exp(JS^{c'})$. It follows that the general element N in $Sp(2n, \mathbb{R})$ can be written uniquely in the form

$$N = M(Z) \exp(JS^{c'}) \quad (5.12.34)$$

for some (unique) point Z in the guhp and some $S^{c'}$. Since the general element in $U(n)$ can also be written in the form (3.9.19) with m unitary, it follows that N can just as well be written uniquely in the form

$$N = M(Z)M(m) \quad (5.12.35)$$

for some (unique) point Z in the guhp and some (unique) $n \times n$ unitary matrix m . The factorization (12.34) or (12.35) provides what we will call a *partial Iwasawa decomposition* or *factorization* for $Sp(2n, \mathbb{R})$. For a discussion of the associated partial Iwasawa decomposition of the Lie algebra $sp(2n, \mathbb{R})$, see Exercise 7.2.12. For a discussion of what is usually called the (full) Iwasawa decomposition, see Section *. For a variant of the partial Iwasawa factorization, see Exercise 12.11.

Let us try to write $M(Z)$ in factorized Lie form. Look again at the two factors in (12.12). The second factor can be written in the form

$$\begin{pmatrix} I & Y^{-1/2}XY^{-1/2} \\ 0 & I \end{pmatrix} = \exp \begin{pmatrix} 0 & Y^{-1/2}XY^{-1/2} \\ 0 & 0 \end{pmatrix} = \exp(JS), \quad (5.12.36)$$

where S is given by the relation

$$S = \begin{pmatrix} 0 & 0 \\ 0 & Y^{-1/2}XY^{-1/2} \end{pmatrix}. \quad (5.12.37)$$

The first factor in (12.12) can be written in the form

$$\begin{pmatrix} Y^{1/2} & 0 \\ 0 & Y^{-1/2} \end{pmatrix} = \exp \begin{pmatrix} (1/2) \log Y & 0 \\ 0 & (-1/2) \log Y \end{pmatrix} = \exp(JS) \quad (5.12.38)$$

where S is given by the relation

$$S = \begin{pmatrix} 0 & (1/2) \log Y \\ (1/2) \log Y & 0 \end{pmatrix}. \quad (5.12.39)$$

For an explanation of the meaning of $\log Y$, see Exercise 12.9. Of course, we also know that $M(Z)$ has a factorization of the form

$$M(Z) = \exp(JS^a) \exp(JS^{c''}), \quad (5.12.40)$$

where each of the factors on the right side of (12.40) is unique, and hence uniquely determined by Z . Upon combining (12.34) and (12.40) we find the result

$$\begin{aligned} N &= \exp(JS^a) \exp(JS^{c''}) \exp(JS^{c'}) \\ &= \exp(JS^a) \exp(JS^c), \end{aligned} \quad (5.12.41)$$

which should be compared with (3.8.24).

5.12.7 Action of $Sp(2n, \mathbb{R})$ on the Generalized Real Axis

We have seen that points Z in the guhp \mathcal{Z} are in one-to-one correspondence with the cosets in $Sp(2n, \mathbb{R})/U(n)$. There is second coset/homogeneous space construction that will be of future use. Consider the space of all real $n \times n$ symmetric matrices X . That is, consider all matrices of the form (12.1) with $Y = 0$. This space may be viewed as a *generalized real axis* (gra). Moreover, if we let any $Sp(2n, \mathbb{R})$ element M act on X by the rule

$$X' = T_M(X) = (AX + B)(CX + D)^{-1}, \quad (5.12.42)$$

then we know from the previous discussion that X' will also be symmetric. Thus, Möbius transformations T_M , with M symplectic and real, send the gra into itself. Moreover, if M is of the form (3.3.9), then we have the transformation

$$X' = X + B. \quad (5.12.43)$$

Consequently, since B can be any symmetric matrix, we see that the action of T_M on the gra is transitive. Therefore the gra is a homogeneous space.

In analogy with our previous discussion, take as a representative element in the gra the matrix X^0 defined by the equation

$$X^0 = 0, \quad (5.12.44)$$

and consider all matrices L in $Sp(2n, \mathbb{R})$ that leave X^0 fixed under the action (12.42). The relation

$$T_L(X^0) = X^0, \quad (5.12.45)$$

with X^0 given by (12.44), is equivalent to the relation

$$(A0 + B)(C0 + D)^{-1} = 0, \quad (5.12.46)$$

which gives the relation

$$B = 0. \quad (5.12.47)$$

We have already learned at the end of Section 3.10 that symplectic matrices with $B = 0$ form a subgroup. See (3.10.19) and (3.10.20). This subgroup does not seem to have an established name, but let us call it $H(2n, \mathbb{R})$ or $H(2n, \mathbb{C})$ depending on the field that is being employed. Then we know from the standard construction discussed earlier that elements in the gra are in one-to-one correspondence with cosets in $Sp(2n, \mathbb{R})/H(2n, \mathbb{R})$.

As a sanity check, let us compare dimensions. First compute the dimension of $H(2n, \mathbb{R})$. Since the block A in (3.10.20) is an arbitrary $n \times n$ matrix, its dimension is n^2 . Since the block C is also $n \times n$, and symmetric, its dimension is $n(n+1)/2$. Finally, the block D is completely specified by (3.3.8), and therefore does not contribute to the dimension count. We conclude that the dimension of $H(2n, \mathbb{R})$ is given by the relation

$$\dim H(2n, \mathbb{R}) = n^2 + n(n+1)/2 = n(3n+1)/2. \quad (5.12.48)$$

We already know that the dimension of $Sp(2n, \mathbb{R})$ is $n(2n+1)$. Therefore we have the count

$$\begin{aligned} \dim[Sp(2n, \mathbb{R})/H(2n, \mathbb{R})] &= \dim[Sp(2n, \mathbb{R})] - \dim[H(2n, \mathbb{R})] \\ &= n(2n+1) - n(3n+1)/2 = n(n+1)/2. \end{aligned} \quad (5.12.49)$$

However, since the gra consists of $n \times n$ symmetric matrices, its dimension must also be $n(n + 1)/2$,

$$\dim \text{gra} = n(n + 1)/2. \quad (5.12.50)$$

Comparison of (12.49) and (12.50) gives the result

$$\dim[Sp(2n, \mathbb{R})/H(2n, \mathbb{R})] = \dim \text{gra}, \quad (5.12.51)$$

as expected.

5.12.8 Symplectic Modular Groups

We close this section with a final remark. Generally, the entries in a matrix belonging to $Sp(2n, \mathbb{R})$ can be any real numbers subject only to the symplectic condition. We might wonder whether there are subgroups of $Sp(2n, \mathbb{R})$ for which all the entries in the various matrices in a given subgroup are *integers* (positive, negative, or zero). Such subgroups do indeed exist, and are called symplectic *modular* groups. The symplectic modular groups and their associated generalized Möbius transformations are important for the theory of automorphic, theta, and elliptic functions.²⁰ Automorphic and theta functions are among the most important tools of analytic number theory. Moreover, theta functions and the elliptic functions they generate are key to many soluble problems in nonlinear dynamics.

Exercises

5.12.1. For the transformations (12.2), show that $T_{-M} = T_M$. See (11.11). Consequently, the group of Möbius transformations described by M is only homomorphic to $Sp(2n, \mathbb{R})$, and does not provide a faithful representation. [It does provide a faithful representation of the quotient group G/H where $G = Sp(2n, \mathbb{R})$ and H is the invariant subgroup consisting of $\pm I$. This quotient group is called the *projective* symplectic group, and is denoted by the symbols $PSp(2n, \mathbb{R})$.]

5.12.2. Verify the relations (12.3) through (12.5). With regard to (12.5), also show that Z' is symmetric if Z is, and vice versa.

5.12.3. Suppose Z is given by (12.1) with X real symmetric, and Y real symmetric and positive definite. That is, suppose Z is in the guhp. Also, assume that M is real symplectic.

- a) Show that Z' given by (12.3) is also in the guhp.
- b) Show that Z' given by (12.4) is also in the guhp.
- c) Show that Z is invertible, and that Z' given by (12.5) is also in the guhp.
- d) Show that Z' given by (12.2) is also in the guhp.

²⁰Poincaré's thesis (he was a student of Hermite) was devoted to what he called Fuchsian functions, but are now called automorphic functions.

Hint for part c: Since Y is real symmetric positive definite, there is a real orthogonal matrix O such that

$$OYO^T = D, \quad (5.12.52)$$

where D is diagonal and has positive entries. Define $D^{1/2}$ to be a diagonal matrix with entries equal to the positive square root of the corresponding entries in D . Then we have the relation

$$(D^{1/2})^{-1}OZO^T(D^{1/2})^{-1} = X' + iI, \quad (5.12.53)$$

where X' is given by the relation

$$X' = (D^{1/2})^{-1}OXO^T(D^{1/2})^{-1}. \quad (5.12.54)$$

Verify that X' is real symmetric. Since X' is real symmetric, there is a real orthogonal matrix R such that

$$RX'R^T = D', \quad (5.12.55)$$

where D' is diagonal. Thus, we have the result

$$R(D^{1/2})^{-1}OZO^T(D^{1/2})^{-1}R^T = D' + iI. \quad (5.12.56)$$

Show that $(D' + iI)$ is invertible and that $-(D' + iI)^{-1}$ is in the guph. Finally, show that

$$-Z^{-1} = -O^T(D^{1/2})^{-1}R^T(D' + iI)^{-1}R(D^{1/2})^{-1}O \quad (5.12.57)$$

is in the guhp.

5.12.4. Suppose Y is a real symmetric positive definite matrix. Study the hint to part c of Exercise 12.3. Show that $Y^{1/2}$ defined by the relation

$$Y^{1/2} = O^T D^{1/2} O \quad (5.12.58)$$

satisfies

$$(Y^{1/2})^2 = Y, \quad (5.12.59)$$

and is real symmetric positive definite and invertible, and that its inverse $Y^{-1/2}$ defined by

$$Y^{-1/2} = O^T (D^{1/2})^{-1} O \quad (5.12.60)$$

is also real symmetric positive definite.

5.12.5. Verify (12.12) through (12.15).

5.12.6. Let H^0 be the subgroup that keeps Z^0 fixed, and H^1 be the subgroup that keeps Z^1 fixed. What (12.22) really shows is that all elements of the form $g_1 h g_1^{-1}$, with h in H^0 , are in H^1 . We write this inclusion relation, using set theoretic notation, in the form

$$g_1 H^0 g_1^{-1} \subset H^1. \quad (5.12.61)$$

Show that there is also the relation

$$g_1 H^0 g_1^{-1} \supset H^1, \quad (5.12.62)$$

and therefore

$$g_1 H^0 g_1^{-1} = H^1. \quad (5.12.63)$$

5.12.7. Let X be some (possibly abstract) set, and let \sim be some relation (something that can be true or false) among pairs of elements in X . The relation \sim is said to be an *equivalence* relation if it satisfies three properties:

- a) $x \sim x$ for all x in X (reflexive property).
- b) $x_1 \sim x_2$ implies $x_2 \sim x_1$ for all x_1, x_2 in X (symmetric property).
- c) $x_1 \sim x_2$ and $x_2 \sim x_3$ implies $x_1 \sim x_3$ for all x_1, x_2, x_3 in X (transitive property).

The set of all elements in X that are equivalent (under some given equivalence relation \sim) to a given x in X is called the *equivalence class* of x . Given an equivalence relation \sim on some set X , show that each x in X belongs to one and only one equivalence class. Thus, under an equivalence relation, a set divides up in a natural way into disjoint subsets. Show that both conjugacy and symplectic conjugacy are equivalence relations. See Exercise 3.5.7. Let G be a group having a subgroup H . Show that (12.23) defines (satisfies the properties of) an equivalence relation among the elements of G .

5.12.8. Verify (12.36) and (12.37).

5.12.9. Let D be the diagonal matrix of Exercise 12.3. Define $\log(D)$ to be a diagonal matrix whose entries are the logarithms of the diagonal entries of D . Since the entries of D are positive, these logarithms can all be taken to be real. Define $\log(Y)$ by the rule

$$\log(Y) = O^T \log(D)O, \quad (5.12.64)$$

where O is the real orthogonal matrix of Exercise 12.3. Show that this matrix satisfies the relation

$$\exp[\log(Y)] = Y. \quad (5.12.65)$$

Show also that $\log(Y)$ is real and symmetric.

5.12.10. In Exercise 3.9.10 you should have found that the dimension of the vector space spanned by all $2n \times 2n$ real matrices of the form JS^a is $n(n+1)$. Use (7.17) and (7.18) to obtain this result. If $Z = X + iY$ is $n \times n$ and symmetric (with X and Y real), show that this space also has real dimension $n(n+1)$. Use (12.34) or (12.35) to derive the relation

$$NN^T = M(Z)M^T(Z). \quad (5.12.66)$$

Use (12.41) to derive the relation

$$NN^T = \exp(2JS^a). \quad (5.12.67)$$

Show from (12.66) and (12.67) that a knowledge of Z completely determines JS^a . Use (12.8), (12.13), and (12.34) to derive the relation

$$T_N(Z^0) = Z. \quad (5.12.68)$$

[Here we are working in the guhp with Z^0 given by (12.7).] Show that a knowledge of JS^a also completely determines Z . That is, show that the $\exp(JS^c)$ part of N in (12.41) makes no contribution to (12.68).

5.12.11. The representation (12.35) might be called a *lower left* partial Iwasawa decomposition or factorization of N because the lower left block of $M(Z)$ is empty. See (12.12). The purpose of this exercise is to show that the general symplectic matrix N also has what we will call an *upper right* partial Iwasawa decomposition or factorization of the form

$$N = \bar{M}(\bar{Z})M(\bar{m}), \quad (5.12.69)$$

where $\bar{M}(\bar{Z})$ is a matrix of the form

$$\bar{M}(\bar{Z}) = \begin{pmatrix} \bar{Y}^{-1/2} & 0 \\ 0 & \bar{Y}^{1/2} \end{pmatrix} \begin{pmatrix} I & 0 \\ -\bar{Y}^{-1/2}\bar{X}\bar{Y}^{1/2} & I \end{pmatrix} = \begin{pmatrix} \bar{Y}^{-1/2} & 0 \\ -\bar{X}\bar{Y}^{-1/2} & \bar{Y}^{1/2} \end{pmatrix}. \quad (5.12.70)$$

(Here, as a test of the reader's mental agility, the overbar does not denote complex conjugation, but rather is used only as a distinguishing mark.) To prove (12.69) and (12.70), consider the matrix \bar{N} defined by the relation

$$\bar{N} = JNJ^{-1}. \quad (5.12.71)$$

Since \bar{N} is symplectic, it must have the factorization (12.35),

$$\bar{N} = M(\bar{Z})M(\bar{m}). \quad (5.12.72)$$

Suppose that N and \bar{N} are written in $n \times n$ block form,

$$N = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \quad (5.12.73)$$

$$\bar{N} = \begin{pmatrix} \bar{A} & \bar{B} \\ \bar{C} & \bar{D} \end{pmatrix}. \quad (5.12.74)$$

Use (12.71) to find the relation between A, B, C, D and $\bar{A}, \bar{B}, \bar{C}, \bar{D}$. Show that

$$\begin{aligned} \bar{Z} &= \bar{X} + i\bar{Y} = (i\bar{A} + \bar{B})(i\bar{C} + \bar{D})^{-1} \\ &= -(C - iD)(A - iB)^{-1} = -Z^{-1}. \end{aligned} \quad (5.12.75)$$

Now solve (12.71) for N and use (12.72) to find the relation

$$N = J^{-1}\bar{N}J = J\bar{N}J^{-1} = JM(\bar{Z})J^{-1}JM(\bar{m})J^{-1}. \quad (5.12.76)$$

Use (12.12) and (3.9.19) to find the results

$$JM(\bar{Z})J^{-1} = \bar{M}(\bar{Z}), \quad (5.12.77)$$

$$JM(\bar{m})J^{-1} = M(\bar{m}). \quad (5.12.78)$$

5.12.12. In the theory of a single complex variable z , the domain $z\bar{z} < 1$ [or, equivalently, $(1 - z\bar{z}) > 0$] is the (open) unit disk. Let Z be a matrix of the form (12.1) with both X and Y real and symmetric. In the space of such matrices we may define a *generalized unit disk* (gud) by the relation

$$(I - ZZ^\dagger) > 0, \quad (5.12.79)$$

where here > 0 means positive definite. (Note that since Z is symmetric, $Z^\dagger = \bar{Z}$.) Suppose that Z is in the guhp. In analogy with the case of a single complex variable, it can be shown that W given by

$$W = (Z - iI)(Z + iI)^{-1} \quad (5.12.80)$$

is then in the gud. Conversely, it can be shown that if W is in the gud, then Z given by the inverse of (12.80),

$$Z = i(I + W)(I - W)^{-1}, \quad (5.12.81)$$

is in the guhp. Show that (12.80) sends the point Z^0 given by (12.7) to the origin of the gud. Note that (12.80) and (12.81) are transformations of the form (12.2) with complex entries in M . Indeed, they can be written in the form

$$W = [(2i)^{-1/2}Z - i(2i)^{-1/2}I][(2i)^{-1/2}Z + i(2i)^{-1/2}I]^{-1}, \quad (5.12.82)$$

$$Z = [i(2i)^{-1/2}W + i(2i)^{-1/2}I][-(2i)^{-1/2}W + (2i)^{-1/2}I]^{-1}. \quad (5.12.83)$$

Show that the matrices M associated with the transformations (12.82) and (12.83), see (12.2), are inverses of each other, and are both in $Sp(2n, \mathbb{C})$.

5.12.13. Perform for the guhp a dimension sanity check analogous to that given by (12.51) for the gra. That is, verify the relation

$$\dim[Sp(2n, \mathbb{R})/U(n)] = \dim \text{guhp}. \quad (5.12.84)$$

5.12.14. Suppose M is a symplectic matrix with integer entries. Then the same is true of its powers. Moreover, if N is any other such matrix, products made from M and N have integer entries. Also, the matrices I and J are symplectic matrices with integer entries. Finally, according to (3.1.9), inverse powers of M then also have integer entries. Thus, any set of symplectic matrices with integer entries must form or be part of some group. As described in Subsection 12.8, such groups are called symplectic modular groups. Show that the matrices J_2 [see (3.2.11)] and M and M^T with M given by

$$M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad (5.12.85)$$

generate (by multiplication) a symplectic modular subgroup. Show that this group has an *infinite* number of elements, and exhibit some of them. Consider $n \times n$ orthogonal matrices with integer entries. Do they form a group? Show that there are only a *finite* number of such matrices. Hint: Use Exercise 10.4.

5.12.15. Subsection 12.4 used the subgroup H to define an equivalence relation among the elements of G . It involved right multiplication of elements g of G by elements h of H . Recall (12.23). One can also set up an equivalence relation among elements of G using left multiplication by elements of H . Suppose g_1 and g_2 are any two elements in G . We may say that g_2 is *equivalent* to g_1 (and write $g_2 \sim g_1$) if there exists an h in H such that

$$g_2 g_1^{-1} = h \text{ or, put another way, } g_2 = hg_1. \quad (5.12.86)$$

Verify that (12.86) is indeed an equivalence relation. See Exercise 12.7. This equivalence relation can also be used to partition the elements of G into disjoint equivalence classes. These equivalence classes are called the *right cosets* of G with respect to H . The collection of all of these right cosets is customarily denoted by the symbols $H \backslash G$.

5.13 Möbius Transformations Relating Symplectic and Symmetric Matrices

5.13.1 Overview

Möbius transformations can also be used to show that there is an intimate connection between symplectic matrices and symmetric matrices. Subsequently, as mentioned before, in Section 6.7 the results of this section will be generalized to show that there is a fundamental connection between symplectic maps and gradient maps.

To proceed, we will have to change notation. In the previous sections M was a $2n \times 2n$ matrix that *characterized* a Möbius transformation. In this section M will be a $2n \times 2n$ symplectic matrix that describes the *outcome* of a Möbius transformation acting on some other $2n \times 2n$ matrix W . Conversely, W will also be the outcome of the inverse Möbius transformation *acting* on M . The Möbius transformation itself will be described by a $4n \times 4n$ matrix.

Inspection of (3.11.5) shows that the Cayley representation for M in terms of the $2n \times 2n$ symmetric matrix W is actually a Möbius transformation, and the inverse relation (3.11.12) is also a Möbius transformation. Moreover, the relation between R and W displayed in (4.8.26), and arising from a F_2 generating function, is a Möbius transformation. In both cases a Möbius transformation provides a relation between symplectic and symmetric matrices.

There is a deep reason why, as evinced by these two examples, there is a connection between symplectic and symmetric matrices. And an understanding of this reason will reveal that there are a great many ways of relating symplectic and symmetric matrices by Möbius transformations. This understanding arises as follows: First, introduce a $4n$ dimensional space and define two different symplectic forms on this space. Next show that these forms are congruent under a Darboux transformation. Then show that one of these symplectic forms is related to symplectic matrices in $2n$ dimensional space, and the other is related to symmetric matrices in $2n$ dimensional space. The congruency of the two forms then leads to a connection between symplectic and symmetric matrices. Finally, we will show that this connection is given by Möbius transformations. Along the way we will make use of Lagrangian planes.

5.13.2 The Cayley Möbius Transformation

Before beginning this trek, we pause to extract a useful matrix from the Cayley transformation. Inspection shows that the Cayley transformation (3.12.5) can be written in the form

$$M = T_\tau(W). \quad (5.13.1)$$

where τ is the $4n \times 4n$ matrix

$$\tau = \begin{pmatrix} A^\tau & B^\tau \\ C^\tau & D^\tau \end{pmatrix} \quad (5.13.2)$$

and the matrices A^τ through D^τ are given by the relations

$$A^\tau = J, \quad (5.13.3)$$

$$B^\tau = I, \quad (5.13.4)$$

$$C^\tau = -J, \quad (5.13.5)$$

$$D^\tau = I. \quad (5.13.6)$$

More compactly, we may write

$$\tau = \begin{pmatrix} J & I \\ -J & I \end{pmatrix}. \quad (5.13.7)$$

Here I is the $2n \times 2n$ identity matrix I^{2n} ; and J , which we will sometimes write as J^{2n} , is the standard $2n \times 2n$ fundamental matrix given by (3.1.1). We note, as is easily checked, that τ has the pleasing property

$$\tau^T \tau = 2I^{4n} \text{ or } \tau^{-1} = \tau^T/2. \quad (5.13.8)$$

Conversely, W can be written as a Möbius transformation of M in the form

$$W = T_{\tau^{-1}}(M). \quad (5.13.9)$$

See (3.12.19), and make use of (13.8) to write

$$\tau^{-1} = \begin{pmatrix} -J/2 & J/2 \\ I/2 & I/2 \end{pmatrix}. \quad (5.13.10)$$

We now define, for future use, a matrix σ given in terms of τ by the equation

$$\sigma = \sqrt{2}\tau^{-1} = \tau^T/\sqrt{2} = (1/\sqrt{2}) \begin{pmatrix} -J^{2n} & J^{2n} \\ I^{2n} & I^{2n} \end{pmatrix}. \quad (5.13.11)$$

By this definition and (13.8), σ is orthogonal,

$$\sigma^{-1} = \sigma^T = (1/\sqrt{2}) \begin{pmatrix} J^{2n} & I^{2n} \\ -J^{2n} & I^{2n} \end{pmatrix}. \quad (5.13.12)$$

It can also be verified that

$$\det \sigma = 1. \quad (5.13.13)$$

See Exercise 13.1. Finally we note that, in terms of σ , the Cayley relations (3.12.19) and (3.12.5) take the form

$$W = T_\sigma(M) \quad (5.13.14)$$

and

$$M = T_{\sigma^{-1}}(W). \quad (5.13.15)$$

[Note that, in view of (13.11) and with use of a scaling relation of the form (11.11) with the substitutions $M \rightarrow \sigma$ or σ^{-1} or τ or τ^{-1} and $U \rightarrow M$ or W , the pair (13.1) and (13.9) and the pair (13.14) and (13.15) are equivalent.] We will call T_σ the *Cayley Möbius transformation*.

5.13.3 Two Symplectic Forms and Their Relation by a Darboux Transformation

Now we are ready to continue. As outlined above, we begin by introducing two different symplectic forms in $4n$ dimensional Euclidean space. The first, which we will denote by J^{4n} , is the standard $4n \times 4n$ antisymmetric matrix given by (3.1.1). The second is the $4n \times 4n$ antisymmetric matrix \tilde{J}^{4n} defined by the equation

$$\tilde{J}^{4n} = \begin{pmatrix} J^{2n} & 0^{2n} \\ 0^{2n} & -J^{2n} \end{pmatrix}. \quad (5.13.16)$$

Here 0^{2n} denotes the $2n \times 2n$ null matrix. Evidently \tilde{J}^{4n} has similar properties to those of J^{4n} . It is nonsingular, and in fact satisfies the relation

$$(\tilde{J}^{4n})^2 = -I^{4n}. \quad (5.13.17)$$

It also satisfies, as direct calculation shows, the relation

$$\det \tilde{J}^{4n} = 1. \quad (5.13.18)$$

Our future discussion will capitalize on the properties of J^{4n} and \tilde{J}^{4n} and their block structure.

Is there a relation between J^{4n} and \tilde{J}^{4n} ? Since both are $4n \times 4n$, antisymmetric, and nonsingular, they must be congruent. That is, they must be related by a Darboux transformation. See Section 3.12. Indeed, it is easily verified that there is the congruency relation

$$\sigma^T(J^{4n})\sigma = \tilde{J}^{4n}. \quad (5.13.19)$$

Because σ is orthogonal, J^{4n} and \tilde{J}^{4n} are also *conjugate (similar)*. We also remark that (13.17) and (13.18) now follow directly from (13.12) and (13.19) and the already established properties of J^{4n} . Finally, we will call σ the *Cayley Darboux* matrix.

5.13.4 The Infinite Family of Darboux Transformations

Moreover, there is a $2n(4n + 1)$ parameter family of Darboux transformations that connect J^{4n} and \tilde{J}^{4n} . Suppose that J^{4n} and \tilde{J}^{4n} are congruent under the action of two Darboux matrices α and β ,

$$\alpha^T(J^{4n})\alpha = \tilde{J}^{4n}, \quad (5.13.20)$$

$$\beta^T(J^{4n})\beta = \tilde{J}^{4n}. \quad (5.13.21)$$

By taking determinants of both sides of (13.20) and (13.21) it is easy to see that both α and β are *invertible*. Indeed, they both have determinant +1. All Darboux matrices have determinant +1. See Exercise 13.1. We can say even more. For example, if (13.20) holds, it follows from (13.20) and (13.17) that there is the relation

$$\alpha^{-1} = -\tilde{J}^{4n}\alpha^T J^{4n}. \quad (5.13.22)$$

To continue, from (13.20) and (13.21) we conclude that

$$\alpha^T(J^{4n})\alpha = \beta^T(J^{4n})\beta, \quad (5.13.23)$$

from which it follows that

$$(\beta\alpha^{-1})^T(J^{4n})\beta\alpha^{-1} = (\alpha^{-1})^T\beta^T(J^{4n})\beta\alpha^{-1} = J^{4n}. \quad (5.13.24)$$

Define a matrix γ by the rule

$$\gamma = \beta\alpha^{-1} \text{ or, equivalently, } \beta = \gamma\alpha. \quad (5.13.25)$$

We then see from the far left and far right sides of (13.24) that γ is an element of the group $Sp(4n)$, $\gamma \in Sp(4n)$. Conversely, if β is any element of the form

$$\beta = \gamma\alpha \quad (5.13.26)$$

where γ is an element of $Sp(4n)$ and α satisfies (13.20), then this β satisfies (13.21). We may, for example, take for α the Cayley Darboux matrix σ and write

$$\beta = \gamma\sigma. \quad (5.13.27)$$

Then we get all possible matrices β satisfying (13.21) by using the representation (13.27) and letting γ range over $Sp(4n)$. Therefore the parameter count cited above, which is the dimension of $sp(4n)$, is correct:

$$\text{dimension of set of } 4n \times 4n \text{ Darboux matrices} = \dim sp(4n) = 2n(4n+1). \quad (5.13.28)$$

See (3.7.35) and Table 3.7.1. Thus, for example, in the simplest case of a two-dimensional phase space, there is a 10 parameter family of Darboux matrices/transformations; and in the case of a six-dimensional phase space there is a 78 parameter family of Darboux matrices/transformations.

There is a variant of the argument just made that is also useful. Rewrite (13.27) in the form

$$\beta = \sigma\mu \quad (5.13.29)$$

where μ is yet to be determined. Now require that β be a Darboux transformation so that (13.21) is satisfied. Then, from (13.19) and (13.29) we see that μ must obey the relation

$$\mu^T(\tilde{J}^{4n})\mu = \tilde{J}^{4n}. \quad (5.13.30)$$

We will describe such μ matrices as being \tilde{J}^{4n} symplectic. They form a group which we will refer to as $\tilde{Sp}(4n)$. According to Section 3.12, this group is related to $Sp(4n)$ by a similarity transformation, and therefore has the same dimension as $sp(4n)$. See also Exercise 13.2. Thus, we also get all possible matrices β satisfying (13.21) by using the representation (13.29) and letting μ range over $\tilde{Sp}(4n)$. Either of the representations (13.27) and (13.29) may be used, but sometimes one is more convenient than the other.

Finally, suppose $\hat{\alpha}$ is any matrix of the form

$$\hat{\alpha} = \hat{\gamma}\sigma\hat{\mu} \quad (5.13.31)$$

where

$$\hat{\gamma} \in Sp(4n) \quad (5.13.32)$$

and

$$\hat{\mu} \in \tilde{Sp}(4n). \quad (5.13.33)$$

Then, we find that

$$\begin{aligned} \hat{\alpha}^T(J^{4n})\hat{\alpha} &= (\hat{\gamma}\sigma\hat{\mu})^T(J^{4n})\hat{\gamma}\sigma\hat{\mu} = \hat{\mu}^T\sigma^T\hat{\gamma}^T(J^{4n})\hat{\gamma}\sigma\hat{\mu} \\ &= \hat{\mu}^T\sigma^T(J^{4n})\sigma\hat{\mu} = \hat{\mu}^T(\tilde{J}^{4n})\hat{\mu} \\ &= \tilde{J}^{4n}. \end{aligned} \quad (5.13.34)$$

Here we have used relations of the forms (3.1.2), (13.19), and (13.30). We conclude that $\hat{\alpha}$ is a Darboux matrix.

5.13.5 Isotropic Vectors and Lagrangian Planes

5.13.5.1 Construction and Definitions

Next we will introduce and employ the concept of a Lagrangian plane. Suppose M is a $2n \times 2n$ symplectic matrix. View M as a collection of $2n$ column vectors by writing it in the form

$$M = (m^1, m^2, m^3, \dots, m^{2n}) \quad (5.13.35)$$

where each vector m^j is the j th column of M ,

$$m_i^j = M_{ij}. \quad (5.13.36)$$

(We will say that each vector m^j is of *length*/dimension $2n$ because each has $2n$ entries.) The vectors m^j form a symplectic basis and are therefore linearly independent. See Section 3.6.3. We will also need the $2n$ column vectors e^j , also of length $2n$, that form the columns of I^{2n} . They have the components

$$e_i^j = \delta_{ij}. \quad (5.13.37)$$

See (3.6.4). Now construct $2n$ column vectors u^j , each of length $4n$, by adjoining the entries of each e^j to the bottom of the entries of each m^j . Thus we have

$$u^1 = (m_1^1, m_2^1, m_3^1, \dots, m_{2n}^1; 1, 0, 0, \dots)^T = (M_{1,1}, M_{2,1}, M_{3,1}, \dots, M_{2n,1}; 1, 0, 0, \dots)^T, \quad (5.13.38)$$

$$u^2 = (m_1^2, m_2^2, m_3^2, \dots, m_{2n}^1; 0, 1, 0, \dots)^T = (M_{1,2}, M_{2,2}, M_{3,2}, \dots, M_{2n,2}; 0, 1, 0, \dots)^T \text{ etc.} \quad (5.13.39)$$

Put another way, the vectors u^j (for $j = 1$ to $2n$) have the components

$$u_i^j = m_i^j \text{ for } i = 1 \text{ to } 2n, \quad (5.13.40)$$

$$u_i^j = \delta_{i-2n,j} \text{ for } i = 2n+1 \text{ to } 4n. \quad (5.13.41)$$

Even more compactly, we may write

$$u^j = (m^j; e^j)^T. \quad (5.13.42)$$

Evidently the u^j are linearly independent. Indeed, the m^j are linearly independent and so are the e^j . Now compute the quantities $(u^i, \tilde{J}^{4n}u^j)$. Because of the form of \tilde{J}^{4n} and the form of the u^k , there is the result

$$(u^i, \tilde{J}^{4n}u^j) = (m^i, J^{2n}m^j) - (e^i, J^{2n}e^j) = J_{ij}^{2n} - J_{ij}^{2n} = 0. \quad (5.13.43)$$

[Here we have used the fact that the m^j form a symplectic basis. See (3.6.34).] Because all the $(u^i, \tilde{J}^{4n}u^j)$ vanish, the vectors u^k are said to be *isotropic* with respect to the symplectic form \tilde{J}^{4n} . More succinctly, we will say that the u^k are \tilde{J}^{4n} isotropic. Finally, a set of $2n$ linearly independent isotropic vectors in a $4n$ dimensional space is said to span a *Lagrangian* plane. In this case we will say that the u^k span a \tilde{J}^{4n} Lagrangian plane. (For the reason why such a plane is called Lagrangian, see Section 6.7.2.)

5.13.5.2 Forming Linear Combinations

Suppose we create a new set of linearly independent vectors \dot{u}^i by forming linear combinations of the u^j . We write

$$\dot{u}^i = \sum_j a_{ji} u^j \quad (5.13.44)$$

where the a_{ji} are various coefficients, not all zero, which we may view as the entries in a $2n \times 2n$ matrix a . It is easily verified that the \dot{u}^k are also \tilde{J}^{4n} isotropic,

$$(\dot{u}^i, \tilde{J}^{4n}\dot{u}^j) = 0, \quad (5.13.45)$$

because they are linear combinations of the u^ℓ . They therefore span the same \tilde{J}^{4n} Lagrangian plane as the u^i .

Suppose we also require that any u^k can be expressed as a linear combination of the \dot{u}^ℓ ,

$$u^k = \sum_\ell b_{\ell k} \dot{u}^\ell. \quad (5.13.46)$$

Inserting (13.44) into (13.46) gives the relation

$$u^k = \sum_\ell b_{\ell k} \sum_j a_{j\ell} u^j = \sum_j \sum_\ell a_{j\ell} b_{\ell k} u^j = \sum_j (ab)_{jk} u^j. \quad (5.13.47)$$

Upon comparing both sides of (13.47), and recalling that the u^i are linearly independent, we conclude that there must be the relation

$$(ab)_{jk} = \delta_{jk}. \quad (5.13.48)$$

That is, we must require that a be invertible so that we may write

$$b = a^{-1}. \quad (5.13.49)$$

5.13.6 Connection between Symplectic Matrices and Lagrangian Planes for the Symplectic Form \tilde{J}^{4n}

We will now see that there is a close connection between symplectic matrices and \tilde{J}^{4n} Lagrangian planes. Suppose we view the first $2n$ entries of the u^i as column vectors of a $2n \times 2n$ matrix E and the last $2n$ entries as column vectors of a $2n \times 2n$ matrix F . Then we have the relations

$$E = M \quad (5.13.50)$$

and

$$F = I^{2n}. \quad (5.13.51)$$

We will call the collection $\{E, F\} = \{M, I^{2n}\}$ a *standard symplectic pair*. The first matrix in the pair is symplectic, and the second is the identity, which is also symplectic.

Recall the vectors \acute{u}^i given by (13.44). If we use the \acute{u}^i to construct $2n \times 2n$ matrices \acute{E} and \acute{F} in the same way E and F were constructed from the u^i , we find from (13.44) the relations

$$\acute{E} = Ea = Ma \quad (5.13.52)$$

and

$$\acute{F} = Fa = I^{2n}a. \quad (5.13.53)$$

We will call the collection $\{\acute{E}, \acute{F}\} = \{Ea, Fa\} = \{Ma, I^{2n}a\}$ an *equivalent symplectic pair* and write

$$\{Ea, Fa\} \sim \{E, F\} \quad (5.13.54)$$

because multiplication on the right of a pair of matrices by a nonsingular matrix can be shown to set up an equivalence relations among pairs of matrices. See Exercise 13.3. And, because such multiplication does set up an equivalence relation and we have assumed a is nonsingular, we may also write

$$\{Ma, I^{2n}a\} \sim \{M, I^{2n}\}. \quad (5.13.55)$$

Moreover, suppose we are given any set of $2n$ linearly independent \tilde{J}^{4n} isotropic vectors \acute{u}^i in a $4n$ dimensional space and from them we form the associated matrices \acute{E} and \acute{F} . Then it is easy to verify from the definition (13.16) and the block structure of \tilde{J}^{4n} that the \tilde{J}^{4n} isotropy condition (13.45) is equivalent to the matrix relation

$$\acute{E}^T J^{2n} \acute{E} = \acute{F}^T J^{2n} \acute{F}. \quad (5.13.56)$$

If \acute{F} is invertible, (13.56) can be rewritten in the equivalent form

$$(\acute{E} \acute{F}^{-1})^T J^{2n} (\acute{E} \acute{F}^{-1}) = J^{2n}, \quad (5.13.57)$$

and we conclude that the matrix M defined by

$$M = \acute{E} \acute{F}^{-1} \quad (5.13.58)$$

is symplectic. In terms of our equivalence relation, we may write

$$\{\acute{E}, \acute{F}\} \sim \{\acute{E} \acute{F}^{-1}, \acute{F} \acute{F}^{-1}\} = \{M, I^{2n}\} \quad (5.13.59)$$

with M given by (13.58). Thus, any set of basis vectors spanning a \tilde{J}^{4n} Lagrangian plane whose associated “ F ” matrix is invertible produces an equivalent standard symplectic pair. Conversely, we have already seen that any symplectic matrix M produces a set of basis vectors spanning a \tilde{J}^{4n} Lagrangian plane and a standard symplectic pair. In this latter case, their associated F matrix is trivially invertible because it is the identity.

5.13.7 Connection between Symmetric Matrices and Lagrangian Planes for the Symplectic Form J^{4n}

There is an analogous construction that can be carried out for the symplectic form J^{4n} , but now using symmetric matrices W . Suppose W is a $2n \times 2n$ symmetric matrix. View W as a collection of $2n$ column vectors, each of length $2n$, by writing it in the form

$$W = (w^1, w^2, w^3, \dots, w^{2n}) \quad (5.13.60)$$

where each vector w^j is the j th column of W ,

$$w_i^j = W_{ij}. \quad (5.13.61)$$

Again we will also employ the $2n$ column vectors e^j , also of length $2n$, that form the columns of I^{2n} . Now construct $2n$ column vectors v^j , each of length $4n$, by adjoining the entries of each e^j to the bottom of the entries of each w^j . This procedure will again yield $2n$ linearly independent vectors because the e^j are linearly independent. Using the compact notation introduced earlier, we may write the v^j in the form

$$v^j = (w^j; e^j)^T. \quad (5.13.62)$$

Let us now compute the quantities $(v^i, J^{4n}v^j)$. From the block form of J^{4n} and (13.62) it is easily checked that the result is given by the relation

$$(v^i, J^{4n}v^j) = (w^i, e^j) - (e^i, w^j) = (e^j, w^i) - (e^i, w^j). \quad (5.13.63)$$

But from (13.61) we have the result

$$(e^i, w^j) = w_i^j = W_{ij}. \quad (5.13.64)$$

Combining these results gives the relation

$$(v^i, J^{4n}v^j) = W_{ji} - W_{ji} = 0. \quad (5.13.65)$$

Here we have used the fact that W is assumed to be symmetric. We conclude that the v^j are J^{4n} isotropic, and span a J^{4n} Lagrangian plane.

As before, from the v^j construct two $2n \times 2n$ matrices, call them G and H . Construct G using the first $2n$ entries in the v^j , and construct H using the last $2n$ entries. In this case we evidently get the results

$$G = W \quad (5.13.66)$$

and

$$H = I^{2n}. \quad (5.13.67)$$

We will call the collection $\{G, H\} = \{W, I^{2n}\}$ a *standard symmetric pair*. The first matrix in the pair is symmetric, and the second is the identity, which is also symmetric.

We can also form vectors \dot{v}^i by taking linear combinations of the v^j . These vectors will also be J^{4n} isotropic,

$$(\dot{v}^i, J^{4n} \dot{v}^j) = 0, \quad (5.13.68)$$

and span the same J^{4n} Lagrangian plane. Moreover, their associated $2n \times 2n$ matrices are given by the relations

$$\dot{G} = Ga = Wa \quad (5.13.69)$$

and

$$\dot{H} = Ha = I^{2n}a. \quad (5.13.70)$$

We will call the collection $\{\dot{G}, \dot{H}\} = \{Ga, Ha\} = \{Wa, I^{2n}a\}$ an equivalent symmetric pair and write

$$\{Ga, Ha\} \sim \{G, H\} = \{W, I^{2n}\}. \quad (5.13.71)$$

Finally, suppose we are given any set of $2n$ linearly independent J^{4n} isotropic vectors \dot{v}^i in a $4n$ dimensional space and from them we form the associated matrices \dot{G} and \dot{H} . Then it is easy to verify from the definition (3.1.1) and the block structure of J^{4n} that the J^{4n} isotropy condition (13.68) is equivalent to the matrix relation

$$\dot{G}^T \dot{H} - \dot{H}^T \dot{G} = 0. \quad (5.13.72)$$

If \dot{H} is invertible, (13.72) can be rewritten in the equivalent form

$$\dot{G} \dot{H}^{-1} = (\dot{H}^{-1})^T \dot{G}^T. \quad (5.13.73)$$

Therefore the matrix W defined by the equation

$$W = \dot{G} \dot{H}^{-1} \quad (5.13.74)$$

is symmetric,

$$W^T = W. \quad (5.13.75)$$

In terms of our equivalence relation, we may write

$$\{\dot{G}, \dot{H}\} \sim \{\dot{G} \dot{H}^{-1}, \dot{H} \dot{H}^{-1}\} = \{W, I^{2n}\} \quad (5.13.76)$$

with W given by (13.74). Thus, any set of basis vectors spanning a J^{4n} Lagrangian plane whose associated “ H ” matrix is invertible produces an equivalent standard symmetric pair. Conversely, we have already seen that any symmetric matrix W produces a set of basis vectors spanning a J^{4n} Lagrangian plane and a standard symmetric pair. In this latter case, their associated H matrix is trivially invertible because it is the identity.

5.13.8 Relation between Symplectic and Symmetric Matrices and the Role of Darboux Möbius Transformations

The stage is set to discover the relation between symplectic and symmetric matrices. Suppose we are given some set of $2n$ vectors u^i that span a \tilde{J}^{4n} Lagrangian plane. Form associated vectors v^i by the rule

$$v^i = \alpha u^i \quad (5.13.77)$$

where α is any $4n \times 4n$ matrix. If we now require that α be a Darboux matrix that satisfies the relation (13.20), then we find the result

$$(v^i, J^{4n} v^j) = (\alpha u^i, J^{4n} \alpha u^j) = (u^i, \alpha^T J^{4n} \alpha u^j) = (u^i, \tilde{J}^{4n} u^j) = 0. \quad (5.13.78)$$

That is, the vectors v^i are J^{4n} isotropic, and span a J^{4n} Lagrangian plane. Next, construct $2n \times 2n$ matrices G and H from the first $2n$ and the last $2n$ entries in the v^i , respectively. Suppose that the matrix H turns out to be invertible. Then we can write

$$\{G, H\} \sim \{GH^{-1}, I^{2n}\}. \quad (5.13.79)$$

From the previous discussion we know that the W given by

$$W = GH^{-1}, \quad (5.13.80)$$

will be symmetric.

What do Möbius transformations have to do with this discussion? Watch. Suppose we are given a symplectic matrix M and from it construct the vectors u^i by (13.42). That is, the u^i are the vectors associated with the standard symplectic pair $\{M, I^{2n}\}$. Define associated vectors v^i using (13.77). Let E and F be the matrices associated with the u^i , and let G and H be the matrices associated with the v^i . Suppose also that we write α in the block form

$$\alpha = \begin{pmatrix} A^\alpha & B^\alpha \\ C^\alpha & D^\alpha \end{pmatrix}. \quad (5.13.81)$$

Then, in terms of the matrices A^α through D^α and the matrices E through G , the relation (13.77) is equivalent to the relations

$$G = A^\alpha E + B^\alpha F, \quad (5.13.82)$$

$$H = C^\alpha E + D^\alpha F. \quad (5.13.83)$$

Therefore we have the result

$$W = GH^{-1} = (A^\alpha E + B^\alpha F)(C^\alpha E + D^\alpha F)^{-1}. \quad (5.13.84)$$

Now use the explicit forms of E and F given by (13.50) and (13.51) to rewrite (13.84). So doing gives the result

$$W = GH^{-1} = (A^\alpha M + B^\alpha)(C^\alpha M + D^\alpha)^{-1}. \quad (5.13.85)$$

5.13.8.1 Mapping of Symplectic Matrices into Symmetric Matrices

We see that W is related to M by the Möbius transformation associated with α ,

$$W = T_\alpha(M). \quad (5.13.86)$$

Thus, W has been expressed as the Möbius transformation of a symplectic matrix. Of course, for (13.86) to be well defined, the matrix $(C^\alpha M + D^\alpha)$ must be invertible,

$$\det(C^\alpha M + D^\alpha) \neq 0. \quad (5.13.87)$$

[Note that (13.87) is precisely the condition for H to be invertible. See (13.83), (13.50), and (13.51).] That is, given M , we must find some Darboux matrix α satisfying (13.20) such that its associated C^α and D^α also satisfy (13.87). Once this is achieved (and we will verify subsequently that it can be achieved), we know from our previous work that the W given by (13.80) and hence by (13.86) will be symmetric.

Suppose we now hold α fixed (thereby holding its associated C^α and D^α fixed), and vary M . It can be verified by continuity that, for small enough variations in M , (13.87) will continue to hold. Correspondingly, again based on our previous work, we know that the varied W associated with the varied M will continue to be symmetric. Thus we get a local mapping of symplectic matrices into symmetric matrices. Since the α appearing in T_α is a Darboux transformation, we might call T_α a Darboux Möbius transformation.

5.13.8.2 Mapping of Symmetric Matrices into Symplectic Matrices

Conversely, suppose we are given a symmetric matrix W , which may be the W of (13.86). From it construct the vectors v^i using (13.61) and (13.62). That is, the v^i are the vectors associated with the standard symmetric pair $\{W, I^{2n}\}$. Define associated vectors u^i in terms of the v^i by the rule

$$u^i = \alpha^{-1} v^i \quad (5.13.88)$$

where, as before, α is any $4n \times 4n$ Darboux matrix that satisfies (13.20). [Note that (13.88) is equivalent to (13.77).] For the u^i we find the relation

$$(u^i, \tilde{J}^{4n} u^j) = (\alpha^{-1} v^i, \tilde{J}^{4n} \alpha^{-1} v^j) = (v^i, (\alpha^T)^{-1} \tilde{J}^{4n} \alpha^{-1} v^j) = (v^i, J^{4n} v^j) = 0. \quad (5.13.89)$$

We see that the vectors u^i are \tilde{J}^{4n} isotropic. Now construct $2n \times 2n$ matrices E and F from the first $2n$ and the last $2n$ entries in the u^i , respectively. Suppose that the matrix F turns out to be invertible. Then we can write

$$\{E, F\} \sim \{EF^{-1}, I^{2n}\}. \quad (5.13.90)$$

From the previous discussion we know that the M given by

$$M = EF^{-1}, \quad (5.13.91)$$

will be symplectic. Also, let G and H be the matrices associated with the v^i and write the matrix α^{-1} in the block form form

$$\alpha^{-1} = \begin{pmatrix} A^{\alpha^{-1}} & B^{\alpha^{-1}} \\ C^{\alpha^{-1}} & D^{\alpha^{-1}} \end{pmatrix}. \quad (5.13.92)$$

In terms of the matrices $A^{\alpha^{-1}}$ through $D^{\alpha^{-1}}$ and the matrices E through G , the relation (13.88) is equivalent to the relations

$$E = A^{\alpha^{-1}}G + B^{\alpha^{-1}}H, \quad (5.13.93)$$

$$F = C^{\alpha^{-1}}G + D^{\alpha^{-1}}H. \quad (5.13.94)$$

Therefore we have the result

$$M = EF^{-1} = (A^{\alpha^{-1}}G + B^{\alpha^{-1}}H)(C^{\alpha^{-1}}G + D^{\alpha^{-1}}H)^{-1}. \quad (5.13.95)$$

Now use the explicit forms of G and H given by (13.66) and (13.67) to rewrite (13.95) in the form

$$M = (A^{\alpha^{-1}}W + B^{\alpha^{-1}})(C^{\alpha^{-1}}W + D^{\alpha^{-1}})^{-1}. \quad (5.13.96)$$

We see, consistent with (13.86), that M is related to W by the Möbius transformation associated with α^{-1} ,

$$M = T_{\alpha^{-1}}(W). \quad (5.13.97)$$

Thus, M has been expressed as the Möbius transformation of a symmetric matrix. Of course, for this relation to make sense, the matrix $(C^{\alpha^{-1}}W + D^{\alpha^{-1}})$ must be invertible,

$$\det(C^{\alpha^{-1}}W + D^{\alpha^{-1}}) \neq 0. \quad (5.13.98)$$

Observe that (13.98) is exactly the condition for F to be invertible.

For fixed α , and hence fixed $C^{\alpha^{-1}}$ and fixed $D^{\alpha^{-1}}$, the relation (13.98) describes an open set in W space. Therefore $T_{\alpha^{-1}}$ provides a local mapping of symmetric matrices into symplectic matrices. Finally we know from the work of Subsection 11.3 that if (13.87) holds (thereby making it possible to find a symmetric W given a symplectic M), then (13.98) also holds (thereby making it possible to find a symplectic M given a symmetric W), and vice versa. That is, there is the logical equivalence

$$\det(C^{\alpha^{-1}}W + D^{\alpha^{-1}}) \neq 0 \Leftrightarrow \det(C^\alpha M + D^\alpha) \neq 0. \quad (5.13.99)$$

To verify this claim, make in the first line of (11.42) the substitutions $M \rightarrow \alpha$, $U' \rightarrow W$, and $U \rightarrow M$.

We close this subsection with the observation that to find α^{-1} it is not actually necessary to carry out the inversion of a $4n \times 4n$ matrix. Instead one can use the inversion relation (13.22), which only involves matrix multiplication. Indeed, it is easily verified that its use gives the results

$$A^{\alpha^{-1}} = J^{2n}(C^\alpha)^T, \quad (5.13.100)$$

$$B^{\alpha^{-1}} = -J^{2n}(A^\alpha)^T, \quad (5.13.101)$$

$$C^{\alpha^{-1}} = -J^{2n}(D^\alpha)^T, \quad (5.13.102)$$

$$D^{\alpha^{-1}} = J^{2n}(B^\alpha)^T. \quad (5.13.103)$$

5.13.9 Completion of Tasks

5.13.9.1 Verification of Möbius Transformation Invertibility Conditions

Several uncompleted tasks remain. The first is to verify that, given a symplectic matrix M , a Darboux matrix α satisfying (13.20) can be found such that the conditions (13.87) and (13.98) are also satisfied. We have already seen, as stated in (13.99), that these conditions are logically equivalent. Now we will learn more. Actually, for notational convenience, we will find a Darboux matrix β with these desired properties.

Suppose L is a symplectic matrix near M so that we may write

$$M = LN \quad (5.13.104)$$

where N is a symplectic matrix near the identity. Inspection of the Cayley Möbius transformation T_σ given by (13.14), see also (3.11.12), shows that it is ideally suited to matrices M near the identity I . What we would like to find is a choice of β such that the Darboux Möbius transformation T_β is ideally suited to matrices near L . This is easily done using group properties. We first find a Möbius transformation that sends L to I and then follow it by a Cayley Möbius transformation. Of course, in so doing, we must ensure that the resulting β is also a Darboux transformation. Let μ be the $4n \times 4n$ matrix defined by the rule

$$\mu = \begin{pmatrix} L^{-1} & 0 \\ 0 & I^{2n} \end{pmatrix}. \quad (5.13.105)$$

Then, analogous to the relations (11.47) and (11.48), we have the result

$$T_\mu(L) = I. \quad (5.13.106)$$

Furthermore, we have the relation

$$T_\mu(M) = N. \quad (5.13.107)$$

Also we observe that μ is an element of $\tilde{Sp}(4n)$ since I is symplectic and L^{-1} is symplectic (because L is assumed to be symplectic). Therefore the β given by (13.29) will be a Darboux matrix. Its associated Möbius transformation will have the property

$$W = T_\beta(M) = T_{\sigma\mu}(M) = T_\sigma(T_\mu(M)) = T_\sigma(N) = (-JN + J)(N + I)^{-1}. \quad (5.13.108)$$

Here we have used the group property of Möbius transformations. Evidently the matrix $(N + I)$ will be invertible for N sufficiently near the identity. Indeed, all that is required is that -1 not be an eigenvalue of N . Correspondingly, $T_\beta(M)$ is well defined. Finally we see from (13.11), (13.29), and (13.105) that β has the explicit form

$$\beta = (1/\sqrt{2}) \begin{pmatrix} -J^{2n}L^{-1} & J^{2n} \\ L^{-1} & I^{2n} \end{pmatrix}. \quad (5.13.109)$$

Therefore, if we write out $T_\beta(M)$ explicitly, we find the result

$$W = T_\beta(M) = (A^\beta M + B^\beta)(C^\beta M + D^\beta)^{-1} \quad (5.13.110)$$

with

$$(A^\beta M + B^\beta) = (1/\sqrt{2})(-J^{2n}L^{-1}M + J^{2n}) = (1/\sqrt{2})(-JN + J), \quad (5.13.111)$$

and

$$(C^\beta M + D^\beta) = (1/\sqrt{2})(L^{-1}M + I^{2n}) = (1/\sqrt{2})(N + I). \quad (5.13.112)$$

We see that

$$\det(C^\beta M + D^\beta) \neq 0 \quad (5.13.113)$$

provided N is sufficiently near I .

The result inverse to (13.108) is given by the relation

$$\begin{aligned} M &= T_{\beta^{-1}}(W) = T_{(\sigma\mu)^{-1}}(W) = T_{\mu^{-1}\sigma^{-1}}(W) \\ &= T_{\mu^{-1}}(T_{\sigma^{-1}}(W)) = T_{\mu^{-1}}(N) = LN. \end{aligned} \quad (5.13.114)$$

Here we have again used the group property of Möbius transformations and the fact that, consistent with (13.108), there is the relation

$$N = T_{\sigma^{-1}}(W) = (JW + I)(-JW + I)^{-1}. \quad (5.13.115)$$

Note that (13.115) is well defined provided

$$\det(-JW + I) \neq 0. \quad (5.13.116)$$

This condition is met for W sufficiently near 0. The matrix W will, in turn, be near 0 if N is sufficiently near I . See (13.108). We can also evaluate $T_{\beta^{-1}}(W)$ directly. For β^{-1} we find the result

$$\beta^{-1} = (1/\sqrt{2}) \begin{pmatrix} LJ^{2n} & L \\ -J^{2n} & I^{2n} \end{pmatrix}. \quad (5.13.117)$$

See Exercise 13.4. Consequently, we obtain the result

$$M = T_{\beta^{-1}}(W) = (LJW + L)(-JW + I)^{-1}. \quad (5.13.118)$$

Again we see that the condition (13.116) arises and is satisfied.

In summary, for an arbitrary symplectic matrix L , and for $M = L$ or M in the neighborhood of L , the Möbius transformations $W = T_\beta(M)$ and $M = T_{\beta^{-1}}(W)$ are well defined when β is the Darboux transformation defined by (13.11), (13.29), and (13.105).

5.13.9.2 Verification of Full Family of Darboux Transformations

The second task is to verify that, given a symplectic matrix M , we have the full advertised $2n(4n + 1)$ degrees of freedom in the choice of α . Suppose we hold M fixed and replace α by the β given by (13.26). Write (13.86) more explicitly as

$$W(\alpha) = T_\alpha(M) \quad (5.13.119)$$

to indicate that W depends on α as well as on M . Then we have the relation

$$W(\beta) = T_\beta(M) = T_{\gamma\alpha}(M) = T_\gamma(T_\alpha(M)) = T_\gamma(W(\alpha)). \quad (5.13.120)$$

Note that T_γ is a Möbius transformation associated with a symplectic transformation since γ is assumed to be in $Sp(4n)$. From the work of Subsection 12.7 and earlier we know that such Möbius transformations send symmetric matrices into symmetric matrices. Therefore the matrix $W(\beta)$ will also be symmetric. Of course we must again worry about the inversion of the matrix occurring in the second factor of the Möbius transformation. In the case where T_γ is applied to $W(\alpha)$ this matrix will be $[C^\gamma W(\alpha) + D^\gamma]$. It is easy to check that, for γ sufficiently near the identity, the matrix C^γ is small and the matrix D^γ is near the identity. Thus the required inverse exists, and we get a full $2n(4n+1)$ parameter family of Möbius transformations T_β that send the symplectic matrix M to the symmetric matrix $W(\beta)$.

To finish this aspect of our discussion, we should also explore what happens in the map (13.97) when W is held fixed, and α is varied. Again we will replace α by β with β given by (13.26). Then we may view M as depending on β and write

$$M(\beta) = T_{\beta^{-1}}(W) = T_{(\gamma\alpha)^{-1}}(W) = T_{\alpha^{-1}\gamma^{-1}}(W) = T_{\alpha^{-1}}(T_{\gamma^{-1}}(W)) = T_{\alpha^{-1}}(W') \quad (5.13.121)$$

where

$$W' = T_{\gamma^{-1}}(W). \quad (5.13.122)$$

Moreover we know that W' will be symmetric because W is symmetric, γ^{-1} is in $Sp(4n)$, and Möbius transformations corresponding to symplectic matrices send symmetric matrices into symmetric matrices. Again see Subsection 12.7. For γ near the identity W' will be near W and consequently, by the argument of the previous paragraph, $M(\beta)$ will be well defined and symplectic. To verify this claim, examine W' . Writing out (13.122) in detail gives the result

$$W' = (A^{\gamma^{-1}}W + B^{\gamma^{-1}})(C^{\gamma^{-1}}W + D^{\gamma^{-1}})^{-1}. \quad (5.13.123)$$

For γ near the identity, The matrices $A^{\gamma^{-1}}$ and $D^{\gamma^{-1}}$ will be near the identity, and the matrices $B^{\gamma^{-1}}$ and $C^{\gamma^{-1}}$ will be small. Therefore W' will be well defined, and indeed will be near W . Thus we get a full $2n(4n+1)$ parameter family of Möbius transformations $T_{\beta^{-1}}$ that send the symmetric matrix W to the symplectic matrix $M(\beta)$.

5.13.9.3 Freedom in the Choice of Darboux Transformation

Finally, for some semblance of completeness, we should address the question of what freedom exists in the choice of α for the relations (13.86) and (13.97) when both M and W are held fixed. Suppose we require that

$$M(\beta) = M(\alpha) \quad (5.13.124)$$

or, equivalently,

$$T_{\beta^{-1}}(W) = T_{\alpha^{-1}}(W). \quad (5.13.125)$$

Then we conclude that

$$W = T_I(W) = T_\beta(T_{\beta^{-1}}(W)) = T_\beta(T_{\alpha^{-1}}(W)) = T_{\beta\alpha^{-1}}(W) = T_\gamma(W). \quad (5.13.126)$$

Here we have used the definition (13.25). Upon comparing the far left and far right sides of (13.126) we see that W must be a fixed point of T_γ . From the work of Subsection 12.4.1 we know that such γ form a subgroup, the stability group of W .

Even more can be said. Suppose δ is the $Sp(4n)$ element

$$\delta = \begin{pmatrix} I^{2n} & W \\ 0^{2n} & I^{2n} \end{pmatrix}. \quad (5.13.127)$$

From the discussion of Section 3.3 we know that δ is indeed in $Sp(4n)$ because W is symmetric. Moreover, by direct calculation or from the discussion surrounding (12.43), we know that this δ has the property

$$T_\delta(0^{2n}) = W. \quad (5.13.128)$$

Consequently, (13.126) can be rewritten in the form

$$T_\delta(0^{2n}) = T_\gamma(T_\delta(0^{2n})) \text{ or, equivalently, } T_\gamma(T_\delta(0^{2n})) = T_\delta(0^{2n}) \quad (5.13.129)$$

from which it follows that

$$T_{\delta^{-1}}(T_\gamma(T_\delta(0^{2n}))) = 0^{2n} \text{ or, equivalently, } T_{(\delta^{-1}\gamma\delta)}(0^{2n}) = 0^{2n}. \quad (5.13.130)$$

Let ϵ be the $Sp(4n)$ element specified by the definition

$$\epsilon = \delta^{-1}\gamma\delta \text{ or, equivalently, } \gamma = \delta\epsilon\delta^{-1}. \quad (5.13.131)$$

From (13.130) we see that ϵ must be in the stability group of the zero matrix,

$$T_\epsilon(0^{2n}) = 0^{2n}. \quad (5.13.132)$$

We have encountered this group before in Subsection 12.7. Using the notation introduced there, it is the group $H(4n, \mathbb{R})$ with dimension $(6n^2 + n)$. Combining (13.26) and (13.131) gives the relation

$$\beta = \delta\epsilon\delta^{-1}\alpha. \quad (5.13.133)$$

We conclude there is a $(6n^2 + n)$ parameter set of matrices β that satisfy, for fixed M and fixed W , the relation

$$M = T_{\beta^{-1}}(W) \quad (5.13.134)$$

and its inverse

$$W = T_\beta(M). \quad (5.13.135)$$

5.13.9.4 Explicit Construction of the Most General Darboux Transformation

To explore the implications of the relations (13.133) through (13.135) in a concrete case, let us begin by constructing a particular Darboux transformation ϕ such that its associated Möbius transformation T_ϕ sends any specified symplectic matrix L into any specified symmetric matrix V :

$$V = T_\phi(L) \quad (5.13.136)$$

and

$$L = T_{\phi^{-1}}(V). \quad (5.13.137)$$

This is easily done. Let θ be the $4n \times 4n$ matrix defined by the rule

$$\theta = \begin{pmatrix} I^{2n} & V \\ 0 & I^{2n} \end{pmatrix}. \quad (5.13.138)$$

Its associated Möbius transformation has the property

$$T_\theta(0^{2n}) = V. \quad (5.13.139)$$

See (12.3) or the V analog of (13.127) and (13.128). Moreover, θ is J^{4n} symplectic. See (3.3.9). Now define a $4n \times 4n$ matrix ϕ by the rule

$$\phi = \theta\sigma\mu. \quad (5.13.140)$$

Here σ and μ are defined by (13.11) and (13.105), respectively. Since θ is J^{4n} symplectic and μ is \tilde{J}^{4n} symplectic, ϕ will be a Darboux transformation. See the discussion associated with (13.31) through (13.34). Also, by construction and the group property, we have the relation

$$T_\phi(L) = T_{\theta\sigma\mu}(L) = T_\theta(T_\sigma(T_\mu(L))) = T_\theta(T_\sigma(I^{2n})) = T_\theta(0^{2n}) = V. \quad (5.13.141)$$

Here we have used (13.106), (13.139), and the Cayley transformation property

$$T_\sigma(I^{2n}) = 0^{2n}. \quad (5.13.142)$$

Evaluation of (13.140) gives the explicit result

$$\phi = (1/\sqrt{2}) \begin{pmatrix} [-JL^{-1} + VL^{-1}] & [J + V] \\ L^{-1} & I \end{pmatrix}. \quad (5.13.143)$$

Let us continue by setting $\alpha = \phi$ in (13.133) and $W = V$ in (13.127) so that $\delta = \theta$. Then we find for β the result

$$\beta = \delta\epsilon\delta^{-1}\theta\sigma\mu = \theta\epsilon\sigma\mu. \quad (5.13.144)$$

This β will be the most general Darboux transformation such that

$$V = T_\beta(L) \quad (5.13.145)$$

and

$$L = T_{\beta^{-1}}(V). \quad (5.13.146)$$

In view of the discussion at the beginning of Section 3.3 and (3.10.20), the general ϵ in the group $H(4n, \mathbb{R})$ can be written in the form

$$\epsilon = \begin{pmatrix} A & 0 \\ 0 & (A^T)^{-1} \end{pmatrix} \begin{pmatrix} I^{2n} & 0 \\ C & I^{2n} \end{pmatrix}. \quad (5.13.147)$$

Carrying out the indicated multiplications (13.144) gives for β the explicit form

$$\beta = (1/\sqrt{2}) \begin{pmatrix} -AJL^{-1} + V(A^T)^{-1}(-CJ + I)L^{-1} & AJ + V(A^T)^{-1}(CJ + I) \\ (A^T)^{-1}(-CJ + I)L^{-1} & (A^T)^{-1}(CJ + I) \end{pmatrix}, \quad (5.13.148)$$

and for its inverse the explicit form

$$\beta^{-1} = (1/\sqrt{2}) \begin{pmatrix} LJ A^{-1} - LCA^{-1} & -LJA^{-1}V + L(CA^{-1}V + A^T) \\ -JA^{-1} - CA^{-1} & JA^{-1}V + (CA^{-1}V + A^T) \end{pmatrix}. \quad (5.13.149)$$

(See Exercise 13.5.) Here J stands for J^{2n} . It is readily verified by direct calculation that (13.145) and (13.146) are satisfied. We observe, since $H(4n, \mathbb{R})$ is a $(6n^2 + n)$ dimensional group, that there is a $(6n^2 + n)$ dimensional family of Darboux matrices that relate a specified L to a specified V . Note also that we have written (parameterized) the general $4n \times 4n$ Darboux matrix β in terms of a general $2n \times 2n$ symplectic matrix L , two general $2n \times 2n$ symmetric matrices C and V , and a general $GL(2n)$ matrix A . Exercise 13.8 shows that this parameterization must in fact have some redundancy because the parameter count for this parameterization exceeds the dimensionality of $sp(4n)$.

5.13.9.5 Two Convenient Simpler Choices

There are two convenient simpler choices for Darboux matrices whose associated Möbius transformations relate any specified symplectic matrix L to any specified symmetric matrix V . The first, call it $\tilde{\beta}$, is the Darboux matrix obtained by setting $A = I$ in (13.148) to yield the result

$$\tilde{\beta} = (1/\sqrt{2}) \begin{pmatrix} [-JL^{-1} + V(-CJ + I)L^{-1}] & [J + V(CJ + I)] \\ (-CJ + I)L^{-1} & (CJ + I) \end{pmatrix} \quad (5.13.150)$$

with the inverse

$$\tilde{\beta}^{-1} = (1/\sqrt{2}) \begin{pmatrix} LJ - LC & -LJV + L(CV + I) \\ -J - C & JV + (CV + I) \end{pmatrix}. \quad (5.13.151)$$

[That $\tilde{\beta}$ is a Darboux matrix follows from the fact β as given by (13.148) is a Darboux matrix for all choices of A .] It is easily verified by direct calculation that

$$V = T_{\tilde{\beta}}(L) \quad (5.13.152)$$

and

$$L = T_{\tilde{\beta}^{-1}}(V). \quad (5.13.153)$$

A still simpler choice, call it $\tilde{\tilde{\beta}}$, is the Darboux matrix obtained by setting $A = I$ and $C = 0$ in (13.148) to yield the result

$$\tilde{\tilde{\beta}} = (1/\sqrt{2}) \begin{pmatrix} [-JL^{-1} + VL^{-1}] & [J + V] \\ L^{-1} & I \end{pmatrix} = \phi \quad (5.13.154)$$

with the inverse

$$\tilde{\tilde{\beta}}^{-1} = (1/\sqrt{2}) \begin{pmatrix} LJ & -LJV + L \\ -J & JV + I \end{pmatrix}. \quad (5.13.155)$$

[Note that this choice amounts to setting $\epsilon = I$. That $\tilde{\beta}$ is also a Darboux matrix follows from the fact β as given by (13.148) is a Darboux matrix for all choices of A and C .] It is easily verified by direct calculation that

$$V = T_{\tilde{\beta}}(L) \quad (5.13.156)$$

and

$$L = T_{\tilde{\beta}^{-1}}(V). \quad (5.13.157)$$

Exercises

5.13.1. Verify the matrix identity

$$\begin{pmatrix} -J & J \\ I & I \end{pmatrix} \begin{pmatrix} I & I \\ 0 & I \end{pmatrix} = \begin{pmatrix} -J & 0 \\ I & 2I \end{pmatrix}. \quad (5.13.158)$$

Use this identity to prove (13.13). Verify (13.19). Use the representation (13.27) to show that all Darboux matrices, i.e. all matrices that satisfy (13.20) or (13.21), must have determinant +1.

5.13.2. Show, using the representations (13.27) and (13.29), that there are the relations

$$\mu = \sigma^{-1}\gamma\sigma \text{ or } \gamma = \sigma\mu\sigma^{-1} \quad (5.13.159)$$

which demonstrate that $\tilde{Sp}(4n)$ and $Sp(4n)$ are related by a similarity transformation.

5.13.3. Review Exercise 12.7. Verify that (13.54) is an equivalence relation.

5.13.4. Verify by direct calculation that β as given by (13.109) satisfies (13.21). Verify (13.117) both by direct calculation and by use of (13.11), (13.29), and (13.105).

5.13.5. Verify (13.143) by working out the product (13.140). Verify by direct calculation that ϕ satisfies (13.136) and (13.137). Verify that ϕ is a Darboux matrix/transformation.

5.13.6. Verify (13.148) and (13.149) using (13.144) and (13.147).

5.13.7. Verify by direct calculation that β and β^{-1} as given by (13.148) and (13.149) satisfy (13.145) and (13.146).

5.13.8. The Darboux matrix β given by (13.148) is parameterized in terms of a general $2n \times 2n$ symplectic matrix L , two general $2n \times 2n$ symmetric matrices C and V , and a general $GL(2n)$ matrix A . Verify that the dimensionality of the space of all $2n \times 2n$ symmetric matrices is $n(2n+1)$, which is also the dimensionality of $sp(2n)$. Also, the dimensionality of $GL(2n)$ is evidently $(2n)^2$. Verify that the dimension count for the parameterization (13.148) of Darboux matrices in terms of a $2n \times 2n$ symplectic matrix, two $2n \times 2n$ symmetric matrices, and a general $GL(2n)$ matrix is given by the sum

$$n(2n+1) + 2[n(2n+1)] + 4n^2 = 10n^2 + 3n. \quad (5.13.160)$$

By comparison, the dimensionality of the set of $4n \times 4n$ Darboux matrices is the same as the dimensionality of $sp(4n)$, which is given by the relation

$$\dim sp(4n) = 2n(4n + 1) = 8n^2 + 2n. \quad (5.13.161)$$

Thus, the parameterization (13.148) must have some redundancy. Determine what this redundancy is in the simplest case of 4×4 Darboux matrices. Verify that the number of parameters in $\tilde{\beta}$ as given by (13.150) is $6n^2 + 3n$. Verify that the number of parameters in $\tilde{\tilde{\beta}}$ as given by (13.154) is $4n^2 + 2n$.

5.13.9. Suppose N is a J^{2n} symplectic matrix and μ is a \tilde{J}^{4n} symplectic matrix. Show, using (13.159), that M given by

$$M = T_\mu(N) \quad (5.13.162)$$

is also a J^{2n} symplectic matrix.

5.13.10. Verify that the relation (12.80) between the guhp and the gud can be rewritten in the form

$$-W = (iZ + I)(-iZ + I)^{-1} = T_\phi(Z) \quad (5.13.163)$$

where

$$\phi = \begin{pmatrix} iI & I \\ -iI & I \end{pmatrix}. \quad (5.13.164)$$

Verify that the Cayley relation (3.11.5) between symmetric and symplectic matrices can be written in the form

$$M = [(-J)(-W) + I][(J)(-W) + I]^{-1} = T_\psi(-W) \quad (5.13.165)$$

where

$$\psi = \begin{pmatrix} -J & I \\ J & I \end{pmatrix}. \quad (5.13.166)$$

Relate ϕ and ψ . Hint: Review Exercise 3.2.6.

5.13.11. Let ν be the matrix defined by the relation

$$\nu = (1/\sqrt{2}) \begin{pmatrix} -I & I \\ I & I \end{pmatrix}. \quad (5.13.167)$$

Verify that the Cayley relation (3.11.5) between a symplectic matrix M and a Hamiltonian matrix JW , a matrix in the symplectic Lie algebra, can be written in the form

$$M = T_\nu(-JW). \quad (5.13.168)$$

Verify that ν has the property

$$\nu^2 = I, \quad (5.13.169)$$

from which it follows that T_ν is an *involution*.²¹ That is, by the composition law (11.6), there is the relation

$$T_\nu T_\nu = T_{\nu^2} = T_I = I. \quad (5.13.170)$$

²¹In this context, an involution is a map whose square is the identity map.

Consequently, show that the relation (13.168) has the inverse relation

$$-JW = T_\nu(M), \quad (5.13.171)$$

in agreement with (3.11.12). We have learned that T_ν provides a map between the *group* $Sp(2n, \mathbb{R})$ and its *Lie algebra* $sp(2n, \mathbb{R})$. Show that it does the same for $Sp(2n, \mathbb{C})$ and $sp(2n, \mathbb{C})$.

Show that T_ν has analogous properties for the orthogonal and unitary (but not special unitary) groups. For example, if A is antisymmetric, show that $M = T_\nu(-A)$ is orthogonal, etc.

5.14 Uniqueness of Cayley Möbius Transformation

The Cayley Möbius transformation has three properties that make it essentially unique. The first is that

$$T_\sigma(M^{-1}) = -T_\sigma(M). \quad (5.14.1)$$

The second, consistent with the first, is that

$$T_\sigma(I) = 0. \quad (5.14.2)$$

The third is the relation

$$JT_\sigma(N^{-1}MN) = N^{-1}JT_\sigma(M)N, \quad (5.14.3)$$

from which it follows that

$$T_\sigma(N^{-1}MN) = -JN^{-1}JT_\sigma(M)N. \quad (5.14.4)$$

Here $J = J^{2n}$ and N is any invertible matrix. Now suppose that N is symplectic. From the symplectic condition written in the form

$$N J N^T = J \quad (5.14.5)$$

we infer the relation

$$-JN^{-1}J = N^T, \quad (5.14.6)$$

so that we also have for symplectic N the result

$$T_\sigma(N^{-1}MN) = N^T T_\sigma(M)N. \quad (5.14.7)$$

These properties are easily shown to follow from the form of σ as given in (13.11), and lead to the inversion and symplectic similarity invariance properties of Cayley matrix symplectification described by (4.7.14) and (4.7.15). They will also be important for the work of Chapter 34 on Optimal Evaluation of Symplectic Maps.

We now verify that essentially only the Cayley Möbius transformation, among all Darboux Möbius transformations, has these properties. To begin suppose, in analogy with (14.1), we seek Darboux Möbius transformations β with the property

$$T_\beta(M^{-1}) = -T_\beta(M). \quad (5.14.8)$$

Also assume that $T_\beta(I)$ is well defined. Then it follows from (14.8) that there is the condition

$$T_\beta(I) = 0, \quad (5.14.9)$$

and hence $L = I$ and $V = 0$ in (13.145) so that β as given by (13.148) takes the form

$$\beta = (1/\sqrt{2}) \begin{pmatrix} -AJ & AJ \\ (A^T)^{-1}(-CJ + I) & (A^T)^{-1}(CJ + I) \end{pmatrix}. \quad (5.14.10)$$

Correspondingly $\tilde{\beta}$, which is obtained from (14.10) by setting $A = I$, is given by the relation

$$\tilde{\beta} = (1/\sqrt{2}) \begin{pmatrix} -J & J \\ (-CJ + I) & (CJ + I) \end{pmatrix}. \quad (5.14.11)$$

So far, we have actually only used the fact that (14.9) is a consequence of (14.8) to put restrictions on β . Now let us make further direct use of (14.8). To do so it is useful to compute $T_{\tilde{\beta}}(M)$ and $T_\beta(M)$. Begin by computing $T_{\tilde{\beta}}(M)$. From (14.11) we find the result

$$T_{\tilde{\beta}}(M) = [-JM + J][(-CJ + I)M + (CJ + I)]^{-1}. \quad (5.14.12)$$

We will also need the result

$$T_{\tilde{\beta}}(M^{-1}) = [-JM^{-1} + J][(-CJ + I)M^{-1} + (CJ + I)]^{-1} \quad (5.14.13)$$

which follows from (14.12) upon replacing M by M^{-1} . Next compute $T_\beta(M)$ using (14.10) and manipulate the result to find the relation

$$\begin{aligned} T_\beta(M) &= \{-AJM + AJ\}\{(A^T)^{-1}(-CJ + I)M + (A^T)^{-1}(CJ + I)\}^{-1} \\ &= A\{-JM + J\}\{(A^T)^{-1}[(-CJ + I)M + (CJ + I)]\}^{-1} \\ &= A\{-JM + J\}\{(-CJ + I)M + (CJ + I)\}^{-1}A^T \\ &= A[T_{\tilde{\beta}}(M)]A^T. \end{aligned} \quad (5.14.14)$$

It follows from (14.14) that

$$T_{\tilde{\beta}}(M) = A^{-1}[T_\beta(M)](A^T)^{-1}. \quad (5.14.15)$$

Similarly, there is the result

$$T_{\tilde{\beta}}(M^{-1}) = A^{-1}[T_\beta(M^{-1})](A^T)^{-1}. \quad (5.14.16)$$

Now add (14.15) and (14.16) to obtain the relation

$$T_{\tilde{\beta}}(M) + T_{\tilde{\beta}}(M^{-1}) = A^{-1}[T_\beta(M) + T_\beta(M^{-1})](A^T)^{-1}. \quad (5.14.17)$$

Upon making use of (14.8) in (14.17) we find the result

$$T_{\tilde{\beta}}(M) + T_{\tilde{\beta}}(M^{-1}) = 0 \text{ or, equivalently, } T_{\tilde{\beta}}(M^{-1}) = -T_{\tilde{\beta}}(M). \quad (5.14.18)$$

We are almost done with this part of the argument. In (14.13) multiply both the numerator and denominator on the right by M to obtain the result

$$T_{\tilde{\beta}}(M^{-1}) = [JM - J][(CJ + I)M + (-CJ + I)]^{-1}. \quad (5.14.19)$$

Upon employing (14.12) and (14.19) in the second version of (14.18) we now find the result

$$[JM - J][(CJ + I)M + (-CJ + I)]^{-1} = [JM - J][(-CJ + I)M + (CJ + I)]^{-1}, \quad (5.14.20)$$

from which it follows that

$$(CJ + I)M + (-CJ + I) = (-CJ + I)M + (CJ + I). \quad (5.14.21)$$

(Here we have assumed $M \neq I$.) Finally, upon canceling like terms on the left and right sides of (14.21), we find the relation

$$CJM - CJ = -CJM + CJ \text{ or, equivalently } 2CJ(M - I) = 0 \quad (5.14.22)$$

from which it follows that

$$C = 0. \quad (5.14.23)$$

Employing (14.23) in (14.11) gives the result

$$\tilde{\beta} = (1/\sqrt{2}) \begin{pmatrix} -J & J \\ I & I \end{pmatrix} = \sigma. \quad (5.14.24)$$

Correspondingly (14.14) can be rewritten as

$$T_{\beta}(M) = A[T_{\sigma}(M)]A^T. \quad (5.14.25)$$

In analogy to (14.7) let us now invoke the further requirement that

$$T_{\beta}(N^{-1}MN) = N^T T_{\beta}(M)N. \quad (5.14.26)$$

With the aid of (14.25) and (14.7) we find for the left side of (14.26) the result

$$T_{\beta}(N^{-1}MN) = A[T_{\sigma}(N^{-1}MN)]A^T = AN^T[T_{\sigma}(M)]NA^T. \quad (5.14.27)$$

For the right side of (14.26), again using (14.25), we find the result

$$N^T T_{\beta}(M)N = N^T AT_{\sigma}(M)A^T N. \quad (5.14.28)$$

Therefore (14.26) is equivalent to the condition

$$AN^T[T_{\sigma}(M)]NA^T = N^T A[T_{\sigma}(M)]A^T N. \quad (5.14.29)$$

In order for (14.29) to hold for all matrices M , there must be the relation

$$AN^T = N^T A. \quad (5.14.30)$$

That is, N^T and A must commute. See Exercise 15.3.

Moreover, since N is an arbitrary symplectic matrix, N^T is also an arbitrary symplectic matrix. It can be shown that a matrix A that commutes with all symplectic matrices must be a multiple of the identity. See Exercise 21.14.1. Therefore, the requirement (14.26) yields the conclusion that A must be of the form

$$A = \lambda I. \quad (5.14.31)$$

Correspondingly, β takes the form

$$\beta = (1/\sqrt{2}) \begin{pmatrix} -\lambda J & \lambda J \\ (1/\lambda)I & (1/\lambda)I \end{pmatrix}. \quad (5.14.32)$$

We see that, apart from a scaling factor λ , the matrix β is essentially σ . If the scaling factor is set to one, β becomes σ . Another possible choice is to set $\lambda = \sqrt{2}$. When this is done, β takes the rational form

$$\beta = \begin{pmatrix} -J & J \\ I/2 & I/2 \end{pmatrix}. \quad (5.14.33)$$

However, for this λ choice β is not orthogonal, and the simplicity of the orthogonality feature possessed by σ is lost.

Exercises

5.14.1. Verify the relations (14.1) through (14.7) for the Cayley Möbius transformation.

5.14.2. Verify that (14.21) follows from (14.20).

5.14.3. The purpose of this exercise is to verify that (14.29) implies (14.30). Define matrices X and Y by the relations

$$X = AN^T, \quad (5.14.34)$$

$$Y = N^TA. \quad (5.14.35)$$

Verify that with these definitions (14.29) can be rewritten in the form

$$X[T_\sigma(M)]X^T = Y[T_\sigma(M)]Y^T. \quad (5.14.36)$$

Next show that, for some finite (but perhaps small) number δ and *any* symmetric matrix S , there is a symplectic M such that

$$T_\sigma(M) = \delta S. \quad (5.14.37)$$

Thus, show that (14.36) is equivalent to the relation

$$XSX^T = YSY^T. \quad (5.14.38)$$

Define a matrix Λ by the equation

$$Y = X\Lambda. \quad (5.14.39)$$

Verify, because A and N are invertible, that this equation actually defines Λ . Using this definition, show that (14.38) is equivalent to the relation

$$\Lambda S \Lambda^T = S. \quad (5.14.40)$$

If we put $S = I$, which is a possibility, we conclude that Λ must satisfy the relation

$$\Lambda \Lambda^T = I, \quad (5.14.41)$$

and therefore Λ is orthogonal. Verify that (14.40) and (14.41) entail the relation

$$\Lambda S = S \Lambda. \quad (5.14.42)$$

Show that (14.41), together with (14.42) holding for all symmetric matrices S , requires that

$$\Lambda = \pm I. \quad (5.14.43)$$

Hint: First show that Λ must be diagonal by considering the cases for which S is diagonal and has only one nonzero entry. Next show that all diagonal entries in Λ must be equal by considering the cases for which S has only two nonzero entries located at symmetric places above and below the diagonal. Finally, use (14.41).

Show that taking the minus sign in (14.43) leads to the condition $AN^T = -N^T A$, and that if this condition holds for all N as it must, it also holds for $N = I$ leading to the conclusion $A = 0$, which is not possible since A is assumed to be nonsingular. Show that taking the plus sign in (14.43) yields the advertised relation (14.30).

5.14.4. Review Exercise 3.11.4. Suppose (13.86) is rewritten in the form

$$W = J(-J)T_\alpha(M) \quad (5.14.44)$$

and we define $g(M)$ by the rule

$$g(M) = -JT_\alpha(M). \quad (5.14.45)$$

Then we have the relation

$$W = Jg(M). \quad (5.14.46)$$

Define a quadratic form $Q_\alpha(z)$ by the rule

$$Q_\alpha(z) = (z, Wz). \quad (5.14.47)$$

Exercise 3.11.4 showed that $Q_\alpha(z)$ is invariant when $\alpha = \sigma$. Show that many other choices of α do not yield a Q_α that has this property. Are there any other choices that do?

5.15 Matrix Symplectification Revisited

Section 4.7 described the use of the Cayley representation to carry out matrix symplectification, and Section 4.8 described the use of generating functions for the same purpose. As pointed out earlier, both procedures are examples of the use of Möbius transformations. Moreover, it was also remarked that there were cases for which the Cayley representation could not be used, and cases for which none of the generating functions F_1 through F_4 could be used. The purpose of this section is to show that, given any nearly symplectic matrix M , there is a symplectification procedure employing Möbius transformations that will succeed. (Subsequently, Exercise 6.7.1 shows that there is an associated quadratic generating function that produces any such Möbius transformation.)

Suppose M is a matrix that is nearly symplectic. Let β be some appropriate Darboux matrix. Use it to define a matrix U in terms of M by the rule

$$U = T_\beta(M). \quad (5.15.1)$$

Since M is nearly symplectic, and by the properties of Darboux Möbius transformations, U will be nearly symmetric. Define a matrix W in terms of U by the rule

$$W = (U + U^T)/2. \quad (5.15.2)$$

Since U is nearly symmetric, W will be near U . Finally, define a matrix R in terms of W by the rule

$$R = T_{\beta^{-1}}(W). \quad (5.15.3)$$

Since W is symmetric by construction, and by the properties of Darboux Möbius transformations, R will be symplectic. Moreover, because W is near U , R will be near M . Note also that $R = M$ if M is symplectic. Therefore R may be viewed as a symplectification of M .

We still have to demonstrate that there is a choice of β such that (15.1) and (15.3) are well defined. Suppose we write M in the form

$$M = LN \quad (5.15.4)$$

where L is symplectic and N is a matrix in the vicinity (in a sense to be made more precise shortly) of the identity. Use for β the Darboux matrix given by (13.148) with the same L that appears in (15.4). From (15.1) we then find for U a result of the form

$$U = T_\beta(M) = \{\text{Numerator}\} \times \{\text{Denominator}\}^{-1} \quad (5.15.5)$$

where the denominator is given by the equation

$$\text{Denominator} = (A^T)^{-1} \{ [(-CJ + I)L^{-1}]M + (CJ + I) \}. \quad (5.15.6)$$

Inserting the representation (15.4) for M into (15.6) gives the result

$$\text{Denominator} = (A^T)^{-1} \{ [(-CJ + I)N + (CJ + I)] \}, \quad (5.15.7)$$

which can be rewritten in the form

$$\begin{aligned}\text{Denominator} &= (A^T)^{-1}\{[(-CJ + I)(N - I) + 2I]\} \\ &= 2(A^T)^{-1}\{[I + (1/2)(-CJ + I)(N - I)]\}.\end{aligned}\quad (5.15.8)$$

Compute the determinant of the denominator to find the result

$$\det(\text{Denominator}) = \det[2(A^T)^{-1}] \times \det\{[I + (1/2)(-CJ + I)(N - I)]\}. \quad (5.15.9)$$

We see from the second factor in (15.9) that the determinant of the denominator cannot vanish as long as N is reasonably near I . Correspondingly we conclude that there is a choice of β such that, for any given M that is nearly symplectic, (15.1) is well defined for this M and for all matrices near this M . Conversely, from the work of Section 11, we know that $T_{\beta^{-1}}(U)$ will then be well defined for the U given by (15.1) and for all matrices near this U . We have already seen that W is near U . Therefore R as given by (15.3) is well defined. The symplectification procedure has succeeded.

The last item to be considered in this section is the extent to which the symplectification procedure given by (15.1) through (15.3) depends on the choice of the Darboux transformation β . We might suspect some redundancy because the procedure (15.1) through (15.3) involves the use of both β and β^{-1} , and therefore there is some possibility for compensation or cancellation.

To explore this question, we will need to study the properties of β as given by (13.148) in some more detail. As the result of some preliminary monkeying around, we observe that the terms in the two upper blocks of β can be rewritten in the form

$$-AJL^{-1} + V(A^T)^{-1}(-CJ + I)L^{-1} = A[-JL^{-1} + A^{-1}V(A^T)^{-1}(-CJ + I)L^{-1}], \quad (5.15.10)$$

$$AJ + V(A^T)^{-1}(CJ + I) = A[J + A^{-1}V(A^T)^{-1}(CJ + I)]. \quad (5.15.11)$$

Define a new matrix \acute{V} by the rule

$$\acute{V} = A^{-1}V(A^{-1})^T. \quad (5.15.12)$$

The matrix \acute{V} will also be symmetric because V is symmetric. Moreover, since A is assumed invertible and V is an arbitrary symmetric matrix, the matrix \acute{V} may be taken to be an arbitrary symmetric matrix. With this definition, (15.10) and (15.11) can be written in the more compact forms

$$-AJL^{-1} + V(A^T)^{-1}(-CJ + I)L^{-1} = A[-JL^{-1} + \acute{V}(-CJ + I)L^{-1}], \quad (5.15.13)$$

$$AJ + V(A^T)^{-1}(CJ + I) = A[J + \acute{V}(CJ + I)]. \quad (5.15.14)$$

Now β can be expressed in the form

$$\beta = (1/\sqrt{2}) \begin{pmatrix} A[-JL^{-1} + \acute{V}(-CJ + I)L^{-1}] & A[J + \acute{V}(CJ + I)] \\ (A^T)^{-1}(-CJ + I)L^{-1} & (A^T)^{-1}(CJ + I) \end{pmatrix}. \quad (5.15.15)$$

When the form for β given by (15.15) is used to compute $T_\beta(M)$, we find the result (15.5) with the denominator given by (15.6) and the numerator given by

$$\text{Numerator} = A\{[-JL^{-1} + \acute{V}(-CJ + I)L^{-1}]M + [J + \acute{V}(CJ + I)]\}. \quad (5.15.16)$$

Note that the numerator has a common factor of A and, as (15.6) shows, the denominator has a common factor of $(A^T)^{-1}$. Therefore we have the *identity* that $T_\beta(M)$ for any M can be written in the factored form

$$T_\beta(M) = A[T_{\tilde{\beta}}(M)]A^T. \quad (5.15.17)$$

Here $\tilde{\beta}$ is a Darboux matrix defined in terms of β by writing

$$\tilde{\beta} = (1/\sqrt{2}) \begin{pmatrix} [-JL^{-1} + \dot{V}(-CJ + I)L^{-1}] & [J + \dot{V}(CJ + I)] \\ (-CJ + I)L^{-1} & (CJ + I) \end{pmatrix}. \quad (5.15.18)$$

[That $\tilde{\beta}$ is a Darboux matrix follows from the fact β as given by (15.15) is a Darboux matrix for all choices of A , and $\tilde{\beta}$ is simply the result of putting $A = I$ in (15.15).]

We note in passing that $\tilde{\beta}$ has the property

$$T_{\tilde{\beta}}(L) = \dot{V}. \quad (5.15.19)$$

That is, $T_{\tilde{\beta}}$ sends the arbitrary symplectic matrix L to the arbitrary symmetric matrix \dot{V} .

We will now learn that for fixed L , \dot{V} , and C in (15.15), the symplectification procedure given by (15.1) through (15.3) employing the β of (15.15) yields a result that is *independent* of the choice of the matrix A . To see this, suppose that the matrix $\tilde{\beta}$ is used to symplectify M using a procedure analogous to (15.1) through (15.3). As just pointed out, this amounts to setting $A = I$ in (15.15). Then we find the results

$$\tilde{U} = T_{\tilde{\beta}}(M), \quad (5.15.20)$$

$$\tilde{W} = (\tilde{U} + \tilde{U}^T)/2, \quad (5.15.21)$$

$$\tilde{R} = T_{\tilde{\beta}^{-1}}(\tilde{W}). \quad (5.15.22)$$

Now carry out some manipulations using previous results. From (15.1), (15.17), and (15.20) it follows that

$$U = T_\beta(M) = A[T_{\tilde{\beta}}(M)]A^T = A\tilde{U}A^T. \quad (5.15.23)$$

Consequently we have the relations

$$U^T = A\tilde{U}^T A^T, \quad (5.15.24)$$

$$W = (U + U^T)/2 = A[(\tilde{U} + \tilde{U}^T)/2]A^T = A\tilde{W}A^T. \quad (5.15.25)$$

But we also have, from (15.3) and application of the identity (15.17), the result

$$W = T_\beta(R) = A[T_{\tilde{\beta}}(R)]A^T. \quad (5.15.26)$$

Upon comparing (15.25) and (15.26) we conclude there is the relation

$$T_{\tilde{\beta}}(R) = \tilde{W}, \quad (5.15.27)$$

and therefore there is also the inverse relation

$$R = T_{\tilde{\beta}^{-1}}(\tilde{W}). \quad (5.15.28)$$

Finally (15.22) and (15.28) taken together show that

$$R = \tilde{R}. \quad (5.15.29)$$

Thus R is indeed independent of the choice of A . Since the first factor in (13.147), the factor that involves A , produces an *arbitrary* linear transformation on the coordinate-space variables, we may say that the Darboux Möbius symplectification procedure is *invariant* under linear transformations of the coordinate-space variables.

Exercises

5.15.1. Verify (15.1) through (15.9).

5.15.2. Verify (15.10) through (15.18).

5.15.3. Verify (15.20) through (15.29).

5.15.4. By studying various examples, explore how the choice of L , \acute{V} , and C in (15.15) affects the outcome of the symplectification procedure. Study, for example, the use of β to symplectify matrices of the form λI where λ is a parameter near 1.

Bibliography

General Non-Integrability of Hamiltonian Systems

- [1] A. Dragt and J. Finn, “Insolubility of Trapped Particle Motion in a Magnetic Dipole Field”, *J. Geophys. Res.* **81** pp. 2327-2340 (1976).
- [2] Maxine Ennata Alves de Almeida, Lideu Moreira, and Haruo Yoshida, “On the Non-Integrability of the Størmer Problem”, *Journal of Physics A: Mathematical and General* **25** (March 1992).
- [3] A. Sáenz and M. Kummer, “Non-Integrability of the Størmer Problem”, *Physica D: Nonlinear Phenomena* (September 1995).
- [4] H. Yoshida, “A Criterion for the Non-Existence of an Additional Integral in Hamiltonian Systems with a Homogeneous Potential”, *Physica D* **29** (1987).
- [5] J. Morales Ruiz, *Differential Galois Theory and Non-Integrability of Hamiltonian Systems*, Springer Basel (1999).
- [6] M. Audin, *Hamiltonian Systems and Their Integrability*, American Mathematical Society and Société Mathématique de France (2008).

Group Theory, $U(3)$, and $SU(3)$

- [7] H. Weyl, *The Classical Groups: Their Invariants and Representations*, Princeton University Press (1946).
- [8] H. Georgi, *Lie Algebras in Particle Physics*, Perseus Books (1999).
- [9] A. Zee, *Group Theory in a Nutshell for Physicists*, Princeton University Press (2016).
- [10] M. Gell-Mann and Y. Ne’eman, *The Eightfold Way*, pp. 49-50, Benjamin (1964).
- [11] R.E. Behrends et al., *Rev. of Mod. Phys.* **34**, p. 1 (1962).
- [12] S. Gasiorowicz, *Elementary Particle Physics*, p. 257, John Wiley (1966).
- [13] S. Gasiorowicz, “A Simple Graphical Method in the Analysis of $SU(3)$ ”, Argonne National Laboratory Report, ANL-6729 (1963).
- [14] S. Coleman, *Aspects of Symmetry*, Cambridge University Press (1985).

- [15] W. Greiner and B. Muller, *Quantum Mechanics—Symmetries*, Springer-Verlag (1994).
- [16] A. Dragt, “Classification of Three-Particle States According to $SU(3)$ ”, *Journal of Mathematical Physics* **6**, 533 (1965).
- [17] P. J. Olver, *Equivalence, Invariants, and Symmetry*, Cambridge University Press (1995).
- Topology of $Sp(2, \mathbb{R})$ and $Sp(2n, \mathbb{R})$
- [18] M. Levi, “Stability of the Inverted Pendulum - a Topological Explanation”, *SIAM Review* **30** 639 (1988).
- [19] A. Abbondandolo, *Morse theory for Hamiltonian systems*, Chapman & Hall/CRC (2001).
- Metaplectic Group and Fourier Optics
- [20] G.B. Folland, *Harmonic Analysis in Phase space*, Annals of Mathematics Studies Number 122, Princeton University Press (1989).
- [21] R.G. Littlejohn, “The Semiclassical Evolution of Wave Packets”, *Physics Reports* (1985). See also the Web link <https://escholarship.org/content/qt8p6601jj/qt8p6601jj.pdf>
- [22] J.W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill (1996). See also the Web link https://docs.google.com/file/d/0B78A_rsP6RDSS3VRWk12Y2FUcVk/edit?resourcekey=0-EdJQY3UFbqEiJnqV8YDPNA
- Quaternions, Octonions, and Lie Groups
- [23] J.B. Kuipers, *Quaternions and Rotation Sequences: A Primer with Application to Orbits, Aerospace and Virtual Reality*, Princeton University Press (2002).
- [24] S. Altmann, *Rotations, Quaternions, and Double Groups*, Dover (2005).
- [25] B.L. van der Waerden, *Modern Algebra, Volume I*, Frederick Ungar (1953).
- [26] I.L. Kantor and A.S. Solodovnik, *Hypercomplex Numbers*, Springer-Verlag (1989).
- [27] C. Chevalley, *Theory of Lie Groups*, Princeton University Press (1946). There is also a Dover reprint/edition (2018).
- [28] C. Chevalley, *The Algebraic Theory of Spinors and Clifford Algebras: Collected Works of Claude Chevalley (v. 2)*, Springer (1996).
- [29] H.-D. Ebbinghaus et al., *Numbers*, Springer Verlag (1991).
- [30] J.C. Baez, “The octonions”, *Bull. American Mathematical Society* **39**, p. 145 (2002) and **42**, p. 213 (2005).

- [31] J.H. Conway and D.A. Smith, *On Quaternions and Octonions: Their Geometry, Arithmetic, and Symmetry*, A.K. Peters (2003).
- [32] J.C. Baez, *My favorite Numbers: 8*, The Rankin Lectures (2008), <http://theoryoforder.com/img/8.pdf>.
- [33] J.C. Baez, *Bull. Amer. Math. Soc.* **42**, p. 229, (2005). Also available at http://math.ucr.edu/home/baez/octonions/conway_smith/.
- [34] I. Porteous, *Clifford Algebras and the Classical Groups*, Cambridge (1995).
- [35] P. Lounesto, *Clifford Algebras and Spinors*, 2nd ed., Cambridge (2001).
- [36] F. Reese Harvey, *Spinors and Calibrations*, Academic Press (1990).
- [37] P. Cvitanović, *Group Theory: Birdtracks, Lie's, and Exceptional Groups*, Princeton University Press (2008).
- [38] D. Hestenes, *Space-Time Algebra*, second edition, Birkhäuser (2015).

Group Theory, Möbius Transformations, Theta Functions, Etc.

(See also the Lie Group Theory sections of the Bibliographies for Chapters 3 and 27.)

- [39] H. Bateman, A. Erdelyi, W. Magnus, F. Oberhettinger, F.G. Tricomi, *Higher Transcendental Functions*, Vol. III, Chapter 14: Automorphic Functions, McGraw-Hill (1955).
- [40] C.L. Siegel, *Symplectic Geometry*, Academic Press (New York, 1964).
- [41] C.L. Siegel, *Topics in Complex Function Theory, Vols. I-III*, Wiley-Interscience (New York, 1971).
- [42] L.K. Hua, “On the theory of automorphic functions of a matrix variable I,II”, *Amer. J. Math.* **66**, 470-488, 531-563 (1944).
- [43] Feng Kang, Wu Hua-mo, Qin Meng-shao, and Wang Dao-liu, “Construction of Canonical Difference Schemes for Hamiltonian Formalism via Generating Functions”, *Journal of Computational Mathematics* **11**, p. 71 (1989).
- [44] Feng Kang, “The Calculus of Generating Functions and the Formal Energy for Hamiltonian Algorithms”, *Journal of Computational Mathematics* **16**, p. 481 (1998).
- [45] M. Eichler, *Introduction to the Theory of Algebraic Numbers and Functions*, Academic Press (1966).
- [46] J. Lehner, *Discontinuous Groups and Automorphic Functions*, *Mathematical Surveys and Monographs* # 8, American Mathematical Society (1964).
- [47] D. Mumford, *Tata Lectures on Theta I-III*, Birkhäuser (1984).
- [48] D. Mumford, C. Series, and D. Wright, *Indra's Pearls: The Vision of Felix Klein*, Cambridge University Press (2002).

- [49] R.E. Bellman, *A Brief Introduction to Theta Functions*, Holt, Rinehart, and Winston (1961).
- [50] I.M. Gel'fand, M.I. Graev, and I.I. Pyatetskii-Shapiro, *Representation Theory and Automorphic Functions*, W.B. Saunders Co. (1969).
- [51] N. Koblitz, *Introduction to Elliptic Curves and Modular Forms*, (Springer-Verlag 1984).
- [52] S. Lang, *SL(2,R)*, Springer-Verlag (1985).
- [53] R. Howe and E.C. Tan, *Non-Abelian Harmonic Analysis*, Springer-Verlag (1992).
- [54] A. Terras, *Harmonic Analysis on Symmetric Spaces and Applications I and II*, Springer-Verlag (1985 and 1988).
- [55] A. Perelomov, *Generalized Coherent States and Their Applications*, Springer-Verlag (1986).
- [56] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*, Academic Press (1978).
- [57] S. Helgason, *Groups and Geometric Analysis: Integral Geometry, Invariant Differential Operators, and Spherical Functions*, Second Edition, American Mathematical Society (2002).
- [58] S. Helgason, *Geometric Analysis on Symmetric Spaces*, American Mathematical Society (2008).
- [59] L. Ahlfors, “Clifford Numbers and Möbius Transformations in R^n ”, published in *Clifford Algebras and their Applications in Mathematical Physics*, J. Chisholm and A. Common, Edit., Proceedings of NATO and SERC Workshop, Canterbury, Kent, 1985, NATO ASI Series (Reidel 1986).
- [60] J. Gray, *Linear Differential Equations and Group Theory from Riemann to Poincaré*, Birkhäuser (1986).

Lorentz Group and Laser Optics

- [61] I.M. Gel'fand, R.A. Minlos, and Z.Y. Shapiro, *Representations of the rotation and Lorentz groups and their applications*, Pergamon Press and Macmillan Co. (New York, 1963).
- [62] V. Bargmann, “Irreducible Unitary Representations of the Lorentz Group”, *Annals of Math.* **48**, no. 3, p. 568 (1947).
- [63] A. Yariv, *Quantum Electronics*, John Wiley (New York, 1989); *Optical Electronics*, Holt, Rinehart, and Winston (1985).
- [64] A.J. Dragt, “Lie Algebraic Methods for Ray and Wave Optics” (University of Maryland, 1995).

Lie Series and Lie Transformations

- [65] W. Gröbner, *Die Lie-Reihen und Ihre Anwendungen*, Deutscher Verlag der Wissenschaften (Berlin 1960).
- [66] W. Gröbner and H. Knapp, *Contributions to the Methods of Lie Series*, Bibliographisches Institut, (Manheim 1967).
- [67] G. Hori, “Theory of general perturbations with unspecified canonical variables”, *Publications of the Astronomical Society of Japan* **18**, p. 287 (1966).
- [68] A. Deprit, *Celest. Mech.* **1**, 12 (1969).
- [69] A. A. Kamel, “Perturbation Method in the Theory of Nonlinear Oscillations”, *Celest. Mech.* **3**, 90 (1970).
- [70] A. A. Kamel, “Lie Transforms and the Hamiltonization of Non-Hamiltonian Systems”, *Celest. Mech.* **4**, 397 (1971).
- [71] J. Henrard, “On Perturbation Theory Using Lie Transforms”, *Celest. Mech.* **3**, 107 (1970).
- [72] A.H. Nayfeh, *Perturbation Methods*, Wiley (New York, 2000).
- [73] E. Leimanis, *The General Problem of the Motion of Coupled Rigid Bodies About a Fixed Point*, p. 121, Springer (1965).
- [74] G.E.O. Giacaglia, *Perturbation Methods in Non-Linear Systems*, Springer-Verlag (1972).
- [75] J.R. Cary, “Lie Transform Perturbation Theory for Hamiltonian Systems”, *Physics Reports* **79**, p. 129 (North Holland 1981).
- [76] K. Kowalski and W. Steeb, *Nonlinear Dynamical Systems and Carleman Linearization*, (World Scientific 1991).
- [77] K. Meyer, G. Hall, and D. Offin, *Introduction to Hamiltonian Dynamical Systems and the N-Body Problem*, Second Edition, Springer (2009).
- [78] D. Boccaletti and G. Pucacco, *Theory of Orbits*, 2 vols., Springer-Verlag (1996).
- [79] G.J. Sussman, J. Wisdom, and M.E. Meyer, *Structure and Interpretation of Classical Mechanics*, MIT Press (2001).
- [80] A.J. Lichtenberg and M.A. Lieberman, *Regular and Stochastic Motion*, Springer-Verlag (1983).
- [81] K. Meyer, “Lie transform tutorial — II”, *Computer Aided Proofs in Analysis*, K. Meyer and D. Schmidt, Eds., Springer-Verlag (1991). Or see the Web site <http://math.uc.edu/~meyer/capa91.pdf>.

- [82] S. Coffey, A. Deprit, E. Deprit, L. Healy, and B. Miller, “A toolbox for nonlinear dynamics”, *Computer Aided Proofs in Analysis*, K. Meyer and D. Schmidt, Eds., Springer-Verlag (1991).
- [83] A. Dragt, “A Lie Algebraic Theory of Geometrical Optics and Optical Aberrations”, *J. Opt. Sci. Am.* **72**, p. 372 (1982).
- [84] A. Dragt, E. Forest, and K. Wolf, “Foundations of a Lie Algebraic Theory of Geometrical Optics”, *Lie Methods in Optics*, J.S. Mondragon and K.B. Wolf, Edit., Springer-Verlag (1986).
- [85] A. Dragt and E. Forest, “Lie Algebraic Theory of Charged Particle Optics and Electron Microscopes”, *Advances in Electronics and Electron Physics* **67**, P. Hawkes, edit., Academic Press (1986). NB: The journals *Advances in Electronics and Electron Physics* and *Advances in Optical and Electron Microscopy* have been merged to form the journal *Advances in Imaging and Electron Physics*.
- [86] A. Dragt and J. Finn, “Normal Form for Mirror Machine Hamiltonians”, *J. Math. Physics* **20**, 2649-2660 (1979).
- [87] L. E. Fried and G. S. Ezra, “PERTURB: A Special-Purpose Algebraic Manipulation Program for Classical Perturbation Theory”, *Journal of Computational Chemistry* **8**, 397-411 (1987).
- [88] L. E. Fried and G. S. Ezra, “PERTURB: A program for calculation of vibrational energies using generalized algebraic quantization”, *Comp. Phys. Comm.* **51**, 103-114 (1988).
- [89] L. E. Fried and G. S. Ezra, “Generalized Algebraic Quantization: Corrections to Arbitrary Order in Planck’s Constant”, *J. Phys. Chem.* **92**, 3144-3154 (1988).

Chapter 6

Symplectic Maps

This chapter defines symplectic maps and explores some of their properties. They form an infinite dimensional Lie group whose Lie algebra (as will become clear in Chapter 7) is the Poisson bracket Lie algebra of all phase-space functions. It is shown that Hamiltonian flows produce symplectic maps, and essentially any family of symplectic maps arises from an associated Hamiltonian. Thus, Hamiltonian Dynamics *is* the study of symplectic maps, and vice versa. It is also shown that, just as symplectic and symmetric matrices are closely related, symplectic and gradient maps are closely related, and this relation provides a general theory of generating functions. Finally, an introductory discussion is given of symplectic invariants.

6.1 Preliminaries and Definitions

Let $z_1 \cdots z_{2n}$ be a set of canonical coordinates for a $2n$ -dimensional space. By *canonical* we mean that we wish to view the $2n$ -dimensional space as a *phase* space and, as in (1.7.9), have identified the first n of the z 's as being q 's and the remaining n as being p 's. Suppose a transformation is made that sends the point z with coordinates $z_1 \cdots z_{2n}$ to some other point \bar{z} with coordinates $\bar{z}_1(z, t) \cdots \bar{z}_{2n}(z, t)$. Such a transformation will be called a mapping, and will be denoted by the symbol \mathcal{M} ,

$$\mathcal{M} : z \rightarrow \bar{z}(z, t). \quad (6.1.1)$$

See Figure 1.1. In this discussion, the time t simply plays the role of a parameter. It is included in the notation to indicate that the transformation may depend on the time. That is, the map \mathcal{M} may be different at different times.

Let $M(z, t)$ be the *Jacobian matrix* of the map \mathcal{M} . It is defined by the equation

$$M_{ab}(z, t) = \partial \bar{z}_a / \partial z_b. \quad (6.1.2)$$

The Jacobian matrix describes the small changes produced in the *final* quantities \bar{z}_a when small changes are made in the *initial* quantities z_b . See Exercise 1.4.6.

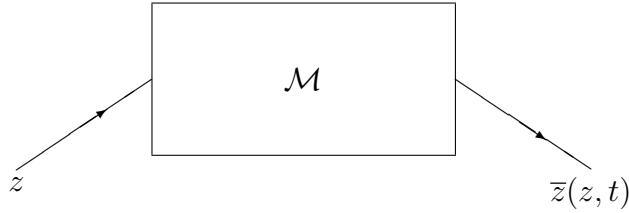


Figure 6.1.1: The map \mathcal{M} sends z to $\bar{z}(z, t)$.

6.1.1 Gradient Maps

As the title to this chapter indicates, it is mostly about symplectic maps. However, we shall subsequently need *gradient* maps as well. Indeed, in Section 6.7 we will learn that there is an intimate connection between gradient and symplectic maps. Therefore, this is a convenient place to make a detour to define gradient maps.

Suppose $g(u, t)$ is some function of the $2n$ variables $u_1 \cdots u_{2n}$ and possibly some parameter t . Use g to define a map \mathcal{G} by the rule

$$\mathcal{G} : u \rightarrow \bar{u}(u, t), \quad (6.1.3)$$

with

$$\bar{u}(u, t)_a = \partial g / \partial u_a. \quad (6.1.4)$$

Note that a gradient is involved in the definition of \mathcal{G} , hence the name *gradient* map. We also note that a *single* function, namely $g(u, t)$, has been used to produce the $2n$ functions $\bar{u}(u, t)$. We will refer to g as a *source* function.¹

Let $G(u, t)$ be the Jacobian matrix of the map \mathcal{G} . In accord with the spirit of (1.2), it is given by the equation

$$G_{ab}(u, t) = \partial \bar{u}_a / \partial u_b. \quad (6.1.5)$$

If we now make use of (1.4), we find the relation

$$G_{ab}(u, t) = \partial^2 g / \partial u_b \partial u_a = \partial^2 g / \partial u_a \partial u_b. \quad (6.1.6)$$

That is, G is the Hessian of g . We observe that G is *symmetric* because the order of partial differentiation is immaterial for functions with continuous derivatives,

$$[G(z, t)]^T = G(z, t). \quad (6.1.7)$$

Conversely, for any map sending u to \bar{u} , consider the differential form

$$\sum_a \bar{u}(u, t)_a du_a. \quad (6.1.8)$$

It will be *closed* if

$$\partial \bar{u}_a / \partial u_b = \partial \bar{u}_b / \partial u_a. \quad (6.1.9)$$

¹Since (1.4) involves a gradient, some authors refer to g as a *potential* function.

See Exercise 1.1. But (1.9) is simply the condition that G be symmetric. Thus if the Jacobian matrix of a map is symmetric, there is a function g such that

$$dg = \sum_a \bar{u}(u, t)_a du_a. \quad (6.1.10)$$

Indeed, g is given by the path integral

$$g(u, t) = \int^u \sum_a \bar{u}(u', t)_a du'_a, \quad (6.1.11)$$

where the integral is to be taken over any path with some fixed initial point and variable end point u . Moreover, it is evident from (1.10) that (1.4) holds. We conclude that a necessary and sufficient condition for a map \mathcal{G} to be a gradient map is that its Jacobian matrix G be symmetric.

We also note that, although we have been working with an even number of variables, namely $2n$, gradient maps are also defined for an odd number of variables. Finally we note that a necessary and sufficient condition for a gradient map to be (locally) invertible is that $\det G \neq 0$, in which case it can be shown that the inverse map is also a gradient map. See Exercise 2.9.²

6.1.2 Symplectic Maps

With our detour complete, let us return to the main subject of symplectic maps. The map $\mathcal{M}(t)$ is said to be *symplectic* if its Jacobian matrix M is a symplectic matrix for all values of z (and all values of t if \mathcal{M} does indeed depend on t),

$$M^T JM = J \quad \text{or} \quad M J M^T = J, \quad \forall z, t. \quad (6.1.12)$$

Note that in general M depends on z and t . However, the particular combinations $M^T JM$ or $M J M^T$ must be z and t *independent*. Therefore, a symplectic map must have very special properties.

To appreciate the significance of a symplectic mapping, consider the Poisson brackets of the various \bar{z} 's with each other. Using (5.1.3), we find the result

$$[\bar{z}_a, \bar{z}_b] = \sum_{c,d} (\partial \bar{z}_a / \partial z_c) J_{cd} (\partial \bar{z}_b / \partial z_d). \quad (6.1.13)$$

By using the definition (1.2) of the Jacobian matrix M , (1.13) can also be written in the form

$$\begin{aligned} [\bar{z}_a, \bar{z}_b] &= \sum_{c,d} M_{ac} J_{cd} M_{bd} \\ &= \sum_{c,d} M_{ac} J_{cd} (M^T)_{db} = (M J M^T)_{ab}. \end{aligned} \quad (6.1.14)$$

²For example in the context of Lagrangian/Hamiltonian dynamics, (1.5.7) is a gradient map from velocity space to momentum space with the Lagrangian L serving as source function, and the first relation in (1.5.11) is the inverse gradient map from momentum space to velocity space with the Hamiltonian H serving as source function. Finally, H and L are Legendre transforms of each other.

Finally, upon using the symplectic condition (1.12), we find the result

$$[\bar{z}_a, \bar{z}_b] = (MJM^T)_{ab} = J_{ab} = [z_a, z_b]. \quad (6.1.15)$$

Consequently, a necessary and sufficient condition for a map \mathcal{M} to be symplectic is that it preserve the fundamental Poisson brackets (1.7.10). As will be shown in Subsection 3, this statement is equivalent, in turn, to the condition that the map \mathcal{M} must preserve the Poisson bracket Lie algebra of all dynamical variables.

Symplectic mappings also have a geometrical aspect. Let z^0 be some point in phase space, and suppose it is sent to the point \bar{z}^0 under the action of a symplectic map \mathcal{M} . Also, let dz and δz be two small vectors originating at the point z^0 . Under the action of \mathcal{M} , they are sent to two vectors $d\bar{z}$ and $\delta\bar{z}$. See Figure 1.2.

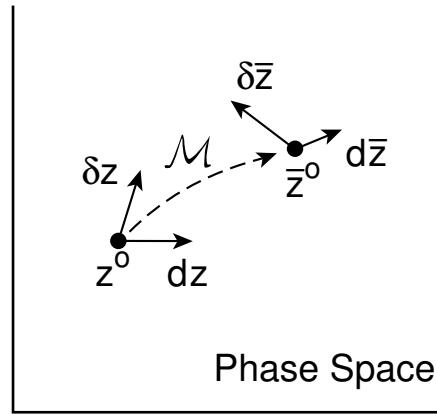


Figure 6.1.2: The action of a symplectic map \mathcal{M} on phase space. The general point z^0 is mapped to the point \bar{z}^0 , and the small vectors dz and δz are mapped to the small vectors $d\bar{z}$ and $\delta\bar{z}$. The figure is only schematic since in general phase space has a large number of dimensions.

From calculus, we have the relation

$$d\bar{z}_a = \sum_b (\partial\bar{z}_a / \partial z_b) dz_b, \quad (6.1.16)$$

or more compactly, using (1.2),

$$d\bar{z} = M dz. \quad (6.1.17)$$

Similarly, the vectors δz and $\delta\bar{z}$ are related by the equation

$$\delta\bar{z} = M \delta z. \quad (6.1.18)$$

Now use the two vectors $\delta\bar{z}, d\bar{z}$ and the matrix J to form the quantity $(\delta\bar{z}, J d\bar{z})$. As described in Section 3.2, this quantity is called the *fundamental symplectic 2-form*. Suppose the relations (1.17) and (1.18) are inserted into the 2-form $(\delta\bar{z}, J d\bar{z})$. Then, using matrix manipulation and the symplectic condition (1.12), we find the relation

$$(\delta\bar{z}, J d\bar{z}) = (M \delta z, J M dz) = (\delta z, M^T J M dz) = (\delta z, J dz). \quad (6.1.19)$$

That is, the value of the fundamental symplectic 2-form is *unchanged* by a symplectic map. Evidently, a necessary and sufficient condition for a map to be symplectic is that it preserve the fundamental symplectic 2-form at all points of phase space and, if time is involved, for all time.³ There is a third aspect of symplectic mappings that should already be familiar. In the usual treatments of Classical Mechanics, an important topic is that of canonical transformations. Canonical transformations are usually defined as those transformations that either

- a. preserve the Hamiltonian form of the equations of motion for all Hamiltonian dynamical systems, or
- b. preserve the fundamental Poisson brackets.

In case *b*, according to the previous discussion, canonical transformations and symplectic maps are the same thing. In case *a*, it can be shown that the most general canonical transformation is a map \mathcal{M} whose Jacobian matrix satisfies the condition

$$M^T JM = \lambda J, \quad (6.1.20)$$

where λ is some real nonzero constant *independent* of z and t . Furthermore, it can be shown that \mathcal{M} in this case consists of a symplectic map followed or preceded by a simple scaling of phase-space variables. See Appendix D. Therefore, in either case, the central object of interest is a symplectic map.

From our perspective, and as will be shown in Subsections 4.1 through 4.3, the most important property of symplectic maps is that Hamiltonian flows produce symplectic maps, and vice versa. Thus, the study of Hamiltonian Dynamics is equivalent to the study of Symplectic Maps.

Exercises

6.1.1. This is an exercise on key properties of *differential forms*. Consider the differential form

$$\sum_{b=1}^m C_b(z) dz_b \quad (6.1.21)$$

where the $C_b(z)$ are specified functions of the m variables z_1, z_2, \dots, z_m . Before going any further, it is convenient to give a differential form a name so that it is not necessary to always write it out in full. We, as is common, will use the symbol ω to denote the differential form (1.21) and write

$$\omega = \sum_{b=1}^m C_b(z) dz_b. \quad (6.1.22)$$

³The fundamental symplectic 2-form is also sometimes called the *Lagrange invariant*. We also remark that, in the language of differential geometry, the vectors $d\bar{z}$ and $\delta\bar{z}$ are said to be the result of *pushing forward* the vectors dz and δz . Conversely, the vectors dz and δz are said to be the result of *pulling back* the vectors $d\bar{z}$ and $\delta\bar{z}$. According to (1.17) and (1.18) the push-forward operation is accomplished by application of the matrix M . And, assuming M^{-1} exists, pulling back would be accomplished by application of M^{-1} . For the case of symplectic maps we know that M^{-1} exists because then $\det(M) = 1$.

We are now prepared to make some definitions and demonstrate some results about differential forms:

- a) A differential form is called *exact* or *perfect* if there exists a function $f(z)$ such that

$$\omega = df. \quad (6.1.23)$$

We know that for any differentiable function f there is the relation

$$df = \sum_{b=1}^m (\partial f / \partial z_b) dz_b. \quad (6.1.24)$$

Now suppose the differential form ω is exact. Then, from (1.23), and comparing (1.22) and (1.24), we find the result

$$C_b = \partial f / \partial z_b. \quad (6.1.25)$$

Show, using the equality of mixed partial derivatives, that (1.25) implies the result

$$\partial C_b / \partial z_a - \partial C_a / \partial z_b = \partial^2 f / \partial z_a \partial z_b - \partial^2 f / \partial z_b \partial z_a = 0. \quad (6.1.26)$$

A differential form ω that satisfies this relation is called *closed*. Thus, being exact implies being closed.

- b) Conversely, suppose (1.26) holds in some *simply-connected* region \mathcal{R} . That is, assume the form ω is closed in \mathcal{R} . Let z^i and z^f be two arbitrary points in \mathcal{R} , and let P be some path in \mathcal{R} joining them. Consider the integral

$$I[P] = \int_{z^i}^{z^f} \sum_b C_b(z) dz_b \quad (6.1.27)$$

evaluated over the path P . In view of (1.22), we may also employ the notation

$$I[P] = \int_{z^i}^{z^f} \omega. \quad (6.1.28)$$

We may regard (1.27) as a *functional* on paths, and write

$$I[z(\tau)] = \int_{\tau^i}^{\tau^f} \left\{ \sum_b C_b(z) \dot{z}_b \right\} d\tau \quad (6.1.29)$$

where $z(\tau)$ is some parameterization of the path and

$$\dot{z}_b = dz_b / d\tau. \quad (6.1.30)$$

Define a “Lagrangian” L by writing

$$L(z, \dot{z}) = \sum_b C_b(z) \dot{z}_b. \quad (6.1.31)$$

With this definition, (1.29) takes the form

$$I = \int_{\tau^i}^{\tau^f} L(z, \dot{z}) d\tau. \quad (6.1.32)$$

Show, using standard variational calculus, that

$$\delta I = \int_{\tau^i}^{\tau^f} d\tau \left\{ \sum_a \left(-\frac{d}{d\tau} \frac{\partial L}{\partial \dot{z}_a} + \frac{\partial L}{\partial z_a} \right) \delta z_a \right\} \quad (6.1.33)$$

for a varied path with the same end points. Show from its definition (1.31) that L satisfies Lagrange's equation,

$$\frac{d}{d\tau} \frac{\partial L}{\partial \dot{z}_a} - \frac{\partial L}{\partial z_a} = 0, \quad (6.1.34)$$

if (1.26) holds, and therefore

$$\delta I = 0 \text{ for all } \delta z_a. \quad (6.1.35)$$

The relation (1.35) shows that I is unchanged in first order when infinitesimal variations (with end points fixed) are made in the path.

From this result, show that I is in fact path independent. In particular, suppose $z(\tau)$ and $\tilde{z}(\tau)$ are two paths in \mathcal{R} with the same end points. Consider the family of paths $z(\tau, \lambda)$ defined by

$$z(\tau, \lambda) = (1 - \lambda)z(\tau) + \lambda\tilde{z}(\tau). \quad (6.1.36)$$

Evidently there are the relations

$$z(\tau, 0) = z(\tau), \quad z(\tau, 1) = \tilde{z}(\tau). \quad (6.1.37)$$

Verify that all the paths in the family have the same end points. Assuming that $z(\tau, \lambda)$ remains in \mathcal{R} for $\tau \in [\tau^i, \tau^f]$ and $\lambda \in [0, 1]$, show from (1.35) that

$$(\partial/\partial\lambda)I[z(\tau, \lambda)] = 0, \quad (6.1.38)$$

and therefore $I[z(\tau, \lambda)]$ is independent of λ so that

$$I[z(\tau)] = I[\tilde{z}(\tau)]. \quad (6.1.39)$$

Now that it has been established that the integral (1.27) is path independent, and therefore depends only on the end points, show that one can define a function $f(z)$ by the rule

$$f(z) = \int_{z^i}^z \sum_b C_b(z') dz'_b. \quad (6.1.40)$$

Show, by selecting and sketching a suitable path, that

$$\partial f / \partial z_a = C_a(z). \quad (6.1.41)$$

Hint: To verify (1.41), select a path such that only z'_a varies near the upper integration limit. That is, near and at the final end of this path, the z'_b for $b \neq a$ have already taken on the values $z'_b = z_b$.

Show that

$$df = \sum_b (\partial f / \partial z_b) dz_b = \sum_b C_b(z) dz_b. \quad (6.1.42)$$

Therefore, (1.26) is both necessary and sufficient for a differential to be exact: An exact form is closed, and a form that is closed in a simply-connected region is exact. This result is sometimes called the *Poincaré lemma*. Note that it is an m -dimensional generalization of the familiar 3-dimensional theorem that a vector field can be written as the gradient of a scalar field if and only if the vector field has vanishing curl.

- c) Finally, show that if ω is exact, then

$$\int_{\Gamma} \omega = \int_{\Gamma} \sum_b C_b(z) dz_b = 0 \quad (6.1.43)$$

where Γ is any closed path in \mathcal{R} , and vice versa.

6.1.2. Consider a two-dimensional phase space consisting of the variables q, p . Evaluate the quantity $(\delta z, Jdz)$ and show that it is related to the area formed by the small parallelogram with sides δz and dz . Note that $(\delta z, Jdz)$ can be either positive or negative. Thus, the area is “signed”. Consider a $2n$ dimensional phase space. Show that the points $z(\sigma, \tau)$ given by the relation

$$z(\sigma, \tau) = z^0 + \sigma dz + \tau \delta z \text{ with } \sigma, \tau \in [0, 1] \quad (6.1.44)$$

form a two dimensional surface in phase space that can be viewed as a generalized parallelogram with sides δz and dz . Show that this generalized parallelogram has projections into the z_a, z_b planes that are “ordinary” parallelograms (each z_a, z_b plane is two dimensional). In particular, the projections of the generalized parallelogram into the q_i, p_i planes are parallelograms. Finally, show that $(\delta z, Jdz)$ is related to the sum of the *signed* areas of the parallelograms in the q_i, p_i planes. Hint: Use (3.2.3).

6.2 Group Properties

6.2.1 The General Case

Let \mathcal{M} be a symplectic mapping of z to \bar{z} , and suppose it has an inverse \mathcal{M}^{-1} ,

$$\mathcal{M} : z \rightarrow \bar{z}, \quad (6.2.1)$$

$$\mathcal{M}^{-1} : \bar{z} \rightarrow z. \quad (6.2.2)$$

According to (1.17), the relation between a small change dz in z , and the associated small change $d\bar{z}$ in \bar{z} , is given by the Jacobian matrix M of \mathcal{M} . Since M is symplectic, it has an inverse M^{-1} . Therefore, (1.17) can be inverted to give the relation

$$dz = M^{-1} d\bar{z}. \quad (6.2.3)$$

But now, comparison of (2.2) and (2.3) shows that the Jacobian matrix of \mathcal{M}^{-1} is M^{-1} . Note also that the local existence of \mathcal{M}^{-1} did not really have to be assumed, but follows instead from the inverse function theorem since M^{-1} is known to exist from the symplectic condition. Finally, the matrix M^{-1} is symplectic since the inverse of a symplectic matrix is also a symplectic matrix. It follows that \mathcal{M}^{-1} is a symplectic map. What has been shown is that if \mathcal{M} is a symplectic map, then \mathcal{M}^{-1} exists (at least locally) and is also a symplectic map.

Next suppose that $\mathcal{M}^{(1)}$ is a symplectic mapping of z to \bar{z} and $\mathcal{M}^{(2)}$ is a symplectic mapping of \bar{z} to another set of variables $\bar{\bar{z}}$. Now consider the composite mapping $\mathcal{M} = \mathcal{M}^{(2)}\mathcal{M}^{(1)}$, which sends z to $\bar{\bar{z}}$.

$$\mathcal{M} = \mathcal{M}^{(2)}\mathcal{M}^{(1)}, \quad (6.2.4)$$

$$\mathcal{M}^{(1)} : z \rightarrow \bar{z}, \quad (6.2.5)$$

$$\mathcal{M}^{(2)} : \bar{z} \rightarrow \bar{\bar{z}}, \quad (6.2.6)$$

$$\mathcal{M}^{(2)}\mathcal{M}^{(1)} : z \rightarrow \bar{\bar{z}}. \quad (6.2.7)$$

According to the chain rule, the Jacobian matrix M of the composite mapping \mathcal{M} is the product of the Jacobian matrices of $\mathcal{M}^{(2)}$ and $\mathcal{M}^{(1)}$,

$$M = M^{(2)}M^{(1)}. \quad (6.2.8)$$

However, the matrices $M^{(2)}$ and $M^{(1)}$ are symplectic since they are the Jacobian matrices of symplectic maps. It follows from (2.8) and the group property for symplectic matrices that M is also a symplectic matrix. Consequently, the composite mapping \mathcal{M} is also a symplectic map. What has been shown is that if $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ are symplectic maps, so is their product $\mathcal{M}^{(2)}\mathcal{M}^{(1)}$.

It is also obvious that the identity mapping, which sends each z into itself, is a symplectic map because the Jacobian matrix of this map is evidently the identity matrix, and the identity matrix is symplectic.

The previous discussion has shown that the set of symplectic maps has properties very analogous to the group properties of the group of symplectic matrices. As defined earlier, the concept of a group applied only to matrices. However, it is clear that the concept of a group can be enlarged to include the possibility of general mappings. When this is done, the set of all symplectic maps is entitled to be called a group. The set of all differentiable maps forms a group called the group of all diffeomorphisms. Because of the symplectic restriction, the set of all symplectic maps is a subgroup of the group of all diffeomorphisms.

6.2.2 Various Subgroups and Their Names

We found in Section 3.6.1 that the set of all real $2n \times 2n$ symplectic matrices forms a group. We have denoted this group and its Lie algebra by the symbols $Sp(2n, \mathbb{R})$ and $sp(2n, \mathbb{R})$. Equivalently, in the present context, the subset of all symplectic maps that send the origin into itself (preserve the origin) and are *linear* is a subgroup of the group of all symplectic maps, and this subgroup is $Sp(2n, \mathbb{R})$. See Exercise 2.1. Evidently, the subset of all symplectic maps that send the origin into itself, but are not necessarily linear, is also

a subgroup of the group of all symplectic maps. For lack of any standard terminology, we will refer to this group as $SpM(2n, \mathbb{R})$. Here the M in the name stands either for the word *map* or, to please the French, the word *morphism* because those of Gallic bent often refer to maps as *morphisms*. The underlying Lie algebra of $SpM(2n, \mathbb{R})$, see Section 7.7, will be referred to as $spm(2n, \mathbb{R})$.

Next consider mappings of the form

$$\bar{z}_a = z_a + c_a, \quad (6.2.9)$$

where the quantities c_a are constants. It is easily verified that such maps are symplectic, and form a group. This group is called the phase-space *translation* group. Now consider phase-space mappings of the form

$$\bar{z}_a = c_a + \sum_b M_{ab} z_b, \quad (6.2.10)$$

where the matrices M are symplectic. Such maps are also symplectic and form a group. This group is called the *inhomogeneous* symplectic group, and will be referred to by the symbols $ISp(2n, \mathbb{R})$. The underlying Lie algebra of $ISp(2n, \mathbb{R})$, see Sections 7.7 and 9.2, will be referred to as $isp(2n, \mathbb{R})$.

Finally, consider the group of all symplectic maps (also called *symplectomorphisms*) that do not necessarily preserve the origin and are not necessarily linear. This group will be referred to as $ISpM(2n, \mathbb{R})$ and its Lie algebra, see Exercise 7.7.2, will be referred to as $ispm(2n, \mathbb{R})$.⁴

We close this section by noting that gradient maps do *not* form a group. In the case of gradient maps there is again a relation like (2.8) for the Jacobian of the product of two maps. However, the product of two symmetric matrices is generally not a symmetric matrix. Therefore, the product of two gradient maps is generally not a gradient map. Gradient maps belong to the group of all diffeomorphisms, but do not form a subgroup. However, it can be shown that the identity map is a gradient map; and the inverse of a gradient map, if the inverse exists, is also a gradient map. See Exercise 2.9.

Exercises

6.2.1. Consider phase-space mappings of the form

$$\bar{z} = Mz \quad (6.2.11)$$

where M is a symplectic matrix. Show that such maps are symplectic, and form a group. Show that the symplectic map for $M = J$ interchanges (with a minus sign) coordinates and momenta.

6.2.2. Consider phase-space mappings of the form (2.9). Show that such maps are symplectic, and form a group. Consider phase-space mappings of the form (2.10). Show that such maps are symplectic, and also form a group.

⁴Some authors refer to $ISpM(2n, \mathbb{R})$ simply as *Symp(n)* and to $Sp(2n, \mathbb{R})$ as *Sp(n)*.

6.2.3. Use (1.2) and the chain rule to verify (2.8).

6.2.4. Consider the nonrelativistic motion of a particle of mass m described by Cartesian coordinates $\mathbf{q}(t)$. In the usual way, define the momentum $\mathbf{p}(t)$ by the relation

$$\mathbf{p} = m\dot{\mathbf{q}}. \quad (6.2.12)$$

The *Euclidean* group consists of spatial transformations of the form

$$\bar{\mathbf{q}} = R\mathbf{q} + \mathbf{d} \quad (6.2.13)$$

where R is a 3×3 rotation matrix and \mathbf{d} is a fixed vector. It describes rotations and translations (displacements) in 3-dimensional space. These transformations are extended to phase space by the rule

$$\bar{\mathbf{p}} = R\mathbf{p}. \quad (6.2.14)$$

Setting $z = (\mathbf{q}; \mathbf{p})$, verify that (2.13) and (2.14) specify a symplectic map. That is, verify that

$$[\bar{z}_a, \bar{z}_b] = J_{ab}. \quad (6.2.15)$$

Thus, the Euclidean group is a subgroup of the group of all symplectic maps.

To the Euclidean group add the further spatial transformations

$$\bar{\mathbf{q}}(t) = \mathbf{q}(t) + \mathbf{u}t. \quad (6.2.16)$$

These transformations describe the (nonrelativistic) coordinate relation between two inertial frames moving with (fixed) relative velocity \mathbf{u} . In accord with (2.12), these transformations may be extended to phase space by the rule

$$\bar{\mathbf{p}} = \mathbf{p} + m\mathbf{u}. \quad (6.2.17)$$

Show that the transformations described by (2.16) and (2.17) are also symplectic maps. Together the transformations described by (2.13), (2.14) and (2.16),(2.17) form the group of all *Galilean* transformations. You have shown that the Galilean group is a subgroup of the group of all symplectic maps.⁵

Suppose we extend phase space to include t as a coordinate and p_t as its conjugate momentum, in which case some parameter τ becomes the independent variable. See Exercises 1.6.4 and 1.6.5. Can the Galilean group be extended to act on this extended phase space? Implicit in the nonrelativistic approach is the assumption that time is the same in all inertial frames,

$$\bar{t} = t. \quad (6.2.18)$$

How should we define \bar{p}_t ? As motivation, consider the free particle case for which we have the relation

$$p_t = -\mathbf{p} \cdot \mathbf{p}/(2m). \quad (6.2.19)$$

⁵Note that all these transformations are in fact a subset of the inhomogeneous symplectic group, and are therefore automatically symplectic. See (2.9) and (2.10).

If we write

$$\bar{p}_t = -\bar{\mathbf{p}} \cdot \bar{\mathbf{p}}/(2m) \quad (6.2.20)$$

and use (2.17), we find the result

$$\bar{p}_t = p_t - \mathbf{u} \cdot \mathbf{p} - (m/2)\mathbf{u} \cdot \mathbf{u}, \quad (6.2.21)$$

which we take to be the rule for how p_t transforms.⁶

Show that the relations (2.13),(2.14) and (2.16),(2.17) and (2.18),(2.21) also yield a group, which we might call the extended Galilean group. Are these transformations symplectic maps on the extended phase space? Show that they are, and therefore the extended Galilean group is a subgroup of the group of all symplectic maps on extended phase space. To make this demonstration, write $z = (\mathbf{q}, t; \mathbf{p}, p_t)$ and again set up the usual rules

$$[z_a, z_b] = J_{ab}. \quad (6.2.22)$$

Show it then follows that (2.15) also holds on the extended phase space. In particular, you will need to verify that

$$[\bar{p}_t, \bar{\mathbf{q}}] = 0. \quad (6.2.23)$$

Finally, note that the fact that a particular group can be realized as a set of phase-space transformations does not necessarily say anything about the invariance properties of the dynamics of any particular system. What is needed for invariance is for trajectories to be sent into trajectories under the action of the group. For example, see Exercise 1.6.9.

6.2.5. Read Exercise 2.4. The reader may be dubious about the use of the free particle case to motivate the transformation rule (2.21). Here is another approach. For simplicity, consider the case in which phase space is two dimensional. Suppose that the transformation rule for p_t is of the form

$$\bar{p}_t = p_t - up + \alpha(u) \quad (6.2.24)$$

where $\alpha(u)$ is a function yet to be determined. Verify that the $-up$ term in (2.24) is necessary to satisfy (2.23), but that the symplectic condition is satisfied for any choice of α . Now make the requirement that the extended Galilean transformations form a group. Make two successive transformations with relative velocities u_1 and u_2 to obtain the relations

$$\bar{q} = q + u_1 t, \quad (6.2.25)$$

$$\bar{t} = t, \quad (6.2.26)$$

$$\bar{p} = p + mu_1, \quad (6.2.27)$$

$$\bar{p}_t = p_t - u_1 p + \alpha(u_1); \quad (6.2.28)$$

$$\bar{\bar{q}} = \bar{q} + u_2 \bar{t}, \quad (6.2.29)$$

$$\bar{\bar{t}} = \bar{t}, \quad (6.2.30)$$

$$\bar{\bar{p}} = \bar{p} + mu_2, \quad (6.2.31)$$

⁶See also Exercises 2.5 and 2.7 below.

$$\bar{\bar{p}}_t = \bar{p}_t - u_2 \bar{p} + \alpha(u_2). \quad (6.2.32)$$

Show that combining the relations (2.25) through (2.32) yields the net relations

$$\bar{q} = q + (u_1 + u_2)t, \quad (6.2.33)$$

$$\bar{\bar{t}} = t, \quad (6.2.34)$$

$$\bar{\bar{p}} = p + m(u_1 + u_2), \quad (6.2.35)$$

$$\bar{\bar{p}}_t = p_t - (u_1 + u_2)p - mu_1 u_2 + \alpha(u_1) + \alpha(u_2). \quad (6.2.36)$$

We see that (2.33) through (2.35) are of the standard Galilean transformation form corresponding to a relative velocity $u_1 + u_2$. Therefore, if we wish (2.36) to also be of the standard form (2.24), we must require the relation

$$-mu_1 u_2 + \alpha(u_1) + \alpha(u_2) = \alpha(u_1 + u_2). \quad (6.2.37)$$

Show that (2.37) implies the relation

$$\alpha(0) = 0. \quad (6.2.38)$$

To make further progress, assume that α is differentiable.⁷ Set

$$u_1 = u \quad (6.2.39)$$

and

$$u_2 = \epsilon. \quad (6.2.40)$$

Then, assuming differentiability, show that there are the relations

$$\alpha(u_1 + u_2) = \alpha(u + \epsilon) = \alpha(u) + \alpha'(u)\epsilon + O(\epsilon)^2, \quad (6.2.41)$$

$$\alpha(u_2) = \alpha(\epsilon) = \alpha(0) + \alpha'(0)\epsilon + O(\epsilon)^2. \quad (6.2.42)$$

Next, show that inserting (2.38) through (2.42) into (2.37) and equating like powers of ϵ yields the differential equation

$$\alpha'(u) = \alpha'(0) - mu \quad (6.2.43)$$

with the solution

$$\alpha(u) = ua'(0) - (1/2)mu^2. \quad (6.2.44)$$

Here, in solving (2.43), we have taken into account the boundary condition (2.38). Finally, let us apply the transformation (2.25) through (2.28) to the phase-space origin $q = p = 0$. Doing so gives the result

$$\bar{p}_t = p_t + \alpha(u) = p_t + ua'(0) - (1/2)u^2. \quad (6.2.45)$$

If we now require that \bar{p}_t be independent of the sign of u , which seems reasonable since there is no preferred direction when $q = p = 0$, we conclude that we should demand the further condition

$$\alpha'(0) = 0. \quad (6.2.46)$$

Thus, under reasonable assumptions, we again arrive at (2.21).

⁷Actually, it is sufficient to assume continuity. It is a remarkable property of Lie groups that the assumption of continuity implies differentiability, and indeed, also the far stronger condition of analyticity.

6.2.6. Study Exercises 1.6.7, 1.6.8, 1.6.16, 1.6.17, and 1.7.5. Suppose x and y are any two space-time points. The *interval* $I(x, y)$ between them is given by the relation

$$I(x, y) = g_{\mu\nu}(x - y)^\mu(x - y)^\nu = ([x - y], g[x - y]) = (x - y) \cdot (x - y). \quad (6.2.47)$$

Here we use the metric g given by (1.6.45) and $(*, *)$ denotes the usual/ordinary scalar product. Consider the set of all transformations that send space-time into itself. Special Relativity asserts that if two events can occur at the points x, y (i.e. the events are consonant with physical law), then they can also occur at the points \tilde{x}, \tilde{y} provided

$$I(\tilde{x}, \tilde{y}) = I(x, y). \quad (6.2.48)$$

Transformations that satisfy (2.48) are called *Poincaré* transformations.

In studying Poincaré transformations it is often assumed from the outset that they are of the form

$$\tilde{x}^\alpha = \sum_{\beta=1}^4 \Lambda^{\alpha\beta} x^\beta + d^\alpha \Leftrightarrow \tilde{x} = \Lambda x + d. \quad (6.2.49)$$

That is, they are assumed to consist of a *linear* transformation described by the matrix Λ followed by a space-time translation described by the 4-vector d . Such an assumption is not necessary. It can be proved that the most general transformation satisfying (2.48) for all pairs of points must be of the form (2.49). See Exercise 7.3.26. Assuming the form (2.49), show from (2.48) that the matrix Λ must satisfy the relation

$$\Lambda^T g \Lambda = g. \quad (6.2.50)$$

Show that the matrices Λ form a group (called the *Lorentz* group).⁸ Show that Poincaré transformations also form a group (called the Poincaré group).⁹ Show that there are the logical implications

$$\Lambda^T g \Lambda = g \Leftrightarrow \Lambda g \Lambda^T = g \Leftrightarrow (\Lambda^T)^T g \Lambda^T = g. \quad (6.2.51)$$

Suppose that Λ is an element of the Lorentz group. Show that then Λ^{-1} and Λ^T and $(\Lambda^T)^{-1} = (\Lambda^{-1})^T$ are elements of the Lorentz group, and vice versa.

Suppose space-time coordinates are transformed according to (2.49) and the action of the Lorentz (and Poincaré) group is *extended* to act on momenta by the rule

$$\tilde{p} = \Lambda p \Leftrightarrow \tilde{p}^\alpha = \sum_{\beta=1}^4 \Lambda^{\alpha\beta} p^\beta. \quad (6.2.52)$$

⁸The finite dimensional representations of the Lorentz group formed by the matrices Λ are described in Exercise 7.3.27. Remarkably, as shown in Exercise 7.3.29, the identity component of the Lorentz group is homomorphic to the group $SL(2, \mathbb{C})$. Indeed, $SL(2, \mathbb{C})$ is the covering group of the Lorentz group.

We also take this occasion to make a comment about nomenclature and notation. Under a Lorentz transformation space-time transforms according to the rule $\tilde{x} = \Lambda x$ from which it follows that $d\tilde{x} = \Lambda dx$. There is also the chain-rule relation $d\tilde{x}^\alpha = \sum_\beta (\partial \tilde{x}^\alpha / \partial x^\beta) dx^\beta$ and therefore $\Lambda^{\alpha\beta} = \partial \tilde{x}^\alpha / \partial x^\beta$. Any collection of four elements V^α is defined to be a *four-vector* if there is the transformation rule $\tilde{V} = \Lambda V$. Similarly, any set of sixteen elements $T^{\alpha\beta}$ is defined to be a *second-rank tensor* if there is the transformation rule $\tilde{T}^{\alpha\beta} = \sum_{\mu\nu} \Lambda^{\alpha\mu} \Lambda^{\beta\nu} T^{\mu\nu}$. Note that these transformation rules can also be written in the less compact forms $\tilde{V}^\alpha = \sum_\mu (\partial \tilde{x}^\alpha / \partial x^\mu) V^\mu$ and $\tilde{T}^{\alpha\beta} = \sum_{\mu\nu} (\partial \tilde{x}^\alpha / \partial x^\mu) (\partial \tilde{x}^\beta / \partial x^\nu) T^{\mu\nu}$, etc., as is frequently done.

⁹The Poincaré group could also be called the *inhomogeneous* Lorentz group.

Define canonical coordinates in an eight-dimensional phase space to consist of the pairs (x^μ, p_ν) . It can be shown that (2.49) and (2.52) produce a symplectic map in this phase space. See Exercise 2.13. That is, (extended) Poincaré transformations are symplectic maps, and therefore form a subgroup of the group of all symplectic maps. Indeed, since Lorentz transformations are linear, the (extended) Lorentz group is a subgroup of $Sp(8, \mathbb{R})$. And, according to (2.49), the (extended) Poincaré group is a subgroup of $ISp(8, \mathbb{R})$.

6.2.7. Read Exercises 2.4 and 2.6 above if you have not already done so. The Poincaré group involves the parameter c . The aim of this exercise is to show that in the limit $c \rightarrow \infty$ the Poincaré group becomes the extended Galilean group plus translations in time. Consider, for simplicity, a velocity transformation along the z axis with velocity u . In this case Λ is the matrix

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \gamma(u) & \beta(u)\gamma(u) \\ 0 & 0 & \beta(u)\gamma(u) & \gamma(u) \end{pmatrix} \quad (6.2.53)$$

where

$$\beta(u) = u/c \quad (6.2.54)$$

and

$$\gamma(u) = 1/\sqrt{1 - [\beta(u)]^2}. \quad (6.2.55)$$

Then, from (2.49), we find for the space-time coordinate variables x^μ the relations

$$\tilde{x}^1 = x^1, \quad (6.2.56)$$

$$\tilde{x}^2 = x^2, \quad (6.2.57)$$

$$\tilde{x}^3 = \gamma(u)x^3 + \gamma(u)\beta(u)x^4, \quad (6.2.58)$$

$$\tilde{x}^4 = \gamma(u)\beta(u)x^3 + \gamma(u)x^4. \quad (6.2.59)$$

Show that, with the aid of (1.6.41), these last two relations can be rewritten in the form

$$\tilde{z} = \gamma(u)z + \gamma(u)\beta(u)ct = \gamma(u)z + \gamma(u)ut, \quad (6.2.60)$$

$$\tilde{t} = \gamma(u)\beta(u)z/c + \gamma(u)t. \quad (6.2.61)$$

Show that in the limit $c \rightarrow \infty$ the relations (2.60) and (2.61) become

$$\tilde{z} = z + ut, \quad (6.2.62)$$

$$\tilde{t} = t, \quad (6.2.63)$$

which, along with (2.56) and (2.57), are a special case of the Galilean transformation given by (2.16) and (2.18).

The limiting case of the momentum relations (2.52) is a bit more delicate because the quantities p^μ are c dependent. Verify that for Λ given by (2.53) the relations (2.52) have the component form

$$\tilde{p}^1 = p^1, \quad (6.2.64)$$

$$\tilde{p}^2 = p^2, \quad (6.2.65)$$

$$\tilde{p}^3 = \gamma(u)p^3 + \gamma(u)\beta(u)p^4, \quad (6.2.66)$$

$$\tilde{p}^4 = \gamma(u)\beta(u)p^3 + \gamma(u)p^4. \quad (6.2.67)$$

Show that use of (1.6.82), (1.6.96), and (1.6.97) in (2.66) gives the result

$$\begin{aligned} \gamma(\tilde{v})m\tilde{v}_z + q\tilde{A}_z &= \gamma(u)[\gamma(v)mv_z + A_z] + \gamma(u)\beta(u)[qA^4 + \gamma(v)mc] \\ &= \gamma(u)[\gamma(v)mv_z + A_z] + \gamma(u)\gamma(v)mu + \gamma(u)\beta(u)qA^4. \end{aligned} \quad (6.2.68)$$

Show that in the limit $c \rightarrow \infty$ (2.68) becomes

$$m\tilde{v}_z + q\tilde{A}_z = mv_z + A_z + mu, \quad (6.2.69)$$

which can be written in the form

$$\tilde{p}_z^{\text{nr}} = p_z^{\text{nr}} + mu \quad (6.2.70)$$

where p^{nr} is the nonrelativistic canonical momentum. Observe that (2.64), (2.65), and (2.70) are a special case of (2.17). Thus we have obtained in the $c \rightarrow \infty$ limit the Galilean transformations for the spatial components of p^μ .

What remains is the temporal component of p^μ . Show that use of (1.6.82), (1.6.96), and (1.7.20) in (2.67) gives the result

$$\tilde{p}_t = -\gamma(u)\beta(u)cp^3 + \gamma(u)p_t, \quad (6.2.71)$$

which can be rewritten in the form

$$\tilde{p}_t = -\gamma(u)up^3 + \gamma(u)p_t. \quad (6.2.72)$$

Next we need to expand p_t as given by (1.7.22). Show that

$$\begin{aligned} p_t &= -q\psi - \gamma(v)mc^2 = -q\psi - mc^2 - (1/2)mv^2 + c^2O(v/c)^4 \\ &= p_t^{\text{nr}} - mc^2 + c^2O(v/c)^4 \end{aligned} \quad (6.2.73)$$

where we have defined a nonrelativistic p_t by the rule

$$p_t^{\text{nr}} = -q\psi - (1/2)mv^2 = p_t + mc^2 + c^2O(v/c)^4. \quad (6.2.74)$$

Now insert (2.73) into (2.72) to obtain the result

$$\begin{aligned} \tilde{p}_t^{\text{nr}} &= -\gamma(u)up^3 + \gamma(u)[p_t^{\text{nr}} - mc^2] + mc^2 + c^2O(\tilde{v}/c)^4 + c^2O(v/c)^4 \\ &= -\gamma(u)up^3 + \gamma(u)p_t^{\text{nr}} + mc^2[1 - \gamma(u)] + c^2O(\tilde{v}/c)^4 + c^2O(v/c)^4. \end{aligned} \quad (6.2.75)$$

Next verify that

$$p^3 = p_z^{\text{nr}} + O(v/c)^2 \quad (6.2.76)$$

and

$$mc^2[1 - \gamma(u)] = -(1/2)mu^2 + c^2O(u/c)^4. \quad (6.2.77)$$

Now we are ready to take the $c \rightarrow \infty$ limit of (2.75). Verify that this limit is

$$\tilde{p}_t^{\text{nr}} = p_t^{\text{nr}} - up_z^{\text{nr}} - (m/2)u^2. \quad (6.2.78)$$

Observe that (2.78) is a special case of (2.21).

We have achieved our goal. We have seen that under a suitable limiting process the Lorentz group reduces to the extended Galilean group.¹⁰ Correspondingly, the Poincaré group reduces to the extended Galilean group plus translations in time, $\tilde{t} = t + a^4$.

6.2.8. Study Exercises 1.6.7 through 1.6.10 and Exercise 1.7.5, and adopt the phase-space coordinates of Exercise 2.6. Show that gauge transformations produce symplectic maps. Note that if it is necessary to append a gauge transformation to the Lorentz transformation (2.52), as described in a footnote to Exercise 1.6.7, the net result is still a symplectic map. Let $\phi(x)$ be any scalar field, and let \mathcal{A} be the symplectic (see Section 7.1) map

$$\mathcal{A} = \exp(-q : \phi :). \quad (6.2.79)$$

Let \mathcal{A} act on the H_R given by (1.6.92), and demonstrate that \mathcal{A} produces gauge transformations.

6.2.9. Let \mathcal{S}_0 be the set of all diffeomorphisms (in $2n$ dimensions), \mathcal{S}_1 be the set of all orientation preserving diffeomorphisms (see Exercise 1.4.6), and \mathcal{S}_2 be the set of all symplectic maps. Show that each of these sets forms a group, and that there is the inclusion relation

$$\mathcal{S}_0 \supset \mathcal{S}_1 \supset \mathcal{S}_2. \quad (6.2.80)$$

Review Section 6.1.1. Show, by an example, that the product of two symmetric matrices need not itself be symmetric, and thereby demonstrate, in view of (2.4) through (2.8), that gradient maps do not form a subgroup of the group of all diffeomorphisms. Review Exercise 5.3.7. Show that the maps produced by integrating gradient vector fields over some time interval are diffeomorphisms, but generally do not form a subgroup of the group of all diffeomorphisms. Show that, despite the use of the adjective *gradient* to describe the underlying vector fields, such maps are also generally not gradient maps.

Show that the identity map \mathcal{I} defined by

$$\bar{u} = \mathcal{I}u = u \quad (6.2.81)$$

has the identity matrix I for its Jacobian matrix, and therefore $\mathcal{G} = \mathcal{I}$ is a gradient map. Using (1.11), show that the associated source function $g(u)$ that produces this \mathcal{G} is given by

$$g(u) = (u, u)/2, \quad (6.2.82)$$

¹⁰This reduction of the Lorentz group to the Galilean group is an example of a process that can be applied to many groups and is called *Inönü-Wigner contraction*. The inverse process to contraction is called *deformation*. For example, it can be shown that the Quantum Mechanical commutator Lie algebra of functions of the quantum variables Q and P is a deformation of the Poisson bracket Lie algebra of functions of the associated classical variables q and p . We may say that Quantum Mechanics is a deformation of Classical Mechanics, and Classical Mechanics is a contraction of Quantum Mechanics in the limit $\hbar \rightarrow 0$. See Appendix Y. Similarly, ray optics is a contraction of wave optics in the limit that the wavelength $\lambda \rightarrow 0$, and wave optics is a deformation of ray optics.

and verify that use of this g in (1.4) yields $\mathcal{G} = \mathcal{I}$.

Suppose \mathcal{G} is a gradient map. Write

$$\bar{u} = \mathcal{G}u, \quad (6.2.83)$$

and require that its Jacobian matrix G be symmetric as in (1.7) so that \mathcal{G} is indeed a gradient map. Suppose further that \mathcal{G} is invertible, and let \mathcal{H} be its inverse so that we may write

$$u = \mathcal{H}\bar{u}, \quad (6.2.84)$$

or

$$\mathcal{H} = \mathcal{G}^{-1}. \quad (6.2.85)$$

Your first task is to show that \mathcal{H} is also a gradient map.

Show from (1.5) that there is the relation

$$d\bar{u} = G(u)du. \quad (6.2.86)$$

By the inverse function theorem, the condition for \mathcal{G} to be invertible is

$$\det G \neq 0. \quad (6.2.87)$$

Show under the assumption (2.87) that (2.86) can be rewritten in the form

$$du = H(\bar{u})d\bar{u} \quad (6.2.88)$$

with

$$H(\bar{u}) = [G(u)]^{-1}. \quad (6.2.89)$$

Thus, H is the Jacobian matrix for \mathcal{H} . Verify it follows that there are the series of relations

$$HG = I, \quad (6.2.90)$$

$$G^T H^T = I, \quad (6.2.91)$$

$$GH^T = I, \quad (6.2.92)$$

$$H^T = G^{-1} = H. \quad (6.2.93)$$

You have shown that \mathcal{H} is also a gradient map.

What is the source function h for \mathcal{H} ? Let g be the source function for \mathcal{G} . Show that the Ansatz

$$h(\bar{u}) = (u, \bar{u}) - g(u) \quad (6.2.94)$$

produces a well-defined function $h(\bar{u})$ when u is viewed as a function of \bar{u} ,

$$u = \mathcal{G}^{-1}\bar{u}. \quad (6.2.95)$$

Show that the differential of h is given by the relation

$$dh = \sum_a [\bar{u}_a(du_a) + u_a(d\bar{u}_a) - (\partial g / \partial u_a)(du_a)]. \quad (6.2.96)$$

Next use (1.4) to show that dh as given by (2.96) can also be written in the form

$$dh = \sum_a u_a(d\bar{u}_a), \quad (6.2.97)$$

and thereby conclude that

$$u_a = \partial h / \partial \bar{u}_a. \quad (6.2.98)$$

Comparison of (2.84) and (2.98) shows that h is the source function for \mathcal{H} .

Show from (2.88) and (2.98) that

$$H_{ab} = \partial^2 h / \partial \bar{u}_a \partial \bar{u}_b. \quad (6.2.99)$$

The map \mathcal{H} will be invertible iff

$$\det H \neq 0. \quad (6.2.100)$$

Show from (2.90) that

$$(\det H)(\det G) = 1. \quad (6.2.101)$$

Therefore \mathcal{H} is invertible if \mathcal{G} is invertible, and vice versa. Moreover, the relation (2.94) can also be written in the form

$$g(u) = (u, \bar{u}) - h(\bar{u}) \quad (6.2.102)$$

where

$$\bar{u} = \mathcal{H}^{-1}u. \quad (6.2.103)$$

Equation (2.94) shows that h is the Legendre transform of g and, conversely, (2.102) shows that g is the Legendre transform of h . In this context, a Legendre transformation is the relation between the source function of a gradient map and the source function of its inverse.

The production of one function from another by performing a Legendre transformation may be viewed as the result of some operator \mathcal{O} acting on *function* space. Observe that g is associated with \mathcal{G} and h is associated with \mathcal{G}^{-1} . Since $(\mathcal{G}^{-1})^{-1} = \mathcal{G}$, it follows that \mathcal{O}^2 is the *identity* operator on function space. An operator whose square is the identity is called an *involution*.¹¹ We have learned that the act of performing a Legendre transformation is an involution.

As a sanity check on this claim, work out an example. Suppose $f(x)$ is a function of a single variable x given by the rule

$$f(x) = \lambda x^n \quad (6.2.104)$$

where λ is some positive constant, and let $g(\bar{x})$ be its Legendre transform. Show that

$$g(\bar{x}) = \bar{\lambda}(\bar{x})^{\bar{n}} \quad (6.2.105)$$

where

$$\bar{n} = n/(n-1) \quad (6.2.106)$$

and

$$\bar{\lambda} = (n-1)\lambda(n\lambda)^{-\bar{n}}. \quad (6.2.107)$$

¹¹Note: this use of the word *involution* is not to be confused with that in Section 5.2.

Let $h(\bar{x})$ be the Legendre transform of $g(\bar{x})$. Find $h(\bar{x})$ and verify that

$$h(\bar{x}) = f(\bar{x}). \quad (6.2.108)$$

Here is an alternate, but equivalent, definition of the Legendre transform. Given the function $g(u)$, show that $h(\bar{u})$ can be defined by the rule

$$h(\bar{u}) = \max_u [(u, \bar{u}) - g(u)]. \quad (6.2.109)$$

This relation holds when g is *convex*, i.e. the Hessian of g is a positive definite matrix at each point u . More generally, one should look for an extremum rather than a maximum. Extrema will exist and be locally isolated as long as the Hessian of g is nonsingular.

In the case of a function $f(x)$ of a single variable x with Legendre transform $g(\bar{x})$, how are the graphs $y = f(x)$ and $\bar{y} = g(\bar{x})$ related geometrically? Suggestion: See the Legendre transformation references listed in the bibliography at the end of this chapter.

Review the passage from a Lagrangian to a Hamiltonian employed in Section 1.5. Observe that the relation (1.5.7) between p and \dot{q} is a gradient map produced by using L as a source function, with the remaining variables q and t simply going along for the ride. Also, (1.5.8) is a Legendre transformation that produces H from L , again with the variables q and t going along for the ride. Finally, the first of the equations (1.5.11), the one yielding the \dot{q}_i , is simply the inverse of the gradient map (1.5.7). Evidently, the only additional “physics” in the relations (1.5.11) is that given by the second equation which yields the \dot{p}_i .

6.2.10. Let \mathcal{M} be a map and let M be its Jacobian matrix. By the inverse function theorem, \mathcal{M} is invertible in the vicinity of a point if M is invertible at that point. Suppose that M is invertible everywhere. Does it then follow that \mathcal{M} is globally invertible? Consider the following two-dimensional counter example: Suppose the map \mathcal{M} sends the points x, y to the points u, v by the rule

$$u = e^x \cos y, \quad (6.2.110)$$

$$v = e^x \sin y. \quad (6.2.111)$$

Show that in this case M is the matrix

$$M = \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix}. \quad (6.2.112)$$

Verify that

$$\det M = e^{2x} \neq 0, \quad (6.2.113)$$

and therefore M is globally invertible. Show that nevertheless \mathcal{M} is not globally invertible. Hint: Introduce complex variables z, w by writing the relations

$$z = x + iy, \quad (6.2.114)$$

$$w = u + iv. \quad (6.2.115)$$

Show that \mathcal{M} as given by (2.110) and (2.111) is equivalent to the complex relation

$$w = e^z, \quad (6.2.116)$$

and therefore \mathcal{M}^{-1} is given by the *multivalued* relation

$$z = \log w. \quad (6.2.117)$$

6.2.11. Consider a two-dimensional phase space with Cartesian coordinates q, p . Define new coordinates Q, P implicitly by the rules

$$q = (2P)^{1/2} \cos(Q), \quad (6.2.118)$$

$$p = -(2P)^{1/2} \sin(Q). \quad (6.2.119)$$

Verify that (2.118) and (2.119) have the inverse relations

$$Q = \tan^{-1}(-p/q) = -\tan^{-1}(p/q), \quad (6.2.120)$$

$$P = (1/2)(p^2 + q^2). \quad (6.2.121)$$

Evidently the quantities $(2P)^{1/2}$ and Q assign what are essentially cylindrical polar coordinates to the phase-space point having Cartesian coordinates q, p . The only difference is that increasing Q produces a clockwise rotation whereas, in the usual convention, increasing the polar angle θ produces a counterclockwise rotation. We also remark that commonly the symbols J, ϕ , called *action-angle variables*, are used instead of P, Q .

Let C be the closed circular path of radius R about the origin of q, p phase space obtained by using (2.118) and (2.119) with $(2P)^{1/2} = R$ and $Q \in [0, 2\pi]$. Verify that the *action* A associated with this circular path, defined by the relation

$$A = \oint_C p \, dq = \int_{p^2+q^2 \leq R^2} dp \, dq, \quad (6.2.122)$$

has the value

$$A = \pi R^2 = 2\pi P. \quad (6.2.123)$$

Show that there is the relation

$$[Q, P] = [q, p] = 1 \quad (6.2.124)$$

so that the quantities Q and P may be thought of as position-like and momentum-like coordinates, respectively. Correspondingly, the relations (2.120) and (2.121) describe a symplectic map \mathcal{M} , and the relations (2.118) and (2.119) describe its inverse. Verify that \mathcal{M} and \mathcal{M}^{-1} are *not* analytic at the origin. This is to be expected because polar coordinates are ill defined at the origin.

Let L be the harmonic oscillator Lagrangian

$$L = (1/2)(\dot{q}^2 - q^2). \quad (6.2.125)$$

Show that the associated Hamiltonian is

$$H = (1/2)(p^2 + q^2). \quad (6.2.126)$$

Show that under the symplectic map \mathcal{M} the Hamiltonian H becomes the transformed Hamiltonian K given by the relation

$$K = P. \quad (6.2.127)$$

Show that there is no Lagrangian whose associated Hamiltonian is K . See Exercise 2.9 and Section 1.5. In particular, see Exercises 1.5.13 and 1.5.14. Thus, under the action of symplectic maps, it is possible that one may move beyond the realm of Lagrangian mechanics.

6.2.12. Exercise to find Λ from F and \bar{F} given that $I_1(\bar{F}) = I_1(F)$, etc. See Exercise 1.6.17.

Suppose \mathbf{E} and \mathbf{B} are electric and magnetic fields having (at some point x^μ in space-time) arbitrary magnitude and direction.

6.2.13. Review the last paragraph of Exercise 2.6. The purpose of this exercise is to show that extended Lorentz/Poincaré transformations are symplectic maps and to study various features of the relation (1.6.287) that connects contravariant and covariant transformation properties. Hints: Show that the Poisson bracket relations (1.7.17) are preserved by (extended) Lorentz/Poincaré transformations. Show that the linear part M of the phase-space map is of the form (3.3.11) and that the condition (3.3.13) is satisfied. Given a Lorentz transformation, what is the f_2 for the corresponding symplectic map associated with the (extended) Lorentz transformation? Also see Exercise 3.7.36.

6.3 Preservation of General Poisson Brackets

Let \mathcal{M} be a symplectic mapping of z to \bar{z} , and let \mathcal{M}^{-1} be its inverse,

$$\mathcal{M} : z \rightarrow \bar{z} = \bar{z}(z, t), \quad (6.3.1)$$

$$\mathcal{M}^{-1} : \bar{z} \rightarrow z = z(\bar{z}, t). \quad (6.3.2)$$

Suppose we view these relations as a transformation of variables. Let $f(z, t)$ be any dynamical variable. Then the map \mathcal{M}^{-1} given by (3.2) produces a *transformed* dynamical variable $f^*(\bar{z}, t)$ (a function of the transformed phase-space variables \bar{z} and perhaps the time t) by the rule

$$\mathcal{M}^{-1} : f(z, t) \rightarrow f^*(\bar{z}, t) = f(z(\bar{z}, t), t). \quad (6.3.3)$$

Conversely, if $f^*(\bar{z}, t)$ is any function of the transformed phase-space variables \bar{z} and perhaps the time t , then the map \mathcal{M} given by (3.1) produces the dynamical variable $f(z, t)$, involving the original phase-space variables, by the rule

$$\mathcal{M} : f^*(\bar{z}, t) \rightarrow f(z, t) = f^*(\bar{z}(z, t), t). \quad (6.3.4)$$

In either case, we have the common relation

$$f^*(\bar{z}, t) = f(z, t). \quad (6.3.5)$$

The matter of transforming dynamical variables can also be viewed from a somewhat different perspective. Suppose $f^{\text{old}}(z, t)$ is some dynamical variable involving the original phase-space variables and perhaps the time t . Then the map \mathcal{M} given by (3.1) produces a *new* dynamical variable $f^{\text{new}}(z, t)$ of the *original* phase-space variables by the rule

$$\mathcal{M} : f^{\text{old}}(z, t) \rightarrow f^{\text{new}}(z, t) = f^{\text{old}}(\bar{z}(z, t), t). \quad (6.3.6)$$

In this case, \bar{z} is to be regarded as a transformed point in the same phase space as the original point z . By contrast, in the relations (3.3) through (3.5), the points \bar{z} and z may be regarded as members of two different phase spaces.

We now examine the relation between original and transformed dynamical variables and their Poisson brackets. Suppose f and g are any two dynamical variables. For clarity, it is convenient to introduce the notation

$$[f, g]_z = (\partial_z f, J \partial_z g), \quad (6.3.7)$$

$$[f, g]_{\bar{z}} = (\partial_{\bar{z}} f, J \partial_{\bar{z}} g). \quad (6.3.8)$$

See (5.1.4). Then the relation (1.15), and the fact that \mathcal{M}^{-1} is a symplectic map if (and only if) \mathcal{M} is also a symplectic map, can be written in the more precise form

$$J_{ab} = [\bar{z}_a, \bar{z}_b]_z = [z_a, z_b]_z = [\bar{z}_a, \bar{z}_b]_{\bar{z}} = [z_a, z_b]_{\bar{z}}. \quad (6.3.9)$$

Now consider $f^*(\bar{z}, t)$, and its counterpart $g^*(\bar{z}, t)$, as defined by (3.3). We claim that corresponding to the relations (3.3) and (3.4) there are the relations

$$[f^*, g^*]_{\bar{z}} = [f, g]_z|_{z=z(\bar{z}, t)}, \quad (6.3.10)$$

$$[f, g]_z = [f^*, g^*]_{\bar{z}}|_{\bar{z}=\bar{z}(z, t)}, \quad (6.3.11)$$

respectively. We will prove (3.10) in a moment, and the proof of (3.11) is similar. Note that the relations (3.10) and (3.11) indicate that the operations of transforming variables and Poisson bracketing are interchangeable. That is, we may first Poisson bracket two functions and then change variables, or we may first change variables, and then Poisson bracket with respect to the transformed variables. In this sense, Poisson brackets in general are preserved under symplectic maps.

The proof of (3.10) makes use of the chain rule and the symplectic condition. From (3.3) and (3.8) we have the relations

$$[f^*, g^*]_{\bar{z}} = (\partial_{\bar{z}} f^*, J \partial_{\bar{z}} g^*) = (\partial_{\bar{z}} f, J \partial_{\bar{z}} g). \quad (6.3.12)$$

By the chain rule there is the relation

$$\partial f / \partial \bar{z}_a = \sum_b (\partial f / \partial z_b) (\partial z_b / \partial \bar{z}_a). \quad (6.3.13)$$

However, (2.3) can be rewritten in the form

$$dz_b = \sum_c (M^{-1})_{bc} d\bar{z}_c, \quad (6.3.14)$$

and it follows that there is the relation

$$\partial z_b / \partial \bar{z}_a = (M^{-1})_{ba}. \quad (6.3.15)$$

Consequently, (3.13) can be written also in the form

$$\partial f / \partial \bar{z}_a = \sum_b [(M^T)^{-1}]_{ab} (\partial f / \partial z_b). \quad (6.3.16)$$

This relation has the compact form

$$\partial_{\bar{z}} f = (M^T)^{-1} \partial_z f. \quad (6.3.17)$$

Upon inserting (3.17) and its counterpart for g into (3.12), we find the advertised result,

$$\begin{aligned} [f^*, g^*]_{\bar{z}} &= (\partial_{\bar{z}} f, J \partial_{\bar{z}} g) = ([M^T]^{-1} \partial_z f, J [M^T]^{-1} \partial_z g) \\ &= (\partial_z f, [M^{-1} J (M^T)^{-1}] \partial_z g) = (\partial_z f, J \partial_z g) \\ &= [f, g]_z. \end{aligned} \quad (6.3.18)$$

Here we have used the symplectic condition for M^{-1} in the form

$$M^{-1} J (M^T)^{-1} = J. \quad (6.3.19)$$

The reader is urged to prove (3.11) in an analogous fashion. Finally, she or he should also prove the related result for *old* and *new* functions as given by (3.6),

$$[f^{\text{new}}(z, t), g^{\text{new}}(z, t)]_z = [f^{\text{old}}(\bar{z}, t), g^{\text{old}}(\bar{z}, t)]_{\bar{z}}|_{\bar{z}=\bar{z}(z, t)}. \quad (6.3.20)$$

Exercises

6.3.1. Prove the relations (3.11) and (3.20).

6.3.2. Suppose that $h(z, t)$ is any function of the phase-space variables z and the time t . Let z be related to \bar{z} by the symplectic map \mathcal{M}^{-1} as in (3.2). Prove the relation

$$[\bar{z}_a(z, t), h(z, t)]_z = [\bar{z}_a, h(z(\bar{z}, t), t)]_{\bar{z}}. \quad (6.3.21)$$

Prove also the relation

$$[z_a(\bar{z}, t), h(\bar{z}, t)]_{\bar{z}} = [z_a, h(\bar{z}(z, t), t)]_z. \quad (6.3.22)$$

Hint: Use (5.1.4), (1.2), and (3.15) to show that the left side of (3.21) can be written in the form

$$[\bar{z}_a(z, t), h(z, t)]_z = (M J \partial_z h)_a, \quad (6.3.23)$$

and the right side can be written in the form

$$[\bar{z}_a, h(z(\bar{z}, t), t)]_{\bar{z}} = (J(M^T)^{-1} \partial_z h)_a. \quad (6.3.24)$$

Then demonstrate and use the relation

$$M J = J(M^T)^{-1}, \quad (6.3.25)$$

which is a consequence of the symplectic condition. Alternatively, use the identity

$$\bar{z}_a^*(\bar{z}, t) = \bar{z}_a(z(\bar{z}, t), t) = \bar{z}_a \quad (6.3.26)$$

and the relation (3.11), etc.

6.4 Relation to Hamiltonian Flows

Let $H(z, t)$ be the Hamiltonian for some dynamical system. Consider a large Euclidean space with $2n + 1$ axes labeled by the phase-space variables $z_1 \cdots z_{2n}$ and the time t . We will call this construction *augmented phase space*.¹² See Figure 4.1. Suppose the $2n$ quantities $z_1(t^i) \cdots z_{2n}(t^i)$ are specified at some *initial* time t^i . Then the quantities $z_1(t) \cdots z_{2n}(t)$ at some other time t are uniquely determined by the initial conditions $z_1(t^i) \cdots z_{2n}(t^i)$ and Hamilton's equations of motion (5.2.2). Recall Theorem 1.3.1 and review Section 1.4. The set of all trajectories in augmented phase space for all possible initial conditions will be called a *Hamiltonian flow*. Indeed we know that no two trajectories can intersect. (See Exercise 1.3.6.) Therefore the behavior of the trajectories in augmented phase space is analogous to fluid flow in a high dimensional space.

6.4.1 Hamiltonian Flows Generate Symplectic Maps

Let t^i be some *initial* time, and let t^f be some other *final* time. Also, let z^i denote the set of quantities $z_1(t^i) \cdots z_{2n}(t^i)$, and let z^f denote the corresponding set $z_1(t^f) \cdots z_{2n}(t^f)$. We have already seen in Section 1.4 that the relation between z^i and z^f can be viewed as a transfer map $\mathcal{M}(t^i, t^f)$ depending on the parameters t^i and t^f . [Indeed, since the set of trajectories in augmented phase space is equivalent to a knowledge of $\mathcal{M}(t^i, t^f)$ for variable t^f , a flow for a differential equation may be equally well, and often is, *defined* to be the family of such maps.] What we will now see is that \mathcal{M} is a *symplectic* map.

Theorem 4.1 Let $H(z, t)$ be the Hamiltonian for some dynamical system, and let z^i denote a set of initial conditions at some initial time t^i . Also, let z^f denote the coordinates at some final time t^f of the trajectory with initial conditions z^i . Finally, let \mathcal{M} denote the mapping from z^i to z^f obtained by following the Hamiltonian trajectory specified by H ,

$$\mathcal{M} : z^i \rightarrow z^f. \quad (6.4.1)$$

Then the mapping \mathcal{M} is symplectic.

Proof Suppose the flow takes place for a time interval of duration T so that t^i and t^f are related by the equation

$$t^f = t^i + T. \quad (6.4.2)$$

Divide the interval T into N small steps each of duration h . Evidently, T, N , and h are related by the equation

$$T = Nh. \quad (6.4.3)$$

Also, define intermediate times t^m at each step by the rule

$$t^0 = t^i,$$

¹²Some authors, following Cartan, call it *state space*. But other authors use *state space* and *phase space* interchangeably. Still other authors call this construction *extended phase space*. However, we have already used that term to describe ordinary phase space augmented by the two additional variables t and p_t . Review Exercise 1.6.5.

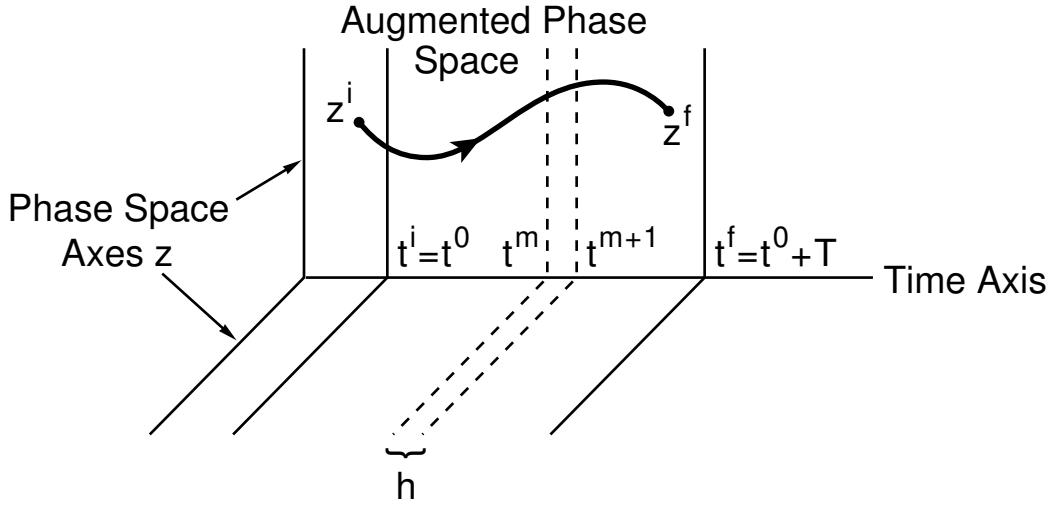


Figure 6.4.1: A trajectory in augmented phase space. Under the Hamiltonian flow specified by a Hamiltonian H , the general phase-space point z^i is mapped into the phase-space point z^f . The mapping \mathcal{M} is symplectic for any Hamiltonian.

$$\begin{aligned} t^m &= t^0 + mh \quad , \quad m = 0, 1, \dots, N, \\ t^N &= t^f. \end{aligned} \quad (6.4.4)$$

Suppose that the mapping \mathcal{M} is viewed as a composite of mappings between adjacent times t^m and t^{m+1} . That is, \mathcal{M} is written in the form

$$\mathcal{M} = \mathcal{M}^{t^f \leftarrow t^{N-1}} \dots \mathcal{M}^{t^{m+1} \leftarrow t^m} \dots \mathcal{M}^{t^1 \leftarrow t^i} \quad (6.4.5)$$

with the notation that $\mathcal{M}^{t^{m+1} \leftarrow t^m}$ denotes the mapping between the quantities

$$z^m = \{z_1(t^m), \dots, z_{2n}(t^m)\}$$

and

$$z^{m+1} = \{z_1(t^{m+1}), \dots, z_{2n}(t^{m+1})\}. \quad (6.4.6)$$

Corresponding to the relation (4.5), the Jacobian matrix M of the mapping \mathcal{M} can be written using the chain rule in the product form

$$M = M^{t^f \leftarrow t^{N-1}} \dots M^{t^{m+1} \leftarrow t^m} \dots M^{t^1 \leftarrow t^i}, \quad (6.4.7)$$

where, as the notation is meant to indicate, $M^{t^{m+1} \leftarrow t^m}$ is the Jacobian matrix for the map $\mathcal{M}^{t^{m+1} \leftarrow t^m}$.

Next it will be shown that each matrix in the product (4.7) is symplectic at least through terms of order h . According to Taylor's series, the relation between z^{m+1} and z^m can be written in the form

$$\begin{aligned} z_a^{m+1} &= z_a(t^{m+1}) = z_a(t^m + h) \\ &= z_a(t^m) + h\dot{z}_a(t^m) + O(h^2) \\ &= z_a^m + h(J\partial_z H)_a + O(h^2). \end{aligned} \quad (6.4.8)$$

Here use has also been made of the equations of motion (5.2.2). Suppose (4.8) is used to compute the associated Jacobian matrix. The result of this computation is the relation

$$\begin{aligned} M_{ab}^{t^{m+1} \leftarrow t^m} &= \partial z_a^{m+1} / \partial z_b^m \\ &= \delta_{ab} + h \sum_c J_{ac} \partial^2 H / \partial z_c \partial z_b + O(h^2). \end{aligned} \quad (6.4.9)$$

Using matrix notation, (4.9) can be written more compactly in the form

$$M^{t^{m+1} \leftarrow t^m} = I + hJS + O(h^2), \quad (6.4.10)$$

where S is the *symmetric* matrix

$$S_{cb} = \partial^2 H / \partial z_c \partial z_b. \quad (6.4.11)$$

Now compare (4.10) with (3.7.28) and (3.7.34). Evidently, the Jacobian matrix (4.10) is a *symplectic* matrix at least through terms of order h .

The desired proof is almost complete. Since symplectic matrices form a group, the product matrix M given by (4.7) differs from a symplectic matrix by terms at most of order Nh^2 because each of the N terms in the product differs from a symplectic matrix by terms at most of order h^2 . Now take the limit $h \rightarrow 0$ and $N \rightarrow \infty$. In this limit terms proportional to Nh^2 vanish since, using (4.3),

$$Nh^2 = (T/h)h^2 = Th, \quad (6.4.12)$$

and the quantity Th vanishes as h goes to zero. It follows that M is a symplectic matrix, and \mathcal{M} is a symplectic map.

What has been shown is that the problem of describing and following Hamiltonian trajectories, which is one of the fundamental aspects of classical mechanics, is equivalent to the problem of representing and calculating symplectic maps. We remark that there is another proof of the result just obtained based on the use of variational equations. It is shorter, but perhaps less instructive. See Exercise 4.3.

In the proof just given, suppose we regard the final time t^f as a general time t . What we have found is that following trajectories specified by H produces a symplectic map $\mathcal{M}(t^i, t)$ for each value of t . Thus, we have produced a *one-parameter family* of symplectic maps $\mathcal{M}(t^i, t)$. Moreover, we have the initial condition

$$\mathcal{M}(t^i, t^i) = \mathcal{I} \quad (6.4.13)$$

where \mathcal{I} denotes the identity map. We describe this state of affairs by saying that the family $\mathcal{M}(t^i, t)$ is *generated* by the Hamiltonian $H(z, t)$ starting from the identity map \mathcal{I} when $t = t^i$.

It can be verified that the set of symplectic maps generated by Hamiltonians forms a group which may be regarded as a subgroup of the set of all symplectic maps. This group is sometimes referred to as $Ham(n)$ where $2n$ is the dimensionality of the phase space under consideration.

6.4.2 Any Family of Symplectic Maps Is Hamiltonian Generated

Consider the space of all symplectic maps. We may regard the $\mathcal{M}(t^i, t)$, for variable t , as a path in this space. And, according to (4.13), the starting point of this path, namely $\mathcal{M}(t^i, t^i)$, is the identity map. Therefore the $\mathcal{M}(t^i, t)$ form a one-parameter family continuously connected to the identity. Is there a converse result? There is.

Theorem 4.2 Suppose we are given a one-parameter family of symplectic maps $\mathcal{N}(t)$ for $t \in [t^i, t^f]$. Let \mathcal{N}_i denote the map

$$\mathcal{N}_i = \mathcal{N}(t^i). \quad (6.4.14)$$

Then there is a *generating Hamiltonian* G that generates this family starting from the map \mathcal{N}_i .¹³ See Figure 4.2. It depicts *augmented symplectic map space*, which consists of a time axis and multiple additional axes that provide coordinates for points in the space of all symplectic maps. Let $\bar{z}(z, t)$ be the result of $\mathcal{N}(t)$ acting on the general phase-space point z ,

$$\mathcal{N}(t) : z \rightarrow \bar{z}(z, t). \quad (6.4.15)$$

What we want to show is that there is a function $G(\bar{z}; t)$ such that

$$(\partial \bar{z}_a / \partial t)|_z = [\bar{z}_a, G(\bar{z}; t)]_{\bar{z}}. \quad (6.4.16)$$

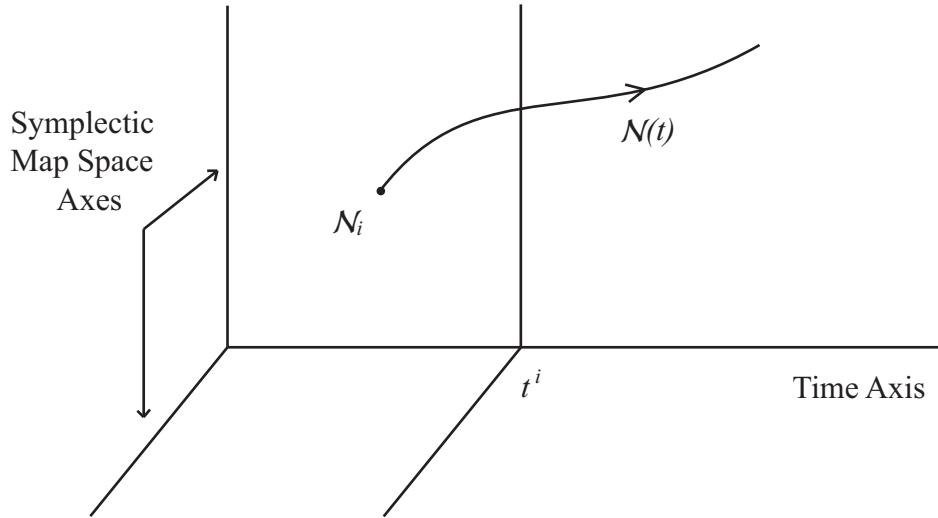


Figure 6.4.2: The symplectic map family $\mathcal{N}(t)$ in augmented symplectic map space.

Proof The proof proceeds by construction. Express the mapping (4.15) in the explicit component form

$$\bar{z}_a(t) = u_a(z, t) \quad (6.4.17)$$

¹³Here we apologize that the symbol G has also been used in Subsection 1.1, and again will be used subsequently, to denote the Jacobian of the gradient map \mathcal{G} . There are not always enough letters to go around. The reader should be able to determine from the context what is meant in any particular case.

where the u_a with $a = 1$ to $2n$ are assumed to be known functions of z and t . Next form the functions $w_a(z, t)$ defined by the relations

$$w_a(z, t) = \partial u_a(z, t)/\partial t. \quad (6.4.18)$$

By these definitions we have the equivalent statements

$$(\partial \bar{z}_a / \partial t)|_z = w_a(z, t). \quad (6.4.19)$$

Suppose that \mathcal{N} has an inverse, as will be the case if \mathcal{N} is symplectic. Then, the relations (4.15) or (4.17) can be inverted to give relations of the form

$$z_b = v_b(\bar{z}, t). \quad (6.4.20)$$

Now form the functions $g_a(\bar{z}, t)$ by using (4.20) in the arguments of the w_a and writing

$$g_a(\bar{z}, t) = w_a(z(\bar{z}, t), t). \quad (6.4.21)$$

The net result of these steps is the set of relations

$$\partial \bar{z}_a / \partial t = g_a(\bar{z}, t). \quad (6.4.22)$$

That is, we have produced a *vector field* $\mathcal{L}\mathbf{g}$ defined by the relation

$$\mathcal{L}\mathbf{g} = \sum_a g_a(\partial / \partial \bar{z}_a) \quad (6.4.23)$$

and having the property

$$\dot{\bar{z}}_a = \mathcal{L}\mathbf{g} \bar{z}_a. \quad (6.4.24)$$

We will now show that this vector field is Hamiltonian. (See Section 5.3.). Consider the quantities $\bar{z}_a(t + \epsilon)$ where ϵ is small. According to Taylor there is the expansion

$$\bar{z}_a(t + \epsilon) = \bar{z}_a(t) + \epsilon(\partial \bar{z}_a / \partial t) + O(\epsilon^2) \quad (6.4.25)$$

or, in view of (4.19),

$$\bar{z}_a(t + \epsilon) = \bar{z}_a(t) + \epsilon w_a(z, t) + O(\epsilon^2). \quad (6.4.26)$$

Let us compute $[\bar{z}_a(t + \epsilon), \bar{z}_b(t + \epsilon)]$. Using (4.26) we find the result

$$[\bar{z}_a(t + \epsilon), \bar{z}_b(t + \epsilon)]_z = [\bar{z}_a(t), \bar{z}_b(t)]_z + \epsilon [\bar{z}_a, w_b]_z + \epsilon [w_a, \bar{z}_b]_z + O(\epsilon^2). \quad (6.4.27)$$

Since \mathcal{N} is symplectic for all $t \in [t^i, t^f]$, there must be the relations

$$[\bar{z}_a(t + \epsilon), \bar{z}_b(t + \epsilon)]_z = [\bar{z}_a(t), \bar{z}_b(t)]_z = J_{ab}. \quad (6.4.28)$$

See (1.15). Therefore, upon equating powers of ϵ , (4.27) provides the relation

$$[\bar{z}_a, w_b]_z + [w_a, \bar{z}_b]_z = 0. \quad (6.4.29)$$

Since symplectic maps preserve Poisson brackets, see Section 3, we may also write (4.29) in the form

$$[\bar{z}_a, g_b(\bar{z}, t)]_{\bar{z}} = [\bar{z}_b, g_a(\bar{z}, t)]_{\bar{z}}. \quad (6.4.30)$$

Here we have used (4.21) and the antisymmetry of the Poisson bracket. Finally, expand out the Poisson brackets using (5.1.3). So doing, for example, gives the result

$$\begin{aligned} [\bar{z}_a, g_b]_{\bar{z}} &= \sum_{cd} (\partial \bar{z}_a / \partial \bar{z}_c) J_{cd} (\partial g_b / \partial \bar{z}_d) \\ &= \sum_{cd} \delta_{ac} J_{cd} (\partial g_b / \partial \bar{z}_d) = \sum_d J_{ad} (\partial g_b / \partial \bar{z}_d). \end{aligned} \quad (6.4.31)$$

The net result is that (4.30) is equivalent to the relation

$$\sum_d J_{ad} (\partial g_b / \partial \bar{z}_d) = \sum_d J_{bd} (\partial g_a / \partial \bar{z}_d). \quad (6.4.32)$$

To make sense of (4.32), multiply both sides by $J_{ac} J_{be}$, sum over a and b , and manipulate to produce the relations

$$\sum_{abd} J_{ac} J_{be} J_{ad} (\partial g_b / \partial \bar{z}_d) = \sum_{abd} J_{ac} J_{be} J_{bd} (\partial g_a / \partial \bar{z}_d), \quad (6.4.33)$$

$$\sum_{abd} (J^T)_{ca} J_{ad} J_{be} (\partial g_b / \partial \bar{z}_d) = \sum_{abd} (J^T)_{eb} J_{bd} J_{ac} (\partial g_a / \partial \bar{z}_d), \quad (6.4.34)$$

$$\sum_{bd} (J^T J)_{cd} J_{be} (\partial g_b / \partial \bar{z}_d) = \sum_{ad} (J^T J)_{ed} J_{ac} (\partial g_a / \partial \bar{z}_d), \quad (6.4.35)$$

$$\sum_{bd} \delta_{cd} J_{be} (\partial g_b / \partial \bar{z}_d) = \sum_{ad} \delta_{ed} J_{ac} (\partial g_a / \partial \bar{z}_d), \quad (6.4.36)$$

$$\sum_b J_{be} (\partial g_b / \partial \bar{z}_c) = \sum_a J_{ac} (\partial g_a / \partial \bar{z}_e), \quad (6.4.37)$$

$$(\partial / \partial \bar{z}_c) \sum_b J_{be} g_b = (\partial / \partial \bar{z}_e) \sum_a J_{ac} g_a. \quad (6.4.38)$$

Here use has been made of (3.1.6). Now introduce quantities η_c by the rule

$$\eta_c = \sum_a J_{ac} g_a. \quad (6.4.39)$$

In terms of these quantities (4.38) yields the relations

$$\partial \eta_e / \partial \bar{z}_c = \partial \eta_c / \partial \bar{z}_e. \quad (6.4.40)$$

[Note that this condition is a restatement of (5.3.26).] It follows that the quantity $\sum_a \eta_a d\bar{z}_a$ is an *exact* differential. See Exercise 1.1.

Now define G by the phase-space path integral

$$G(\bar{z}; t) = \int^{\bar{z}} \sum_a \eta_a dz'_a. \quad (6.4.41)$$

Since the integrand is an exact differential, the integral is path independent and satisfies the relation

$$\partial G / \partial \bar{z}_a = \eta_a. \quad (6.4.42)$$

Let us put everything together. The relations (4.39) can be solved for the g_a to give the result

$$g_a = \sum_b J_{ab} \eta_b. \quad (6.4.43)$$

Upon combining (4.22), (4.42), and (4.43), we find the net result

$$\partial \bar{z}_a / \partial t = \sum_b J_{ab} \partial G / \partial \bar{z}_b, \quad (6.4.44)$$

or, more compactly,

$$\partial \bar{z} / \partial t = J \partial_{\bar{z}} G = [\bar{z}, G(\bar{z}; t)], \quad (6.4.45)$$

which is the desired result (4.16).

Even a bit more can be said. Consider the straight-line path $\bar{z}'_a(\tau)$ in phase space that connects the origin to \bar{z} . It has the parametric form

$$\bar{z}'_a(\tau) = \tau \bar{z}_a, \quad \tau \in [0, 1]. \quad (6.4.46)$$

Suppose the integrability condition (4.40) holds in a simply-connected region that surrounds this path. Then we may employ this path in (4.41) to obtain the result

$$G(\bar{z}; t) = \int_0^1 d\tau \sum_a \eta_a(\tau \bar{z}, t) \bar{z}_a = - \int_0^1 d\tau \sum_{ab} J_{ab} \bar{z}_a g_b(\tau \bar{z}, t). \quad (6.4.47)$$

Let us recapitulate what has been done. By differentiating the map $\mathcal{N}(t)$ with respect to t , the steps involved in (4.17) through (4.21) produced the vector field $\mathcal{L}_{\mathbf{g}}$ with components g_a . These steps can be carried out for any invertible one-parameter family of maps $\mathcal{N}(t)$. Then we used the symplectic condition in the steps associated with (4.25) through (4.41) to show that the vector field was Hamiltonian and to explicitly construct the Hamiltonian.

There are a few more steps that can be made. In analogy to the work of Subsection 4.1, let $\mathcal{M}(t^i, t)$ be the map generated by the $G(z; t)$ constructed in this subsection and with the initial condition (4.13). Then, using the methods of Section 10.1, it can be verified that there is the relation

$$\mathcal{N}(t) = \mathcal{N}_i \mathcal{M}(t^i, t). \quad (6.4.48)$$

We have found an explicit expression for \mathcal{N} in terms of the map generated by its associated Hamiltonian. Moreover, differentiating (4.48), and again using the methods of Section 10.1, gives the result

$$\dot{\mathcal{N}}(t) = \mathcal{N}_i \dot{\mathcal{M}}(t^i, t) = \mathcal{N}_i \mathcal{M}(t^i, t) : -G := \mathcal{N}(t) : -G :, \quad (6.4.49)$$

from which we conclude that

$$: -G := \mathcal{N}^{-1} \dot{\mathcal{N}}(t). \quad (6.4.50)$$

In summary, given any family of symplectic maps, we have shown that this family is generated by a Hamiltonian G starting from an initial map \mathcal{N}_i , have explicitly constructed G , and have found a representation for \mathcal{N} .

6.4.3 Almost All Symplectic Maps Are Hamiltonian Generated

In Section 7.9 we will learn that under rather general low-order differentiability conditions any symplectic map can be connected to the identity map by a one-parameter family of symplectic maps. This one-parameter family can then be used to construct a Hamiltonian, and we can also set

$$\mathcal{N}_i = \mathcal{I}. \quad (6.4.51)$$

We conclude that, under mild assumptions, any symplectic map can be generated by a Hamiltonian starting from the identity map. Thus, for our purposes, we will generally not make a distinction between $ISpM(2n, \mathbb{R}) [= Symp(n)]$ and $Ham(n)$.

6.4.4 Transformation of a Hamiltonian Under the Action of a Symplectic Map

Suppose $H(z; t)$ is the Hamiltonian governing the motion of some system described by the canonical coordinates z . Suppose we wish to introduce new canonical coordinates $\bar{z}(z, t)$ that are related to the old coordinates z by a (possibly time dependent) symplectic map $\mathcal{N}(t)$ as in (4.15). The purpose of this subsection is to show that the motion of the system, when described by the new coordinates \bar{z} , is also governed by a new Hamiltonian that we will call $K(\bar{z}; t)$, and to find the relation between K and the old Hamiltonian H .

We proceed as follows: View the quantities $\bar{z}_a(z, t)$ as dynamical variables. Then, by the work of Section 1.7 that defined the Poisson bracket, there is the result

$$d\bar{z}_a(z, t)/dt = \partial\bar{z}_a(z, t)/\partial t + [\bar{z}_a(z, t), H(z; t)]_z. \quad (6.4.52)$$

Here we have placed a subscript z on the Poisson bracket to make it clear that the Poisson bracket is taken with respect to the variable z . But, since \bar{z} and z are related by the *symplectic* map \mathcal{N} , it follows from the invariance property of the Poisson bracket that

$$[\bar{z}_a(z, t), H(z; t)]_z = [\bar{z}_a, H(z(\bar{z}, t); t)]_{\bar{z}}. \quad (6.4.53)$$

See Section 3. Moreover, from the work of Subsection 4.2, we know that there is a generating Hamiltonian $G(\bar{z}; t)$ for \mathcal{N} such that

$$\partial\bar{z}_a(z, t)/\partial t = [\bar{z}_a, G(\bar{z}; t)]_{\bar{z}}. \quad (6.4.54)$$

Upon combining (4.52) through (4.54) we see that there is the relation

$$d\bar{z}_a/dt = [\bar{z}_a, K(\bar{z}; t)]_{\bar{z}} \quad (6.4.55)$$

where

$$K(\bar{z}; t) = H(z(\bar{z}, t); t) + G(\bar{z}; t) \quad (6.4.56)$$

so that $K(\bar{z}; t)$ is the desired new Hamiltonian. We note that if \mathcal{N} is time independent, then $G = 0$. See (4.50). In this case the new Hamiltonian is simply the old Hamiltonian expressed in terms of the new variables,

$$K(\bar{z}; t) = H(z(\bar{z}); t). \quad (6.4.57)$$

Exercises

6.4.1. Use (1.2) and the chain rule to verify (4.7).

6.4.2. Show that the matrix S defined by (4.11) is indeed symmetric.

6.4.3. The purpose of this exercise is to provide another proof of Theorem 4.1: Hamiltonian flows generate symplectic maps. Refer to Exercise 1.4.6. From Hamilton's equations of motion written in the form (5.2.3), show, for the associated variational equations, that the A matrix of (1.4.51) is given by

$$A = JS \quad (6.4.58)$$

with S given by (4.11). Next show that in terms of a general final time t the Jacobian matrix M satisfies the differential equation

$$\dot{M}(t) = JS(t)M(t) \quad (6.4.59)$$

with the initial condition

$$M(t^i) = I. \quad (6.4.60)$$

Now consider the matrix product M^TJM . Because of (4.59), it satisfies the differential equation

$$\begin{aligned} (d/dt)[M^T(t)JM(t)] &= \dot{M}^TJM + M^TJ\dot{M} \\ &= [JSM]^TJM + M^TJJSM = -M^TSJ JM + M^TJJSM \\ &= M^TSM - M^TSM = 0. \end{aligned} \quad (6.4.61)$$

Thus, in view of (4.60), this equation has the unique solution

$$M^T(t)JM(t) = J, \quad (6.4.62)$$

and we conclude that the Jacobian matrix must be symplectic.

6.4.4. Show that the maps between q, p and Q, P given by (1.4.9) is symplectic. Show that the map given by (1.4.13) between Q^i, P^i and Q^f, P^f is symplectic. In both cases, find the associated Jacobian matrix M and verify that it is symplectic.

6.4.5. Show that the maps given by (1.4.22), (1.4.23) and (1.4.24), (1.4.25) are symplectic. In both cases, find the associated Jacobian matrix M and verify that it is symplectic.

6.4.6. Suppose $H(z, t)$ is a possibly time-dependent quadratic Hamiltonian written, without loss of generality, in the form

$$H(z, t) = (1/2)(z, Sz) \quad (6.4.63)$$

where S is a symmetric and possibly time-dependent matrix. Verify that the equations of motion generated by this H are linear, and therefore that the associated transfer map is linear and can be described by a matrix M . Use the machinery of Exercise 4.3 above to show that $M(t, t_0)$ is symplectic. Here t is a general time and t_0 is some initial time such that

$$M(t_0, t_0) = I. \quad (6.4.64)$$

Let u_0 and v_0 be two initial conditions. Then, for these initial conditions, verify that the associated solutions to the equations of motion are given by the relations

$$u(t) = M(t, t_0)u_0, \quad (6.4.65)$$

$$v(t) = M(t, t_0)v_0. \quad (6.4.66)$$

Form the quantity $C(u, v)$ by the rule

$$C(u, v) = (u, Jv). \quad (6.4.67)$$

Show that

$$C(u, v) = C(u_0, v_0), \quad (6.4.68)$$

and therefore C is *constant* (time independent) and depends only on the initial conditions.

Consider the set of differential equations arising from any Hamiltonian and, for any particular trajectory, form the associated variational equations. Show that any two solutions u and v to the variational equations also satisfy (4.68).

6.4.7. Consider the one-parameter family of maps

$$\bar{z}_1(z, t) = z_1 \cos t - z_2 \sin t, \quad (6.4.69)$$

$$\bar{z}_2(z, t) = z_1 \sin t + z_2 \cos t. \quad (6.4.70)$$

Verify that these maps are symplectic. Find the Hamiltonian that generates this family of maps.

6.4.8. Consider the two-parameter family of maps (called the general Hénon map, see Section 19.7) given by the relations

$$\bar{q} = 1 + p - aq^2, \quad (6.4.71)$$

$$\bar{p} = bq. \quad (6.4.72)$$

Show that the inverse of this map is given by the relations

$$q = \bar{p}/b, \quad (6.4.73)$$

$$p = \bar{q} - 1 + a(\bar{p}/b)^2. \quad (6.4.74)$$

Show that if b is held fixed and a is treated as a variable parameter, then the resulting one-parameter family of maps is generated by the vector field

$$\mathcal{L} = -(\bar{p}/b)^2(\partial/\partial\bar{q}) =: [1/(3b^2)]\bar{p}^3 : . \quad (6.4.75)$$

Note that this vector field is Hamiltonian even though the general Hénon map is symplectic only when $b = -1$. Show that if a is held fixed and b is treated as a variable parameter, then the resulting one-parameter family of maps is generated by the vector field

$$\mathcal{L} = (1/b)\bar{p}(\partial/\partial\bar{p}). \quad (6.4.76)$$

This vector field is not Hamiltonian. See Section 18.3.

6.4.9. Newton's equation of motion for an harmonic oscillator consisting of a mass m and a spring with spring constant k is given by the relation

$$d^2x/dt^2 + (k/m)x = 0 \quad (6.4.77)$$

where x is the difference between the actual and natural lengths of the spring. Introduce the notation

$$K = k/m, \quad (6.4.78)$$

and consider the possibility that K is time dependent so that (4.77) becomes

$$d^2x/dt^2 + K(t)x = 0. \quad (6.4.79)$$

If K is in fact time dependent, the harmonic oscillator is said to be *parametrically driven*.

The purpose of this exercise is to explore some aspects of the behavior of a parametrically driven harmonic oscillator. The behavior of a parametrically driven harmonic oscillator can be very complicated, and there is a vast literature on the subject. If K is *periodic*, (4.79) is a form of *Hill's equation*. If K is periodic and consists of only a constant term and a square wave, (4.79) becomes *Meissner's equation*. If K is periodic and consists of only a constant term and a rectangular wave, (4.79) becomes the *Kronig-Penney model*. If K is periodic and consists of only a constant term and a string of equally spaced delta function spikes, (4.79) becomes the *Dirac comb* or *periodic delta function model*. If K is periodic and consists of only a constant term and a sinusoidal term, (4.79) becomes a form of *Mathieu's equation*. Mathieu functions will play an important role in Section 17.4. The general periodic case, in essence Hill's equation, is important for the subject of *strong focussing* in Accelerator Physics. It is also important for many other areas of physics including band theory in Condensed Matter Physics, the motion of the Moon (the context in which Hill formulated and studied his equation), and wave-guide theory.¹⁴

¹⁴ It is interesting to note that George William Hill (1838-1914) did not hold any permanent academic appointment. For ten years of his life he was a clerk at the U. S. National Bureau of Standards (NBS, now NIST, the National Institute of Standards and Technology) working long hours and doing his own research at home at night. Much of the rest of his life was spent working only at home. He went unappreciated by his colleagues for many years. When Poincaré (along with Darboux, Picard, and Boltzmann) visited the United States in 1904 to lecture at the St. Louis Mathematics Congress held in connection with St. Louis

Your first task is to show that the equation of motion (4.79) arises from a Hamiltonian. Define p by the rule

$$p = dx/dt. \quad (6.4.80)$$

Show that (4.79) and (4.80) are generated by the Hamiltonian

$$H = (1/2)[p^2 + K(t)x^2] \quad (6.4.81)$$

where p and x are taken to be canonically conjugate.

Consider, as a specific example, the Mathieu case for which the Fourier series for $K(t)$ only has two terms,

$$K(t) = K_0 + K_1 \cos(\Omega t + \phi). \quad (6.4.82)$$

In this case (4.79) takes the form

$$d^2x/dt^2 + [K_0 + K_1 \cos(\Omega t + \phi)]x = 0. \quad (6.4.83)$$

The quantity K_0 describes the natural frequency of the oscillator,

$$\omega = \sqrt{K_0}, \quad (6.4.84)$$

where it is assumed that $K_0 > 0$, and K_1 describes the parametric driving strength. Compare (4.83) with the standard form of the Mathieu equation given by (17.4.22). Make the change of variable

$$\tau = \Omega t + \phi, \quad (6.4.85)$$

and verify that this change of variable brings (4.83) to the form

$$d^2x/d\tau^2 + [\bar{K}_0 + \bar{K}_1 \cos(\tau)]x = 0 \quad (6.4.86)$$

where

$$\bar{K}_0 = K_0/\Omega^2 = \omega^2/\Omega^2, \quad (6.4.87)$$

$$\bar{K}_1 = K_1/\Omega^2. \quad (6.4.88)$$

In the case that $K(t)$ is periodic, the solution to (4.79), and hence also to (4.86), can be described in terms of a stroboscopic map. See Section 1.4.3. Moreover, since the equations of motion are linear and are generated by a Hamiltonian, namely (4.81), the stroboscopic map will be linear and symplectic. See Exercise 4.6 above. Introduce the notation

$$z = (x, p). \quad (6.4.89)$$

hosting a World's Fair, the one American mathematician he sought out was Hill. After the congress these four foreign speakers boarded a train to Washington D.C. (where NBS was located) to visit Hill and attend a reception hosted by President Theodore Roosevelt, followed by subsequent stops at Harvard and Columbia Universities, before sailing back to Europe. Ernest Brown, in his 1915 National Academy of Sciences Biographical Memoir of Hill, wrote "Hill's 1877 publication 'Researches in the Lunar Theory' of but fifty quarto pages has become fundamental for the development of celestial mechanics in three different directions. It would be difficult to say as much for any other publication of its length in the whole range of modern mathematics, pure or applied. Poincaré's remark that in it we may perceive the germ of all the progress which has been made in celestial mechanics since its publication is doubtless fully justified".

Let z^i be the initial condition at the beginning of a drive period ($\tau = 0$) and let z^f be the final condition at the end of a drive period ($\tau = 2\pi$). Then we may write, in the case of the equation of motion (4.86), the relation

$$z^f = Mz^i \quad (6.4.90)$$

where M is a 2×2 symplectic matrix to be determined. In writing (4.90), because of the variable change (4.85), we take the associated Hamiltonian to be that given by (4.81) with K replaced by \bar{K} where $\bar{K} = K/\Omega^2$. Also, (4.80) is replaced by $p = dx/d\tau$.

In the case that $\bar{K}_1 = 0$, the equation of motion (4.86) can be solved in terms of trigonometric functions. Show that in this case the matrix M , which describes the stroboscopic map, is given by the relation

$$M = \begin{pmatrix} \cos 2\pi\bar{\omega} & (1/\bar{\omega}) \sin 2\pi\bar{\omega} \\ -\bar{\omega} \sin 2\pi\bar{\omega} & \cos 2\pi\bar{\omega} \end{pmatrix} \quad (6.4.91)$$

where

$$\bar{\omega} = \sqrt{\bar{K}_0} = \omega/\Omega. \quad (6.4.92)$$

Verify that the eigenvalues of M lie on the unit circle and have the values

$$\lambda_{\pm} = \exp(\pm 2\pi i\bar{\omega}). \quad (6.4.93)$$

Now suppose that \bar{K}_1 takes on small nonzero values. Then the eigenvalues of M will remain on the unit circle provided they were originally not too close to the values ± 1 . On the other hand, they could leave the unit circle if originally they were close to or had the values ± 1 . Recall Figures 3.4.1 and 3.4.3. The eigenvalues have the value $+1$ when

$$2\pi\bar{\omega} = 2n\pi \iff \bar{\omega} = n, \quad (6.4.94)$$

and have the value -1 when

$$2\pi\bar{\omega} = \pi + 2n\pi \iff \bar{\omega} = n + 1/2. \quad (6.4.95)$$

Here $n = 0, 1, 2, \dots$. Finally, verify that combining (4.92), (4.94), and (4.95) yields the conditions

$$\omega = n\Omega \text{ or } \Omega = \omega/n \text{ with } n = 1, 2, \dots, \quad (6.4.96)$$

$$\omega = (n + 1/2)\Omega \text{ or } \Omega = \omega/(n + 1/2) \text{ with } n = 0, 1, 2, \dots. \quad (6.4.97)$$

Note that in (4.96) we have excluded the case $n = 0$ since the case $\omega = 0$ requires more refined analysis.

When its eigenvalues are off the unit circle, repeated application of the matrix M leads to exponential growth. See Subsections 3.4.5 and 3.5.8. Verify that the conditions (4.96) and (4.97) for possible instability can be combined to yield the *parametric resonance* conditions

$$\Omega = 2\omega/m \Leftrightarrow \omega = (m/2)\Omega \Leftrightarrow 1/\Omega = m(1/2)(1/\omega) \text{ with } m = 1, 2, \dots. \quad (6.4.98)$$

Verify in this latter formulation that odd values of m correspond to the possibility of the eigenvalues leaving the unit circle through the value -1 , and even values of m correspond to the possibility of the eigenvalues leaving the unit circle through the value $+1$.

A pendulum of length ℓ in a gravitational field g has a small-amplitude natural frequency $\omega = (g/\ell)^{1/2}$. Show that, according to (4.98), the trapeze artist Jules Léotard (1838–1870) could increase the amplitude of his swing by alternatively crouching down and then standing up with frequency $\Omega = 2\omega$.¹⁵ Also, like a child on a swing, he could do so with frequency $\Omega = \omega$. Remarkably, he could also do so with the subharmonic frequencies $\Omega = (2/3)\omega$, $\Omega = (2/4)\omega$, \dots . The first choice $\Omega = 2\omega$ is used by professionals and is the most effective. We know from childhood experience with pumping swings that the second choice also works pretty well, and is easier for mortals. The other choices produce successively slower amplitude growths.

6.4.10. The purpose of this exercise is to explore the difference between *forcefully* and parametrically driven harmonic oscillators. Review Exercise 4.9 above. By forcefully driven we mean an oscillator described by an equation of motion of the form

$$\frac{d^2x}{dt^2} + \beta \frac{dx}{dt} + \omega^2 x = d \cos(\Omega t + \psi). \quad (6.4.99)$$

When $\beta > 0$ the motion of this oscillator is bounded for all values of Ω . It is also bounded when $\beta = 0$ provided $\Omega \neq \omega$. See Section 28.2. Verify, when $\beta = 0$ and $\Omega = \omega$, that (4.99) has the solution

$$x(t) = [d/(2\omega)]t \sin(\omega t + \psi). \quad (6.4.100)$$

Thus, *exactly* at resonance and in the absence of damping, the amplitude of a forcefully driven harmonic oscillator grows *linearly* in time. By contrast it can be shown, in accord with the results of Exercise 4.9, that the amplitude of a parametrically driven oscillator described by the Mathieu equation grows *exponentially* in time when K_1 is small and any of the parametric resonance conditions (4.98) is approximately satisfied.

Moreover, even if the parametrically driven oscillator is damped by adding a term of the form $\beta dx/dt$ (with $\beta > 0$) to the left side of (4.83), it can be shown that there is still a range of K_1 and Ω values for which the amplitude grows exponentially in time. Thus, parametric driving can overcome damping, and can do so even for a range of K_1 and Ω values. That is, there is no resonance condition that needs to be met exactly. Rather, there is a whole band of parameter values for which there is exponential growth. Alternatively, suppose the parametrically driven oscillator is *anti-damped* by adding a term of the form $\beta dx/dt$ with $\beta < 0$ to the left side of (4.83), or suppose $K_0 < 0$. Then the solution would grow exponentially when $K_1 = 0$. However, there is now a range of K_1 and Ω values for which parametric driving can *stabilize* the oscillator. That is, when $\beta < 0$ or $K_0 < 0$, it can be shown that there is a range of K_1 and Ω values for which $x(t)$ is nevertheless bounded.

Finally we remark, as seems plausible from the arguments made in Exercise 4.9, that the behavior we have found/claimed for the Mathieu case will occur quite generally for other cases of Hill's equation.

6.4.11. Consider the motion of a charged particle in an electromagnetic field. With time as the independent variable, suppose one integrates the first-order set of differential equations

¹⁵“He'd fly through the air with the greatest of ease, a daring young man on the flying trapeze. His movements were graceful, all girls he could please. And my love he purloined away”. The *leotard* garment is named after Léotard who invented and first wore it in his performances.

(1.6.69) and (1.6.70) for the quantities \mathbf{r} and \mathbf{p} from $t = t^{in}$ to $t = t^{fin}$. Recall that here the quantity \mathbf{p} is the *mechanical* momentum. Therefore, to be more precise, we will use the notation \mathbf{r}^{mech} , \mathbf{p}^{mech} and \mathbf{r}^{can} , \mathbf{p}^{can} to refer to mechanical and canonical quantities, respectively. With this notation in mind, is the relation between the initial conditions $(\mathbf{r}^{\text{mech}})^{in}$, $(\mathbf{p}^{\text{mech}})^{in}$ and the final conditions $(\mathbf{r}^{\text{mech}})^{fin}$, $(\mathbf{p}^{\text{mech}})^{fin}$ a symplectic map? You are to show that in general the answer is *no*.

More precisely, let $d(\mathbf{r}^{\text{mech}})^{in}$, $d(\mathbf{p}^{\text{mech}})^{in}$ denote small changes in the initial conditions, and let $d(\mathbf{r}^{\text{mech}})^{fin}$, $d(\mathbf{p}^{\text{mech}})^{fin}$ be the corresponding changes in the final conditions. By definition they are connected by the Jacobian matrix relation

$$\begin{pmatrix} d(\mathbf{r}^{\text{mech}})^{fin} \\ d(\mathbf{p}^{\text{mech}})^{fin} \end{pmatrix} = N \begin{pmatrix} d(\mathbf{r}^{\text{mech}})^{in} \\ d(\mathbf{p}^{\text{mech}})^{in} \end{pmatrix}. \quad (6.4.101)$$

Your task is to show that in general N is *not* a symplectic matrix.

The mechanical and canonical quantities are connected by the relations

$$\mathbf{r}^{\text{mech}} = \mathbf{r}^{\text{can}}, \quad (6.4.102)$$

$$\mathbf{p}^{\text{mech}} = \mathbf{p}^{\text{can}} - q\mathbf{A}. \quad (6.4.103)$$

Recall (1.5.30). Use (4.102) and (4.103) to obtain the relations

$$d(\mathbf{r}^{\text{mech}})^{in} = d(\mathbf{r}^{\text{can}})^{in}, \quad (6.4.104)$$

$$d(\mathbf{r}^{\text{mech}})^{fin} = d(\mathbf{r}^{\text{can}})^{fin}, \quad (6.4.105)$$

$$d(p_j^{\text{mech}})^{in} = d(p_j^{\text{can}})^{in} - q \sum_k A_{j,k}[(\mathbf{r}^{\text{can}})^{in}, t^{in}]d(x_k^{\text{can}})^{in}, \quad (6.4.106)$$

$$d(p_j^{\text{mech}})^{fin} = d(p_j^{\text{can}})^{fin} - q \sum_k A_{j,k}[(\mathbf{r}^{\text{can}})^{fin}, t^{fin}]d(x_k^{\text{can}})^{fin}, \quad (6.4.107)$$

where

$$A_{j,k}(\mathbf{r}^{\text{can}}, t) = \partial A_j(\mathbf{r}^{\text{can}}, t) / \partial x_k^{\text{can}}. \quad (6.4.108)$$

Next verify that (4.104), (4.106) and (4.105), (4.107) can be written in the more compact matrix form

$$\begin{pmatrix} d(\mathbf{r}^{\text{mech}})^{in} \\ d(\mathbf{p}^{\text{mech}})^{in} \end{pmatrix} = V^{\text{in}} \begin{pmatrix} d(\mathbf{r}^{\text{can}})^{in} \\ d(\mathbf{p}^{\text{can}})^{in} \end{pmatrix}, \quad (6.4.109)$$

$$\begin{pmatrix} d(\mathbf{r}^{\text{mech}})^{fin} \\ d(\mathbf{p}^{\text{mech}})^{fin} \end{pmatrix} = V^{\text{fin}} \begin{pmatrix} d(\mathbf{r}^{\text{can}})^{fin} \\ d(\mathbf{p}^{\text{can}})^{fin} \end{pmatrix}, \quad (6.4.110)$$

where V is the matrix

$$V(\mathbf{r}^{\text{can}}, t) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -qA_{1,1} & -qA_{1,2} & -qA_{1,3} & 1 & 0 & 0 \\ -qA_{2,1} & -qA_{2,2} & -qA_{2,3} & 0 & 1 & 0 \\ -qA_{3,1} & -qA_{3,2} & -qA_{3,3} & 0 & 0 & 1 \end{pmatrix}, \quad (6.4.111)$$

and V^{in} and V^{fin} are the matrices

$$V^{in} = V[(\mathbf{r}^{\text{can}})^{in}, t^{in}], \quad (6.4.112)$$

$$V^{fin} = V[(\mathbf{r}^{\text{can}})^{fin}, t^{fin}]. \quad (6.4.113)$$

Finally, explain why there is a relation of the form

$$\begin{pmatrix} d(\mathbf{r}^{\text{can}})^{fin} \\ d(\mathbf{p}^{\text{can}})^{fin} \end{pmatrix} = M \begin{pmatrix} d(\mathbf{r}^{\text{can}})^{in} \\ d(\mathbf{p}^{\text{can}})^{in} \end{pmatrix} \quad (6.4.114)$$

where M is a symplectic matrix. Recall the Hamiltonian (1.5.31).

We are now ready for some matrix manipulation. For calculational convenience write (4.111) in the block form

$$V = \begin{pmatrix} I & 0 \\ C & I \end{pmatrix} \quad (6.4.115)$$

where C is the matrix with entries

$$C_{jk} = -qA_{j,k}. \quad (6.4.116)$$

Verify that V is invertible, and its inverse is given by the formula

$$V^{-1} = \begin{pmatrix} I & 0 \\ -C & I \end{pmatrix}. \quad (6.4.117)$$

Verify that (4.101), (4.109), and (4.110) can be combined to yield the relations

$$V^{fin} \begin{pmatrix} d(\mathbf{r}^{\text{can}})^{fin} \\ d(\mathbf{p}^{\text{can}})^{fin} \end{pmatrix} = NV^{in} \begin{pmatrix} d(\mathbf{r}^{\text{can}})^{in} \\ d(\mathbf{p}^{\text{can}})^{in} \end{pmatrix}, \quad (6.4.118)$$

or, equivalently,

$$\begin{pmatrix} d(\mathbf{r}^{\text{can}})^{fin} \\ d(\mathbf{p}^{\text{can}})^{fin} \end{pmatrix} = (V^{fin})^{-1}NV^{in} \begin{pmatrix} d(\mathbf{r}^{\text{can}})^{in} \\ d(\mathbf{p}^{\text{can}})^{in} \end{pmatrix}. \quad (6.4.119)$$

Verify that comparison of (4.114) and (4.119) yields the matrix relations

$$(V^{fin})^{-1}NV^{in} = M, \quad (6.4.120)$$

or, equivalently,

$$N = V^{fin}M(V^{in})^{-1}. \quad (6.4.121)$$

You are ready for the final steps. Begin by showing that in general V is not symplectic. In particular verify, using the results of Section 3.3.2, that the condition for V to be symplectic is that

$$C - C^T = 0. \quad (6.4.122)$$

Show that in fact for the present case there is the result

$$C - C^T = q\mathbf{B} \cdot \mathbf{L} \quad (6.4.123)$$

where $\mathbf{B}(\mathbf{r}^{\text{can}}, t)$ is the magnetic field and \mathbf{L} denotes the collection of matrices given by (3.7.177) through (3.7.179). We see that V is not symplectic unless the magnetic field vanishes, which should not be too surprising in view of (1.7.18). Show that in fact V satisfies the relation

$$V^T JV = J \begin{pmatrix} I & 0 \\ q\mathbf{B} \cdot \mathbf{L} & I \end{pmatrix}. \quad (6.4.124)$$

Verify that the product of a symplectic and a nonsymplectic matrix is nonsymplectic. Since V^{in} and V^{fin} are in general different, it follows from (4.121) that in general N is not symplectic. To strengthen the argument further, suppose that M is of the form

$$M = \begin{pmatrix} \lambda I & 0 \\ 0 & \lambda^{-1} I \end{pmatrix} \quad (6.4.125)$$

where λ is any scalar. According to Section 3.3.2 such an M is symplectic. Verify in this case that

$$N = M \begin{pmatrix} I & 0 \\ C' & I \end{pmatrix} \quad (6.4.126)$$

where

$$C' = \lambda^2 C^{fin} - C^{in}. \quad (6.4.127)$$

Show that

$$C' - (C')^T = q(\lambda^2 \mathbf{B}^{fin} - \mathbf{B}^{in}) \cdot \mathbf{L}, \quad (6.4.128)$$

and therefore in general the second matrix on the right side of (4.126) is not symplectic. Correspondingly, in this case N is generally not symplectic even if V^{in} and V^{fin} are the same. Verify that in this case

$$N^T J N = J \begin{pmatrix} I & 0 \\ q\mathbf{B}' \cdot \mathbf{L} & I \end{pmatrix} \quad (6.4.129)$$

where

$$\mathbf{B}' = \lambda^2 \mathbf{B}^{fin} - \mathbf{B}^{in}. \quad (6.4.130)$$

6.4.12. Recall Section 4.3 and review Exercise 4.3.24. Find the symplectic polar decomposition for the matrix V given by (4.115).

6.5 Mixed-Variable Generating Functions

It is well known that canonical transformations/symplectic maps can be produced by the use of mixed-variable *generating* functions, the most familiar of which are traditionally referred to as F_1 through F_4 . The generating functions are called *mixed* because they involve both “old” and “new” variables.¹⁶

In this section we will verify that generating functions of types 1 through 4 can be used to produce symplectic maps. Conversely, given a symplectic map \mathcal{M} , we will find if generating

¹⁶In the field of light ray optics, where the use of generating functions was first introduced in the seminal work of Hamilton, generating functions are sometimes referred to as *characteristic* functions.

function of types 1 through 4 can possibly be used to produce \mathcal{M} . A time-dependent generating function F_j produces a one-parameter family of symplectic maps. In that case we will find the associated generating Hamiltonian.

In Section 7 we will find that the function types 1 through 4 are but four examples of an *infinite* set of types of generating functions. Until then, the term *generating functions* will refer simply to the function types F_1 through F_4 .

6.5.1 Generating Functions Produce Symplectic Maps

6.5.1.1 Background

Since the use of generating functions does not treat coordinate and momentum variables on a common footing, it is convenient (as in Section 4.8) to introduce the notation

$$z = (q_1 \cdots q_n, p_1 \cdots p_n), \quad (6.5.1)$$

$$Z = (Q_1 \cdots Q_n, P_1 \cdots P_n). \quad (6.5.2)$$

In this notation the symplectic map \mathcal{M} sends z to Z ,

$$\mathcal{M} : z \rightarrow Z. \quad (6.5.3)$$

We begin with the mixed-variable generating function types $F_1(q, Q, t)$, $F_2(q, P, t)$, $F_3(p, Q, t)$, and $F_4(p, P, t)$. These four function types produce maps \mathcal{M} by the (implicit) relation pairs

$$\begin{aligned} p_k &= \partial F_1 / \partial q_k, \quad P_k = -\partial F_1 / \partial Q_k; \text{ requires } \det(\beta) \neq 0 \Leftrightarrow \det(B) \neq 0 \\ \text{where } \beta_{k\ell} &= \partial^2 F_1 / \partial q_k \partial Q_\ell \text{ and } B = \beta^{-1}, \end{aligned} \quad (6.5.4)$$

$$\begin{aligned} p_k &= \partial F_2 / \partial q_k, \quad Q_k = \partial F_2 / \partial P_k; \text{ requires } \det(\beta) \neq 0 \Leftrightarrow \det(D) \neq 0 \\ \text{where } \beta_{k\ell} &= \partial^2 F_2 / \partial q_k \partial P_\ell \text{ and } D = \beta^{-1}, \end{aligned} \quad (6.5.5)$$

$$\begin{aligned} q_k &= -\partial F_3 / \partial p_k, \quad P_k = -\partial F_3 / \partial Q_k; \text{ requires } \det(\beta) \neq 0 \Leftrightarrow \det(A) \neq 0 \\ \text{where } \beta_{k\ell} &= \partial^2 F_3 / \partial p_k \partial Q_\ell \text{ and } A = \beta^{-1}, \end{aligned} \quad (6.5.6)$$

$$\begin{aligned} q_k &= -\partial F_4 / \partial p_k, \quad Q_k = \partial F_4 / \partial P_k; \text{ requires } \det(\beta) \neq 0 \Leftrightarrow \det(C) \neq 0 \\ \text{where } \beta_{k\ell} &= \partial^2 F_4 / \partial p_k \partial P_\ell \text{ and } C = \beta^{-1}. \end{aligned} \quad (6.5.7)$$

We will find that the matrices A through D are related to M , the linear part of \mathcal{M} , by writing M in the block form

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}. \quad (6.5.8)$$

As described above, the matrices A through D in each case are the *inverse* of a particular block of the *Hessian* matrix of the associated F_j . It can be verified that each relation pair produces a symplectic map subject to only *mild* restrictions on the functional behavior of

the associated mixed-variable generating function (e.g. *differentiability*) and the *stringent* restriction/requirement that various matrices in the collection A through D be invertible (have nonvanishing determinant). These invertibility requirements are the compatibility conditions for the case of mixed-variable generating functions. As mentioned at the end of Section 4.8 and will be proved subsequently, there are symplectic maps that cannot be produced by (are incompatible with) any of the generating function types F_j . For these maps all four of the determinants $\det(A)$ through $\det(D)$ vanish.

6.5.1.2 Use of $F_2(q, P, t)$

Consider, for example, the use of F_2 . The equations given by the relation pair (5.5) are implicit,

$$p_k = \partial F_2 / \partial q_k = p_k(q, P, t), \quad (6.5.9)$$

$$Q_k = \partial F_2 / \partial P_k = Q_k(q, P, t), \quad (6.5.10)$$

and have to be brought to the explicit form

$$Q_k = Q_k(q, p, t), \quad (6.5.11)$$

$$P_k = P_k(q, p, t). \quad (6.5.12)$$

(The fact that the map equations are initially implicit and subsequently, often with considerable effort, have to be made explicit is one of the drawbacks of using generating functions to produce symplectic maps. By contrast, as we will see in Chapter 7, Lie transformations can be used to produce symplectic maps that are immediately in explicit form.)

Take differentials of both sides of (5.9) and (5.10), and use (5.5), to get the relations

$$\begin{aligned} dp_k &= \sum_{\ell} [(\partial p_k / \partial q_{\ell}) dq_{\ell} + (\partial p_k / \partial P_{\ell}) dP_{\ell}] \\ &= \sum_{\ell} [(\partial^2 F_2 / \partial q_k \partial q_{\ell}) dq_{\ell} + (\partial^2 F_2 / \partial q_k \partial P_{\ell}) dP_{\ell}], \end{aligned} \quad (6.5.13)$$

$$\begin{aligned} dQ_k &= \sum_{\ell} [(\partial Q_k / \partial q_{\ell}) dq_{\ell} + (\partial Q_k / \partial P_{\ell}) dP_{\ell}] \\ &= \sum_{\ell} [(\partial^2 F_2 / \partial P_k \partial q_{\ell}) dq_{\ell} + (\partial^2 F_2 / \partial P_k \partial P_{\ell}) dP_{\ell}]. \end{aligned} \quad (6.5.14)$$

These relations can be written in the matrix form

$$dp = \alpha dq + \beta dP, \quad (6.5.15)$$

$$dQ = \gamma dq + \delta dP = \beta^T dq + \delta dP, \quad (6.5.16)$$

where α through δ are the matrices

$$\alpha_{k\ell} = \partial p_k / \partial q_{\ell} = \partial^2 F_2 / \partial q_k \partial q_{\ell}, \quad (6.5.17)$$

$$\beta_{k\ell} = \partial p_k / \partial P_{\ell} = \partial^2 F_2 / \partial q_k \partial P_{\ell}, \quad (6.5.18)$$

$$\gamma_{k\ell} = \partial Q_k / \partial q_\ell = \partial^2 F_2 / \partial P_k \partial q_\ell, \quad (6.5.19)$$

$$\delta_{k\ell} = \partial Q_k / \partial P_\ell = \partial^2 F_2 / \partial P_k \partial P_\ell. \quad (6.5.20)$$

(Note here that the matrix δ is not to be confused with the Kronecker delta.) By inspection, these matrices have the properties

$$\alpha^T = \alpha, \quad \beta^T = \gamma, \quad \delta^T = \delta; \quad (6.5.21)$$

and we have used the second property in (5.21) to write the terms on the far right side of (5.16). Observe that, according to (5.18), the matrix β is a particular block of the Hessian of F_2 . The other blocks are α , γ , and δ .

Solve (5.15) for dP to find the result

$$dP = -\beta^{-1}\alpha dq + \beta^{-1}dp, \quad (6.5.22)$$

and insert this result in (5.16) to get the complementary relation

$$dQ = (\gamma - \delta\beta^{-1}\alpha)dq + \delta\beta^{-1}dp. \quad (6.5.23)$$

Note that these manipulations require that β be invertible. That is, we require that

$$\det(\beta) \neq 0. \quad (6.5.24)$$

And, as we have already observed, β is a particular block of the Hessian of F_2 . [See (5.18) and the requirement made in (5.5).] By the inverse function theorem, this invertibility is equivalent to requiring that the first set of equations in (5.5), see (5.9), can be solved for the P_ℓ to find $P_\ell(q, p, t)$.

Next write (5.22) and (5.23) in the compact matrix form

$$dZ = M dz. \quad (6.5.25)$$

Comparison of (5.25) with (5.22) and (5.23) shows that M has the block form

$$M = \begin{pmatrix} \gamma - \delta\beta^{-1}\alpha & \delta\beta^{-1} \\ -\beta^{-1}\alpha & \beta^{-1} \end{pmatrix}. \quad (6.5.26)$$

As in Section 3.3, it is convenient to employ the notation

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}. \quad (6.5.27)$$

Thus, we have the identifications

$$A = \gamma - \delta\beta^{-1}\alpha = \beta^T - \delta\beta^{-1}\alpha, \quad (6.5.28)$$

$$B = \delta\beta^{-1}, \quad (6.5.29)$$

$$C = -\beta^{-1}\alpha, \quad (6.5.30)$$

$$D = \beta^{-1}. \quad (6.5.31)$$

At this point we remark that the relations (4.8.9) and (4.8.10) resemble the relations (5.15) and (5.16); and the relations (4.8.14) through (4.8.17) are identical to the relations (5.28) through (5.31). This is as it should be because linear maps are a special case of general maps. We also see from (5.31) that

$$\det(D) = \det(\beta^{-1}) = [\det(\beta)]^{-1} \neq 0, \quad (6.5.32)$$

in accord with (5.5).

Finally, we must verify that M is symplectic. With the aid of (5.21) we find the relations

$$A^T = \gamma^T - \alpha^T(\beta^{-1})^T\delta^T = \beta - \alpha\gamma^{-1}\delta, \quad (6.5.33)$$

$$B^T = (\beta^{-1})^T\delta^T = \gamma^{-1}\delta, \quad (6.5.34)$$

$$C^T = -\alpha^T(\beta^{-1})^T = -\alpha\gamma^{-1}, \quad (6.5.35)$$

$$D^T = (\beta^{-1})^T = \gamma^{-1}. \quad (6.5.36)$$

Compute the various combinations of matrices that appear in (3.3.3) through (3.3.5). We find the results

$$A^T C = (\beta - \alpha\gamma^{-1}\delta)(-\beta^{-1}\alpha) = -\alpha + \alpha\gamma^{-1}\delta\beta^{-1}\alpha, \quad (6.5.37)$$

$$C^T A = -\alpha - \gamma^{-1}(\gamma - \delta\beta^{-1}\alpha) = -\alpha + \alpha\gamma^{-1}\delta\beta^{-1}\alpha, \quad (6.5.38)$$

$$B^T D = \gamma^{-1}\delta\beta^{-1}, \quad (6.5.39)$$

$$D^T B = \gamma^{-1}\delta\beta^{-1}, \quad (6.5.40)$$

$$A^T D = (\beta - \alpha\gamma^{-1}\delta)\beta^{-1} = I - \alpha\gamma^{-1}\delta\beta^{-1}, \quad (6.5.41)$$

$$C^T B = -\alpha\gamma^{-1}\delta\beta^{-1}. \quad (6.5.42)$$

By looking at these results we see that the relations (3.3.3) through (3.3.5) are satisfied, and therefore M is symplectic. Correspondingly, when the implicit relations (5.9) and (5.10) are solved to yield Z in terms of z , the result is a symplectic map \mathcal{M} .

6.5.1.3 Use of $F_1(q, Q, t)$

We have examined the use of F_2 . As a second example, we will consider the use of F_1 .¹⁷ The cases of F_3 and F_4 proceed similarly.

For the case of F_1 the equations given by the relation pair (5.4) have the implicit form

$$p_k = \partial F_1 / \partial q_k = p_k(q, Q, t), \quad (6.5.43)$$

$$P_k = -\partial F_1 / \partial Q_k = P_k(q, Q, t), \quad (6.5.44)$$

and have to be brought to the explicit form

$$Q_k = Q_k(q, p, t), \quad (6.5.45)$$

¹⁷Some authors refer to F_1 type generating functions as *Lagrangian* generating functions. Perhaps that is because this type of generating function, somewhat like Lagrangians, involves only *coordinate (configuration space)* variables. Or perhaps it is because, in the context of Lagrangian dynamics, generating functions of this type are related to the integral over time t of the Lagrangian. See the end of Exercise 5.7.

$$P_k = P_k(q, p, t). \quad (6.5.46)$$

Take differentials of both sides of (5.43) and (5.44), and use (5.4), to get the relations

$$\begin{aligned} dp_k &= \sum_{\ell} (\partial p_k / \partial q_{\ell}) dq_{\ell} + (\partial p_k / \partial Q_{\ell}) dQ_{\ell} \\ &= \sum_{\ell} (\partial^2 F_1 / \partial q_k \partial q_{\ell}) dq_{\ell} + (\partial^2 F_1 / \partial q_k \partial Q_{\ell}) dQ_{\ell}, \end{aligned} \quad (6.5.47)$$

$$\begin{aligned} dP_k &= \sum_{\ell} (\partial P_k / \partial q_{\ell}) dq_{\ell} + (\partial P_k / \partial Q_{\ell}) dQ_{\ell} \\ &= - \sum_{\ell} (\partial^2 F_1 / \partial Q_k \partial q_{\ell}) dq_{\ell} - (\partial^2 F_1 / \partial Q_k \partial Q_{\ell}) dQ_{\ell}. \end{aligned} \quad (6.5.48)$$

These relations can be written in the matrix form

$$dp = \alpha dq + \beta dQ, \quad (6.5.49)$$

$$dP = \gamma dq + \delta dQ, \quad (6.5.50)$$

where α through δ are the matrices

$$\alpha_{k\ell} = \partial p_{\ell} / \partial q_k = \partial^2 F_1 / \partial q_k \partial q_{\ell}, \quad (6.5.51)$$

$$\beta_{k\ell} = \partial p_k / \partial Q_{\ell} = \partial^2 F_1 / \partial q_k \partial Q_{\ell}, \quad (6.5.52)$$

$$\gamma_{k\ell} = \partial P_k / \partial q_{\ell} = -\partial^2 F_1 / \partial Q_k \partial q_{\ell}, \quad (6.5.53)$$

$$\delta_{k\ell} = \partial P_k / \partial Q_{\ell} = -\partial^2 F_1 / \partial Q_k \partial Q_{\ell}. \quad (6.5.54)$$

By inspection, these matrices have the properties

$$\alpha^T = \alpha, \quad \beta^T = -\gamma, \quad \delta^T = \delta. \quad (6.5.55)$$

Solve (5.49) for dQ to find the result

$$dQ = -\beta^{-1} \alpha dq + \beta^{-1} dp, \quad (6.5.56)$$

and insert this result in (5.50) to get the complementary relation

$$dP = (\gamma - \delta \beta^{-1} \alpha) dq + \delta \beta^{-1} dp. \quad (6.5.57)$$

Note that these manipulations require that β be invertible. [See (5.52) and the requirement made in (5.4).] By the inverse function theorem, this invertibility is equivalent to requiring that the first set of equations in (5.4), see (5.43), can be solved for the Q_{ℓ} to find $Q_{\ell}(q, p, t)$.

As before, write (5.56) and (5.57) in the compact matrix form (5.25) and employ (5.27). Comparison of (5.56) and (5.57) with (5.25) and (5.27) yields the relations

$$A = -\beta^{-1} \alpha, \quad (6.5.58)$$

$$B = \beta^{-1}, \quad (6.5.59)$$

$$C = \gamma - \delta\beta^{-1}\alpha, \quad (6.5.60)$$

$$D = \delta\beta^{-1}. \quad (6.5.61)$$

We see from (5.59) that B must be invertible, $\det(B) \neq 0$, in accord with (5.4).

Finally, we must verify that M is symplectic. With the aid of (5.55) we find the relations

$$A^T = -\alpha^T(\beta^{-1})^T = \alpha\gamma^{-1}, \quad (6.5.62)$$

$$B^T = (\beta^{-1})^T = -\gamma^{-1}, \quad (6.5.63)$$

$$C^T = \gamma^T - \alpha^T(\beta^{-1})^T\delta^T = -\beta + \alpha\gamma^{-1}\delta, \quad (6.5.64)$$

$$D^T = (\beta^{-1})^T\delta^T = -\gamma^{-1}\delta. \quad (6.5.65)$$

Compute the various combinations of matrices that appear in (3.3.3) through (3.3.5). We find the results

$$A^T C = \alpha\gamma^{-1}(\gamma - \delta\beta^{-1}\alpha) = \alpha - \alpha\gamma^{-1}\delta\beta^{-1}\alpha, \quad (6.5.66)$$

$$C^T A = (\beta - \alpha\gamma^{-1}\delta)\beta^{-1}\alpha = \alpha - \alpha\gamma^{-1}\delta\beta^{-1}\alpha, \quad (6.5.67)$$

$$B^T D = -\gamma^{-1}\delta\beta^{-1}, \quad (6.5.68)$$

$$D^T B = -\gamma^{-1}\delta\beta^{-1}, \quad (6.5.69)$$

$$A^T D = \alpha\gamma^{-1}\delta\beta^{-1}, \quad (6.5.70)$$

$$C^T B = (-\beta + \alpha\gamma^{-1}\delta)\beta^{-1} = -I + \alpha\gamma^{-1}\delta\beta^{-1}. \quad (6.5.71)$$

By looking at these results we see that the relations (3.3.3) through (3.3.5) are satisfied, and therefore M is symplectic. Correspondingly, when the implicit relations (5.43) and (5.44) are solved to yield Z in terms of z , the result is a symplectic map \mathcal{M} .

6.5.1.4 What Maps Can Be Produced by What F_j ?

We have seen that the F_j produce symplectic maps, but now wonder what maps can be produced in this fashion. Here we will make a few observations. A more complete exploration of this question is made in Subsection 7.4.

Suppose \mathcal{M} is the identity map so that

$$M = I^{2n}. \quad (6.5.72)$$

In this case

$$\det(A) = \det(D) = \det(I^n) = 1. \quad (6.5.73)$$

Therefore, according to (5.5) and (5.6), we expect the use of F_2 and F_3 to succeed. Indeed, it is easy to verify that $F_2(q, P)$ and $F_3(p, Q)$ given by

$$F_2(q, P) = \sum_k q_k P_k \quad (6.5.74)$$

and

$$F_3(p, Q) = - \sum_k p_k Q_k \quad (6.5.75)$$

do indeed produce the identity map. By contrast,

$$\det(B) = \det(C) = 0 \quad (6.5.76)$$

for the identity map. Therefore, according to (5.4) and (5.7), use of either F_1 or F_4 cannot produce the identity map; attempted use of either F_1 or F_4 fails.

What about the linear symplectic map \mathcal{M} for which $M = J$? In this case we see from (3.1.1) that

$$\det(A) = \det(D) = 0 \quad (6.5.77)$$

and

$$\det(B) \neq 0 \text{ and } \det(C) \neq 0. \quad (6.5.78)$$

Examination of (5.4) through (5.7) shows that attempted use of F_2 and F_3 are expected to fail, and attempted use of F_1 and F_4 are expected to succeed. Indeed, it is easily verified that

$$F_1(q, Q, t) = \sum_k q_k Q_k \quad (6.5.79)$$

and

$$F_4(p, P, t) = \sum_k p_k P_k \quad (6.5.80)$$

produce $M = J$ when employed in (5.4) and (5.7), respectively.

What about the linear symplectic map \mathcal{M} for which $M = R$ where R is the symplectic matrix given by (4.8.31)? If you worked Exercise 4.8.4, you verified that in this case all the submatrices A through D fail to have inverses. Therefore none of the procedures listed in (5.4) through (5.7) can be implemented if one attempts to produce this R using any F_j . That is, one cannot produce this R using any of the F_j ; attempted use of any of the F_j fails.

6.5.1.5 Differentials and Differential Forms associated with the F_j

Associated with each of the F_j are both a differential dF_j and a *differential form* which we will call ω_j . For example, we may write the differential

$$dF_1(q, Q, t) = \sum_k [(\partial F_1 / \partial q_k) dq_k + (\partial F_1 / \partial Q_k) dQ_k]. \quad (6.5.81)$$

Correspondingly, making use of the relations for p_k and P_k in (5.4), we define an associated differential form ω_1 by the rule

$$\omega_1 = \sum_k p_k dq_k - P_k dQ_k. \quad (6.5.82)$$

Note that ω_1 involves the $2n+2n = 4n$ variables z and Z . Similarly, there are the differentials and associated differential forms

$$dF_2(q, P, t) = \sum_k (\partial F_2 / \partial q_k) dq_k + (\partial F_2 / \partial P_k) dP_k, \quad (6.5.83)$$

$$\omega_2 = \sum_k p_k dq_k + Q_k dP_k; \quad (6.5.84)$$

$$dF_3(p, Q, t) = \sum_k (\partial F_3 / \partial p_k) dp_k + (\partial F_3 / \partial Q_k) dQ_k, \quad (6.5.85)$$

$$\omega_3 = \sum_k -q_k dp_k - P_k dQ_k; \quad (6.5.86)$$

$$dF_4(p, P, t) = \sum_k (\partial F_4 / \partial p_k) dp_k + (\partial F_4 / \partial P_k) dP_k, \quad (6.5.87)$$

$$\omega_4 = \sum_k -q_k dp_k + Q_k dP_k. \quad (6.5.88)$$

We have seen, if a particular block of the Hessian of a given F_j is invertible, then there is an associated symplectic map which we will call \mathcal{M}_j , and the related $n \times n$ block in the associated Jacobian matrix M_j as given by (5.27) will be invertible. [In the case of F_1 , for example, according to (5.4) the related block is the matrix B .] Conversely, given a symplectic map \mathcal{M} and the desire of find an associated generating function F_j , and after verifying that the nature of \mathcal{M} is such that all the remaining variables among the z and Z can be found in terms of the variables on which F_j is supposed to depend, then it can be shown that the differential form ω_j is exact in terms of these variables, and correspondingly the desired F_j can be constructed. Moreover, the requirement that all remaining variables can be found in terms of the variables on which F_j is supposed to depend is equivalent to assuming that the related $n \times n$ block in M is *invertible*. In this case we say that form/type of the desired F_j is *compatible* with the nature of \mathcal{M} .) Note that once the form/type of F_j has been specified, the question of whether a given map \mathcal{M} is compatible with a generating function of type F_j depends *only* on M , the *linear* part of \mathcal{M} . See, for example, Subsubsection 5.2.1 where F_2 is constructed from a knowledge of \mathcal{M} .

But what happens if the form of the desired F_j is *not* compatible with the nature of \mathcal{M} ? Then an attempted construction of F_j will fail. Suppose, for example, that \mathcal{M} is the identity map. In this case, since $Q_k = q_k$ and $P_k = p_k$, there are, according to (5.82), (5.84), (5.86), and (5.88), the results

$$\omega_1 = \sum_k p_k dq_k - P_k dQ_k = \sum_k p_k dq_k - p_k dq_k = 0, \quad (6.5.89)$$

$$\omega_2 = \sum_k p_k dq_k + Q_k dP_k = \sum_k p_k dq_k + q_k dp_k = d\left(\sum_k q_k p_k\right) = d\left(\sum_k q_k P_k\right), \quad (6.5.90)$$

$$\omega_3 = \sum_k -q_k dp_k - P_k dQ_k = \sum_k -q_k dp_k - p_k dq_k = d\left(-\sum_k p_k q_k\right) = d\left(-\sum_k p_k Q_k\right), \quad (6.5.91)$$

$$\omega_4 = \sum_k -q_k dp_k + Q_k dP_k = \sum_k -q_k dp_k + q_k dp_k = 0. \quad (6.5.92)$$

Note that (5.89) and (5.92) are in accord with the fact that attempted use of F_1 or F_4 fails for the identity map. Also, (5.90) and (5.74), and (5.91) and (5.75), are in agreement for the identity map. That is, $\omega_2 = dF_2$ and $\omega_3 = dF_3$.

6.5.2 Finding a Generating Function from a Map or a Generating Hamiltonian

We have seen that, modulo the invertibility of certain matrices, the mixed-variable generating functions F_1 through F_4 can be used to produce symplectic maps \mathcal{M} . What about the converse: given a symplectic map \mathcal{M} , can we find a mixed-variable generating function that produces it? Or, given the Hamiltonian H that generates a family of symplectic maps $\mathcal{M}(t)$, can we find an associated time-dependent generating function? We shall see that, again modulo the invertibility of certain matrices which amounts to the question of compatibility, the answer is *yes*.

6.5.2.1 Finding a Generating Function Directly from a Map

As an example, we will consider the problem of constructing $F_2(q, P, t)$ given a symplectic map \mathcal{M} . Begin by writing the relation (5.3) in the component form

$$Q_k = S_k(q, p, t), \quad (6.5.93)$$

$$P_k = T_k(q, p, t), \quad (6.5.94)$$

and assume that the S_k and T_k are known functions. Next assume that the relations (5.91) can be inverted to give the p_k as functions of q , P , and t ,

$$p_k = p_k(q, P, t). \quad (6.5.95)$$

By the inverse function theorem, this inversion is possible if the Jacobian matrix

$$\partial P_k / \partial p_\ell = \partial T_k / \partial p_\ell \quad (6.5.96)$$

is invertible. Next substitute the relations (5.92) into (5.90) to obtain the Q_k as functions of q , P , and t ,

$$Q_k = Q_k(q, P, t). \quad (6.5.97)$$

Now consider the differential form

$$\omega_2 = \sum_k (p_k dq_k + Q_k dP_k). \quad (6.5.98)$$

Recall (5.81). We shall soon see that the assumption that \mathcal{M} is symplectic implies that this differential form is exact with regard to the variables q_k , P_k . Taking this assertion as granted, we may define a function $F_2(q, P, t)$ by the path integral

$$F_2(q, P, t) = \int^{q, P} \omega_2 = \int^{q, P} \sum_k [p_k(q', P', t) dq'_k + Q_k(q', P', t) dP'_k]. \quad (6.5.99)$$

By construction F_2 will have the properties

$$\partial F_2 / \partial q_k = p_k(q, P, t), \quad (6.5.100)$$

$$\partial F_2 / \partial P_k = Q_k(q, P, t), \quad (6.5.101)$$

and we see that the desired relations (5.5) have been obtained.

We still must show that ω_2 given by (5.95) is exact. According to Exercise 1.1, we must verify the relations (1.26). In the present context these relations take the form

$$\frac{\partial p_m}{\partial q_n} = \frac{\partial p_n}{\partial q_m}, \quad (6.5.102)$$

$$\frac{\partial Q_m}{\partial q_n} = \frac{\partial p_n}{\partial P_m}, \quad (6.5.103)$$

$$\frac{\partial p_m}{\partial P_n} = \frac{\partial Q_n}{\partial q_m}, \quad (6.5.104)$$

$$\frac{\partial Q_m}{\partial P_n} = \frac{\partial Q_n}{\partial P_m}. \quad (6.5.105)$$

Note that (5.100) and (5.101) say the same thing.

Take differentials of both sides of (5.90) and (5.91) and use the notation of (5.1), (5.2), (5.23), and (5.25) to find the relations

$$dQ = Adq + Bdp, \quad (6.5.106)$$

$$dP = Cdq + Ddp, \quad (6.5.107)$$

where A through D are the matrices

$$A_{k\ell} = \frac{\partial Q_k}{\partial q_\ell} = \frac{\partial S_k}{\partial q_\ell}, \quad B_{k\ell} = \frac{\partial Q_k}{\partial p_\ell} = \frac{\partial S_k}{\partial p_\ell}, \quad (6.5.108)$$

$$C_{k\ell} = \frac{\partial P_k}{\partial q_\ell} = \frac{\partial T_k}{\partial q_\ell}, \quad D_{k\ell} = \frac{\partial P_k}{\partial p_\ell} = \frac{\partial T_k}{\partial p_\ell}. \quad (6.5.109)$$

We now want to take q and P as independent variables. Solve (5.104) and (5.103) for dp and dQ in terms of dq and dP to find the results

$$dp = -D^{-1}Cdq + D^{-1}dP, \quad (6.5.110)$$

$$dQ = (A - BD^{-1}C)dq + BD^{-1}dP. \quad (6.5.111)$$

Note that in finding these results we assumed the existence of D^{-1} . But comparison of (5.93) and (5.106) shows that D is the Jacobian matrix whose invertibility has already been assumed. From (5.107) and (5.108) we obtain the results

$$\frac{\partial p_m}{\partial q_n} = (-D^{-1}C)_{mn}, \quad (6.5.112)$$

$$\frac{\partial p_m}{\partial P_n} = (D^{-1})_{mn}, \quad (6.5.113)$$

$$\frac{\partial Q_m}{\partial q_n} = (A - BD^{-1}C)_{mn}, \quad (6.5.114)$$

$$\frac{\partial Q_m}{\partial P_n} = (BD^{-1})_{mn}. \quad (6.5.115)$$

With these results before us, we see that establishing the relations (5.99) through (5.102) is equivalent to verifying the conjectures

$$(D^{-1}C) \stackrel{?}{=} (D^{-1}C)^T, \quad (6.5.116)$$

$$A - BD^{-1}C \stackrel{?}{=} (D^{-1})^T, \quad (6.5.117)$$

$$(BD^{-1}) \stackrel{?}{=} (BD^{-1})^T. \quad (6.5.118)$$

But, thanks to the symplectic condition, (5.113) is a consequence of (3.3.7), (5.115) is a consequence of (3.3.4), and (5.114) is a consequence of (3.3.8) and (3.37). Thus we have proved that ω_2 is an exact differential, and have verified that F_2 can be constructed using (5.96). Note that in this construction the time t played no role and, if present at all, appeared only as a parameter.

6.5.2.2 Finding a Generating Function from a Generating Hamiltonian

Given a Hamiltonian H , we know that integrating Hamilton's equations of motion produces a time-dependent symplectic map $\mathcal{M}(t)$. Conversely, given a time-dependent symplectic map $\mathcal{M}(t)$, we know that there is an underlying generating Hamiltonian H . Recall Subsection 4.2. Here we explore how the generating Hamiltonian H can be used to construct the $F_2(q, P, t)$ generating function associated with $\mathcal{M}(t)$. Similar constructions can be made for the F_1 , F_3 , and F_4 generating functions.

To see how F_2 can be constructed, it is convenient, in analogy with (1.7.9), to introduce the phase-space variables

$$\zeta = (\xi, \eta). \quad (6.5.119)$$

Here the ξ 's play the role of coordinates and the η 's are conjugate momenta. Let q, p be initial conditions at $t = t^i$, and let Q, P be the final conditions reached by following to time t the trajectories generated by $H(\zeta, t)$ starting with these initial conditions. We know that trajectories can be labeled by specifying either the initial conditions q, p or the final conditions Q, P . Assume that the trajectories are such that they can also be labeled by specifying q and P . See Figure 5.1. This means that there are relations of the form

$$Q_j = Q_j(q, P, t), \quad (6.5.120)$$

$$p_j = p_j(q, P, t). \quad (6.5.121)$$

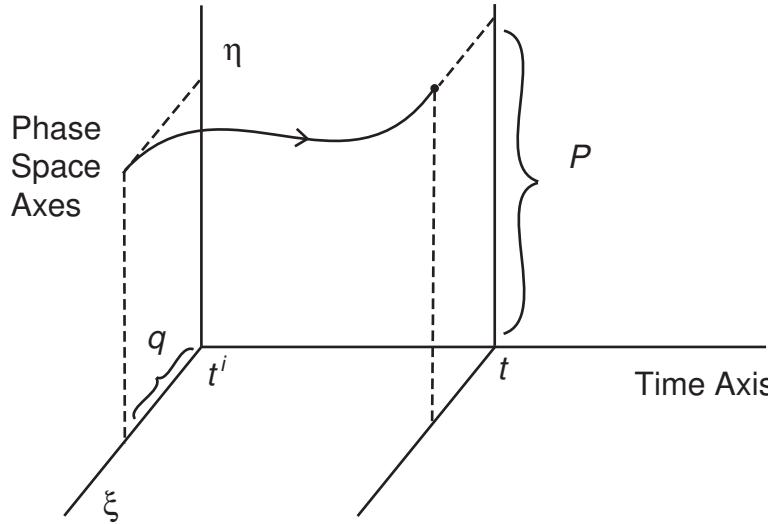


Figure 6.5.1: A trajectory of $H(\zeta, t)$ in the augmented ξ, η, t phase space having initial coordinates q and final momenta P .

With these assumptions in mind, define/construct the function F_2 by the rule

$$F_2(q, P, t) = \sum_k P_k Q_k - \int_{t^i}^t d\tau \left[\left(\sum_j \eta_j \dot{\xi}_j \right) - H(\zeta, \tau) \right]. \quad (6.5.122)$$

Here the integral on the right side is to be evaluated over the trajectory generated by H whose initial coordinates are q and final momenta are P . In actual practice, this trajectory

may have to be found by some kind of *shooting* method: One integrates a variety of trajectories all having the same initial q and various initial p until one finds a trajectory that has the desired final momenta P . The search for this trajectory may be facilitated by also integrating the variational equations, see Exercise 1.4.6, to determine how changes in the initial conditions produce changes in the final conditions.

We will want to see how F_2 changes when changes are made in q, P, t . As a first step, let us study how the integral on the right side of (5.119) depends on the variables q, P . We make the definition

$$A(q, P, t) = \int_{t^i}^t d\tau \left[\left(\sum_j \eta_j \dot{\xi}_j \right) - H(\zeta, \tau) \right], \quad (6.5.123)$$

and recognize that A is the *action*. See (1.6.11). Define \mathcal{A} by the rule

$$\mathcal{A}(\zeta, \dot{\zeta}, \tau) = \left(\sum_j \eta_j \dot{\xi}_j \right) - H(\zeta, \tau) \quad (6.5.124)$$

so that we may write

$$A(q, P, t) = \int_{t^i}^t \mathcal{A}(\zeta, \dot{\zeta}, \tau) d\tau = \int_{t^i}^t L d\tau. \quad (6.5.125)$$

In writing (5.122) we have used the fact that \mathcal{A} , as seen from looking at (5.121), is the Lagrangian L associated with the Hamiltonian H .

Changing the q 's and P 's changes the trajectory. Consequently, from variational calculus, we find that the change in A is given by the relation

$$\delta A = \int_{t^i}^t d\tau \left[\sum_j \left(\frac{\partial \mathcal{A}}{\partial \xi_j} \delta \xi_j + \frac{\partial \mathcal{A}}{\partial \dot{\xi}_j} \delta \dot{\xi}_j + \frac{\partial \mathcal{A}}{\partial \eta_j} \delta \eta_j + \frac{\partial \mathcal{A}}{\partial \dot{\eta}_j} \delta \dot{\eta}_j \right) \right]. \quad (6.5.126)$$

The integrand in (5.123) can be manipulated in the standard way to rewrite δA in the form

$$\begin{aligned} \delta A &= \int_{t^i}^t d\tau \left\{ \sum_j \left[\left(\frac{\partial \mathcal{A}}{\partial \xi_j} - \frac{d}{d\tau} \left(\frac{\partial \mathcal{A}}{\partial \dot{\xi}_j} \right) \right) \delta \xi_j \right. \right. \\ &\quad + \sum_j \left[\left(\frac{\partial \mathcal{A}}{\partial \eta_j} - \frac{d}{d\tau} \left(\frac{\partial \mathcal{A}}{\partial \dot{\eta}_j} \right) \right) \delta \eta_j \right. \\ &\quad \left. \left. + \left(\frac{d}{d\tau} \left[\sum_j \left(\frac{\partial \mathcal{A}}{\partial \dot{\xi}_j} \delta \xi_j + \frac{\partial \mathcal{A}}{\partial \dot{\eta}_j} \delta \eta_j \right) \right] \right) \right] \right\}. \end{aligned} \quad (6.5.127)$$

For the various ingredients in the integrand of (5.124) we find the results

$$\frac{\partial \mathcal{A}}{\partial \xi_j} - \frac{d}{d\tau} \left(\frac{\partial \mathcal{A}}{\partial \dot{\xi}_j} \right) = -\frac{\partial H}{\partial \xi_j} - \frac{d}{d\tau} \eta_j = -\frac{\partial H}{\partial \xi_j} - \dot{\eta}_j = 0, \quad (6.5.128)$$

$$\frac{\partial \mathcal{A}}{\partial \dot{\xi}_j} = \eta_j, \quad (6.5.129)$$

$$\frac{\partial \mathcal{A}}{\partial \dot{\eta}_j} = 0, \quad (6.5.130)$$

$$\frac{\partial \mathcal{A}}{\partial \eta_j} - \frac{d}{d\tau} \left(\frac{\partial \mathcal{A}}{\partial \dot{\eta}_j} \right) = \frac{\partial \mathcal{A}}{\partial \eta_j} = \dot{\xi}_j - \frac{\partial H}{\partial \eta_j} = 0. \quad (6.5.131)$$

Here (5.127) follows from the fact that \mathcal{A} does not actually depend on the $\dot{\eta}_j$. See (5.121). And (5.125) and (5.128) follow from the stipulation that the $\zeta(\tau)$ are trajectories of H . As a consequence of these results, δA becomes

$$\delta A = \int_{t^i}^t d\tau (d/d\tau) [\sum_j \eta_j \delta \xi_j] = [\sum_j \eta_j \delta \xi_j]|_{t^i}^t = \sum_j P_j \delta Q_j - p_j \delta q_j. \quad (6.5.132)$$

We are now ready to study F_2 . In terms of the definition (5.120) the expression (5.119) for F_2 can be rewritten in the form

$$F_2(q, P, t) = -A(q, P, t) + \sum_k P_k Q_k. \quad (6.5.133)$$

It follows that the change in F_2 produced by changes in q, P is given by the relation

$$\begin{aligned} \delta F_2 &= -\delta A + \delta (\sum_k P_k Q_k) \\ &= \sum_j (-P_j \delta Q_j + p_j \delta q_j + P_j \delta Q_j + Q_j \delta P_j) \\ &= \sum_j (p_j \delta q_j + Q_j \delta P_j). \end{aligned} \quad (6.5.134)$$

Here we have used (5.129). Evidently (5.131) yields the relations

$$\partial F_2 / \partial q_j = p_j, \quad \partial F_2 / \partial P_j = Q_j, \quad (6.5.135)$$

which are the desired results (5.5).

As a final step, and in anticipation of results to be established in the next section, let us take the total time derivative of both sides of (5.130). From the chain rule we find the result

$$dF_2/dt = \partial F_2 / \partial t + \sum_j (\partial F_2 / \partial P_j) \dot{P}_j = \partial F_2 / \partial t + \sum_j Q_j \dot{P}_j. \quad (6.5.136)$$

Here we have also used (5.132). For A as given by (5.120) we find the result

$$dA/dt = [(\sum_j \eta_j \dot{\xi}_j) - H(\zeta, \tau)]|_{\tau=t} = (\sum_j P_j \dot{Q}_j) - H(Q, P, t). \quad (6.5.137)$$

Also, there is the simple result

$$(d/dt)(\sum_j P_j Q_j) = \sum_j (\dot{P}_j Q_j + P_j \dot{Q}_j). \quad (6.5.138)$$

It follows that the total time derivative of (5.130) is given by the relation

$$\begin{aligned} dF_2/dt = (d/dt)[-A + (\sum_j P_j Q_j)] &= [\sum_j (\dot{P}_j Q_j + P_j \dot{Q}_j - P_j \dot{Q}_j)] + H(Q, P, t) \\ &= (\sum_j Q_j \dot{P}_j) - H(Q, P, t). \end{aligned} \quad (6.5.139)$$

Comparison of (5.133) and (5.136) gives the final result

$$\partial F_2 / \partial t = H(Q, P, t). \quad (6.5.140)$$

6.5.3 Finding the Generating Hamiltonian from a Generating Function; Hamilton-Jacobi Theory/Equations

If a generating function F_j is time dependent, then its use in the appropriate associated relation selected from (5.4) through (5.7) will produce a *family* of symplectic maps $\mathcal{M}(t)$. Thanks to the work of Section 6.4, we know that any family of symplectic maps is generated by a Hamiltonian. In this subsection we will find the Hamiltonian associated with a time dependent F_j .

6.5.3.1 Derivation

Consider, for example, the case where $F_2(q, P, t)$ is employed. Since our derivation will involve a flurry of partial differentiations with respect to various variables, it is convenient to introduce the notation

$$F_2(q, P, t; \ , \ , 1) = \partial F_2 / \partial t, \quad (6.5.141)$$

$$F_2(q, P, t; k, \ , 1) = \partial^2 F_2 / \partial q_k \partial t, \quad (6.5.142)$$

$$F_2(q, P, t; k\ell, \ ,) = \partial^2 F_2 / \partial q_k \partial q_\ell, \quad (6.5.143)$$

$$F_2(q, P, t; k, \ell,) = \partial^2 F_2 / \partial q_k \partial P_\ell. \quad (6.5.144)$$

With this notation in mind, define the function $F_2^t(q, P, t)$ by the rule

$$F_2^t(q, P, t) = F_2(q, P, t; \ , \ , 1). \quad (6.5.145)$$

We know that use of the map produced by F_2 yields relations of the form (5.10) and (5.11). Moreover, since the map is symplectic, these relations can be inverted to yield relations of the form

$$q_k = q_k(Q, P, t), \quad (6.5.146)$$

$$p_k = p_k(Q, P, t). \quad (6.5.147)$$

Now substitute (5.143) into the first argument of (5.142) to produce the function $H_2(Q, P, t)$ defined by the rule

$$H_2(Q, P, t) = F_2^t(q(Q, P, t), P, t), \quad (6.5.148)$$

which we write more compactly, but with less precision, as

$$H_2 = \partial F_2 / \partial t. \quad (6.5.149)$$

We claim that H_2 is the Hamiltonian that generates the family of maps $\mathcal{M}(t)$ produced by the use of $F_2(q, P, t)$.

To see that this claim is correct, write (5.5) in the form

$$p_k = F_2(q, P, t; k, \ ,), \quad (6.5.150)$$

$$Q_k = F_2(q, P, t; \ , k,). \quad (6.5.151)$$

Now suppose the q, p are held *fixed*, and t is changed by an amount dt . So doing will change the Q, P by the amounts dQ, dP given by the relations

$$0 = dp_k = \sum_{\ell} F_2(q, P, t; k, \ell,)dP_{\ell} + F_2(q, P, t; k, , 1)dt, \quad (6.5.152)$$

$$dQ_k = \sum_{\ell} F_2(q, P, t; , \ell k,)dP_{\ell} + F_2(q, P, t; , k, 1)dt. \quad (6.5.153)$$

Note that the zero on the left side of (5.149) indicates that the p_k remain fixed, as desired. Recall the matrices α and δ given in (5.16) and (5.19). In terms of these matrices (5.149) and (5.150) can be written in the form

$$0 = \sum_{\ell} \beta_{k\ell} dP_{\ell} + F_2(q, P, t; , k, 1)dt, \quad (6.5.154)$$

$$dQ_k = \sum_{\ell} \delta_{k\ell} dP_{\ell} + F_2(q, P, t; , k, 1)dt. \quad (6.5.155)$$

Solve (5.151) for the dP to find the result

$$dP_m = -dt \sum_n (\beta^{-1})_{mn} F_2(q, P, t; n, , 1). \quad (6.5.156)$$

Also, insert (5.153) into (5.152) to give an expression for the dQ ,

$$dQ_m = dt \left[- \sum_n (\delta \beta^{-1})_{mn} F_2(q, P, t; n, , 1) \right] + dt F_2(q, P, t; , m, 1). \quad (6.5.157)$$

Finally, dividing through by dt gives the results

$$dQ_m/dt = - \left[\sum_n (\delta \beta^{-1})_{mn} F_2(q, P, t; n, , 1) \right] + F_2(q, P, t; , m, 1), \quad (6.5.158)$$

$$dP_m/dt = - \sum_n (\beta^{-1})_{mn} F_2(q, P, t; n, , 1). \quad (6.5.159)$$

Note that, as before, these manipulations require that β be invertible.

Next let us work out $(\partial H_2/\partial Q)$ and $(\partial H_2/\partial P)$. From (5.145) and (5.142) we find the result

$$(\partial H_2/\partial Q_m) = \sum_n F_2(q, P, t; n, , 1) (\partial q_n/\partial Q_m). \quad (6.5.160)$$

However, if we solve (5.15) and (5.14) for dq and dp , we find the relations

$$dq = \gamma^{-1} dQ - \gamma^{-1} \delta dP, \quad (6.5.161)$$

$$dp = \alpha \gamma^{-1} dQ + (\beta - \alpha \gamma^{-1} \delta) dP. \quad (6.5.162)$$

Note that, according to (5.20), the invertibility of γ is guaranteed by the invertibility of β . From (5.158) and (5.20) we find the relation

$$(\partial q_n/\partial Q_m) = (\gamma^{-1})_{nm} = (\beta^{-1})_{mn}. \quad (6.5.163)$$

Therefore (5.157) can also be written in the form

$$(\partial H_2 / \partial Q_m) = \sum_n (\beta^{-1})_{mn} F_2(q, P, t; n, , 1). \quad (6.5.164)$$

For $(\partial H_2 / \partial P_m)$ we find from (5.145) and (5.142) the more complicated result

$$(\partial H_2 / \partial P_m) = [\sum_n F_2(q, P, t; n, , 1) (\partial q_n / \partial P_m)] + F_2(q, P, t; , m, 1). \quad (6.5.165)$$

From (5.158) and (5.20) we find the relation

$$(\partial q_n / \partial P_m) = -(\gamma^{-1} \delta)_{nm} = -[\delta^T (\gamma^{-1})^T]_{mn} = -(\delta \beta^{-1})_{mn}. \quad (6.5.166)$$

Therefore (5.162) can also be written in the form

$$(\partial H_2 / \partial P_m) = -[\sum_n (\delta \beta^{-1})_{mn} F_2(q, P, t; n, , 1)] + F_2(q, P, t; , m, 1). \quad (6.5.167)$$

Now we are essentially done. Comparison of the right sides of (5.155) and (5.164) shows that they agree; and comparison of the right sides of (5.156) and (5.161) shows that they agree except for a minus sign. We therefore have demonstrated the desired results

$$dQ_m / dt = \partial H_2 / \partial P_m, \quad (6.5.168)$$

$$dP_m / dt = -\partial H_2 / \partial Q_m. \quad (6.5.169)$$

Finally we remark that similar calculations for all the F_j show (again after a transformation to the variables Q, P, t has been made) that there is the general result

$$H_j = \partial F_j / \partial t. \quad (6.5.170)$$

Note that (5.137) is a special case of (5.167). The relations (5.167) are closely related to the *Hamilton-Jacobi* equations. See the discussion below. We will revisit this subject in Subsection 7.3.

6.5.3.2 Transformation of Hamiltonians and Application to Hamilton-Jacobi Theory

Subsection 4.2 showed that any family of symplectic maps $\mathcal{N}(t)$ is Hamiltonian generated, and the associated Hamiltonian was called G . Subsection 4.4 described the transformation of an old Hamiltonian to a new Hamiltonian under the action of a symplectic map. Here we study the relation between the old and new Hamiltonians in the case that the symplectic map \mathcal{N} arises from some specified mixed-variable generating function F_j , and apply the results to Hamilton-Jacobi theory for this case.

If we make the identification

$$Z = \bar{z}, \quad (6.5.171)$$

the relation (4.56) between old and new Hamiltonians can be rewritten in the form

$$K(Z; t) = H(z(Z, t); t) + G(Z; t). \quad (6.5.172)$$

In the special case that $\mathcal{N}(t)$ arises from the use of an F_j , we found in Subsection 5.3.1 that the associated generating Hamiltonian, which we called H_j , was given by the relation (5.167). Therefore, if we make the identification

$$G = H_j = \partial F_j / \partial t, \quad (6.5.173)$$

we see that (5.169) can be rewritten in the form

$$K(Z; t) = H(z(Z, t); t) + \partial F_j / \partial t \quad (6.5.174)$$

when \mathcal{N} arises from the use of an F_j .

Suppose an $\mathcal{N}(t)$ can be found such that

$$K(Z; t) = 0. \quad (6.5.175)$$

This is, in principle, always possible because we can take the Z to be the initial conditions and take $\mathcal{N}(t)$ to be the symplectic map that transforms final conditions into initial conditions. If an F_j can be found such that $\mathcal{N}(t)$ arises from the use of this F_j , then combining (5.171) and (5.172) gives the Hamilton-Jacobi relation/equation

$$H(z(Z, t); t) + \partial F_j / \partial t = 0. \quad (6.5.176)$$

Exercises

6.5.1. Consider linear symplectic maps of the form (3.3.9) through (3.3.11), (3.10.16), and (3.10.19). Determine which generating functions F_j can be used in these cases, and find explicitly those that are applicable.

6.5.2. Consider the matrices (3.3.9) through (3.3.11). Show that they can all be produced by one of the mixed-variable generating functions F_1 through F_4 , and hence any symplectic matrix can be produced by using a *sequence* of such generating functions.

6.5.3. We have seen that the matrix R given by (4.8.31) cannot be produced by any one of the mixed-variable generating functions F_1 through F_4 . Refer to (4.8.27). Show that there are matrices M near R for which none of the matrices a through d are invertible and hence for these M the method of mixed-variable generating function symplectification using F_1 through F_4 fails.

6.5.4. Use the machinery of Section 4.2 to produce the relations (5.167).

6.5.5. In some situations, for example in passing from Cartesian to curvilinear coordinates in configuration (position) space, it is desirable to make configuration coordinate transformations of the kind

$$Q_k = f_k(q, t). \quad (6.5.177)$$

Transformations of this kind are called *Lagrange point transformations*. Here we assume that the relations (5.174) are invertible so that there are functions $g_k(Q, t)$ such that

$$q_k = g_k(Q, t). \quad (6.5.178)$$

If this change of variables is done in a canonical context, we would like to extend the configuration-space transformation (5.174) into a full phase-space transformation. The purpose of this exercise is to show that this extension can be done symplectically with the aid of the generating function F_2 given by

$$F_2(q, P, t) = \sum_{m=1}^n P_m f_m(q, t). \quad (6.5.179)$$

This symplectic extension is called a *lift* of the configuration coordinate transformation from configuration space to phase space.

Review Subsection 5.1. Show, with the aid of (5.5), that use of the F_2 given by (5.176) yields the desired relation (5.174). Find the matrices α through δ in this case and verify that the matrix β is invertible (as required) if, as has been assumed, (5.174) is invertible. You should find that

$$\alpha_{k\ell} = \partial^2 F_2 / \partial q_k \partial q_\ell = \sum_{m=1}^n P_m \partial^2 f_m(q, t) / \partial q_k \partial q_\ell, \quad (6.5.180)$$

$$\beta_{k\ell} = \partial^2 F_2 / \partial q_k \partial P_\ell = \partial f_\ell / \partial q_k, \quad (6.5.181)$$

$$\gamma_{k\ell} = \partial^2 F_2 / \partial P_k \partial q_\ell = \partial f_k / \partial q_\ell = (\beta^T)_{k\ell}, \quad (6.5.182)$$

$$\delta_{k\ell} = \partial^2 F_2 / \partial P_k \partial P_\ell = 0. \quad (6.5.183)$$

See (5.16) through (5.19).

Verify, using (5.5), that there is the relation

$$p_k = \sum_{m=1}^n P_m \beta_{km} = \sum_{m=1}^n \beta_{km} P_m, \quad (6.5.184)$$

which can be written in the compact matrix-vector form

$$p = \beta P. \quad (6.5.185)$$

It follows that there is the relation

$$P = \beta^{-1} p, \quad (6.5.186)$$

which specifies the transformed momenta associated with the transformed positions (5.174). Verify from (5.174) and (5.178) that there is the differential relation

$$dQ = \beta^T dq. \quad (6.5.187)$$

Canonical transformations given by relations of the form (5.174) and (5.183) are called *Mathieu* transformations, and (5.176) may be called a Mathieu generating function.

Verify that for Mathieu transformations the corresponding A through D matrices are given by the relations

$$A = \gamma - \delta \beta^{-1} \alpha = \gamma = \beta^T, \quad (6.5.188)$$

$$B = \delta \beta^{-1} = 0, \quad (6.5.189)$$

$$C = -\beta^{-1}\alpha, \quad (6.5.190)$$

$$D = \beta^{-1}. \quad (6.5.191)$$

According to Exercise 3.10.5, symplectic matrices with $B = 0$ form a subgroup of the symplectic group. Since the Jacobian of the product of two maps is the product of their Jacobians, it follows that Mathieu transformations form a subgroup of the group of symplectic maps. What is the nature of this subgroup? Evidently invertible Lagrange point transformations form a group which, assuming the underlying topology of configuration space to be Cartesian/Euclidean, is (under differentiability assumptions) the diffeomorphism group $\text{Diff}(\mathbb{R}^n)$. Thus, the subgroup of Mathieu transformations is isomorphic to the group $\text{Diff}(\mathbb{R}^n)$.

Suppose that the transformation (5.174) is in fact *linear* so that it can be written in the form

$$Q = Nq \quad (6.5.192)$$

where N is any real and invertible $n \times n$ matrix. That is, $N \in GL(n, \mathbb{R})$. Find the matrices α through δ and A through D in this case. Show, in particular, that in this case $\alpha = 0$ so that $C = 0$, and that corresponding to (5.189) there is the complementary relation

$$P = (N^T)^{-1}p. \quad (6.5.193)$$

Compare this result to (3.3.13). Verify that all Mathieu transformations for which a relation of the form (5.189) holds constitute a subgroup of the group of all Mathieu transformations, and this subgroup is isomorphic to $GL(n, \mathbb{R})$.

Suppose that the transformation (5.174) is in fact linear and *orthogonal* so that it can be written in the form

$$Q = Oq \quad (6.5.194)$$

where O is an orthogonal matrix. Show that in this case there is the complementary relation

$$P = Op. \quad (6.5.195)$$

Compare this result to (3.3.13) and the discussion of $SO(n, \mathbb{R})$ at the end of Section 7.2.2 and in Exercise 7.2.5. Verify that all Mathieu transformations for which a relation of the form (5.191) holds constitute a subgroup of the group of all Mathieu transformations, and this subgroup is isomorphic to $O(n, \mathbb{R})$.

6.5.6. Consider F_2 generating functions of the form

$$F_2(q, P, t) = -\chi(q, t) + \sum_{m=1}^n P_m q_m. \quad (6.5.196)$$

Show that these F_2 produce symplectic transformations of the form

$$Q_m = q_m, \quad (6.5.197)$$

$$P_m = p_m + \partial\chi/\partial q_m. \quad (6.5.198)$$

These symplectic transformations/maps are sometimes called *gauge* transformations because they arise naturally, for the case $n = 3$, in the context of charged-particle motion in electro-magnetic fields.¹⁸ Indeed, we have already seen in Exercise 2.8 that gauge transformations are symplectic maps.

Show that for gauge transformations the matrices α through δ and A through D are given by the relations

$$\alpha_{k\ell} = \partial^2 F_2 / \partial q_k \partial q_\ell = -\partial^2 \chi / \partial q_k \partial q_\ell, \quad (6.5.199)$$

$$\beta_{k\ell} = \partial^2 F_2 / \partial q_k \partial P_\ell = \bar{\delta}_{k\ell}, \quad (6.5.200)$$

$$\gamma_{k\ell} = \partial^2 F_2 / \partial P_k \partial q_\ell = \bar{\delta}_{k\ell}, \quad (6.5.201)$$

$$\delta_{k\ell} = \partial^2 F_2 / \partial P_k \partial P_\ell = 0; \quad (6.5.202)$$

$$A = \gamma - \delta \beta^{-1} \alpha = \gamma = I, \quad (6.5.203)$$

$$B = \delta \beta^{-1} = 0, \quad (6.5.204)$$

$$C_{k\ell} = -(\beta^{-1} \alpha)_{k\ell} = -\alpha_{k\ell} = \partial^2 \chi / \partial q_k \partial q_\ell, \quad (6.5.205)$$

$$D = \beta^{-1} = I. \quad (6.5.206)$$

Here we have used the symbol $\bar{\delta}_{k\ell}$ to denote the Kronecker delta.

Compare the matrices A through D found above with those for (3.3.10). Show that gauge transformation symplectic maps form a subgroup of the set of all symplectic maps. Hint: See Exercise 3.10.1.

What is the nature of this subgroup? Define a symplectic map \mathcal{M} by the rule

$$\mathcal{M} = \exp : \chi(q, t) : . \quad (6.5.207)$$

Verify that the assertions

$$Q = \mathcal{M}q, \quad (6.5.208)$$

$$P = \mathcal{M}p \quad (6.5.209)$$

yield (5.194) and (5.195). Let $\chi(q, t)$ and $\chi'(q, t)$ be any two gauge functions. Evidently there is the relation

$$[\chi, \chi'] = 0. \quad (6.5.210)$$

It follows that the maps \mathcal{M} and \mathcal{M}' defined by (5.204) and

$$\mathcal{M}' = \exp : \chi'(q, t) : \quad (6.5.211)$$

commute. Also, observe that functions of the form $\chi(q, t)$ comprise an infinite-dimensional vector space. Therefore the set of gauge transformations comprises an infinite-dimensional Abelian group.

¹⁸However note that the symplectic transformations given by (5.194) and (5.195) are defined for all n .

6.5.7. This exercise studies use of the four mixed-variable generating functions of types 1 through 4 and their possible interrelations by Legendre transformations. Consider the diagram below:

$$\begin{array}{ccc} F_1(q, Q) & \leftrightarrow & F_2(q, P) \\ \downarrow & & \downarrow \\ F_3(p, Q) & \leftrightarrow & F_4(p, P) \end{array} \quad (6.5.212)$$

We will see that the F_n pairs connected by two-pointed arrows are possibly related by *simple* Legendre transformations. And the diagonally opposite pairs are possibly related by more complicated Legendre transformations. These simple and more complicated Legendre transformations will be specified shortly.

For ease of exposition, it is best to begin by considering a particular instance: that of a possible relation between generating functions of types 1 and 2 as depicted in the top line of (5.212). Suppose \mathcal{M} is some symplectic map and that a type 2 generating function is compatible with \mathcal{M} . Then there will be a generating function $F_2(q, P)$ which, by employing the standard machinery (5.5), yields the map \mathcal{M} . Make for F_1 the Legendre transformation Ansatz

$$F_1(q, Q) = F_2(q, P) - \sum_k Q_k P_k, \quad (6.5.213)$$

which involves on the right side the known-to-exist generating function F_2 . Form differentials of both sides of (5.213) to find the relations

$$\begin{aligned} dF_1(q, Q) &= \sum_k [(\partial F_2 / \partial q_k) dq_k + (\partial F_2 / \partial P_k) dP_k] - \sum_k [(dQ_k) P_k + Q_k (dP_k)] \\ &= \sum_k [(p_k dq_k + (Q_k dP_k)] - \sum_k [(dQ_k) P_k + Q_k (dP_k)] = \sum_k [p_k dq_k - P_k dQ_k]. \end{aligned} \quad (6.5.214)$$

Here we have used the first line of (5.5) which we assume defines F_2 up to a constant. Verify it follows from (5.214) that

$$\partial F_1(q, Q) / \partial q_k = p_k \text{ and } \partial F_1(q, Q) / \partial Q_k = -P_k, \quad (6.5.215)$$

in accord with the first line of (5.4). Thus, if the relation (5.213) can be realized, then use of the machinery (5.4) applied to the $F_1(q, Q)$ it provides also yields the map \mathcal{M} . Indeed, this will be the case if a type 1 generating function is compatible with \mathcal{M} . But can it happen that a type 1 generating function is *not* compatible with \mathcal{M} ? If so, then the attempted Ansatz (5.213) will fail. The answer is *yes!* For example, a type 2 generating function is compatible with $\mathcal{M} = \mathcal{I}$ but a type 1 is not.

Conversely, suppose a type 1 generating function is compatible with \mathcal{M} and we solve (5.213) for F_2 to yield the Ansatz

$$F_2(q, P) = F_1(q, Q) + \sum_k Q_k P_k, \quad (6.5.216)$$

which involves on the right side the known-to-exist generating function F_1 . Form differentials of both sides of (5.216) to find the relations

$$\begin{aligned} dF_2(q, P) &= \sum_k [(\partial F_1 / \partial q_k) dq_k + (\partial F_1 / \partial Q_k) dQ_k] + \sum_k [(dQ_k) P_k + Q_k (dP_k)] \\ &= \sum_k [p_k dq_k - P_k dQ_k] + \sum_k [(dQ_k) P_k + Q_k (dP_k)] = \sum_k [p_k dq_k] + [Q_k (dP_k)]. \end{aligned} \quad (6.5.217)$$

Here we have used the first line of (5.4), which we assume defines F_1 up to a constant. Verify it follows from (5.217) that

$$\partial F_2(q, P) / \partial q_k = p_k \text{ and } \partial F_2(q, P) / \partial P_k = Q_k, \quad (6.5.218)$$

in accord with the first line of (5.5). Thus, if the relation (5.216) can be realized, then use of the $F_2(q, P)$ it provides with the machinery (5.5) also yields the map \mathcal{M} . Indeed, this will be the case if a type 2 generating function is compatible with \mathcal{M} . But it can happen that a type 2 generating function is *not* compatible with \mathcal{M} . Then the attempted Ansatz (5.216) will fail.

Putting together everything we have found so far, we may completely summarize the relation between F_1 and F_2 by writing the plausible Legendre transformation relations

$$F_2(q, P) = F_1(q, Q) + \sum_k Q_k P_k \Leftrightarrow F_1(q, Q) = F_2(q, P) - \sum_k Q_k P_k. \quad (6.5.219)$$

Both relations will hold providing generating functions of types 1 and 2 are both compatible with \mathcal{M} , and both F_1 and F_2 will then produce the same map \mathcal{M} . And failure will occur if generating functions of types 1 and 2 are *not* both compatible with \mathcal{M} because then either the requisite F_j required to state an Ansatz does not exist or the F_k the Ansatz is supposed to produce does not exist.

Let us complete the list of Legendre transformations depicted in (5.212). Verify the further plausible Legendre transformation results

$$F_4(p, P) = F_2(q, P) + \sum_k q_k p_k \Leftrightarrow F_2(q, P) = F_4(p, P) - \sum_k q_k p_k, \quad (6.5.220)$$

$$F_3(p, Q) = F_4(p, P) - \sum_k Q_k P_k \Leftrightarrow F_4(p, P) = F_3(p, Q) + \sum_k Q_k P_k, \quad (6.5.221)$$

$$F_1(q, Q) = F_3(p, Q) - \sum_k q_k p_k \Leftrightarrow F_3(p, Q) = F_1(q, Q) + \sum_k q_k p_k. \quad (6.5.222)$$

There are also relations between the diagonally opposite entries in (5.212). Look at (5.212) and consider the triplet of generating functions F_1, F_2, F_4 encountered by going around (5.212) clockwise starting with F_1 . Assume that generating functions of types 1, 2, and 4 are compatible with \mathcal{M} . Combine the relations

$$F_4(p, P) = F_2(q, P) + \sum_k q_k p_k \quad (6.5.223)$$

and

$$F_2(q, P) = F_1(q, Q) + \sum_k Q_k P_k, \quad (6.5.224)$$

which are drawn from (5.220) and (5.2.19), to show that there is the more complicated (sometimes called *double*) Legendre transformation relation

$$F_4 = F_1 - \sum_k (q_k p_k - Q_k P_k) \Leftrightarrow F_1 = F_4 + \sum_k (q_k p_k - Q_k P_k). \quad (6.5.225)$$

Similarly, consider the triplet of generating functions F_1, F_3, F_4 encountered by going around (5.212) counterclockwise starting with F_1 . Using the Legendre transformation relations among them, again obtain (5.225). Note that while derivation of the relations (5.225) employed the use of (assumed compatible) generating functions of types 2 or 3 as *stepping stones*, generating functions of these types do not appear in the final results. One might wonder if all that is required for (5.225) to hold is that both generating functions of types 1 and 4 be compatible with \mathcal{M} . Verify that this is indeed the case. For example consider the first relation in (4.225),

$$F_4 = F_1 - \sum_k (q_k p_k - Q_k P_k). \quad (6.5.226)$$

Its right side is well defined under the assumption that a type 1 generating function is compatible with \mathcal{M} . Now form differentials of both sides of (5.226) to find the relations

$$\begin{aligned} dF_4(p, P) &= \sum_k [(\partial F_1 / \partial q_k) dq_k + (\partial F_1 / \partial Q_k) dQ_k] - d \sum_k (q_k p_k - Q_k P_k) \\ &= \sum_k [p_k dq_k - P_k dQ_k] - [(dq_k)p_k + q_k dp_k - (dQ_k)P_k - Q_k dP_k] \\ &= \sum_k [-q_k dp_k + Q_k dP_k]. \end{aligned} \quad (6.5.227)$$

Here, in moving from the first line in (5.227) to the second, we have used (5.4). Verify It follows from (5.227) that

$$\partial F_4(p, P) / \partial p_k = -q_k \text{ and } \partial F_4(p, P) / \partial P_k = Q_k, \quad (6.5.228)$$

in accord with (5.7). We see that the more complicated Legendre transformaton (5.226) succeeds provided a type 4 generating function is compatible with \mathcal{M} . Can it happen that for some \mathcal{M} there is a type 1 generating function that is compatible with \mathcal{M} but there is *no* type 4 generating that is compatible? The answer is *yes*. Consider the map \mathcal{M} whose linear part M has the form

$$M = \begin{pmatrix} I & B \\ C & I \end{pmatrix}, \quad (6.5.229)$$

with

$$\det(B) \neq 0 \quad (6.5.230)$$

and

$$C = 0. \quad (6.5.231)$$

There are symplectic matrices of this form (see Subsection 3.3.2) and correspondingly, the associated \mathcal{M} will be a symplectic map. From (5.4) and (5.230) we see that a type 1 generating function is compatible with this \mathcal{M} . But from (5.7) and (5.231) we see that any type 4 generating function is not compatible with this \mathcal{M} because $\det(C) = 0$. Therefore in this case the Ansatz (5.226) will fail.

Finally, analogous to the relation between F_1 and F_4 , show that for F_2 and F_3 there is the more complicated Legendre transformation relation

$$F_3 = F_2 + \sum_k \Leftrightarrow F_2 = F_3 + \sum_k. \quad (6.5.232)$$

We end this exercise by calling to attention an application of the result (5.213). Review Subsubsection 5.5.2 and verify that, in the context of Lagrangian dynamics, one may write

$$F_2(q, P, t) = \sum_k P_k Q_k - \int_{t^i}^t L(q, \dot{q}, \tau) d\tau. \quad (6.5.233)$$

Compare (5.213) with (5.233) to see that, in the context of Lagrangian dynamics, there is the relation

$$F_1(q, Q, t) = - \int_{t^i}^t L(q, \dot{q}, \tau) d\tau. \quad (6.5.234)$$

6.5.8. The purpose of this exercise is to study, as a sanity check on some of our previous work, a simple example case. It will be the relation between $F_1(q, Q)$ and $F_2(q, P)$ for linear (and symplectic) transformations acting on a two-dimensional phase space.

Evidently linear transformations are produced by quadratic generating functions. Therefore make for $F_2(q, P)$ the Ansatz

$$F_2(q, P) = aq^2 + 2bqP + cP^2. \quad (6.5.235)$$

Verify that using (5.212) yields the results

$$Q = \partial F_2 / \partial P = 2bq + 2cP \quad (6.5.236)$$

and

$$p = \partial F_2 / \partial q = 2aq + 2bP. \quad (6.5.237)$$

Verify that solving (5.225) and (5.226) for Q, P in terms of q, p yields the results

$$P = (p - 2aq)/(2b) = q[-(a/b)] + p[1/(2b)] \quad (6.5.238)$$

and

$$\begin{aligned} Q &= 2bq + 2cP = 2bq + 2c\{q[-(a/b)] + p[1/(2b)]\} \\ &= q[2b - 2a(c/b)] + p[(c/b)]. \end{aligned} \quad (6.5.239)$$

Write these relations in the vector-matrix form

$$\begin{pmatrix} Q \\ P \end{pmatrix} = M \begin{pmatrix} q \\ p \end{pmatrix} \quad (6.5.240)$$

to find the result

$$M = \begin{pmatrix} 2b - 2a(c/b) & c/b \\ -(a/b) & 1/(2b) \end{pmatrix}. \quad (6.5.241)$$

Verify that

$$\det(M) = 1 - ac/b^2 + ac/b^2 = 1, \quad (6.5.242)$$

and therefore M is symplectic. You have found the symplectic matrix M associated with the F_2 given by (5.224). Note that

$$\partial^2 F_2 / \partial q \partial P = 2b \quad (6.5.243)$$

and that M as given by (5.230) is undefined when $b = 0$, in accord with (5.5). Observe also that F_2 involves three parameters, which is consistent with $Sp(2, \mathbb{R})$ being three dimensional. However, the parameterization does not cover *all* of $Sp(2, \mathbb{R})$ because we know that there are matrices in $Sp(2, \mathbb{R})$ that are *incompatible* with generating functions of the type F_2 . For example, J is incompatible with the use of F_2 . See Subsubsection 5.1.4.

Now look at the associated $F_1(q, Q)$ given by (5.213). In the case of a two-dimensional phase space it takes the form

$$F_1(q, Q) = F_2(q, P) - QP. \quad (6.5.244)$$

To execute the calculation it calls for, it is necessary to evaluate the right side of (5.233) in terms of the quantities q and Q . To do so we need to find P in terms of q and Q . Looking at (5.225) we see that

$$P = [1/(2c)](Q - 2bq). \quad (6.5.245)$$

Verify It follows, using (5.224), that

$$\begin{aligned} F_2(q, P) &= aq^2 + 2bq[1/(2c)](Q - 2bq) + c\{[1/(2c)](Q - 2bq)\}^2 \\ &= [a - 2b^2(1/c) + b^2(1/c)]q^2 + 2[b/(2c) - b/(2c)]qQ + [1/(4c)]Q^2 \\ &= [a - b^2(1/c)]q^2 + [1/(4c)]Q^2 \end{aligned} \quad (6.5.246)$$

and

$$QP = Q[1/(2c)](Q - 2bq) = +2[-b/(2c)]qQ + [1/(2c)]Q^2. \quad (6.5.247)$$

Show that using (5.235) and (5.236) in (5.233) gives the result

$$\begin{aligned} F_1(q, Q) &= F_2(q, P) - QP = \\ &= [a - b^2(1/c)]q^2 + [1/(4c)]Q^2 - 2[-b/(2c)]qQ - [1/(2c)]Q^2. = \\ &= [a - b^2(1/c)]q^2 + 2[b/(2c)]qQ + [-1/(4c)]Q^2. \end{aligned} \quad (6.5.248)$$

Consequently F_1 may be written in the form

$$F_1(q, Q) = \alpha q^2 + 2\beta qQ + \gamma Q^2 \quad (6.5.249)$$

with

$$\alpha = [a - b^2(1/c)], \quad (6.5.250)$$

$$\beta = [b/(2c)], \quad (6.5.251)$$

$$\gamma = [-1/(4c)]. \quad (6.5.252)$$

Your next task is to find the symplectic matrix, call it N , associated with the F_1 given by (5.238). Use (5.4) and (5.238) to find that

$$p = \partial F_1 / \partial q = 2\alpha q + 2\beta Q \quad (6.5.253)$$

and

$$P = -\partial F_1 / \partial Q = -2\beta q - 2\gamma Q. \quad (6.5.254)$$

Solve (5.242) for Q to find

$$Q = [1/(2\beta)](p - 2\alpha q) = (-\alpha/\beta)q + [1/(2\beta)]p. \quad (6.5.255)$$

Substitute this result in (5.243) to find

$$\begin{aligned} P &= -2\beta q - 2\gamma Q = -2\beta q - 2\gamma[1/(2\beta)](p - 2\alpha q) = \\ &= [-2\beta + 2\alpha\gamma(1/\beta)]q + [-\gamma/\beta]p. \end{aligned} \quad (6.5.256)$$

Write the relations (5.244) and (5.245) in the vector-matrix form

$$\begin{pmatrix} Q \\ P \end{pmatrix} = N \begin{pmatrix} q \\ p \end{pmatrix}, \quad (6.5.257)$$

and show that

$$N = \begin{pmatrix} -\alpha/\beta & 1/(2\beta) \\ -2\beta + 2\alpha\gamma(1/\beta) & -\gamma/\beta \end{pmatrix}. \quad (6.5.258)$$

Verify that

$$\det(N) = 1 - \alpha\gamma(1/\beta^2) + \alpha\gamma(1/\beta^2) = 1, \quad (6.5.259)$$

and therefore N is symplectic. You have found the symplectic matrix N associated with the F_1 generating function given by (5.238). Note that

$$\partial^2 F_1 / \partial q \partial Q = 2\beta \quad (6.5.260)$$

and that N as given by (5.247) is undefined when $\beta = 0$, in accord with (5.4).

According to Exercise 5.7 there should be the relation

$$N = M. \quad (6.5.261)$$

Verify that (5.250) holds when the values (5.239) through (5.241) are employed in (5.247). That is, verify that

$$N_{11} = -\alpha/\beta = -[a - b^2/c][2c/b] = -2ac/b + 2b = M_{11}, \quad (6.5.262)$$

$$N_{12} = 1/(2\beta) = -[a - b^2/c][2c/b] = -2ac/b + 2b = M_{12}, \quad (6.5.263)$$

$$N_{21} = -2\beta + 2\alpha\gamma(1/\beta) = -[a - b^2/c][2c/b] = -2ac/b + 2b = M_{21}, \quad (6.5.264)$$

$$N_{22} = -\gamma/\beta = -[a - b^2/c][2c/b] = -2ac/b + 2b = M_{22}. \quad (6.5.265)$$

Finally, here is something to ponder: Suppose

$$M = I, \quad (6.5.266)$$

which will be the case when the values

$$b = 1/2 \text{ and } a = c = 0 \quad (6.5.267)$$

are employed in (5.224) and (5.230). But we know from earlier work that $N = I$ is *incompatible* with the use of F_1 . See Subsubsections 5.1.4 and 5.1.5. Therefore in this case we may expect that the Legendre transformation (5.233) will fail. Indeed verify that when (5.252) holds, then all the relations (5.239) through (5.241) are singular (have poles) at $c = 0$. The moral that is to be drawn from this exercise is that if generating functions of types 1 and 2 are both compatible with some symplectic map \mathcal{M} , and both are used to produce generating functions $F_1(q, Q)$ and $F_2(q, P)$ whose use yields the same map \mathcal{M} , then F_1 and F_2 are related by the Legendre transformations (5.218). But if one of the types 1 and 2 is incompatible with \mathcal{M} and the other is not, then an attempt to relate them by Legendre transformations (5.218) will fail. The same will be true for all pairs of generating functions possibly related by the Legendre transformations (5.218) through (5.223).

6.5.9. Let \mathcal{M} be the linear symplectic map described by the symplectic matrix M given by

$$M = (1/\sqrt{2})(I + J). \quad (6.5.268)$$

(See Exercise 3.1.5.) Show that *all* mixed-variable generating functions of types 1 through 4 are compatible with \mathcal{M} , and find the F_j in each case.

6.5.10. Exercise on example of (5.119). Consider a two-dimensional phase space described by a coordinate-like variable ξ and a conjugate momentum-like variable η . Let H be the Hamiltonian

$$H = (1/2)(\eta^2 + \xi^2). \quad (6.5.269)$$

Verify that the associated Lagrangian L is given by

$$L = (1/2)(\dot{\xi}^2 - \dot{\eta}^2). \quad (6.5.270)$$

Let the pair $\xi(t), \eta(t)$ be the solution to the equations of motion generated by H , let $t^i = 0$ be the initial time, and let q, p be the initial conditions so that

$$q = \xi(0) \text{ and } p = \eta(0). \quad (6.5.271)$$

Review Exercise * and apply its results to find that the solution to Hamilton's equations for the specified initial conditions is given by

$$\xi(t) = q \cos(t) + p \sin(t) \quad (6.5.272)$$

and

$$\eta(t) = -q \sin(t) + p \cos(t). \quad (6.5.273)$$

Next, let $A(q, p)$ be the action for this phase-space trajectory,

$$A(q, p) = \int_0^{t^f} L dt, \quad (6.5.274)$$

where t^f is some final time. Let us compute the ingredients of L . From (5.212) we find

$$\begin{aligned} \xi^2 &= q^2 \cos^2(t) + 2qp \cos(t) \sin(t) + p^2 \sin^2(t) \\ &= q^2(1/2)[1 + \cos(2t)] + qp \sin(2t) + p^2(1/2)[1 - \cos(2t)]. \end{aligned} \quad (6.5.275)$$

And from (5.212) we also find

$$\dot{\xi}(t) = -q \sin(t) + p \cos(t) \quad (6.5.276)$$

so that

$$\begin{aligned} \dot{\xi}^2 &= q^2 \sin^2(t) - 2qp \cos(t) \sin(t) + p^2 \cos^2(t) \\ &= q^2(1/2)[1 - \cos(2t)] - qp \sin(2t) + p^2(1/2)[1 + \cos(2t)]. \end{aligned} \quad (6.5.277)$$

Consequently

$$L = \cos(2t)(p^2 - q^2) + qp \sin(2t). \quad (6.5.278)$$

$$A(q, p) = \int_0^{t^f} L dt = \sin(2t^f)(p^2 - q^2) + [1 - \cos(2t^f)]qp. \quad (6.5.279)$$

Also write

$$Q = \xi(t) \text{ and } P = \eta(t) \quad (6.5.280)$$

where t is some general time. Review Exercise * and apply its results to find that

$$Q(t = q \cos(t) + p \sin(t)) \quad (6.5.281)$$

and

$$P = \dots \quad (6.5.282)$$

$Q = \dots$

6.6 Generating Functions Come from an Exact Differential

6.6.1 Overview

So far the discussion of generating functions has been relatively straight forward, but not particularly illuminating. Let us write (5.3) in the form

$$Z = \mathcal{M}z. \quad (6.6.1)$$

By this relation we mean that there are $2n$ functions $K_a(z, t)$ of the $2n$ variables z_b , and perhaps the time t , such that

$$Z_a = K_a(z, t). \quad (6.6.2)$$

That is, in general $2n$ functions are required to specify a map in $2n$ variables.

However in the last section we have seen that, with the use of any one of the generating functions F_1 through F_4 , all the required $2n$ functions come from a *single* master function, namely the generating function being employed. How does it happen that the information required to specify $2n$ functions can come from a single function? Presumably this occurs because in our case the $2n$ functions $K_a(z, t)$ are *not*, in fact, independent. Of course, in principle we know that they are not independent because of the assumption that \mathcal{M} is symplectic. Apparently the symplectic condition is so stringent as to reduce the number of required functions down from $2n$ to a *single* function. In one sense this should not be too surprising, because we know that any family of symplectic maps $\mathcal{M}(t)$ is generated by a single function, namely the Hamiltonian. But that is an infinitesimal statement. How, precisely, could one have guessed that there were functions F_1 through F_4 that could be used in the manner (5.4) through (5.7) to manufacture symplectic maps? And can all symplectic maps be obtained in this fashion? Below we present a partial clue. Still deeper insight is presented in Subsection 7.1. There we will learn that the functions F_1 through F_4 are but 4 members of a $2n(4n + 1)$ parameter family of types of generating functions, all of which can be used to manufacture symplectic maps.¹⁹ The final explanation is given in Subsection 7.2.

6.6.2 A Democratic Differential Form

6.6.2.1 Definition

Consider the differential form

$$\omega_d = (Z, JdZ) - (z, Jdz). \quad (6.6.3)$$

Note that ω_d involves all the $4n$ variables z and Z . It has the beauty that it treats the coordinates and momenta on an equal footing, and is “*democratic*” in its use of z and Z . Also, we will see in Subsection 7.2 that it arises in a natural way.²⁰

6.6.2.2 The Democratic Differential Form Is Exact Iff \mathcal{M} Is Symplectic

Suppose the Z ’s are viewed as functions of the z ’s by using (5.23) to write ω_d as

$$\omega_d(z) = (Z, JMdz) - (z, Jdz). \quad (6.6.4)$$

Then, if the \mathcal{M} in (6.1) is a symplectic map, we will find that ω_d is *exact* with respect to the $2n$ variables z . Similarly, if the z ’s are viewed as functions of the Z ’s, ω_d can be rewritten as

$$\omega_d(Z) = (Z, JdZ) - (z, JM^{-1}dZ). \quad (6.6.5)$$

¹⁹Here $2n$ is the phase-space dimension.

²⁰Despite its attractive appearance, the differential form ω_d given by (6.3) is not commonly employed by (and perhaps unfamiliar to some) other authors. Its existence, exactness, and utility were known, however, to Poincaré.

It can be shown that this form is also exact (with respect to the $2n$ variables Z) if \mathcal{M} is a symplectic map. We will soon verify these claims by brute calculation. Subsection 7.2 will find the same results in an obvious way.

To see that ω_d is exact with respect to the variables z , observe that that it can be written more explicitly as

$$\begin{aligned}\omega_d(z) &= (Z, JM dz) - (z, J dz) = (M^T J^T Z, dz) - (J^T z, dz) \\ &= \sum_b [(M^T J^T Z)_b - (J^T z)_b] dz_b.\end{aligned}\quad (6.6.6)$$

Upon comparing (6.6) with (1.22), we see that the coefficients $C_b(z, t)$ are given by the relation

$$C_b(z, t) = (M^T J^T Z)_b - (J^T z)_b. \quad (6.6.7)$$

Note that there is a possible time dependence since \mathcal{M} may depend on t . However, as before, t only plays the role of a parameter.

We must see if the conditions (1.26) are met. An easy computation gives for the second term in (6.7) the result

$$(\partial/\partial z_a)(J^T z)_b = (\partial/\partial z_a) \sum_c (J^T)_{bc} z_c = \sum_c (J^T)_{bc} \delta_{ac} = (J^T)_{ba} = J_{ab}. \quad (6.6.8)$$

Dealing with the first term in (6.7) is more complicated. We find the preliminary result

$$\begin{aligned}(\partial/\partial z_a)(M^T J^T Z)_b &= (\partial/\partial z_a) \sum_c (M^T J^T)_{bc} Z_c = \sum_c [(\partial/\partial z_a)(M^T J^T)_{bc}] Z_c \\ &\quad + \sum_c (M^T J^T)_{bc} (\partial/\partial z_a) Z_c.\end{aligned}\quad (6.6.9)$$

But, from (5.23), there is the relation

$$(\partial/\partial z_a) Z_c = M_{ca}. \quad (6.6.10)$$

It follows that for the second term on the right side of (6.9) there is the simplification

$$\sum_c (M^T J^T)_{bc} (\partial/\partial z_a) Z_c = \sum_c (M^T J^T)_{bc} M_{ca} = (M^T J^T M)_{ba} = (M^T JM)_{ab}. \quad (6.6.11)$$

For the first term on the right side of (6.9) there is the result

$$\begin{aligned}\sum_c [(\partial/\partial z_a)(M^T J^T)_{bc}] Z_c &= \sum_c Z_c (\partial/\partial z_a) (JM)_{cb} = \sum_{cd} Z_c J_{cd} (\partial/\partial z_a) M_{db} \\ &= \sum_{cd} Z_c J_{cd} (\partial^2 Z_d / \partial z_a \partial z_b).\end{aligned}\quad (6.6.12)$$

Here we have again used a variant of (6.10). Combining (6.8) (6.9), (6.11), and (6.12) gives the net result

$$(\partial/\partial z_a) C_b = [(M^T JM)_{ab} - J_{ab}] + [\sum_{cd} Z_c J_{cd} (\partial^2 Z_d / \partial z_a \partial z_b)]. \quad (6.6.13)$$

Here we have separated the right side of (6.13) into parts that are antisymmetric and symmetric under the interchange of a and b . It follows that there is the relation

$$(\partial/\partial z_a)C_b - (\partial/\partial z_b)C_a = 2[M^TJM - J]_{ab}. \quad (6.6.14)$$

Consequently the differential form $\omega_d(z)$ given by (6.4) is exact if, and only if, the map \mathcal{M} is symplectic.

It can be shown in a similar way that the differential form $\omega_d(Z)$ given by (6.5) is exact if, and only if, the map \mathcal{M} is symplectic.

6.6.3 Information about \mathcal{M} Carried by the Democratic Form

Since (6.3) is exact, there is a function $F(z, t)$ such that

$$dF = \omega_d = (Z, JdZ) - (z, Jdz). \quad (6.6.15)$$

The function $F(z, t)$ may be called the *primitive* function associated with the differential form ω_d .²¹

How much information does $F(z, t)$ carry about \mathcal{M} ? Put another way, by its definition, the differential form ω_d depends on \mathcal{M} . Are there possibly several maps \mathcal{M} that produce the same differential form ω_d ? According to (5.23) and (6.1) we may rewrite write (6.15) in the form

$$dF_{\mathcal{M}} = (\mathcal{M}z, JMdz) - (z, Jdz) \quad (6.6.16)$$

where have have appended the subscript \mathcal{M} to F to indicate that F depends on \mathcal{M} . We will begin our exploration of this uniqueness question by considering various symplectic maps \mathcal{M} .

Suppose \mathcal{M} is a member of the *inhomogeneous* symplectic group $ISp(2n, \mathbb{R})$. See Sub-section 2.2 and Section 9.2. At this point it is convenient to use Lie-algebraic notation and tools. See Chapter 7 for details. Let f_1 be a first-degree polynomial such that

$$\exp(: f_1 :)z = z + \delta. \quad (6.6.17)$$

We define a *translation* operator \mathcal{T} by writing

$$\mathcal{T} = \exp(: f_1 :) \quad (6.6.18)$$

so that \mathcal{T} has the action

$$\mathcal{T}z = z + \delta. \quad (6.6.19)$$

Also, let

$$\mathcal{R}_f = \exp(: f_2^c :) \exp(: f_2^a :), \quad (6.6.20)$$

be a general linear symplectic map. It has the action

$$\mathcal{R}_f z = R_f z \quad (6.6.21)$$

²¹In calculus parlance the terms antiderivative, primitive function, primitive integral, and indefinite integral are used interchangeably.

where R_f is a general symplectic matrix. From the work of Section 9.2 we know that any element in $ISp(2n, \mathbb{R})$ can be written in the factored form

$$\mathcal{M}_f = \mathcal{T}\mathcal{R} \quad (6.6.22)$$

and has the action

$$Z = \mathcal{M}_f z = R_f \delta + R_f z. \quad (6.6.23)$$

Let us employ this result in (6.16). We observe from (6.23) that there is the relation

$$M = R_f. \quad (6.6.24)$$

Therefore, in this case, (6.16) takes the form

$$dF_{\mathcal{M}_f} = (R_f \delta + R_f z, JR_f dz) - (z, Jdz) = (R_f \delta, JR_f dz) + (R_f z, JR_f dz) - (z, Jdz). \quad (6.6.25)$$

But, employing the symplectic condition for R_f yields the results

$$(R_f \delta, JR_f dz) = (\delta, R_f^T JR_f dz) = (\delta, Jdz), \quad (6.6.26)$$

$$(R_f z, JR_f dz) - (z, Jdz) = (z, R_f^T JR_f dz) - (z, Jdz) = (z, Jdz) - (z, Jdz) = 0. \quad (6.6.27)$$

Consequently, (6.25) becomes

$$dF_{\mathcal{M}_f} = (\delta, Jdz) \quad (6.6.28)$$

and therefore

$$F_{\mathcal{M}_f} = (\delta, Jz) + C \quad (6.6.29)$$

where C is an arbitrary additive constant.

What can we conclude from looking at (6.28)? First, suppose there is no translation part so that $\delta = 0$ and $\mathcal{T} = \mathcal{I}$. Then we see from (6.22) that

$$\mathcal{M}_f = \mathcal{R}_f, \quad (6.6.30)$$

and it follows from (6.28) that

$$dF_{\mathcal{R}_f} = 0. \quad (6.6.31)$$

Consequently, up to an inconsequential additive constant, all *linear* symplectic maps produce the *same* primitive function and this primitive function may be taken to be *zero*. Second, (6.29) can be rewritten in the form

$$dF_{\mathcal{M}_f} = dF_{\tau\mathcal{R}_f} = dF_\tau \quad (6.6.32)$$

with

$$dF_\tau = (\delta, Jdz). \quad (6.6.33)$$

That is, in the case of $ISp(2n, \mathbb{R})$ and when the factorization (6.22) is employed, the differential form $dF_{\mathcal{M}_f}$ depends *only* on the *translation* part of \mathcal{M}_f .

The results obtained this far suggest the following exploration: Suppose \mathcal{N} is any symplectic map.²² We will write \mathcal{N} in the Lie form

$$\mathcal{N} = \exp(: f_1 :) \exp(: f_2^c :) \exp(: f_2^a :) \exp(: f_3 :) \exp(: f_4 :) \cdots. \quad (6.6.34)$$

²²That is, using the notation introduced in Subsection 2.2, assume only that $\mathcal{N} \in ISpM(2n, \mathbb{R})$.

Now define a map \mathcal{M} by writing

$$\mathcal{M} = \mathcal{N}\mathcal{L} \quad (6.6.35)$$

where \mathcal{L} is a *linear* symplectic map with the action

$$\mathcal{L}z = Lz. \quad (6.6.36)$$

That is, $L \in Sp(2n, \mathbb{R})$. Let us consider $dF_{\mathcal{M}}$ in this case,

$$dF_{\mathcal{M}} = (\mathcal{M}z, JMdz) - (z, Jdz) \quad (6.6.37)$$

where M is the Jacobian of \mathcal{M} .

Because \mathcal{L} is a linear map, we may write

$$\mathcal{M}z = L\mathcal{N}z. \quad (6.6.38)$$

Also, by the chain rule, we know that M is given by the relation

$$M = LN \quad (6.6.39)$$

where N is the Jacobian of \mathcal{N} . Next make use of (6.38) and (6.39) and the symplectic condition to carry out the series of deductions

$$(\mathcal{M}z, JMdz) = (LNz, JLNdz) = (\mathcal{N}z, L^T JLNdz) = (\mathcal{N}z, JNdz). \quad (6.6.40)$$

From (6.40) we conclude that

$$dF_{\mathcal{M}} = dF_{\mathcal{N}\mathcal{L}} = dF_{\mathcal{N}}. \quad (6.6.41)$$

[Note that (6.32) is a special case of (6.41).] Thus, up to a possible additive constant which is of no consequence, $F(z, t)$ is the *same* for all maps \mathcal{M} obtained by *right* multiplication of \mathcal{N} by linear symplectic maps \mathcal{L} . In coset language, F depends only on the left cosets $ISpM(2n, \mathbb{R})/Sp(2n, \mathbb{R})$. Recall Section 5.12 for a discussion of cosets.

6.6.4 Breaking the Degeneracy

Although the fact that many symplectic maps lead to the same F may seem alarming, we shall soon be adding other functions to F that will break this degeneracy. Let us rewrite (6.15) in terms of the q, p and the Q, P . It is easily verified that

$$(z, Jdz) = \sum_i (q_i dp_i - p_i dq_i) \text{ and } (Z, JdZ) = \sum_i (Q_i dP_i - P_i dQ_i). \quad (6.6.42)$$

It follows that

$$dF = \sum_i [(Q_i dP_i - P_i dQ_i) - (q_i dp_i - p_i dq_i)]. \quad (6.6.43)$$

This is a key result from which all the relations (5.4) through (5.7) follow.

6.6.4.1 F_2 Example

For example, suppose we define F_2 by the rule

$$F_2 = [F + \sum_i (p_i q_i + P_i Q_i)]/2. \quad (6.6.44)$$

That is we have added $\sum_i (p_i q_i + P_i Q_i)$ to F . Then we find the result

$$dF_2 = [dF + \sum_i (p_i dq_i + q_i dp_i + P_i dQ_i + Q_i dP_i)]/2 = \sum_i (p_i dq_i + Q_i dP_i). \quad (6.6.45)$$

Evidently the left side of (6.45) is an exact differential, and comparison with (5.71) provides an independent proof that the differential form ω_2 given by (5.81) is exact. Finally, from (6.45), we immediately derive the relations (5.5).

6.6.4.2 F_1 Example

As a second example, suppose we define F_1 by the rule

$$F_1 = [F + \sum_i (p_i q_i - P_i Q_i)]/2. \quad (6.6.46)$$

Then we find the result

$$dF_1 = [dF + \sum_i (p_i dq_i + q_i dp_i - P_i dQ_i - Q_i dP_i)]/2 = \sum_i (p_i dq_i - P_i dQ_i). \quad (6.6.47)$$

From (6.47) it follows that

$$\partial F_1 / \partial q_i = p_i, \quad \partial F_1 / \partial Q_i = -P_i, \quad (6.6.48)$$

which are the relations (5.4). The relations (5.6) and (5.7) follow in a similar fashion. See Exercise 6.4.

6.6.4.3 Poincaré Generating Function

There are also other generating functions beside F_1 through F_4 that are less familiar. For example, consider the function F_+ defined by the rule

$$F_+ = F + (Z, Jz) = F + \sum_i (p_i Q_i - q_i P_i). \quad (6.6.49)$$

For F_+ we find the differential

$$\begin{aligned} dF_+ &= dF + \sum_i [(p_i dQ_i + Q_i dp_i) - (P_i dq_i + q_i dP_i)] \\ &= \sum_i [(Q_i dP_i - P_i dQ_i) - (q_i dp_i - p_i dq_i) + (p_i dQ_i + Q_i dp_i) - (P_i dq_i + q_i dP_i)] \\ &= \sum_i [(Q_i - q_i)(dp_i + dP_i) - (P_i - p_i)(dq_i + dQ_i)]. \end{aligned} \quad (6.6.50)$$

We may therefore write

$$F_+ = F_+[(q + Q), (p + P), t] \quad (6.6.51)$$

to obtain from (6.50) the relations

$$Q_i - q_i = \partial F_+ / \partial (p_i + P_i), \quad (6.6.52)$$

$$P_i - p_i = -\partial F_+ / \partial (q_i + Q_i). \quad (6.6.53)$$

The function F_+ is sometimes called a *Poincaré generating function*. It is more democratic than the functions F_1 through F_4 in the sense that it involves the old and new variables in a symmetric fashion. Indeed, introduce the quantities Σ and Δ by the rules

$$\Sigma = Z + z, \quad (6.6.54)$$

$$\Delta = Z - z. \quad (6.6.55)$$

Then for (6.51) we may write

$$F_+ = F_+(\Sigma, t). \quad (6.6.56)$$

Correspondingly, the relations (6.52) and (6.53) can be written in the compact form

$$\Delta = J \partial_\Sigma F_+ |_{\Sigma=Z+z}. \quad (6.6.57)$$

Here, by employing J in (6.57), we have viewed $\Delta = \{(Q - q), (P - p)\}$ as composed of a canonical pair. We remark that if F_+ is a quadratic function of Σ , then the use of (6.57) leads to the Cayley transformation. See Exercise 6.5.

There is an important feature of the Poincaré generating function that is sometimes useful. A point Σ^c is called a *critical point* of F_+ if there is the result

$$\partial_\Sigma F_+ |_{\Sigma=\Sigma^c} = 0. \quad (6.6.58)$$

From (6.57) we see that at a critical point $\Delta = 0$ so that

$$Z = \mathcal{M}z = z \quad (6.6.59)$$

and, by (6.54),

$$Z = z = \Sigma^c / 2. \quad (6.6.60)$$

Thus, if we make the definition

$$z^f = \Sigma^c / 2, \quad (6.6.61)$$

we see that z^f is a *fixed point* of \mathcal{M} . We conclude that critical points of F_+ correspond to fixed points of \mathcal{M} and vice versa. This result can be useful in some circumstances because there are theorems (e.g. *Morse theory*) about critical points of smooth functions on various manifolds.

Exercises

6.6.1. Show that the differential form $\omega_d(Z)$ given by (6.5) is exact in terms of the variables Z .

6.6.2. Verify the claims (6.38) and (6.39).

6.6.3. Review Section 1.2.3. Define maps \mathcal{S} and $\mathcal{R}(\phi)$ by the rules

$$\mathcal{S} = \exp(:q^3:), \quad (6.6.62)$$

$$\mathcal{R}(\phi) = \exp[-(\phi/2):p^2 + q^2:] \quad (6.6.63)$$

Let \mathcal{M} be the map

$$\mathcal{M} = \mathcal{S}\mathcal{R}(\phi). \quad (6.6.64)$$

Verify that the maps (1.2.50) and (6.64) are related by a similarity transformation involving a map of the form $\mathcal{R}(\psi)$, which amounts to changing the observation point O .

Show that for the democratic differential form ω_d given by (6.3) there are the results

$$dF_{\mathcal{M}} = dF_{\mathcal{S}\mathcal{R}(\phi)} = dF_{\mathcal{S}} = , \quad (6.6.65)$$

$$F_{\mathcal{M}} = . \quad (6.6.66)$$

Why is there no ϕ dependence?

6.6.4. Consider the functions F_3 and F_4 defined by the relations

$$F_3 = [F - \sum_i (p_i q_i + P_i Q_i)]/2, \quad (6.6.67)$$

$$F_4 = [F - \sum_i (p_i q_i - P_i Q_i)]/2. \quad (6.6.68)$$

where dF is the exact differential (6.15). Show that they satisfy (5.6) and (5.7). Compare (6.67) and (6.68) with (6.44) and (6.46). Observe that all these relations differ only in the signs assigned to the quantities $p_i q_i$ and $P_i Q_i$, and that the four possibilities yield the four functions F_1 through F_4 .

6.6.5. Suppose that the Poincaré generating function F_+ is of the form

$$F_+(\Sigma) = (1/2)(\Sigma, W\Sigma) \quad (6.6.69)$$

where W is a symmetric matrix. It follows from (6.69) that in this case

$$\partial_{\Sigma} F_+ = W\Sigma. \quad (6.6.70)$$

Show that this F_+ , when employed in (6.57), produces the result

$$\Delta = JW\Sigma, \quad (6.6.71)$$

which in turn yields the relation

$$Z - z = JW(Z + z). \quad (6.6.72)$$

Solve this relation for Z in terms of z to yield the linear relation

$$Z = (I - JW)^{-1}(I + JW)z. \quad (6.6.73)$$

Finally write the relation between Z and z in terms of a matrix M ,

$$Z = Mz. \quad (6.6.74)$$

Comparison of (6.73) and (6.74) then gives the result

$$M = (I - JW)^{-1}(I + JW). \quad (6.6.75)$$

Observe that this result is the Cayley representation (3.12.5). It follows, as expected, that M is a symplectic matrix. Verify that the relation (6.75) can be inverted to give the result

$$W = -J(M - I)(M + I)^{-1} \quad (6.6.76)$$

in agreement with (3.12.19). Verify that there are the Cayley Möbius transformation relations

$$W = T_\sigma(M) \quad (6.6.77)$$

and

$$M = T_{\sigma^{-1}}(W) \quad (6.6.78)$$

with σ given by (5.13.11).

Suppose that F_+ is of the form

$$F_+(\Sigma) = (v, \Sigma) + (1/2)(\Sigma, W\Sigma) \quad (6.6.79)$$

where v is any vector. It follows from (6.79) that in this case

$$\partial_\Sigma F_+ = v + W\Sigma. \quad (6.6.80)$$

Show that this F_+ , when employed in (6.57), produces the result

$$Z = Mz + (I - JW)^{-1}Jv. \quad (6.6.81)$$

6.6.6. Verify, using the methods of Subsection 5.1, that the Poincaré generating function F_+ when employed in (6.57) does indeed produce a symplectic map. Suppose that F_+ is time dependent so that its use produces a one-parameter family of symplectic maps. Find the associated generating Hamiltonian for this family. Hint: If you are stuck, see Subsection 7.3.1.

6.6.7. Find a generating function F_- analogous to F_+ . That is, in (6.49), replace $+(Z, Jz)$ by $-(Z, Jz)$.

6.7 Plethora of Generating Functions

We have seen that there are the five generating function types F_1 through F_4 and F_+ . We will now learn that (for a $2n$ -dimensional phase space) there are an *infinite* number of generating functions types, of which the five cited above are but examples, that comprise a full $2n(4n + 1)$ parameter family. Indeed, there is a generating function type for each of the Darboux matrices/transformations of Section 5.13. And we know there is a distinct Darboux matrix corresponding to each element of $Sp(4n)$ whose dimension is $2n(4n + 1)$. See (5.13.28), (3.7.35), and Table 3.7.1. Thus, for example, in the simplest case of a two-dimensional phase space, there is a 10 parameter family of generating function types; and in the case of a four-dimensional phase space there is a 36 parameter family of generating function types; and in the case of a six-dimensional phase space there is a 78 parameter family of generating function types. We will begin with a derivation of this result, and then follow the derivation with a discussion describing how the various results we have found all fit together.

6.7.1 Derivation

Consider again the differential form ω_d given by (6.3). If Z and z are $2n$ dimensional, let \hat{Z} denote the $4n$ dimensional column vector *constructed* by appending the entries of z below those of Z ,

$$\hat{Z} = (Z; z)^T. \quad (6.7.1)$$

With this notation, the differential form ω_d can be rewritten as

$$\omega_d = (Z, JdZ) - (z, Jdz) = (\hat{Z}, \tilde{J}^{4n} d\hat{Z}). \quad (6.7.2)$$

Here, as in section 5.13.3, we have used (5.13.16) to define \tilde{J}^{4n} .

Let α denote a $4n \times 4n$ invertible matrix. Use α to define new variables \hat{U} by the rule

$$\hat{U} = \alpha \hat{Z}, \quad (6.7.3)$$

or

$$\hat{Z} = \alpha^{-1} \hat{U}. \quad (6.7.4)$$

With this change of variables the differential form on the right side of (7.2) becomes

$$(\hat{Z}, \tilde{J}^{4n} d\hat{Z}) = (\alpha^{-1} \hat{U}, \tilde{J}^{4n} \alpha^{-1} d\hat{U}) = (\hat{U}, (\alpha^{-1})^T \tilde{J}^{4n} (\alpha^{-1}) d\hat{U}). \quad (6.7.5)$$

Next, inspired by the discussion of Section 5.13, require that α satisfy the relation

$$(\alpha^{-1})^T \tilde{J}^{4n} (\alpha^{-1}) = J^{4n}. \quad (6.7.6)$$

That is, require that α be a Darboux transformation. Also, in analogy with (7.1), introduce $4n$ dimensional vectors \hat{U} by letting U and u be the first $2n$ entries and last $2n$ entries of \hat{U} , respectively,

$$\hat{U} = (U; u)^T. \quad (6.7.7)$$

Upon combining (7.5) through (7.7), we find the result

$$(\hat{Z}, \tilde{J}^{4n} d\hat{Z}) = (\hat{U}, (\alpha^{-1})^T \tilde{J}^{4n} (\alpha^{-1}) d\hat{U}) = (\hat{U}, J^{4n} d\hat{U}) = (U, du) - (u, dU). \quad (6.7.8)$$

We know that $(\hat{Z}, \tilde{J}^{4n} d\hat{Z})$ is exact. See (6.15). Therefore, from the work so far, we have the relation

$$dF = (U, du) - (u, dU). \quad (6.7.9)$$

From F construct another function g by the rule

$$g = [F + (U, u)]/2. \quad (6.7.10)$$

That is, we have added (U, u) to F . Then, by this construction and the properties of F , g has the differential

$$dg = [dF + (U, du) + (u, dU)]/2 = [(U, du) - (u, dU) + (U, du) + (u, dU)]/2 = (U, du). \quad (6.7.11)$$

By the chain rule, we also have the relation

$$dg = \sum_a [(\partial g / \partial U_a)(dU_a) + (\partial g / \partial u_a)(du_a)]. \quad (6.7.12)$$

Upon comparing (7.11) and (7.12) we deduce the two relations

$$\partial g / \partial U_a = 0, \quad (6.7.13)$$

$$U_a = \partial g / \partial u_a. \quad (6.7.14)$$

The first of these states the remarkable result that g depends *only* on u ,

$$g = g(u). \quad (6.7.15)$$

The second states that U and u are related by the *gradient* map \mathcal{G} produced by the function g playing the role of a source function. More abstractly, we may write (7.14) in the form

$$U = \mathcal{G}u. \quad (6.7.16)$$

See Subsection 1.1.

When written in block form, (7.3) is equivalent to the two relations

$$U = A^\alpha Z + B^\alpha z, \quad (6.7.17)$$

$$u = C^\alpha Z + D^\alpha z. \quad (6.7.18)$$

When expanded in component form, (7.17) reads

$$U_a = \sum_b [(A^\alpha)_{ab} Z_b + (B^\alpha)_{ab} z_b], \quad (6.7.19)$$

and (7.18) reads

$$u_c = \sum_d [(C^\alpha)_{cd} Z_d + (D^\alpha)_{cd} z_d]. \quad (6.7.20)$$

In terms of these components, the relations (7.14) become

$$\sum_b [(A^\alpha)_{ab} Z_b + (B^\alpha)_{ab} z_b] = [\partial g / \partial u_a]|_{u=C^\alpha Z + D^\alpha z}. \quad (6.7.21)$$

In matrix-vector form they can be written more compactly as

$$A^\alpha Z + B^\alpha z = \partial_u g|_{u=C^\alpha Z + D^\alpha z}. \quad (6.7.22)$$

Equations (7.21) provide $2n$ *implicit* relations between Z and z . When made *explicit*, they produce the map \mathcal{M} (which will soon be shown to be symplectic) with

$$Z = \mathcal{M}z. \quad (6.7.23)$$

Equations (7.21) relate the symplectic map \mathcal{M} to the gradient map \mathcal{G} associated with the source function g . Any such function produces a symplectic map, and we will call g , in association with the Darboux matrix α , the *generating* function that produces \mathcal{M} . Note that in principle, although we have not taken note of it until now, g could also depend on the time t ,

$$g = g(u, t). \quad (6.7.24)$$

Here t should again be regarded as a parameter, and its presence in g when employed in (7.14) and (7.21) leads to a parameter dependent gradient map $\mathcal{G}(t)$ and a parameter dependent symplectic map $\mathcal{M}(t)$.

At this point we note that there is another way of viewing the relations (7.21). Suppose we pick a $2n$ -vector u and, if \mathcal{G} depends on t , also specify a value for t . Then, according to (7.16), we can determine $U(u, t)$ by the rule

$$U(u, t) = \mathcal{G}(t)u. \quad (6.7.25)$$

In view of (7.7), we have now also specified $\hat{U}(u, t)$. Indeed, we have the relation

$$\hat{U}(u, t) = (U(u, t); u)^T = (\mathcal{G}(t)u; u)^T. \quad (6.7.26)$$

Next, use the Darboux relation (7.4) to determine $\hat{Z}(u, t)$ by writing

$$\hat{Z}(u, t) = \alpha^{-1} \hat{U}(u, t). \quad (6.7.27)$$

When written in block form, (7.27) yields the relations

$$Z(u, t) = A^{\alpha^{-1}} \mathcal{G}(t)u + B^{\alpha^{-1}} u, \quad (6.7.28)$$

and

$$z(u, t) = C^{\alpha^{-1}} \mathcal{G}(t)u + D^{\alpha^{-1}} u. \quad (6.7.29)$$

We see that (7.28) and (7.29) specify the map $\mathcal{M}(t)$ in *parametric* form with u being a set of $2n$ parameters.

We still have to verify that the implicit relations (7.21), or the parametric relations (7.28) and (7.29), can be made explicit. That is, given the z 's, we need to show that we can solve

for the Z 's, and vice versa. At this point one might ask if there is a choice of α such that the relation (7.21) is already explicit. Exercise 7.3 shows that this is impossible. There is no such α that also satisfies the Darboux requirement (5.13.20). Thus, we must begin with implicit relations. We also note that gradient map $\mathcal{G}(t)$ employed in (7.21) is explicit in form. Therefore, as we will next see, the implicit nature of (7.21) with regard to the z 's and Z 's is controlled primarily by the properties of the associated Darboux matrix α and its related Möbius transformations T_α and $T_{\alpha^{-1}}$.

Suppose we make small variations dz in the variables z thereby producing small variations dZ in the variables Z . Then, by the inverse function theorem, we must show that there is a relation of the form

$$dZ = M dz \quad (6.7.30)$$

where the matrix M is invertible. From (1.5) we find that

$$dU = G du. \quad (6.7.31)$$

And, from (7.19) and (7.20), we have the results

$$dU = A^\alpha dZ + B^\alpha dz, \quad (6.7.32)$$

and

$$du = C^\alpha dZ + D^\alpha dz. \quad (6.7.33)$$

Combining (7.31) through (7.33) gives the series of results

$$A^\alpha dZ + B^\alpha dz = G(C^\alpha dZ + D^\alpha dz), \quad (6.7.34)$$

$$(A^\alpha - GC^\alpha)dZ = (GD^\alpha - B^\alpha)dz, \quad (6.7.35)$$

$$dZ = [(A^\alpha - GC^\alpha)^{-1}(GD^\alpha - B^\alpha)]dz. \quad (6.7.36)$$

Comparison of (7.30) and (7.36) gives the relation

$$M = (A^\alpha - GC^\alpha)^{-1}(GD^\alpha - B^\alpha). \quad (6.7.37)$$

But, by (5.11.23) with the substitutions $M \rightarrow \alpha$ and $U' \rightarrow G$, there is the identity

$$(A^\alpha - GC^\alpha)^{-1}(GD^\alpha - B^\alpha) = (A^{\alpha^{-1}}G + B^{\alpha^{-1}})(C^{\alpha^{-1}}G + D^{\alpha^{-1}})^{-1}. \quad (6.7.38)$$

Therefore we may also write

$$M = (A^{\alpha^{-1}}G + B^{\alpha^{-1}})(C^{\alpha^{-1}}G + D^{\alpha^{-1}})^{-1}. \quad (6.7.39)$$

We conclude that M and G are related by the Möbius transformation $T_{\alpha^{-1}}$,

$$M = T_{\alpha^{-1}}(G). \quad (6.7.40)$$

See also Exercise 7.4.

We already know that G is symmetric. See (1.7). Moreover, from the work of Section 5.13, we know that Möbius transformations of the form $T_{\alpha^{-1}}$ can be found such that (7.40) is

well defined, and we know that these Möbius transformations send symmetric matrices into symplectic matrices. It follows that M is a symplectic matrix, and therefore M is invertible. We have shown that the implicit relations (7.21) can be made explicit so that we may indeed write (7.23). Moreover, since M is a symplectic matrix, we have also verified that \mathcal{M} is a symplectic map.

The discourse so far described how, given the gradient map \mathcal{G} associated with any source function g and a Darboux matrix α , we can construct a symplectic map \mathcal{M} by use of (7.21). In the spirit of Subsection 5.2, suppose we are instead given \mathcal{M} and we wish to construct g . Begin with the relation (7.18), which can be rewritten in the form

$$u = C^\alpha(\mathcal{M}z) + D^\alpha z. \quad (6.7.41)$$

Let us see if this relation can be solved for z . That is, given u , we want to use (7.41) to determine z . Taking differentials of both sides of (7.41) and using (7.30) gives the result

$$du = C^\alpha dZ + D^\alpha dz = C^\alpha M dz + D^\alpha dz = (C^\alpha M + D^\alpha)dz. \quad (6.7.42)$$

By the inverse function theorem, if the matrix $(C^\alpha M + D^\alpha)$ is invertible, we may solve (7.41) for z , in which case there is also the relation

$$dz = (C^\alpha M + D^\alpha)^{-1} du. \quad (6.7.43)$$

Next we observe that (7.17) can be rewritten in the form

$$U = A^\alpha(\mathcal{M}z) + B^\alpha z. \quad (6.7.44)$$

Since we have already found z as a function of u by solving (7.41), equation (7.44) enables us to find U as a function of u .

We claim that the map \mathcal{G} that sends u to U is a gradient map if \mathcal{M} is a symplectic map. Taking differentials of both sides of (7.44), and again using (7.30), give the result

$$dU = A^\alpha dZ + B^\alpha dz = A^\alpha M dz + B^\alpha dz = (A^\alpha M + B^\alpha)dz. \quad (6.7.45)$$

Next insert (7.43) into (7.45) to yield the result

$$dU = (A^\alpha M + B^\alpha)(C^\alpha M + D^\alpha)^{-1} du. \quad (6.7.46)$$

Now compare (7.31) and (7.46) to find the relation

$$G = (A^\alpha M + B^\alpha)(C^\alpha M + D^\alpha)^{-1}. \quad (6.7.47)$$

This result is evidently the Möbius relation

$$G = T_\alpha(M), \quad (6.7.48)$$

which is consistent with (7.40). Also, we know from Section 5.13 that G will be symmetric if M is symplectic, and consequently \mathcal{G} is indeed a gradient map.

Finally, we may determine the source function g associated with \mathcal{G} by performing the path integral

$$g(u, t) = \int^u \sum_a U(u', t)_a du'_a. \quad (6.7.49)$$

Here we have indicated that g may also depend on t if \mathcal{M} depends on t .

We close this subsection by listing the Darboux matrices α associated with the familiar generating function types F_1 through F_4 and F_+ , and the related γ in the representation $\alpha = \gamma\sigma$. They appear in the table below. Note the interesting fact that all these Darboux matrices are orthogonal. In Exercises 7.5 through 7.7 you, dear reader, will have the pleasure of spot checking that the use of these Darboux matrices in (7.21) reproduces the relations (5.4) through (5.7) and (6.57).

Table 6.7.1: Darboux Matrices α for the Generating Function types F_1 through F_4 and F_+ .

$$\text{Here } \alpha = \begin{pmatrix} A^\alpha & B^\alpha \\ C^\alpha & D^\alpha \end{pmatrix} = \gamma\sigma. \quad (6.7.50)$$

 $F_1(q, Q, t)$

$$p_k = \partial F_1 / \partial q_k, \quad P_k = -\partial F_1 / \partial Q_k. \quad (6.7.51)$$

$$A^\alpha = \begin{pmatrix} 0 & 0 \\ 0 & -I^n \end{pmatrix}, \quad B^\alpha = \begin{pmatrix} 0 & I^n \\ 0 & 0 \end{pmatrix}, \quad (6.7.52)$$

$$C^\alpha = \begin{pmatrix} 0 & 0 \\ I^n & 0 \end{pmatrix}, \quad D^\alpha = \begin{pmatrix} I^n & 0 \\ 0 & 0 \end{pmatrix}. \quad (6.7.53)$$

$$\gamma = (1/\sqrt{2}) \begin{pmatrix} I^n & 0 & 0 & I^n \\ I^n & 0 & 0 & -I^n \\ 0 & -I^n & I^n & 0 \\ 0 & I^n & I^n & 0 \end{pmatrix}. \quad (6.7.54)$$

 $F_2(q, P, t)$

$$p_k = \partial F_2 / \partial q_k, \quad Q_k = \partial F_2 / \partial P_k. \quad (6.7.55)$$

$$A^\alpha = \begin{pmatrix} 0 & 0 \\ I^n & 0 \end{pmatrix}, \quad B^\alpha = \begin{pmatrix} 0 & I^n \\ 0 & 0 \end{pmatrix}, \quad (6.7.56)$$

$$C^\alpha = \begin{pmatrix} 0 & 0 \\ 0 & I^n \end{pmatrix}, \quad D^\alpha = \begin{pmatrix} I^n & 0 \\ 0 & 0 \end{pmatrix}. \quad (6.7.57)$$

$$\gamma = (1/\sqrt{2}) \begin{pmatrix} I^n & 0 & 0 & I^n \\ 0 & I^n & I^n & 0 \\ 0 & -I^n & I^n & 0 \\ -I^n & 0 & 0 & I^n \end{pmatrix}. \quad (6.7.58)$$

 $F_3(p, Q, t)$

$$q_k = -\partial F_3 / \partial p_k, \quad P_k = -\partial F_3 / \partial Q_k. \quad (6.7.59)$$

$$A^\alpha = \begin{pmatrix} 0 & 0 \\ 0 & -I^n \end{pmatrix}, \quad B^\alpha = \begin{pmatrix} -I^n & 0 \\ 0 & 0 \end{pmatrix}, \quad (6.7.60)$$

$$C^\alpha = \begin{pmatrix} 0 & 0 \\ I^n & 0 \end{pmatrix}, \quad D^\alpha = \begin{pmatrix} 0 & I^n \\ 0 & 0 \end{pmatrix}. \quad (6.7.61)$$

$$\gamma = (1/\sqrt{2}) \begin{pmatrix} 0 & I^n & -I^n & 0 \\ I^n & 0 & 0 & -I^n \\ I^n & 0 & 0 & I^n \\ 0 & I^n & I^n & 0 \end{pmatrix}. \quad (6.7.62)$$

Table 6.7.1 continued

 $F_4(p, P, t)$

$$q_k = -\partial F_4 / \partial p_k, \quad Q_k = \partial F_4 / \partial P_k. \quad (6.7.63)$$

$$A^\alpha = \begin{pmatrix} 0 & 0 \\ I^n & 0 \end{pmatrix}, \quad B^\alpha = \begin{pmatrix} -I^n & 0 \\ 0 & 0 \end{pmatrix}, \quad (6.7.64)$$

$$C^\alpha = \begin{pmatrix} 0 & 0 \\ 0 & I^n \end{pmatrix}, \quad D^\alpha = \begin{pmatrix} 0 & I^n \\ 0 & 0 \end{pmatrix}. \quad (6.7.65)$$

$$\gamma = (1/\sqrt{2}) \begin{pmatrix} 0 & I^n & -I^n & 0 \\ 0 & I^n & I^n & 0 \\ I^n & 0 & 0 & I^n \\ -I^n & 0 & 0 & I^n \end{pmatrix}. \quad (6.7.66)$$

 $F_+(u, t)$

$$\Delta = J\partial_u F_+|_{u=\Sigma} \text{ where } \Sigma = Z + z \text{ and } \Delta = Z - z. \quad (6.7.67)$$

$$\alpha = \sigma = (1/\sqrt{2}) \begin{pmatrix} -J^{2n} & J^{2n} \\ I^{2n} & I^{2n} \end{pmatrix}. \quad (6.7.68)$$

$$\gamma = I^{4n}. \quad (6.7.69)$$

6.7.2 Discussion

6.7.2.1 Graphs

In this section, and in Sections 5.13 and 5.14, we introduced a $4n$ -dimensional space even though the underlying entities of interest, namely symplectic matrices and symplectic maps, were associated with a $2n$ -dimensional space. How could one have guessed that this would be a good thing to do? Of course, results are what count. But one way to look at the matter is in terms of *graphs*.

Suppose we wish to analyze a function $f(x)$ of a *single* real variable x . One way to do so is to introduce a *two*-dimensional space \mathbb{R}^2 , with axes x and y , and then “darken” those points in \mathbb{R}^2 for which $y = f(x)$. The darkened points form the graph of f ,

$$\text{graph of } f = \{\{x, y\} \in \mathbb{R}^2 \mid y = f(x)\}. \quad (6.7.70)$$

Moreover, the graph of f is a one-dimensional submanifold of \mathbb{R}^2 . Let τ_1 be some parameter. Then we may also write

$$\text{graph of } f = \{\{x, y\} \in \mathbb{R}^2 \mid x = \tau_1, y = f(\tau_1) \text{ with } \tau_1 \in \mathbb{R}^1\}. \quad (6.7.71)$$

All these considerations are so commonplace that we hardly ever think about them, but they do involve a doubling of dimension.²³

Now consider a $4n$ -dimensional space with coordinates $\{Z_1 \cdots Z_{2n}\}$ and $\{z_1 \cdots z_{2n}\}$ or, equivalently, coordinates $\{\hat{Z}_1 \cdots \hat{Z}_{4n}\}$. Let \mathcal{M} be the map (6.1). Then we can describe \mathcal{M} in terms of a graph by writing

$$\text{graph of } \mathcal{M} = \{\hat{Z} \in \mathbb{R}^{4n} \mid Z_a = K_a(z) \text{ for } a = 1, 2n\}. \quad (6.7.72)$$

(Here we have suppressed the possible dependence of K on the parameter t .) We see that the construction (7.72) is completely analogous to (7.70). Moreover, the graph of \mathcal{M} is a $2n$ -dimensional submanifold of \mathbb{R}^{4n} . Let $\{\tau_1 \cdots \tau_{2n}\}$ be a set of $2n$ parameters. Then we may also write

$$\text{graph of } \mathcal{M} = \{\hat{Z} \in \mathbb{R}^{4n} \mid Z_a = K_a(\tau), z_a = \tau_a \text{ for } a = 1, 2n \text{ with } \tau \in \mathbb{R}^{2n}\}. \quad (6.7.73)$$

6.7.2.2 The Graph of \mathcal{M} Is a \tilde{J}^{4n} Lagrangian Submanifold

Let us find the tangent vectors to the graph of \mathcal{M} (now regarded as a $2n$ -dimensional submanifold in a $4n$ -dimensional space). They describe how \hat{Z} varies when τ is varied. Write τ in the form

$$\tau = \tau^0 + \sum_1^{2n} \lambda_i e^i. \quad (6.7.74)$$

Here the vectors e^i are the same as those introduced in Section 5.13, namely those that form the columns of I^{2n} . Employing this notation for τ , we define $2n$ vectors ζ^j tangent to the graph of \mathcal{M} at the point $\hat{Z}(\tau^0)$ by writing the definition

$$\zeta^j(\tau^0) = \partial \hat{Z} / \partial \lambda_j|_{\lambda=0} = (\partial Z / \partial \lambda_j|_{\lambda=0}; \partial z / \partial \lambda_j|_{\lambda=0})^T. \quad (6.7.75)$$

As indicated, the tangent vectors ζ^j are of length $4n$ (have $4n$ entries) as is appropriate for a $4n$ -dimensional space. The last $2n$ entries in each tangent vector ζ^j , those to the right of the semicolon in (7.75), are easy to find. From (7.73) and (7.74) we readily compute the result

$$\partial z / \partial \lambda_j|_{\lambda=0} = e^j. \quad (6.7.76)$$

The calculation of the first $2n$ entries is a bit more involved. From (7.73) (which contains the information that $z = \tau$) and (7.74) we find, in terms of components, the result

$$\partial Z_i / \partial \lambda_j|_{\lambda=0} = \partial Z_i / \partial z_j|_{z=\tau^0} = M_{ij}(\tau^0) = m_i^j. \quad (6.7.77)$$

Here we have used the notation (5.13.36), and $M(\tau^0)$ is the Jacobian matrix $M(z)$ for \mathcal{M} evaluated at $z = \tau^0$.

We see from (7.75) through (7.77) that there is the relation

$$\zeta^j = (m^j; e^j)^T, \quad (6.7.78)$$

²³The use of graphs to portray functions was invented by Descartes. He plotted x (the independent variable) along the vertical axis and y (the dependent variable) along the horizontal axis. Newton turned this around to plot x along the horizontal axis and y along the vertical axis, and humankind have followed his convention ever since.

and recognize that the ζ^j are just the vectors u^j introduced in Section 5.13 and given by (5.13.42). We know that these vectors are \tilde{J}^{4n} isotropic. Thus, we have shown that the tangent vectors of the graph of \mathcal{M} are \tilde{J}^{4n} isotropic at any point $\hat{Z}(\tau^0)$ in the submanifold, and therefore span a \tilde{J}^{4n} Lagrangian plane at every such point. For this reason, the graph of \mathcal{M} is entitled to be called a \tilde{J}^{4n} Lagrangian submanifold.

6.7.2.3 The Graph of \mathcal{G} Is a J^{4n} Lagrangian Submanifold

We can carry out a similar analysis for the graph of \mathcal{G} . The map \mathcal{G} is defined by (7.7) and (7.14) through (7.16). Again let $\{\tau_1 \dots \tau_{2n}\}$ be a set of $2n$ parameters. The graph of \mathcal{G} , as a $2n$ -dimensional submanifold in \mathbb{R}^{4n} , is given by the definition

$$\text{graph of } \mathcal{G} = \{\hat{U} \in \mathbb{R}^{4n} \mid U_a = \partial g(\tau)/\partial \tau_a, u_a = \tau_a \text{ for } a = 1, 2n \text{ with } \tau \in \mathbb{R}^{2n}\}. \quad (6.7.79)$$

Again employing the notation (7.74), the graph of \mathcal{G} will have $2n$ tangent vectors ν^j at the point $\hat{U}(\tau^0)$ given by the definition

$$\nu^j(\tau^0) = \partial \hat{U}/\partial \lambda_j|_{\lambda=0} = (\partial U/\partial \lambda_j|_{\lambda=0}; \partial u/\partial \lambda_j|_{\lambda=0})^T. \quad (6.7.80)$$

The last $2n$ entries in each tangent vector ν^j , those to the right of the semicolon in (7.80), are calculated the same way as in the case of \mathcal{M} , and are therefore given by the relation

$$\partial u/\partial \lambda_j|_{\lambda=0} = e^j. \quad (6.7.81)$$

In terms of components, the first $2n$ entries are given by the relations

$$\partial U_i/\partial \lambda_j|_{\lambda=0} = \partial^2 g/\partial \lambda_i \partial \lambda_j|_{\lambda=0} = G_{ij}(\tau^0). \quad (6.7.82)$$

In analogy to the notation (5.13.60) and (5.13.61) employed in Section 5.13, define vectors w^j , each of length $2n$, by writing

$$w_i^j = G_{ij}. \quad (6.7.83)$$

Then, with this notation, the tangent vectors ν^j become

$$\nu^j = (w^j; e^j)^T. \quad (6.7.84)$$

Since G is a symmetric matrix, these vectors are completely analogous to the vectors v^j constructed in Section 5.13 and given by (5.13.62). Therefore we know that these vectors are J^{4n} isotropic. Thus, we have shown that the tangent vectors of the graph of \mathcal{G} are J^{4n} isotropic at any point $\hat{U}(\tau^0)$ in the submanifold, and therefore span a J^{4n} Lagrangian plane at every such point. Consequently the graph of \mathcal{G} is entitled to be called a J^{4n} Lagrangian submanifold.

6.7.2.4 Relation between the Graphs of \mathcal{M} and \mathcal{G}

We have learned that the $4n$ -dimensional constructions used in Sections 5.13 through 5.15, and in this section, appear to be less ad hoc when one thinks in terms of graphs. The graph of \mathcal{M} is a \tilde{J}^{4n} Lagrangian submanifold, and the graph of \mathcal{G} is a J^{4n} Lagrangian submanifold. Moreover, according to (7.3) and (7.4), these two submanifolds are mapped into each other by a Darboux transformation α . Or, put another way, they are the *same* submanifold, and this submanifold appears to be \tilde{J}^{4n} Lagrangian when the coordinates \hat{Z} are used and J^{4n} Lagrangian when the coordinates \hat{U} are used.

6.7.2.5 Reason for the Term “Lagrangian”

To keep a promise, we still need to describe the origin of the term “Lagrangian” when applied to planes and submanifolds. It has to do with *Lagrange brackets*. Consider a $2n$ -dimensional phase space with coordinates $(q; p)$ as in (1.7.9). Define an n -dimensional submanifold in this phase space (parameterized by the quantities τ_1, \dots, τ_n) by writing $2n$ equations of the form

$$z_a = f_a(\tau) \quad (6.7.85)$$

where the f_a are any functions of the n variables τ . Next form the tangent vectors $\partial z / \partial \tau_i$. [These n vectors are assumed to be linearly independent since (7.85) is assumed to define an n -dimensional submanifold.] The Lagrange bracket $\{\tau_i, \tau_j\}$, which is a function of the variables τ , is defined by the rule

$$\{\tau_i, \tau_j\} = (\partial z / \partial \tau_i, J^{2n} \partial z / \partial \tau_j). \quad (6.7.86)$$

If we use the specific form for J^{2n} given by (3.1.1), we observe that the Lagrange bracket can also be written in what may be the more familiar text-book form

$$\{\tau_i, \tau_j\} = \sum_k (\partial q_k / \partial \tau_i)(\partial p_k / \partial \tau_j) - (\partial p_k / \partial \tau_i)(\partial q_k / \partial \tau_j). \quad (6.7.87)$$

From (7.86) we see that the tangent vectors $\partial z / \partial \tau_i$ for any fixed $\tau = \tau^0$ are J^{2n} isotropic if the Lagrange brackets $\{\tau_i, \tau_j\}$ all vanish (for $\tau = \tau^0$), and we say that the plane spanned by the tangent vectors $\partial z / \partial \tau_i$ is Lagrangian. Correspondingly, the submanifold given by (7.85) is Lagrangian if the Lagrange brackets $\{\tau_i, \tau_j\}$ all vanish for all values of τ . Similar nomenclature carries over to $2n$ -dimensional planes and $2n$ -dimensional submanifolds in $4n$ -dimensional spaces and the use of J^{4n} or \tilde{J}^{4n} .

6.7.2.6 Closing Observation

We close this subsection with the observation that the family of maps produced by the generating/source function $g(u, t)$ and some Darboux matrix α does not necessarily pass through the identity map \mathcal{I} for some value of t . For each value of t there will be a symmetric matrix $G(u, t)$ given by (1.6), and for this value of t the map $\mathcal{M}(t)$ will have a Jacobian matrix M given by (7.40). Suppose $\mathcal{M}(t) = \mathcal{I}$ when $t = t_0$. Then, since the Jacobian matrix of the identity map \mathcal{I} is the identity matrix I , (7.40) becomes the the relation

$$I = T_{\alpha^{-1}}(G) \quad (6.7.88)$$

which requires that

$$G = G_0 \quad (6.7.89)$$

where

$$G_0 = T_\alpha(I) = (A^\alpha + B^\alpha)(C^\alpha + D^\alpha)^{-1}. \quad (6.7.90)$$

Corresponding, we must have

$$g(u, t_0) = (1/2)(u, G_0 u) + (v, u) + g_0 \quad (6.7.91)$$

where v is a fixed vector yet to be determined and g_0 is an immaterial constant.

To determine v , which will turn out to vanish, we need to find the \mathcal{M} associated with the $g(u, t_0)$ given by (7.91). Partial differentiation of (7.91) gives the intermediate result

$$\partial g / \partial u_a = v_a + \sum_b (G_0)_{ab} u_b, \quad (6.7.92)$$

and employing this intermediate result in (7.21) gives the further result

$$A^\alpha Z + B^\alpha z = G_0(C^\alpha Z + D^\alpha z) + v. \quad (6.7.93)$$

Now solve (7.93) to find the result

$$\begin{aligned} Z &= (A^\alpha - G_0 C^\alpha)^{-1}(G_0 D^\alpha - B^\alpha)z + [(A^\alpha - G_0 C^\alpha)^{-1}]v \\ &= [(A^{\alpha^{-1}} G_0 + B^{\alpha^{-1}})(C^{\alpha^{-1}} G_0 + D^{\alpha^{-1}})^{-1}]z + [(A^\alpha - G_0 C^\alpha)^{-1}]v \\ &= [T_{\alpha^{-1}}(G_0)]z + [(A^\alpha - G_0 C^\alpha)^{-1}]v \\ &= z + [(A^\alpha - G_0 C^\alpha)^{-1}]v. \end{aligned} \quad (6.7.94)$$

Here we have used (7.38) and (7.88). We see that we must have $v = 0$ to achieve the identity map, in which case

$$g(u, t_0) = (1/2)(u, G_0 u) + g_0. \quad (6.7.95)$$

Now it may well happen that $g(u, t)$ is never of the form (7.95) for any value of t , in which case the family of maps $\mathcal{M}(t)$ never passes through the identity map. There is also a possible second obstacle. Note that G_0 as given by (7.90) is not defined if the matrix $(C^\alpha + D^\alpha)$ is not invertible,

$$\det(C^\alpha + D^\alpha) = 0. \quad (6.7.96)$$

Thus, there are Darboux matrices for which $\mathcal{M}(t)$ can never pass through the identity map. See, for example, the Darboux matrices associated with F_1 and F_4 given in Table 6.7.1.

To conclude this observation we note that, although we have verified that there are families of symplectic maps that never pass through the identity map, the symplectic maps associated with the Hamiltonian Cauchy initial value problem, see Section 1.3 and Subsection 4.1, pass through the identity map by definition because of the initial condition requirement (4.13).

6.7.2.7 Final Remark

Finally, we remark that there is no reason why the Darboux matrix α cannot also be taken to depend on t . We know that we can always write

$$\alpha = \gamma \sigma \quad (6.7.97)$$

where σ is the matrix (5.13.11) and γ is any matrix in the group $Sp(4n)$. We also know that the symplectic group is connected. See Section 5.9. Therefore the set of Darboux matrices is connected, and we may sensibly write

$$\alpha(t) = \gamma(t)\sigma. \quad (6.7.98)$$

for any path $\gamma(t)$ in $Sp(4n)$. If we now employ $g(u, t)$ and $\alpha(t)$ in (7.21), the result will again be a family of symplectic maps $\mathcal{M}(t)$.

6.7.3 Relating Source Functions and Generating Hamiltonians, Transformation of Hamiltonians, and Hamilton-Jacobi Theory/Equations

Suppose the source function g appearing in (7.24) does indeed depend on the time. Then its use in (7.21) produces a family of symplectic maps which, for our present notational purposes, we will call $\mathcal{N}(t)$ rather than $\mathcal{M}(t)$ and will have Jacobian $N(t)$. We know from Subsection 4.2 that any such family is Hamiltonian generated. Indeed, Subsection 5.3 determined this Hamiltonian for the case of $F_2(q, P, t)$, and (5.167) covers the cases F_1 through F_4 . The relation between $g(u, t)$, the associated symplectic map $\mathcal{N}(t)$, and the associated generating Hamiltonian, which we will here call H^g , is part of Hamilton-Jacobi theory; and Subsection 5.3.2 describes examples of the Hamilton-Jacobi equation for the cases where $\mathcal{N}(t)$ arises from one of the F_j . In this subsection we will solve the general problem of finding the Hamiltonian H^g when $\mathcal{N}(t)$ arises from $g(u, t)$ and the use of some Darboux matrix α . We will also solve the inverse general problem of finding $g(u, t)$ in terms of H^g . Finally, we will relate these results to Hamilton-Jacobi Theory for the general problem. Our results will also be of use for the work of Chapter 34.

6.7.3.1 Finding the Generating Hamiltonian H^g from the Source Function g

Our discussion will be patterned after that of Subsection 5.3, so again we will have to deal with a variety of partial derivatives. Therefore we introduce the notation

$$\begin{aligned} g(u, t; , 1) &= \partial g / \partial t, \\ g(u, t; a, 1) &= \partial^2 g / \partial u_a \partial t, \\ g(u, t; ab,) &= \partial^2 g / \partial u_a \partial u_b. \end{aligned} \quad (6.7.99)$$

Employing this notation, define the function $g^t(u, t)$ by the rule

$$g^t(u, t) = g(u, t; , 1) \quad (6.7.100)$$

According to (7.18), u may be regarded as a function of $Z(t)$ and z . Also, according to (7.22) with the substitution $\mathcal{M} \rightarrow \mathcal{N}$, we may view z as being a function of t and $Z(t)$ by writing

$$z = \mathcal{N}^{-1}(t)Z. \quad (6.7.101)$$

Therefore, u may be regarded as a function of Z and t ,

$$u = u(Z, t). \quad (6.7.102)$$

Now substitute (7.102) into (7.100) to define the function $H^g(Z, t)$ by the rule

$$H^g(Z, t) = g^t(u(Z, t), t). \quad (6.7.103)$$

We claim that $H^g(Z, t)$ is the Hamiltonian that generates the map $\mathcal{N}(t)$ produced by the use of the source function $g(u, t)$ and some Darboux matrix α . Here we assume that α is some *fixed* Darboux matrix, although it would be interesting to also entertain the possibility (7.98).

We will now seek to verify this claim about H^g . Suppose z is held fixed and t is increased by the amount dt . So doing will change Z by the amount dZ . Also, according to (7.18), u will experience a change that we will call du . Look at the relations (7.28) and (7.29). From (7.28) we conclude that

$$\begin{aligned} dZ_a &= \left[\sum_b (A^{\alpha^{-1}})_{ab} g(u, t; b, 1) \right] dt \\ &\quad + \sum_{bc} (A^{\alpha^{-1}})_{ab} g(u, t; bc,) du_c \\ &\quad + \sum_b (B^{\alpha^{-1}})_{ab} du_b \\ &= \left[\sum_b (A^{\alpha^{-1}})_{ab} g(u, t; b, 1) \right] dt \\ &\quad + \sum_b [A^{\alpha^{-1}} G(u, t) + B^{\alpha^{-1}}]_{ab} du_b. \end{aligned} \tag{6.7.104}$$

Here we have used (1.6). That is, here G is the Hessian matrix of g and the Jacobian matrix of \mathcal{G} . From (7.29), since z is to be held fixed, we conclude that

$$\begin{aligned} 0 = dz_a &= \left[\sum_b (C^{\alpha^{-1}})_{ab} g(u, t; b, 1) \right] dt \\ &\quad + \sum_{bc} (C^{\alpha^{-1}})_{ab} g(u, t; bc,) du_c \\ &\quad + \sum_b (D^{\alpha^{-1}})_{ab} du_b \\ &= \left[\sum_b (C^{\alpha^{-1}})_{ab} g(u, t; b, 1) \right] dt \\ &\quad + \sum_b [C^{\alpha^{-1}} G(u, t) + D^{\alpha^{-1}}]_{ab} du_b. \end{aligned} \tag{6.7.105}$$

Let us now eliminate du between (7.104) and (7.105). First solve (7.105) for du to find the result

$$du_b = -dt \sum_c \{ [C^{\alpha^{-1}} G(u, t) + D^{\alpha^{-1}}]^{-1} C^{\alpha^{-1}} \}_{bc} g(u, t; c, 1). \tag{6.7.106}$$

Now substitute (7.106) into (7.104) to obtain the result

$$\begin{aligned} dZ_a &= dt \sum_b (A^{\alpha^{-1}})_{ab} g(u, t; b, 1) \\ &\quad - dt \sum_b \{ [A^{\alpha^{-1}} G(u, t) + B^{\alpha^{-1}}] [C^{\alpha^{-1}} G(u, t) + D^{\alpha^{-1}}]^{-1} C^{\alpha^{-1}} \}_{ab} g(u, t; b, 1) \\ &= dt \sum_b \{ A^{\alpha^{-1}} - [A^{\alpha^{-1}} G(u, t) + B^{\alpha^{-1}}] [C^{\alpha^{-1}} G(u, t) + D^{\alpha^{-1}}]^{-1} C^{\alpha^{-1}} \}_{ab} g(u, t; b, 1). \end{aligned} \tag{6.7.107}$$

We also observe that

$$[A^{\alpha^{-1}}G + B^{\alpha^{-1}}][C^{\alpha^{-1}}G + D^{\alpha^{-1}}]^{-1} = T_{\alpha^{-1}}(G) = N. \quad (6.7.108)$$

Therefore, (7.107) can also be written as

$$dZ_a = dt \sum_b (A^{\alpha^{-1}} - NC^{\alpha^{-1}})_{ab} g(u, t; b, 1), \quad (6.7.109)$$

or

$$dZ_a/dt = \sum_c (A^{\alpha^{-1}} - NC^{\alpha^{-1}})_{ac} g(u, t; c, 1). \quad (6.7.110)$$

Note that here we have renamed the dummy summation index.

Next let us work out the quantities $\partial H^g(Z, t)/\partial Z_a$. From (7.100), (7.103), and the chain rule (and holding t fixed) we have the result

$$dH^g = \sum_a g(u, t; a, 1) du_a. \quad (6.7.111)$$

Also, use of (7.18) provides the relation

$$du_a = \sum_b [(C^\alpha)_{ab} dZ_b + (D^\alpha)_{ab} dz_b] \quad (6.7.112)$$

which, using (7.30) with the substitution $M \rightarrow N$, can be rewritten in the form

$$du_a = \sum_b [C^\alpha + D^\alpha(N^{-1})]_{ab} dZ_b. \quad (6.7.113)$$

When combined, (7.111) and (7.113) yield the relation

$$dH^g = \sum_{ab} g(u, t; a, 1) [C^\alpha + D^\alpha(N^{-1})]_{ab} dZ_b \quad (6.7.114)$$

from which we conclude that

$$\begin{aligned} \partial H^g / \partial Z_b &= \sum_a g(u, t; a, 1) [C^\alpha + D^\alpha(N^{-1})]_{ab} \\ &= \sum_c \{[C^\alpha + D^\alpha(N^{-1})]^T\}_{bc} g(u, t; c, 1). \end{aligned} \quad (6.7.115)$$

We are almost done. Multiply both sides of (7.115) by J_{ab} and sum over b to find the result

$$\sum_b J_{ab} (\partial H^g / \partial Z_b) = \sum_c \{J[C^\alpha + D^\alpha(N^{-1})]^T\}_{ac} g(u, t; c, 1). \quad (6.7.116)$$

Here again we have renamed the dummy summation index. We now claim that there is the relation

$$(A^{\alpha^{-1}} - NC^{\alpha^{-1}}) = J[C^\alpha + D^\alpha(N^{-1})]^T. \quad (6.7.117)$$

If so, then comparison of (7.110) and (7.116) gives the result

$$dZ_a/dt = \sum_b J_{ab}(\partial H^g/\partial Z_b), \quad (6.7.118)$$

which is the expected equations of motion set for Z when H^g is the Hamiltonian.

To complete the proof, we need to verify (7.117). Its right side can be rewritten as

$$J[C^\alpha + D^\alpha(N^{-1})]^T = J(C^\alpha)^T + J(N^{-1})^T(D^\alpha)^T. \quad (6.7.119)$$

According to (3.1.11), the symplectic condition for N gives the relation

$$(N^{-1})^T = -JNJ. \quad (6.7.120)$$

Therefore the right side of (7.117) can also be rewritten as

$$J[C^\alpha + D^\alpha(N^{-1})]^T = J(C^\alpha)^T - JNJ(D^\alpha)^T = J(C^\alpha)^T + NJ(D^\alpha)^T. \quad (6.7.121)$$

Now employ the relations (5.13.100) and (5.13.102) to again rewrite the right side of (7.117) as

$$J[C^\alpha + D^\alpha(N^{-1})]^T = J(C^\alpha)^T + NJ(D^\alpha)^T = A^{\alpha^{-1}} - NC^{\alpha^{-1}}, \quad (6.7.122)$$

which, we see, agrees with the left side of (7.117). Therefore our claim is correct.

In summary, we have shown that in the general case the generating Hamiltonian $H^g(Z, t)$ for the family $\mathcal{N}(t)$ of symplectic maps produced by the source function $g(u, t)$ and the Darboux matrix α is given by the relation

$$H^g(Z, t) = [\partial g(u, t)/\partial t]|_{u=C^\alpha Z+D^\alpha(N^{-1}Z)}. \quad (6.7.123)$$

We also observe that (5.167) is a special case of (7.123)

6.7.3.2 Finding the Source Function g from the Generating Hamiltonian H^g

In Subsection 5.2.2 we showed, as an example, how to find the source function F_2 , which amounts to the choice (7.56) and (7.57) for the Darboux matrix α , in terms of an integral over a trajectory arising from the generating Hamiltonian which we there called H . See (5.119) and (5.130). The purpose of this subsection is to provide an analogous treatment of the general case: Given a Darboux matrix α and a generating Hamiltonian $H^g(\zeta, t)$, find the source function $g(u, t)$. Here, as in Subsection 5.2.2, it is convenient to employ the phase-space variables $\zeta = (\xi, \eta)$.

Let $(q, p) = z$ be initial conditions at $\tau = t^i$, and let $(Q, P) = Z$ be the final conditions reached by following to time $\tau = t$ the trajectories generated by $H^g(\zeta, \tau)$ starting with these initial conditions. We know that trajectories can be labeled by specifying either the initial conditions z or the final conditions Z . Assume that the trajectories are such that they can also be labeled by specifying u as given by (7.18). See Figure 7.1. To do so will generally require a $2n$ -dimensional search: Pick a $2n$ -vector u . Begin by guessing z . Next follow the trajectory with initial conditions

$$\zeta(t^i) = z \quad (6.7.124)$$

and generated by $H^g(\zeta, \tau)$ to the time $\tau = t$ and set

$$Z = \zeta(t). \quad (6.7.125)$$

Now compute the quantity $(C^\alpha Z + D^\alpha z)$ and see if it equals u ,

$$C^\alpha Z + D^\alpha z = u? \quad (6.7.126)$$

If it does, then the desired z (and Z) have been found. If not, guess again. In actual practice, this trajectory may have to be found by some kind of *shooting* method facilitated, perhaps, by a Newton's method search that involves also integrating the variational equations to determine how changes in the initial conditions produce changes in the final conditions.²⁴

Observe that taking differentials of both sides of (7.18) and using (7.30) gives the result

$$du = C^\alpha dZ + D^\alpha dz = (C^\alpha N + D^\alpha)dz \quad (6.7.127)$$

from which it follows that

$$dz = (C^\alpha N + D^\alpha)^{-1}du. \quad (6.7.128)$$

Thus, if

$$\det(C^\alpha N + D^\alpha) \neq 0, \quad (6.7.129)$$

the quantity z (by the inverse function theorem) is indeed specified by the quantity u .

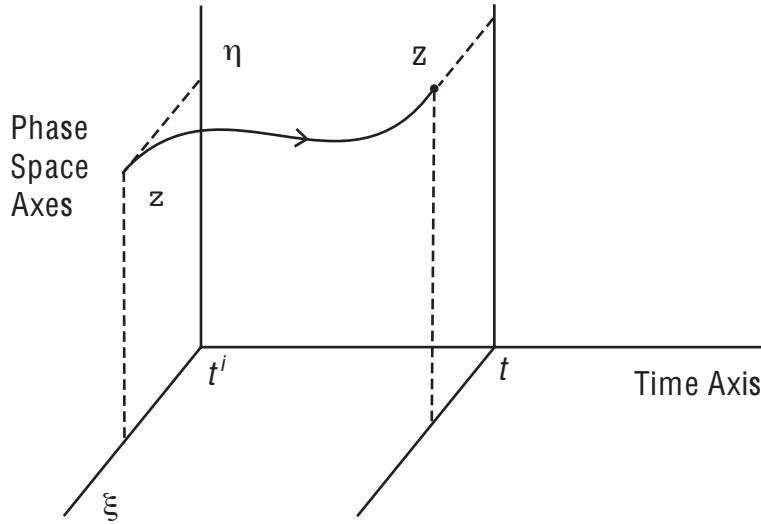


Figure 6.7.1: A trajectory of $H^g(\zeta, \tau)$ in the augmented $(\zeta, t) = (\xi, \eta; t)$ phase space. Given a Darboux matrix α , an initial time t^i , a final time t , and the $2n$ -vector u , the initial condition $\zeta(t^i) = z$ is to be selected such that $C^\alpha Z + D^\alpha z = u$ where $\zeta(t) = Z$.

With these assumptions in mind, define the function $A'(u, t)$ by the rule

$$A'(u, t) = \int_{t^i}^t [(\zeta, J\dot{\zeta}) + 2H^g(\zeta, \tau)]d\tau. \quad (6.7.130)$$

²⁴Note that although in general a $2n$ -dimensional search is required, for some special α 's, such as those for F_1 through F_4 , only an n -dimensional search is required. See Subsection 5.2.2.

Here the integral on the right side is to be evaluated over the trajectory satisfying (7.126). We will want to see how $A'(u, t)$ changes when changes are made in u and/or t .

Write (7.130) in the form

$$A'(u, t) = \int_{t^i}^t \mathcal{A}'(\zeta, \dot{\zeta}, \tau) d\tau \quad (6.7.131)$$

where

$$\mathcal{A}'(\zeta, \dot{\zeta}, \tau) = \left(\sum_{cd} \zeta_c J_{cd} \dot{\zeta}_d \right) + 2H^g(\zeta, \tau). \quad (6.7.132)$$

Changing u (while holding t fixed) changes the initial and final conditions and the trajectory in between. Consequently, from variational calculus, we find that the change in \mathcal{A}' is given by

$$\delta A' = \int_{t^i}^t d\tau \left\{ \sum_a [(\partial \mathcal{A}' / \partial \zeta_a) \delta \zeta_a + (\partial \mathcal{A}' / \partial \dot{\zeta}_a) \delta \dot{\zeta}_a] \right\}. \quad (6.7.133)$$

The integrand in (7.133) can be manipulated in the standard way to rewrite $\delta A'$ in the form

$$\delta A' = \int_{t^i}^t d\tau \left\{ \sum_a [(\partial \mathcal{A}' / \partial \zeta_a) - (d/d\tau)(\partial \mathcal{A}' / \partial \dot{\zeta}_a)] \delta \zeta_a + (d/d\tau) \left[\sum_a (\partial \mathcal{A}' / \partial \dot{\zeta}_a) \delta \dot{\zeta}_a \right] \right\}. \quad (6.7.134)$$

For the various ingredients in the integrand of (7.134) we find the results

$$\partial \mathcal{A}' / \partial \zeta_a = \left(\sum_b J_{ab} \dot{\zeta}_b \right) + 2 \partial H^g / \partial \zeta_a, \quad (6.7.135)$$

$$\partial \mathcal{A}' / \partial \dot{\zeta}_a = - \sum_b J_{ab} \zeta_b, \quad (6.7.136)$$

$$\sum_a (\partial \mathcal{A}' / \partial \dot{\zeta}_a) \delta \zeta_a = \sum_{ab} \zeta_a J_{ab} \delta \zeta_b = (\zeta, J \delta \zeta), \quad (6.7.137)$$

$$-(d/d\tau)(\partial \mathcal{A}' / \partial \dot{\zeta}_a) = \sum_b J_{ab} \dot{\zeta}_b, \quad (6.7.138)$$

$$[(\partial \mathcal{A}' / \partial \zeta_a) - (d/d\tau)(\partial \mathcal{A}' / \partial \dot{\zeta}_a)] = 2 \left(\sum_b J_{ab} \dot{\zeta}_b \right) + 2(\partial H^g / \partial \zeta_a). \quad (6.7.139)$$

But, since ζ is assumed to be a trajectory for $H^g(\zeta, \tau)$, it satisfies Hamilton's equations

$$\dot{\zeta}_a = \sum_b J_{ab} (\partial H^g / \partial \zeta_b) \quad (6.7.140)$$

from which it follows that

$$[(\partial \mathcal{A}' / \partial \zeta_a) - (d/d\tau)(\partial \mathcal{A}' / \partial \dot{\zeta}_a)] = 2 \left(\sum_b J_{ab} \dot{\zeta}_b \right) + 2(\partial H^g / \partial \zeta_a) = 0. \quad (6.7.141)$$

As a consequence of all these results, $\delta A'$ becomes

$$\delta A' = \int_{t^i}^t d\tau (d/d\tau)[(\zeta, J \delta \zeta)] = (\zeta, J \delta \zeta)|_{t^i}^t = (Z, J dZ) - (z, J dz). \quad (6.7.142)$$

As a further step, we observe that the quantity (U, u) can be written in the form

$$(U, u) = (\hat{U}, S\hat{U}) \quad (6.7.143)$$

where S is the $4n \times 4n$ symmetric matrix

$$S = (1/2) \begin{pmatrix} 0 & I^{2n} \\ I^{2n} & 0 \end{pmatrix}. \quad (6.7.144)$$

[Recall the notation (7.1) and (7.7).] In terms of these quantities, and using (7.3), we may also write the relation

$$(U, u) = (\hat{U}, S\hat{U}) = (\alpha\hat{Z}, S\alpha\hat{Z}) = (\hat{Z}, \alpha^T S\alpha\hat{Z}). \quad (6.7.145)$$

We now have the tools to construct $g(u, t)$. It is defined by the rule

$$g(u, t) \stackrel{\text{def}}{=} [A'(u, t) + (\hat{Z}, \alpha^T S\alpha\hat{Z})]/2. \quad (6.7.146)$$

Our task is to verify that this g has the desired properties. We will first show that this g produces \mathcal{N} according to the rule (7.21). Then we will show that it leads back to the specified Hamiltonian.

To see that this g produces \mathcal{N} , suppose t is held fixed and u is varied by an amount du . Then \hat{Z} (that is, Z and z) will vary by an amount $d\hat{Z}$. Correspondingly, we find that g is changed by an amount δg with

$$\delta g = [\delta A'(u, t) + \delta(\hat{Z}, \alpha^T S\alpha\hat{Z})]/2. \quad (6.7.147)$$

Next employ (7.1) through (7.5) and (7.8) and (7.142) to rewrite $\delta A'$ in the form

$$\delta A' = (Z, JdZ) - (z, Jdz) = (\hat{Z}, \tilde{J}^{4n}d\hat{Z}) = (U, du) - (u, dU). \quad (6.7.148)$$

Also, we find that

$$\delta(\hat{Z}, \alpha^T S\alpha\hat{Z}) = 2(\hat{Z}, \alpha^T S\alpha d\hat{Z}) = 2(\alpha\hat{Z}, S\alpha d\hat{Z}) = 2(\hat{U}, Sd\hat{U}) = (U, du) + (u, dU). \quad (6.7.149)$$

Therefore we get the final relation

$$\delta g = [(U, du) - (u, dU) + (U, du) + (u, dU)]/2 = (U, du). \quad (6.7.150)$$

It follows that

$$U_a = \partial g(u, t)/\partial u_a, \quad (6.7.151)$$

which, in view of (7.19), is the desired result (7.21).

To check that this g in turn leads back to the specified Hamiltonian, let us take the total time derivative of both sides of (7.146). By ‘total’ we mean that the trajectory employed in computing A' should simply be extended in time, but otherwise unchanged. This means that z will not change, but Z and consequently also u will change. By the chain rule and using (7.151), we get for the left side of (7.146) the result

$$(d/dt)(\text{left side}) = dg/dt = \partial g/\partial t + \sum_a (\partial g/\partial u_a)(du_a/dt) = \partial g/\partial t + (U, \dot{u}). \quad (6.7.152)$$

For the right side of (7.146) we find

$$\begin{aligned}(d/dt)(\text{right side}) &= (d/dt)[A'(u, t) + (\hat{Z}, \alpha^T S \alpha \hat{Z})]/2 \\ &= (1/2)(d/dt)A'(u, t) + (1/2)(d/dt)(\hat{Z}, \alpha^T S \alpha \hat{Z}).\end{aligned}\quad (6.7.153)$$

The first term on the right side of (7.153) is easily evaluated using the fundamental theorem of calculus,

$$(1/2)(d/dt)[A'(u, t)] = (1/2)\mathcal{A}'(\zeta, \dot{\zeta}, \tau)|_{\tau=t} = (1/2)(Z, J\dot{Z}) + H^g(Z, t). \quad (6.7.154)$$

According to our understanding that z should not change ($dz = 0$) there is also the result

$$(1/2)(z, J\dot{z}) = 0. \quad (6.7.155)$$

Therefore, using (7.155), (7.2), and (7.8), the relation (7.154) can also be written in the form

$$\begin{aligned}(1/2)(d/dt)[A'(u, t)] &= [(1/2)(Z, J\dot{Z}) - (1/2)(z, J\dot{z})] + H^g(Z, t) \\ &= (1/2)(\hat{Z}, \tilde{J}^{4n}(d/dt)\hat{Z}) + H^g(Z, t) \\ &= (1/2)[(U, \dot{u}) - (u, \dot{U})] + H^g(Z, t).\end{aligned}\quad (6.7.156)$$

Also, by (7.145), there is the simple result

$$(1/2)(d/dt)(\hat{Z}, \alpha^T S \alpha \hat{z}) = (1/2)(d/dt)(U, u) = (1/2)(\dot{U}, u) + (1/2)(U, \dot{u}). \quad (6.7.157)$$

Consequently the derivative of the right side of (7.146) can also be written as

$$\begin{aligned}(d/dt)(\text{right side}) &= (1/2)[(U, \dot{u}) - (u, \dot{U})] + H^g(Z, t) + (1/2)[(\dot{U}, u) + (U, \dot{u})] \\ &= (U, \dot{u}) + H^g(Z, t).\end{aligned}\quad (6.7.158)$$

Comparison of (7.152) and (7.158) now gives the final result

$$\partial g/\partial t = H^g(Z, t), \quad (6.7.159)$$

which is in agreement with (7.123).

6.7.3.3 Transformation of Hamiltonians and Application to Hamilton-Jacobi Theory in the General Case

Subsection 4.2 showed that any family of symplectic maps $\mathcal{N}(t)$ is Hamiltonian generated, and the associated Hamiltonian was called G . Subsection 4.4 described the transformation of an old Hamiltonian to a new Hamiltonian under the action of a symplectic map. And in Subsection 5.3.2 we found the the relation between the old and new Hamiltonians in the case that the symplectic map \mathcal{N} arises from some specified mixed-variable generating function F_j , and applied the results to Hamilton-Jacobi theory for this case. Here we study the relation between the old and new Hamiltonian in the general case that the symplectic map \mathcal{N} arises from some specified source function g and some specified Darboux matrix α , and apply the results to Hamilton-Jacobi theory for the general case.

We begin by recalling the relation (5.169), which we copy below,

$$K(Z; t) = H(z(Z, t); t) + G(Z; t), \quad (6.7.160)$$

and again remind ourselves that here G is the generating Hamiltonian for \mathcal{N} . In the general case that \mathcal{N} arises from some specified source function g and some specified Darboux matrix α , we found in Subsections 7.3.1 and 7.3.2 that the associated generating Hamiltonian, which we there called H^g , was given by the relations (7.103) or (7.123) or (7.159). Therefore, if we make the identification

$$G = \partial g / \partial t, \quad (6.7.161)$$

we see that (7.160) can be rewritten in the form

$$K(Z; t) = H(z(Z, t); t) + \partial g / \partial t \quad (6.7.162)$$

when \mathcal{N} arises from some specified g, α pair. We have found the relation between the old and new Hamiltonians in the general case.

Suppose an $\mathcal{N}(t)$ can be found such that

$$K(Z; t) = 0. \quad (6.7.163)$$

This is, in principle, always possible because we can take the Z to be the initial conditions and take $\mathcal{N}(t)$ to be the symplectic map that transforms final conditions into initial conditions. If a g, α pair can be found such that $\mathcal{N}(t)$ arises from their use, then combining (7.162) and (7.163) gives the general Hamilton-Jacobi relation/equation

$$H(z(Z, t); t) + \partial g / \partial t = 0. \quad (6.7.164)$$

In the next subsection we will see that, at least locally, a g, α pair can be found for any $\mathcal{N}(t)$ such that $\mathcal{N}(t)$ arises from their use. Therefore there is always a suitable α such that the associated general Hamilton-Jacobi equation, at least locally, has a solution.

6.7.4 What Kind of Generating Function/Darboux Matrix Should We Choose?

6.7.4.1 Background

The relations (5.86) and (5.89) illustrated that attempted use of the generating functions F_1 and F_4 fails for the identity map. Here is another example of failure: Consider Mathieu transformations given by (5.174) and (5.183). Can they be obtained from an F_1 generating function? According to (5.4) we must have $\det(B) \neq 0$ for this to be possible. But we observe from (5.186) that $\det(B) = 0$ for any Mathieu transformation. Therefore, attempted construction of the desired F_1 must fail. Nevertheless, let us examine ω_1 in this case as given by (5.79). Using (5.79), (5.183), and (5.184) gives for Mathieu transformations the result

$$\begin{aligned} \omega_1 &= \sum_k p_k dq_k - P_k dQ_k = (p, dq) - (P, dQ) = (p, dq) - (\beta^{-1}p, \beta^T dq) \\ &= (p, dq) - (\beta\beta^{-1}p, dq) = (p, dq) - (p, dq) = 0. \end{aligned} \quad (6.7.165)$$

We see that ω_1 vanishes when evaluated for any Mathieu transformation, in accord with the fact that Mathieu transformations cannot be obtained from an F_1 .

We also learned that the linear symplectic map described by the symplectic matrix R given by (4.8.31) cannot be obtained by use of any of the generating functions F_j . See the discussion in the paragraph below Equation (5.77). We will now verify that this troublesome R can be obtained using the Poincaré generating function F_+ .

According to Exercise 6.6.5 the quadratic F_+ given by (6.69), when employed in the Poincaré recipe (6.57), produces the symplectic matrix M given by (6.75). Conversely, the symmetric matrix W specifying the quadratic F_+ is given in terms of M by the relation (6.76). Examination of (6.76) shows that W can be found if $(M+I)^{-1}$ exists or, equivalently, $\det(M+I) \neq 0$. That is, M must not have -1 as an eigenvalue. For the case at hand $M = R$ and therefore

$$M + I = R + I = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}. \quad (6.7.166)$$

Simple calculation yields the result

$$\det(M + I) = 8 \neq 0. \quad (6.7.167)$$

(More extensive calculation reveals that R has the eigenvalues $1, 1, i, -i$.) Therefore W is well defined by (6.76), and produces $M = R$ when employed in (6.57) and (6.75).

Here is a slightly different perspective on the general question: Suppose the relation of a symplectic map \mathcal{M} to a gradient map \mathcal{G} with the aid of a Darboux matrix α succeeds. Then, according to the work of Subsection 7.1, there must be the relations

$$G = T_\alpha(M), \quad (6.7.168)$$

and

$$M = T_{\alpha^{-1}}(G). \quad (6.7.169)$$

Here G is the Jacobian matrix of the gradient map \mathcal{G} , and accordingly must be a symmetric matrix; and M is the Jacobian matrix of the symplectic map \mathcal{M} , and accordingly must be a symplectic matrix. For the relation (7.168) to be a well defined Möbius transformation there must be the invertibility condition

$$\det(C^\alpha M + D^\alpha) \neq 0. \quad (6.7.170)$$

We also know from the work of Section 5.11.3 that if (7.170) is satisfied, then there is also the result

$$\det(C^{\alpha^{-1}} G + D^{\alpha^{-1}}) \neq 0, \quad (6.7.171)$$

and vice versa, so that the Möbius transformation (7.169) is also well defined. That is, there is the logical equivalence

$$\det(C^{\alpha^{-1}} G + D^{\alpha^{-1}}) \neq 0 \Leftrightarrow \det(C^\alpha M + D^\alpha) \neq 0. \quad (6.7.172)$$

To verify this claim, make the substitution $W \rightarrow G$ in (5.13.99).

Let us test the condition (7.170) for some of the examples we have already discussed: First suppose $M = I$ and we choose the Darboux matrix associated with F_2 . Then we have the result

$$\det(C^\alpha M + D^\alpha) = \det(C^\alpha + D^\alpha). \quad (6.7.173)$$

But, from (7.57), we have the result

$$C^\alpha + D^\alpha = \begin{pmatrix} I^n & 0 \\ 0 & I^n \end{pmatrix}, \quad (6.7.174)$$

and therefore

$$\det(C^\alpha M + D^\alpha) = \det(C^\alpha + D^\alpha) = \det(I^{2n}) = 1 \neq 0. \quad (6.7.175)$$

Thus, we expect the use of an F_2 to succeed when $M = I$; and indeed F_2 as given by (5.71) does yield $M = I$.

Next suppose $M = J$ and we again choose the Darboux matrix associated with F_2 . Then we have the result

$$C^\alpha M + D^\alpha = C^\alpha J + D^\alpha. \quad (6.7.176)$$

But, from (7.57), we find the results

$$C^\alpha J = \begin{pmatrix} 0 & 0 \\ 0 & I^n \end{pmatrix} \begin{pmatrix} 0 & I^n \\ -I^n & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ -I^n & 0 \end{pmatrix} \quad (6.7.177)$$

and

$$C^\alpha J + D^\alpha = \begin{pmatrix} I^n & 0 \\ -I^n & 0 \end{pmatrix}. \quad (6.7.178)$$

We see that in this case

$$\det(C^\alpha M + D^\alpha) = \det(C^\alpha J + D^\alpha) = 0. \quad (6.7.179)$$

It follows that attempted use of the Darboux matrix associated with F_2 must fail when $M = J$, as we already know from the work of Subsubsection 5.1.4.

To continue, suppose $M = J$ and we choose the Darboux matrix associated with F_1 . Then we again have the result

$$C^\alpha M + D^\alpha = C^\alpha J + D^\alpha. \quad (6.7.180)$$

But now, from (7.53), we find the results

$$C^\alpha J = \begin{pmatrix} 0 & 0 \\ I^n & 0 \end{pmatrix} \begin{pmatrix} 0 & I^n \\ -I^n & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & I^n \end{pmatrix} \quad (6.7.181)$$

and

$$C^\alpha J + D^\alpha = \begin{pmatrix} I^n & 0 \\ 0 & I^n \end{pmatrix}. \quad (6.7.182)$$

We see that in this case

$$\det(C^\alpha M + D^\alpha) = \det(C^\alpha J + D^\alpha) = \det(I^{2n}) = 1 \neq 0. \quad (6.7.183)$$

Thus, we expect the use of an F_1 to succeed when $M = J$; and indeed F_1 as given by (5.76) does yield $M = J$.

Finally, to complete this set of examples, suppose again that $M = I$ but that we choose the Darboux matrix associated with F_1 . Then we have the result

$$C^\alpha M + D^\alpha = C^\alpha + D^\alpha. \quad (6.7.184)$$

But, from (7.53), we find the result

$$C^\alpha + D^\alpha = \begin{pmatrix} I^n & 0 \\ I^n & 0 \end{pmatrix}. \quad (6.7.185)$$

We see that in this case

$$\det(C^\alpha M + D^\alpha) = \det(C^\alpha + D^\alpha) = 0. \quad (6.7.186)$$

It follows that attempted use of the Darboux matrix associated with F_1 must fail when $M = I$, as we already know from the work of Subsection 5.1.4.

6.7.4.2 The General case

What can be said in general? What we wish to examine is under what conditions there is a Darboux matrix α and a source/generating function $g(u)$ such that the implicit relation (7.21) can be made explicit to become the relation (7.23) with \mathcal{M} being the desired map. We will first consider maps that have only a constant and a linear part, and then we will consider maps that have nonlinear parts.

6.7.4.2.1 Maps for the Inhomogeneous Symplectic Group $ISp(2n, \mathbb{R})$

Let us begin with maps that have only a constant and a linear part. These are the maps for the inhomogeneous symplectic group $ISp(2n, \mathbb{R})$. Our goal will be to find a Darboux matrix α and a source/generating function $g(u)$ such that their use produces maps of the form (2.10). As a first example, let us employ for the Darboux matrix α the $\tilde{\beta}$ given by (5.13.154) evaluated for $L = M$ and $V = 0$. So doing gives the result

$$\alpha = \tilde{\beta}|_{L=M, V=0} = (1/\sqrt{2}) \begin{pmatrix} -JM^{-1} & J \\ M^{-1} & I \end{pmatrix}. \quad (6.7.187)$$

For g make the choice

$$g(u) = (v, u) \quad (6.7.188)$$

where v is some vector yet to be determined. For this choice there is the relation

$$\partial_u g = v, \quad (6.7.189)$$

and use of (7.22) yields the result

$$(1/\sqrt{2})(-JM^{-1}Z + Jz) = v. \quad (6.7.190)$$

Upon solving (7.190) for Z we find the relation

$$Z = Mz + \sqrt{2}MJv. \quad (6.7.191)$$

This relation is equivalent to (2.10), after setting $Z = \bar{z}$, provided a v can be found such that

$$\sqrt{2}MJv = c, \quad (6.7.192)$$

for then (7.191) becomes

$$Z = Mz + c, \quad (6.7.193)$$

and our goal will have been accomplished. Finally, (7.192) can indeed be solved for v to yield the well defined relation

$$v = -(1/\sqrt{2})JM^{-1}c \quad (6.7.194)$$

because M is assumed to be symplectic and therefore invertible. Note that for this example the burden of producing M is borne *entirely* by the Darboux matrix, and the generating function provides only the translation part.

Next suppose we continue to use the Darboux matrix given by (7.187) but consider a more general generating function specified by the Ansatz

$$g(u) = (v, u) + (1/2)(u, Wu) \quad (6.7.195)$$

where v and W are to be determined. For this choice there is the relation

$$\partial_u g = v + Wu, \quad (6.7.196)$$

and use of (7.22) yields the result

$$(1/\sqrt{2})(-JM^{-1}Z + Jz) = v + W(1/\sqrt{2})(M^{-1}Z + z). \quad (6.7.197)$$

And solving (7.197) for Z gives the result

$$\begin{aligned} Z &= -(JM^{-1} + WM^{-1})^{-1}(W - J)z - (JM^{-1} + WM^{-1})^{-1}\sqrt{2}v \\ &= -M(J + W)^{-1}(W - J)z - M(J + W)^{-1}\sqrt{2}v \\ &= -M(J + W)^{-1}J^{-1}J(W - J)z - M(J + W)^{-1}J^{-1}J\sqrt{2}v \\ &= M(I - JW)^{-1}(I + JW)z + M(I - JW)^{-1}J\sqrt{2}v \end{aligned} \quad (6.7.198)$$

Let us define a matrix N by the rule

$$N = (I - JW)^{-1}(I + JW). \quad (6.7.199)$$

We observe from (7.199) that N is the Cayley transform of the symmetric matrix W , and therefore N is symplectic. Moreover, according to the work of Exercise 6.5, there are the relations

$$N = T_{\sigma^{-1}}(W) \quad (6.7.200)$$

and

$$W = T_\sigma(N) = -J(N - I)(N + I)^{-1}. \quad (6.7.201)$$

With the aid of the definition (7.199), the relation (7.198) can be rewritten in the form

$$Z = MNz + M(I - JW)^{-1}J\sqrt{2}v. \quad (6.7.202)$$

Yet a bit more can be accomplished by algebraic manipulation. From (7.199) we see that

$$N + I = (I - JW)^{-1}[(I + JW) + (I - JW)] = (I - JW)^{-1}(2I), \quad (6.7.203)$$

and therefore

$$(I - JW)^{-1} = (1/2)(N + I). \quad (6.7.204)$$

Consequently, (7.202) can also be rewritten in the pleasing form

$$Z = MNz + (1/\sqrt{2})M(N + I)Jv. \quad (6.7.205)$$

To continue this discussion of maps for $ISp(2n, \mathbb{R})$ using the Darboux matrix given by (7.187), let us make the definitions

$$M' = MN \text{ or } N = M^{-1}M' \quad (6.7.206)$$

and

$$c = (1/\sqrt{2})M(N + I)Jv \text{ or } v = -\sqrt{2}J(N + I)^{-1}M^{-1}c. \quad (6.7.207)$$

With these definitions we see that the relation between Z and z can be written in the final form

$$Z = M'z + c. \quad (6.7.208)$$

The general $ISp(2n, \mathbb{R})$ map has been obtained using the fixed Darboux matrix α given by (7.187) and the generating function g given by (7.195) subject only to the caveat that v and W be well defined. From the second form of (7.207) we see that the condition for v to be well defined is that $(N + I)^{-1}$ must exist. That is, there is the requirement

$$\det(N + I) \neq 0. \quad (6.7.209)$$

And from (7.201) we see that the same condition must hold for W to be well defined. We close the discussion of this example by noting that there are also the relations

$$W = T_\alpha(M') \quad (6.7.210)$$

and

$$M' = T_{\alpha^{-1}}(W). \quad (6.7.211)$$

See Exercise 7.2.

As a second example, suppose we use for the Darboux matrix α that given by (7.56) and (7.57), the Darboux matrix for the generating function F_2 , and continue to employ a $g(u)$ of the form (7.195). In this case we find that use of (7.22) yields the result

$$Z = Mz + (A^\alpha - WC^\alpha)^{-1}v \quad (6.7.212)$$

with M and W connected by the relation

$$M = T_{\alpha^{-1}}(W). \quad (6.7.213)$$

Again see Exercise 7.2.

The relation (7.213) has as its inverse the relation

$$W = T_\alpha(M) = (A^\alpha M + B^\alpha)(C^\alpha M + D^\alpha)^{-1}. \quad (6.7.214)$$

Evidently W is well defined in terms of M provided

$$\det(C^\alpha M + D^\alpha) \neq 0. \quad (6.7.215)$$

That is, a specified M can be obtained with the use of α as given by (7.56) and (7.57) and a generating function $g(u)$ provided M is such that (7.215) is satisfied. Let us write M in the block form (5.25). Employing this form gives the result

$$C^\alpha M + D^\alpha = \begin{pmatrix} I^n & 0 \\ C & D \end{pmatrix}, \quad (6.7.216)$$

from which it follows that

$$\det(C^\alpha M + D^\alpha) = \det D. \quad (6.7.217)$$

Therefore, for the use of the Darboux matrix α associated with F_2 , the condition (7.215) becomes

$$\det D \neq 0, \quad (6.7.218)$$

in agreement with (5.5).

As a third example, suppose we use for α the Cayley Darboux matrix σ , the Darboux matrix associated with F_+ . See (7.68). Put another way, suppose we use for α the Darboux matrix $\tilde{\beta}$ given by (5.13.154) evaluated at $L = I$ and $V = 0$. So doing gives the result

$$\alpha = \tilde{\beta}|_{L=I,V=0} = (1/\sqrt{2}) \begin{pmatrix} -J & J \\ I & I \end{pmatrix} = \sigma. \quad (6.7.219)$$

The relations (7.212) through (7.215) continue to hold since they are true for any choice of α . But now (7.216) becomes

$$C^\sigma M + D^\sigma = (1/\sqrt{2})(M + I). \quad (6.7.220)$$

Correspondingly, the condition (2.15) becomes

$$\det(M + I) \neq 0. \quad (6.7.221)$$

That is, M must not have -1 as an eigenvalue, a requirement that we already have learned for the existence of a Cayley representation. Recall Section 3.12.

Let us summarize what we have learned about the use of Darboux matrices and generating functions for the case of $ISp(2n, \mathbb{R})$ maps. From the first example we have seen that for any symplectic matrix M there is an associated Darboux matrix α given by (7.187) such that the full burden of producing M is borne by the Darboux matrix and the generating function $g(u)$ is required only to produce the translation part. We may say that this Darboux matrix is *optimally compatible* with M . Moreover, $ISp(2n, \mathbb{R})$ maps of the form (7.208) with symplectic matrices M' of the form (7.206), can also be produced using the

same Darboux matrix and a suitable $g(u)$ provided M' is sufficiently near M so that (7.209) is satisfied.

Conversely, it is attractive to conjecture that, for any Darboux matrix α , there is an $Sp(2n, \mathbb{R})$ [and also an $ISp(2n, \mathbb{R})$] map with symplectic matrix M for which M and α are incompatible. That is, given any Darboux matrix α , there is a symplectic matrix M such that

$$\det(C^\alpha M + D^\alpha) = 0. \quad (6.7.222)$$

Put another way, there is no fixed/universal Darboux matrix (generating function kind) that is compatible with all symplectic matrices. Correspondingly, there is no fixed/universal Darboux matrix (generating function kind) that is compatible with all symplectic maps. See Exercise 7.17. Examples 2 and 3 of this subsection, as well as examples at the beginning of this section, illustrate instances of incompatibility.

6.7.4.2.2 Maps with Nonlinear Parts

In Section 7.8 we will see that any symplectic map \mathcal{M} can be written in the Lie form

$$\mathcal{M} = \exp(: f_1 :) \mathcal{RN} \quad (6.7.223)$$

where the factor $\exp(: f_1 :)$ produces a translation, the factor

$$\mathcal{R} = \exp(: f_2^c :) \exp(: f_2^a :) \quad (6.7.224)$$

produces a linear transformation described by the symplectic matrix R , and \mathcal{N} is the nonlinear map

$$\mathcal{N} = \exp(: f_3 :) \exp(: f_4 :) \cdots. \quad (6.7.225)$$

It can be shown that any such map \mathcal{M} can be produced with the aid of a suitable Darboux matrix α and generating function $g(u)$ pair. However, if this is done, the choice of α will be constrained by the requirement that it be compatible with R . An alternative, which we will follow, is to represent just the nonlinear part \mathcal{N} by a Darboux matrix-generating function pair. This allows for flexibility in the treatment of \mathcal{N} and causes no undue problem in dealing with the $ISp(2n, \mathbb{R})$ part of \mathcal{M} , namely the $\exp(: f_1 :) \mathcal{R}$ part, since it can be handled separately using the methods of Subsection 7.4.2.1 above and those of Chapter 9.

In the nonlinear case we would like the relation (7.22), and the ability to make it explicit, to hold over as large a phase-space region as possible. In this subsection we will see from some simple examples that the choice of α influences what can be achieved. These examples will also illustrate that the subject of nonlinear symplectic maps is very complicated. Therefore the discussion of this subsection will be limited. A fuller discussion will be undertaken in Chapter 34.

The complexity of nonlinear symplectic maps is already evident at the quadratic level in two variables. Consider the map

$$Q = q - 2qp + O(z^3), \quad (6.7.226)$$

$$P = p + p^2 + O(z^3). \quad (6.7.227)$$

It satisfies the relation

$$\begin{aligned}[Q, P] &= [q - 2qp, p + p^2] + O(z^2) \\ &= [q, p] + [q, p^2] - 2[qp, p] - 2[qp, p^2] + O(z^2) \\ &= 1 + 2p - 2p + O(z^2) = 1 + O(z^2).\end{aligned}\tag{6.7.228}$$

Therefore the terms displayed in (7.226) and (7.227) constitute a *symplectic jet*. Indeed, they can be written in the form

$$Z = z + :f_3:z + O(z^3)\tag{6.7.229}$$

with

$$f_3 = qp^2.\tag{6.7.230}$$

As described in Chapter 34, there are important instances in which a symplectic jet approximation to a symplectic map is *inadequate*. In these cases it is desirable to find an exactly symplectic map whose truncated Taylor expansion matches some specified symplectic jet. Such a symplectic map will be called a *symplectic completion* of the specified symplectic jet.

One way to symplectically complete the symplectic jet (7.229) is to write

$$Z = \mathcal{N}z\tag{6.7.231}$$

with

$$\mathcal{N} = \exp(:f_3:).\tag{6.7.232}$$

We will call this *Lie symplectification*. So doing gives the result

$$Q = q(1 - p)^2,\tag{6.7.233}$$

$$P = p/(1 - p).\tag{6.7.234}$$

See Section 1.4.2. [Note that the Taylor expansion through second order of the map given by (7.233) and (7.234) agrees with that in (7.226) and (7.227).] We observe that the map given by (7.230) through (7.234) is analytic at the origin and has a pole on the surface $p = 1$.

We will next explore two examples of how symplectic completion can be achieved with the use of generating functions. In these examples we will again work with the symplectic jet given by (7.226) and (7.227). Let us begin with the use of an F_2 generating function. Suppose we make the Ansatz

$$F_2(q, P) = qP - qP^2.\tag{6.7.235}$$

Use of this Ansatz in (5.5) produces the implicit equations

$$p = \partial F_2 / \partial q = P - P^2,\tag{6.7.236}$$

$$Q = \partial F_2 / \partial P = q - 2qP.\tag{6.7.237}$$

Since these equations are quadratic, they can be solved exactly, and we will do so shortly. First, however, let us find the first few terms in the Taylor expansions of $Q(q, p)$ and $P(q, p)$ in powers of q and p . Rewrite (7.236) in the form

$$P = p + P^2.\tag{6.7.238}$$

Now we can expand Q and P in powers of q and p by iteration of (7.237) and (7.238). In lowest approximation, they have the solution

$$Q = q + O(z^2), \quad (6.7.239)$$

$$P = p + O(z^2). \quad (6.7.240)$$

Now substitute (7.239) and (7.240) into (7.237) and (7.238) to get the improved solution

$$Q = q - 2qp + O(z^3), \quad (6.7.241)$$

$$P = p + p^2 + O(z^3). \quad (6.7.242)$$

We have verified that the use of the generating function Ansatz (7.235) produces a (symplectic) map whose Taylor expansion through second order yields the jet map given by (7.226) and (7.227).

Let us now solve (7.236) and (7.237) to find $Q(q, p)$ and $P(q, p)$ exactly. Solving (7.236) for P gives the result

$$P = (1/2)[1 - (1 - 4p)^{1/2}], \quad (6.7.243)$$

and substituting (7.243) into (7.237) gives the complementary result

$$Q = q(1 - 4p)^{1/2}. \quad (6.7.244)$$

[Note that the Taylor expansion through second order of the map given by (7.243) and (7.244) agrees with that in (7.226) and (7.227).] We see that the map given by (7.243) and (7.244) is analytic at the origin and has a branch point on the surface $p = 1/4$.

What happens if we use an F_+ generating function instead of F_2 ? According to (7.68) this amounts to choosing the Darboux matrix α to be σ . Also, according to (7.219) and the previous discussion of compatibility, in choosing α to be σ we have chosen α to be the Darboux matrix that is optimally compatible with the symplectic matrix I . Finally, the linear part of the map given by (7.226) and (7.227) is indeed the identity matrix I .

In the two-dimensional case the variable u appearing in (7.22) will have the components

$$u = \{u_1; u_2\}^T. \quad (6.7.245)$$

Make the Ansatz

$$g(u) = -(\sqrt{2}/4)u_1(u_2)^2. \quad (6.7.246)$$

In this case

$$\partial_u g = \{-(\sqrt{2}/4)(u_2)^2; -(\sqrt{2}/2)u_1u_2\}^T, \quad (6.7.247)$$

and use of (7.22) with the Darboux matrix α given by (7.68) yields the implicit equations

$$(1/\sqrt{2})(-JZ + Jz) = \{-(\sqrt{2}/4)(u_2)^2; -(\sqrt{2}/2)u_1u_2\}^T|_{u=(1/\sqrt{2})(Z+z)}. \quad (6.7.248)$$

The equations (7.248) can be rewritten in the vector form

$$\begin{aligned} Z - z &= (\sqrt{2})J\{-(\sqrt{2}/4)(u_2)^2; -(\sqrt{2}/2)u_1u_2\}^T|_{u=(1/\sqrt{2})(Z+z)} \\ &= J\{-(1/4)(P + p)^2; -(1/2)(Q + q)(P + p)\}^T \\ &= \{-(1/2)(Q + q)(P + p); +(1/4)(P + p)^2\}^T. \end{aligned} \quad (6.7.249)$$

Finally, the vector form (7.249) is equivalent to the component equations

$$Q - q = -(1/2)(Q + q)(P + p), \quad (6.7.250)$$

$$P - p = (1/4)(P + p)^2. \quad (6.7.251)$$

The component equations (7.250) and (7.251) are quadratic and can be solved explicitly, which we will do shortly. But first let us seek Taylor expansions for $Q(q, p)$ and $P(q, p)$. Rewrite (7.250) and (7.251) in the forms

$$Q = q - (1/2)(Q + q)(P + p), \quad (6.7.252)$$

$$P = p + (1/4)(P + p)^2 \quad (6.7.253)$$

and iterate them once and then once again to find the results

$$Q = q + O(z^2), \quad (6.7.254)$$

$$P = p + O(z^2); \quad (6.7.255)$$

$$Q = q - 2qp + O(z^3), \quad (6.7.256)$$

$$P = p + p^2 + O(z^3). \quad (6.7.257)$$

We see that (7.256) and (7.257) agree with (7.226) and (7.227) as desired.

Let us now solve (7.252) and (7.253) to find $Q(q, p)$ and $P(q, p)$ exactly. Solving (7.253) for P gives the result

$$P = 2 - p - 2(1 - 2p)^{1/2}, \quad (6.7.258)$$

and substituting (7.258) into (7.252) and solving for Q gives the complementary result

$$Q = q(1 - 2p)^{1/2} / [2 - (1 - 2p)^{1/2}]. \quad (6.7.259)$$

[Note that the Taylor expansion through second order of the map given by (7.258) and (7.259) agrees with that in (7.226) and (7.227).] We see that the map given by (7.258) and (7.259) is analytic at the origin and has a branch point on the surface $p = 1/2$.

6.7.4.2.3 Concluding Discussion

We have studied how the choice of a Darboux matrix-generating function pair affects the representation of maps that have only constant and linear terms, namely $ISp(2n, \mathbb{R})$ maps, and also how the choice affects nonlinear maps. For the $ISp(2n, \mathbb{R})$ case, because of its relative simplicity, the treatment was essentially complete. By contrast, the nonlinear case is much more complicated. Below are some questions/observations that naturally arise about the nonlinear case along with partial responses:

- How did we know to make the generating function Ansätze (7.235) and (7.246)? Chapter 34 describes how, for any choice of Darboux matrix α , the Lie f_n and the generating function g_n are related.

- Why are the resulting maps given by (7.233) and (7.234), by (7.243) and (7.244), and by (7.258) and (7.259) all different even though they symplectify (symplectically complete) the same two-jet given by (7.226) and (7.227)? As found in Chapter 34, even if all the f_n vanish beyond some n value n_{\max} , the same will not be true for the associated g_n . Had these high-order g_n (generally infinite in number) been retained, the resulting maps would agree.
- Because only g_3 (and perhaps g_2) generating functions were involved in our examples, the implicit equations produced by their use were quadratic, and therefore could be solved exactly. What can be done, as occurs for more realistic cases, when the implicit equations are higher order? Chapter 34 describes various iterative methods, including Newton's method, for efficiently solving the implicit equations numerically.
- For the f_n case studied, namely that given by (7.230), it was found that the use of an F_2 generating function produced a map that was singular at $p = 1/4$, and the use of an F_+ generating function produced a map that was singular at $p = 1/2$. We have seen that in the linear case an F_+ generating function is more compatible with the identity symplectic matrix I than is an F_2 generating function, and for nonlinear maps of the form (7.225) the linear part is the identity map. Is it significant that the use of an F_+ generating function produced a map with a *larger* domain of analyticity than the use of an F_2 generating function?

Let us review/reconsider when the Möbius transformations relating gradient and symplectic maps succeed or fail. We recall that G is the Jacobian of the gradient map \mathcal{G} and M is the Jacobian of the symplectic map \mathcal{M} . Let α be the Darboux matrix that produces the Möbius transformations T_α and $T_{\alpha^{-1}}$. According to (7.48) and (7.40) there are the complementary relations

$$G = T_\alpha(M) = (A^\alpha M + B^\alpha)(C^\alpha M + D^\alpha)^{-1} \quad (6.7.260)$$

and

$$M = T_{\alpha^{-1}}(G) = (A^{\alpha^{-1}} G + B^{\alpha^{-1}})(C^{\alpha^{-1}} G + D^{\alpha^{-1}})^{-1}. \quad (6.7.261)$$

For (7.260) to be well defined we must have

$$\det(C^\alpha M + D^\alpha) \neq 0, \quad (6.7.262)$$

and for (7.261) to be well defined we must have

$$\det(C^{\alpha^{-1}} G + D^{\alpha^{-1}}) \neq 0. \quad (6.7.263)$$

Recall that the conditions (7.262) and (7.263) are logically equivalent. See (7.172). Therefore (7.260) is well defined if (7.261) is well defined, and vice versa.

Conversely, we expect that determination of G and therefore construction of the associated generating function $g(u)$ will fail if

$$\det(C^\alpha M + D^\alpha) = 0 : \text{Condition for incompatibility of } \alpha \text{ and } M. \quad (6.7.264)$$

We have already seen examples of this incompatibility. Moreover, we expect that the construction of a satisfactory \mathcal{M} from a generating function $g(u)$ using (7.22) will fail if

$$\det(C^{\alpha^{-1}}G + D^{\alpha^{-1}}) = 0 : \text{ Condition for map construction failure.} \quad (6.7.265)$$

Put another way, complementary to the logical equivalence (7.172), there are the logical implications

$$\det(C^\alpha M + D^\alpha) = 0 \Rightarrow G \text{ and hence } \mathcal{G} \text{ and } g \text{ are not defined,} \quad (6.7.266)$$

$$\det(C^{\alpha^{-1}}G + D^{\alpha^{-1}}) = 0 \Rightarrow M \text{ and hence } \mathcal{M} \text{ are not defined.} \quad (6.7.267)$$

We will see, for our examples, that the appearance of singularities is related to map construction failure, the condition (7.265).

Consider first the use of F_2 . We could treat this case by employing the condition (7.265) with α being the Darboux matrix associated with F_2 . See Exercise 7.18. Instead, and equivalently as shown in Exercise 7.19, we already know that use of F_2 assumes that

$$\det(\partial^2 F_2 / \partial q_k \partial P_\ell) \neq 0. \quad (6.7.268)$$

See (5.5). For the case (7.235) there is the result

$$\partial^2 F_2 / \partial q \partial P = 1 - 2P, \quad (6.7.269)$$

and (7.268) fails when

$$P = 1/2. \quad (6.7.270)$$

From (7.243) we see that (7.270) implies that

$$p = 1/4, \quad (6.7.271)$$

the value for which the map given by (7.243) and (7.244) has a singularity!

Next consider the use of F_+ . For the case of F_+ , the Darboux matrix is σ . See (7.68). Correspondingly, the condition for map construction failure becomes

$$\det(C^{\sigma^{-1}}G + D^{\sigma^{-1}}) = 0. \quad (6.7.272)$$

From (5.13.12) we see that (7.272) has the specific form

$$\det(-JG + I) = 0. \quad (6.7.273)$$

Since G is the Jacobian of the gradient map \mathcal{G} , see (1.6), it follows that in our two-dimensional case G is given by the matrix

$$G = \begin{pmatrix} \partial^2 g / \partial u_1 \partial u_1 & \partial^2 g / \partial u_1 \partial u_2 \\ \partial^2 g / \partial u_2 \partial u_1 & \partial^2 g / \partial u_2 \partial u_2 \end{pmatrix} = \begin{pmatrix} 0 & -(\sqrt{2}/2)u_2 \\ -(\sqrt{2}/2)u_2 & -(\sqrt{2}/2)u_1 \end{pmatrix}. \quad (6.7.274)$$

Here we have used (7.246). Consequently, we find that

$$-JG + I = \begin{pmatrix} 1 + (\sqrt{2}/2)u_2 & (\sqrt{2}/2)u_1 \\ 0 & 1 - (\sqrt{2}/2)u_2 \end{pmatrix}, \quad (6.7.275)$$

and the condition (7.273) yields the relation

$$1 - (1/2)(u_2)^2 = 0 \quad (6.7.276)$$

with the solution

$$u_2 = \pm\sqrt{2}. \quad (6.7.277)$$

Also, when $\alpha = \sigma$, (7.18) becomes

$$u = C^\sigma Z + D^\sigma z = (1/\sqrt{2})(Z + z) = (1/\sqrt{2})\{Q + q; P + p\} \quad (6.7.278)$$

so that

$$u_2 = (1/\sqrt{2})(P + p). \quad (6.7.279)$$

See (7.68). Employing (7.277) with a + sign converts (7.279) to the relation

$$P + p = 2. \quad (6.7.280)$$

Finally, inserting (7.280) into (7.258) yields the result

$$p = 1/2, \quad (6.7.281)$$

the value for which the map given by (7.258) and (7.259) has a singularity!

Let us summarize what has been learned about generating function symplectification: Suppose one is given a symplectic jet of the form

$$Z_a = z_a + \sum_{bc} T_{abc} z_b z_c + \sum_{bcd} U_{abcd} z_b z_c z_d + \dots + O(z^{n_{\max}+1}). \quad (6.7.282)$$

That is, only the terms through degree n_{\max} are given. Select a Darboux matrix α that is compatible with the identity matrix I . From the coefficients in the jet (7.282) and the selected α one can construct a unique polynomial generating function (that will depend on α) of the form

$$g = \sum_{n=2}^{n=n_{\max}+1} g_n \quad (6.7.283)$$

such that its use in (7.22) reproduces the jet (7.282). Then the (symplectic) map obtained by making explicit the implicit relations (7.22), call it $\mathcal{M}(\alpha, g)$, will be analytic about the origin and may be expected to have singularities when (7.265) holds. If we select $\alpha = \sigma$, the Darboux matrix that is optimally compatible with the symplectic matrix I , then $g_2 = 0$. Moreover, based on our examples, we anticipate that $\mathcal{M}(\sigma, g)$ will have optimal analytic properties.

- Lie symplectification (symplectic completion), that given by (7.232), appears from our first example to be superior from the perspective of the size of the analyticity domain because it produced a symplectic map that is singular on the surface $p = 1$. But Lie symplectification can be carried out analytically only in a few special cases, and its implementation by numerical methods requires the summation of a very large (generally infinite) number of terms. Recall the definition (1.2.44). Therefore its use is generally impractical when rapid computation is required.
- For a symplectification (symplectic completion) procedure that produces from a symplectic jet a symplectic map having *no* singularities (save at infinity), see Section 34.2.4.

Exercises

6.7.1. Relate the discussion surrounding (4.8.21) through (4.8.26) in Section 4.8 to the relations (7.55) through (7.58) in Subsection 7.1.

6.7.2. Suppose g is a source function of the form

$$g(u) = (1/2)(u, Wu) \quad (6.7.284)$$

where W is a symmetric $2n \times 2n$ matrix. It follows that in this case

$$\partial_u g = Wu. \quad (6.7.285)$$

Show that the \mathcal{M} produced by this g using (7.21) is linear and is described by the matrix

$$M = T_{\alpha^{-1}}(W). \quad (6.7.286)$$

To do so, use (7.38) with the substitution $G \rightarrow W$.

Suppose g is of the form

$$g(u) = (v, u) + (1/2)(u, Wu) \quad (6.7.287)$$

where v is any vector. It follows that in this case

$$\partial_u g = v + Wu. \quad (6.7.288)$$

Show that now there is the relation

$$Z = Mz + (A^\alpha - WC^\alpha)^{-1}v. \quad (6.7.289)$$

Verify that

$$M = T_{\alpha^{-1}}(W) = (A^{\alpha^{-1}}W + B^{\alpha^{-1}})(C^{\alpha^{-1}}W + D^{\alpha^{-1}})^{-1} \quad (6.7.290)$$

is well defined providing

$$\det(C^{\alpha^{-1}}W + D^{\alpha^{-1}}) \neq 0. \quad (6.7.291)$$

But, for (7.289) to make sense, we must also have

$$\det(A^\alpha - WC^\alpha) \neq 0. \quad (6.7.292)$$

Is this an additional requirement? Show that it is not. Select from the chain of inferences (5.11.42) the inference

$$\det(C^M U + D^M) \neq 0 \Leftrightarrow \det(UC^{M^{-1}} - A^{M^{-1}}) \neq 0 \quad (6.7.293)$$

and verify that one may make the substitutions $M \rightarrow \alpha^{-1}$ and $U \rightarrow W$ to obtain the logical equivalence

$$\det(C^{\alpha^{-1}} W + D^{\alpha^{-1}}) \neq 0 \Leftrightarrow \det(A^\alpha - WC^\alpha) \neq 0. \quad (6.7.294)$$

Therefore (7.292) is a consequence of (7.291), and vice versa.

Verify the correctness of the discussion involving (7.88) through (7.96).

6.7.3. If (7.21) or (7.22) is explicit, α must have the property $B^\alpha = C^\alpha = 0$. See (7.17) and (7.18). Compute γ for such an α using the relation $\gamma = \alpha\sigma^{-1}$ with σ^{-1} given by (5.13.12). Show that the γ so obtained cannot satisfy the symplectic conditions (3.3.3) and (3.3.4). Therefore α cannot be a Darboux matrix.

6.7.4. Verify that the parametric representation of \mathcal{M} given by (7.28) and (7.29) yields (7.39) and (7.40) directly.

6.7.5. The purpose of this exercise is to verify the F_1 contents of Table 7.1. Verify that the α given by (7.52) and (7.53) is a Darboux matrix, is also orthogonal, and its use reproduces the equations (7.51) when employed in (7.21). Verify that γ given by (7.54) satisfies $\alpha = \gamma\sigma$ and is J^{4n} symplectic.

Hint: Show that, for the α given by (7.52) and (7.53), the relations (7.17) and (7.18) take the form

$$U_i = p_i \text{ for } i = 1 \text{ to } n, \quad (6.7.295)$$

$$U_i = -P_{i-n} \text{ for } i = n+1 \text{ to } 2n, \quad (6.7.296)$$

$$u_i = q_i \text{ for } i = 1 \text{ to } n, \quad (6.7.297)$$

$$u_i = Q_{i-n} \text{ for } i = n+1 \text{ to } 2n. \quad (6.7.298)$$

Suppose we partition u into two parts, each of length/dimension n , by writing

$$u = (v; w). \quad (6.7.299)$$

Then the relations (7.297) and (7.298) become

$$v = q, \quad w = Q. \quad (6.7.300)$$

Thus, if we use the partition (7.299), we may write

$$g(u, t) = g(v; w, t). \quad (6.7.301)$$

Show that there is the relation

$$g(v; w, t) = F_1(v, w, t). \quad (6.7.302)$$

6.7.6. The purpose of this exercise is to verify the F_2 contents of Table 7.1. Verify that the α given by (7.56) and (7.57) is a Darboux matrix, is also orthogonal, and its use reproduces the equations (7.55) when employed in (7.21). Verify that γ given by (7.58) satisfies $\alpha = \gamma\sigma$ and is J^{4n} symplectic.

Hint: Show that, for the α given by (7.56) and (7.57), the relations (7.17) and (7.18) take the form

$$U_i = p_i \text{ for } i = 1 \text{ to } n, \quad (6.7.303)$$

$$U_i = Q_{i-n} \text{ for } i = n+1 \text{ to } 2n, \quad (6.7.304)$$

$$u_i = q_i \text{ for } i = 1 \text{ to } n, \quad (6.7.305)$$

$$u_i = P_{i-n} \text{ for } i = n+1 \text{ to } 2n. \quad (6.7.306)$$

Suppose we partition u into two parts, each of length/dimension n , by writing

$$u = (v; w). \quad (6.7.307)$$

Then the relations (7.305) and (7.306) become

$$v = q, \quad w = P. \quad (6.7.308)$$

Thus, if we use the partition (7.307), we may write

$$g(u, t) = g(v; w, t). \quad (6.7.309)$$

Show that there is the relation

$$g(v; w, t) = F_2(v, w, t). \quad (6.7.310)$$

6.7.7. The purpose of this exercise is to verify and work with the F_+ contents of Table 7.1. Your task is to show that use of the α given by (7.68) reproduces the equations (7.67) when employed in (7.21) or (7.22).

First show that, for the α given by (7.68), the relations (7.17) and (7.18) take the form

$$U = A^\sigma Z + B^\sigma z = -(1/\sqrt{2})J\Delta, \quad (6.7.311)$$

$$u = C^\sigma Z + D^\sigma z = (1/\sqrt{2})\Sigma. \quad (6.7.312)$$

Note that (7.312) can be rewritten as

$$\Sigma = \sqrt{2}u. \quad (6.7.313)$$

Next show that the relation (7.22) takes the form

$$-(1/\sqrt{2})J\Delta = \partial_u g|_{u=(1/\sqrt{2})\Sigma}, \quad (6.7.314)$$

from which it follows that

$$\Delta = \sqrt{2}J\partial_u g|_{u=(1/\sqrt{2})\Sigma}. \quad (6.7.315)$$

Let us compare this result with the Poincaré generating function result (7.67) which reads

$$\Delta = J\partial_u F_+|_{u=\Sigma}. \quad (6.7.316)$$

Verify that (7.315) and (7.316) are equivalent when there is the relation

$$g(u, t) = (1/2)F_+(\Sigma, t) = (1/2)F_+(u\sqrt{2}, t). \quad (6.7.317)$$

To do so, apply the chain rule to (7.317) to show that

$$\partial g / \partial u_a = (1/2) \sum_b (\partial F_+ / \partial \Sigma_b) (\partial \Sigma_b / \partial u_a). \quad (6.7.318)$$

But, by (7.313), verify that there is the relation

$$\partial \Sigma_b / \partial u_a = \sqrt{2} \delta_{ba}. \quad (6.7.319)$$

Verify that therefore (7.318) can be rewritten in the component form

$$\partial g / \partial u_a = (1/\sqrt{2}) \partial F_+ / \partial \Sigma_a \quad (6.7.320)$$

or, more compactly, in the vector form

$$\partial_u g = (1/\sqrt{2}) \partial_\Sigma F_+. \quad (6.7.321)$$

Finally, employ (7.321) in (7.315) to find the result

$$\Delta = \sqrt{2} J \partial_u g = \sqrt{2} J (1/\sqrt{2}) \partial_\Sigma F_+ = J \partial_\Sigma F_+, \quad (6.7.322)$$

which agrees with (7.316).

As a simple example, suppose g is of the form

$$g(u) = (v', u) + (1/2)(u, W'u) \quad (6.7.323)$$

where v' is any vector and W' is any symmetric matrix. Then

$$\partial_u g = v' + W'u. \quad (6.7.324)$$

Show that in this case there is the relation

$$\begin{aligned} Z &= Mz + (A^\sigma - W'C^\sigma)^{-1}v' = Mz + [-(1/\sqrt{2})J - (1/\sqrt{2})W']^{-1}v' \\ &= Mz - \sqrt{2}(J + W')^{-1}v' = Mz - \sqrt{2}(J - JW')^{-1}v' \\ &= Mz - \sqrt{2}(I - JW')^{-1}J^{-1}v' = Mz + \sqrt{2}(I - JW')^{-1}Jv' \end{aligned} \quad (6.7.325)$$

with

$$M = T_{\sigma^{-1}}(W') = (I - JW')^{-1}(I + JW'). \quad (6.7.326)$$

Compare this result with that given by (6.75) in the case that F_+ is of the form (6.69). Verify that (6.81) and (7.326) agree provided

$$W' = W, \quad (6.7.327)$$

and

$$v' = (1/\sqrt{2})v. \quad (6.7.328)$$

Is this result consistent with (7.317)?

6.7.8. Verify that (7.87) is equivalent to (7.86).

6.7.9. In a $2n$ -dimensional phase space consider the n -dimensional submanifold parameterized by the equations

$$q_i = \tau_i, \quad (6.7.329)$$

$$p_i = p_i^0. \quad (6.7.330)$$

Show that this submanifold is J^{2n} Lagrangian. Consider the submanifold parameterized by the equations

$$q_i = \tau_i, \quad (6.7.331)$$

$$p_i = \tau_i. \quad (6.7.332)$$

What can be said about it? What can be said about the submanifold parameterized by the equations

$$q_i = \tau_i, \quad (6.7.333)$$

$$p_i = \partial f(\tau)/\partial \tau_i, \quad (6.7.334)$$

where f is any function of τ ?

6.7.10. Verify that the graph of \mathcal{M} is a \tilde{J}^{4n} Lagrangian submanifold using the parametric form of \mathcal{M} given by (7.28) and (7.29). Hint: Write that

$$\text{graph of } \mathcal{M} = \{\hat{Z} \in \mathbb{R}^{4n} \mid \hat{Z} = \alpha^{-1}\hat{U}, \hat{U} = (U; u)^T = (\mathcal{G}u; u)^T \text{ with } u \in \mathbb{R}^{2n}\}. \quad (6.7.335)$$

Make the Ansatz

$$u = u^0 + \sum_1^{2n} \lambda_i e^i. \quad (6.7.336)$$

Show that the tangent vectors ζ^j to the graph of \mathcal{M} are given by the relations

$$\zeta^j(u^0) = \partial \hat{Z}/\partial \lambda_j|_{\lambda=0} = \alpha^{-1} \partial \hat{U}/\partial \lambda_j|_{\lambda=0} = \alpha^{-1} \nu^j(u^0). \quad (6.7.337)$$

Verify that these tangent vectors are \tilde{J}^{4n} isotropic by showing that

$$(\zeta^j, \tilde{J}^{4n} \zeta^k) = (\alpha^{-1} \nu^j, \tilde{J}^{4n} \alpha^{-1} \nu^k) = (\nu^j, (\alpha^{-1})^T \tilde{J}^{4n} \alpha^{-1} \nu^k) = (\nu^j, J^{4n} \nu^k) = 0. \quad (6.7.338)$$

6.7.11. Suppose all the tangent vectors of a submanifold are mutually isotropic at each point in the submanifold. Such a submanifold is called isotropic or *null*. Show that in a $2n$ -dimensional phase space the largest dimension a null submanifold can have is n . Thus a Lagrangian submanifold has the largest possible dimension for a null submanifold. For simplicity, work with J^{2n} isotropic submanifolds.

6.7.12. Suppose, in the relation between symplectic and gradient maps, we wish to arrange to have the identity map \mathcal{I} correspond to the case of a zero (or constant) source function g , and vice versa. What can be said about the Darboux matrix α in this case? Is it unique? Far from it.

If g is zero or constant, we must have

$$\mathcal{G} = 0 \quad (6.7.339)$$

so that

$$G = 0. \quad (6.7.340)$$

Show that, since the linear part of the identity map is the identity matrix I , the relation (7.40) then becomes

$$I = T_{\alpha^{-1}}(0), \quad (6.7.341)$$

or, equivalently,

$$T_\alpha(I) = 0. \quad (6.7.342)$$

Employ the factorization (7.50), namely

$$\alpha = \gamma\sigma, \quad (6.7.343)$$

so that (7.342) becomes

$$T_{\gamma\sigma}(I) = 0. \quad (6.7.344)$$

Show, from the group property of Möbius transformations and (5.14.2), that

$$T_{\gamma\sigma}(I) = T_\gamma(T_\sigma(I)) = T_\gamma(0), \quad (6.7.345)$$

and conclude that

$$T_\gamma(0) = 0. \quad (6.7.346)$$

Review the discussions at the ends of Sections 5.12.7 and 5.13.9.3, and show that one must have

$$\gamma \in H(4n, \mathbb{R}). \quad (6.7.347)$$

Finally, verify that (7.339), (7.343), and (7.347), when employed in (7.28) and (7.29), yield the identity map,

$$Z = z. \quad (6.7.348)$$

6.7.13. The discussion at the end of Subsection 7.4.1 examined what the conditions on M were for W as given by (6.76) to be well defined. What are the conditions on W for M as given by (6.75) to be well defined? Compute W for the case $M = R$ with R given by (4.8.31). Verify that this W satisfies the conditions for M as given by (6.75) to be well defined.

6.7.14. Observe that A as given by (5.120) and A' as given by (7.130) are different. From our previous discussion we know that they both take extrema on the trajectories generated by H . However, A' treats the coordinates ξ and momenta η on an equal footing while A does not. Nevertheless, they are related. Show that

$$A = -(1/2)A' + (1/2) \sum_i (Q_i P_i - q_i p_i). \quad (6.7.349)$$

6.7.15. From (6.15) and (7.142) we see that F and A' are related. Show that if $\mathcal{M}(t^i, t^f)$ is a symplectic map generated by the Hamiltonian $H(\zeta, t)$ using t^i and t^f as initial and final times, then the F of (6.15) is given by the relation

$$F(z, t^i, t^f) = 2 \int_{t^i, z}^{t^f, Z(z)} [(\zeta, J\dot{\zeta})/2 + H(\zeta, t)] dt. \quad (6.7.350)$$

Here the integral is to be evaluated for the trajectory of H satisfying (7.124).

Consider each of the three cases

$$H = (k, \zeta) \quad (6.7.351)$$

where k is a constant vector,

$$H = (1/2)(\zeta, S\zeta) \quad (6.7.352)$$

where S is a constant symmetric matrix, and

$$H = (1/3)(\zeta_1)^3. \quad (6.7.353)$$

In each case find \mathcal{M} , verify that (6.3) when viewed as a function z is an exact differential, and find an F such that (6.15) is satisfied.

Next suppose that H is of the form

$$H(\zeta) = h_m(\zeta) \quad (6.7.354)$$

where $h_m(\zeta)$ is a homogeneous polynomial of degree m in the variables ζ_a . Show that in this case

$$F(z, t^i, t^f) = -(m-2)(t^f - t^i)h_m(z). \quad (6.7.355)$$

Note that if H is quadratic, then F vanishes. Because quadratic Hamiltonians generate linear symplectic maps, this result is consistent with the earlier discussion of the fact that F is the same for all linear symplectic maps.

Finally, if

$$H(\zeta) = \sum_m h_m(\zeta), \quad (6.7.356)$$

show that

$$F(z, t^i, t^f) = - \int_{t^i}^{t^f} dt \sum_m (m-2)h_m(\zeta). \quad (6.7.357)$$

6.7.16. For the map given by (7.233) and (7.234) and the map given by (7.243) and (7.244) show by direct computation/evaluation that

$$[Q, P] = 1, \quad (6.7.358)$$

thereby verifying that these maps are symplectic.

6.7.17. Recall the *incompatibility* condition (7.264), which also appears in (7.222). To free up some symbols for subsequent different use, let us rewrite (7.264) in the form

$$\det(C^\beta M' + D^\beta) = 0. \quad (6.7.359)$$

Our goal is to show that, given any Darboux matrix β , there is a symplectic matrix M' such that the incompatibility condition (7.359) holds.

According to Section 5.13.9.4, the most general Darboux matrix β can be written in the form (5.13.148). Verify that, consequently, there are the relations

$$C^\beta = (1/\sqrt{2})(A^T)^{-1}(-CJ + I)L^{-1} \quad (6.7.360)$$

and

$$D^\beta = (1/\sqrt{2})(A^T)^{-1}(CJ + I); \quad (6.7.361)$$

and therefore (7.359) amounts to the requirement

$$\det[(1/\sqrt{2})(A^T)^{-1}(-CJ + I)L^{-1}M' + (1/\sqrt{2})(A^T)^{-1}(CJ + I)] = 0. \quad (6.7.362)$$

Verify, since A is assumed to be invertible, that the requirement (7.362) is equivalent to the requirement

$$\det[(-CJ + I)L^{-1}M' + (CJ + I)] = 0. \quad (6.7.363)$$

Let us now write M' in the factorized form

$$M' = -LK \quad (6.7.364)$$

where K is yet to be determined. Verify that employing this M' in the argument of (7.363) produces the result

$$\begin{aligned} (-CJ + I)L^{-1}M' + (CJ + I) &= -(-CJ + I)L^{-1}LK + (I + CJ) \\ &= -(-CJ + I)K + (I + CJ). \end{aligned} \quad (6.7.365)$$

Suppose we require that

$$(-CJ + I)L^{-1}M' + (CJ + I) = -(-CJ + I)K + (I + CJ) = 0. \quad (6.7.366)$$

Then (7.363) is automatically satisfied, and accordingly we should examine the implication for K of the requirement

$$-(-CJ + I)K + (I + CJ) = 0. \quad (6.7.367)$$

To do so, begin by verifying the manipulation

$$CJ = JJ^{-1}CJ = JW \quad (6.7.368)$$

with

$$W = J^{-1}CJ = -JCJ. \quad (6.7.369)$$

By assumption C is symmetric. Verify from (7.369) that therefore W is also symmetric. Consequently the requirement (7.367) can also be written in the form

$$-(I - JW)K + (I + JW) = 0. \quad (6.7.370)$$

with W being symmetric.

Suppose (7.370) can be solved for K . This is possible if the *invertibility* condition

$$\det(I - JW) \neq 0 \quad (6.7.371)$$

holds, and doing so gives the result

$$K = (I - JW)^{-1}(I + JW). \quad (6.7.372)$$

Observe that (7.372) is a Cayley representation, and therefore K is symplectic. See (3.12.5). Also, by assumption, L is symplectic, and therefore $-L$ is symplectic. It follows from the group property of symplectic matrices that M' , given by the product (7.364), is symplectic. We have achieved our goal in the generic subcase (7.371). We have found, for the invertible subcase, a symplectic matrix M' such that the incompatibility condition (7.359) holds.

To complete our discussion, we must also explore what can be said when the generic condition (7.371) does not hold and instead there is the *singular/noninvertibility* condition

$$\det(I - JW) = 0. \quad (6.7.373)$$

This appears to be a more difficult subcase. But, considerable progress can be made using group theory.

First verify that there is the logical implication

$$\det(I - JW) = 0 \Leftrightarrow \det(I + JW) = 0 \quad (6.7.374)$$

To see this, check the equality chain

$$\begin{aligned} \det(I - JW) &= \det[(I - JW)^T] = \det(I + WJ) \\ &= \det[J(I + WJ)J^{-1}] = \det(I + JW). \end{aligned} \quad (6.7.375)$$

Next observe that JW is a Hamiltonian matrix. See Sections 3.7.2 and 3.7.3. Let H be any Hamiltonian matrix and N be any symplectic matrix. Verify that \bar{H} defined by

$$\bar{H} = NHN^{-1} \quad (6.7.376)$$

is also a Hamiltonian matrix. To do this, set up and employ some Lie-algebraic machinery. Let A and B be any two matrices of the same dimension. Associated with A introduce an operator $\#A\#$ that maps matrices to matrices by the rule

$$\#A\#B = \{A, B\}. \quad (6.7.377)$$

Note that $\#A\#$ is essentially the *adjoint* operator associated with A . See the discussion in Sections 3.7.7, 5.3, and 8.1 where something similar is described. Next, it can be verified that

$$\begin{aligned} \exp(A)B\exp(-A) &= \exp(\#A\#)B = B + \#A\#B + (1/2!)(\#A\#)^2B + \cdots \\ &= B + \{A, B\} + (1/2!)\{A, \{A, B\}\} + (1/3!)\{A, \{A, \{A, B\}\}\} + \cdots. \end{aligned} \quad (6.7.378)$$

See the discussion in Section 8.1 where again something similar is described. Now write N in the factored form

$$N = \exp(JS^a)\exp(JS^c). \quad (6.7.379)$$

See (3.8.26). Show, using (7.378) and (7.379), that there is the result

$$\begin{aligned} \bar{H} &= NHN^{-1} = \exp(JS^a)\exp(JS^c)H\exp(-JS^c)\exp(-JS^a) \\ &= \exp(JS^a)[\exp(\#JS^c\#)H]\exp(-JS^a) = \exp(\#JS^a\#)[\exp(\#JS^c\#)H]. \end{aligned} \quad (6.7.380)$$

Observe that JS^c is a Hamiltonian matrix, and recall that Hamiltonian matrices form the Lie algebra $sp(2n, \mathbb{R})$. It follows from (7.378) that $\exp(\#JS^c\#)H$ is also a Hamiltonian matrix. And, again by analogous reasoning, it follows that $\exp(\#JS^a\#)[\exp(\#JS^c\#)H]$ is also a Hamiltonian matrix, thereby verifying (7.376).

Show, as a consequence of (7.376), that there are the results

$$N(I \pm JW)N^{-1} = (I \pm J\hat{W}) \quad (6.7.381)$$

where \hat{W} is symmetric iff W is symmetric. Indeed, if we write

$$NJWN^{-1} = J\hat{W}, \quad (6.7.382)$$

verify that

$$\hat{W} = -JNJWN^{-1} = (N^T)^{-1}WN^{-1} = (N^{-1})^TWN^{-1}. \quad (6.7.383)$$

To do so, verify that the symplectic condition

$$N^T J N = J \quad (6.7.384)$$

can be rewritten in the form

$$JNJ = -(N^T)^{-1}. \quad (6.7.385)$$

If two symmetric matrices W and \hat{W} are connected by a relation of the form

$$\hat{W} = (N^{-1})^TWN^{-1} \quad (6.7.386)$$

where N is symplectic, then we write

$$\hat{W} \sim W. \quad (6.7.387)$$

Verify that \sim is an equivalence relation. See Exercise (5.12.7). Finally, suppose further that the noninvertibility condition (7.373) holds. Show that there are the results

$$\det(I \pm J\hat{W}) = \det[N(I \pm JW)N^{-1}] = \det(I \pm JW) = 0. \quad (6.7.388)$$

Thus, the “sandwiching” operation described by (7.381) preserves noninvertibility.

The stage is now set to study the incompatibility requirement (7.363) in the noninvertible subcase. Show, using (7.364), (7.365), and (7.368), that (7.363) is equivalent to the requirement

$$\det[-(I - JW)K + (I + JW)] = 0. \quad (6.7.389)$$

Suppose (7.369) holds. Then it is also true that

$$\det\{N[-(I - JW)K + (I + JW)]N^{-1}\} = 0, \quad (6.7.390)$$

and vice versa. Verify that matrix manipulation gives the result

$$\begin{aligned} N[-(I - JW)K + (I + JW)]N^{-1} &= -N(I - JW)N^{-1}NKN^{-1} + N(I + JW)N^{-1} \\ &= -(I - J\hat{W})\check{K} + (I + J\hat{W}) \end{aligned} \quad (6.7.391)$$

where

$$\check{K} = NKN^{-1}. \quad (6.7.392)$$

Note that \check{K} will be symplectic iff K is symplectic. If two symplectic matrices \check{K} and K are connected by a relation of the form (7.392), then we write

$$\check{K} \approx K. \quad (6.7.393)$$

Verify that \approx is also an equivalence relation.

What you have shown is that there is the logical implication

$$\det[-(I - JW)K + (I + JW)] = 0 \Leftrightarrow \det[-(I - J\hat{W})\check{K} + (I + J\hat{W})] = 0. \quad (6.7.394)$$

Therefore, the incompatibility condition is a *class* condition. If it holds for the W, K pair, then it also holds for the \hat{W}, \check{K} pair, and vice versa.

A possible strategy now comes into view. Suppose we partition the set of symmetric matrices W into equivalence classes using the equivalence relation \sim and select a normal form W^{norm} for each equivalence class. Suppose each normal form is sufficiently simple that we can construct a corresponding matrix \bar{K} such that the W^{norm}, \bar{K} pair is incompatible. Then we will have proved, also in the noninvertible subcase, that for every choice of a Darboux matrix β there exists a symplectic matrix M' such that (7.359) holds.

Let us see how this strategy works in the case of a two-dimensional phase space so that J , W , and K are 2×2 matrices. In the two-dimensional case W has the general form

$$W = \begin{pmatrix} c & a \\ a & b \end{pmatrix}. \quad (6.7.395)$$

Verify that in this case JW takes the form

$$JW = \begin{pmatrix} a & b \\ -c & -a \end{pmatrix} \quad (6.7.396)$$

and $I \pm JW$ take the forms

$$I \pm JW = \begin{pmatrix} 1 \pm a & \pm b \\ \mp c & 1 \mp a \end{pmatrix}. \quad (6.7.397)$$

Next show that

$$\det(I \pm JW) = 1 - a^2 + bc. \quad (6.7.398)$$

Also, verify that from (7.382) that $\det(JW)$ is a class function. That is,

$$\det(J\hat{W}) = \det(JW). \quad (6.7.399)$$

For the parameterization (7.395) we find that

$$d \stackrel{\text{def}}{=} \det(JW) = [\det(J)][\det(W)] = \det(W) = -a^2 + bc, \quad (6.7.400)$$

Verify that, for the case of a two-dimensional phase space, there is the relation

$$\det(I \pm JW) = 1 + d. \quad (6.7.401)$$

Finally, from the noninvertibility condition (7.373), verify that we are interested in any equivalence class for which

$$d = -1. \quad (6.7.402)$$

Serendipitously, the normal forms for 2×2 symmetric matrices are given in Section 32.2.2.1 in the context of finding normal forms for second-order polynomials. There δ plays the role of d ,

$$\delta = d, \quad (6.7.403)$$

and we find that a possible normal form is given by

$$W^{\text{norm}} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad (6.7.404)$$

See (32.2.41) and (32.2.45). Correspondingly, verify that there are the results

$$JW^{\text{norm}} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (6.7.405)$$

$$I + JW^{\text{norm}} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}, \quad (6.7.406)$$

$$I - JW^{\text{norm}} = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}. \quad (6.7.407)$$

Let us evaluate the argument of the right side of (7.394) when

$$\hat{W} = W^{\text{norm}} \quad (6.7.408)$$

and we make the inspired (and symplectic) choice

$$\check{K} = \bar{K} = J. \quad (6.7.409)$$

Show that so doing gives the result

$$\begin{aligned} -(I - JW^{\text{norm}})\bar{K} + (I + JW^{\text{norm}}) &= \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix}. \end{aligned} \quad (6.7.410)$$

Evidently the matrix on the far right end of (7.410) has determinant 0. It follows that the W^{norm}, \bar{K} pair given by (7.408) and (7.409) is incompatible. Consequently we have shown, for the case of a two-dimensional phase space, that for every choice of a Darboux matrix β there exists a symplectic matrix M' such that the incompatibility condition (7.359) holds even in the noninvertible subcase.

What can be said about higher dimensional phase-space cases? We have already treated the invertible subcase for any number of dimensions. What remains is to treat the non-invertible subcase for dimensions four and higher. We can try to proceed in analogy to

the two-dimensional phase-space case. In principle normal forms are known for symmetric matrices in any (even) number of dimensions. See Sections 32.2.2.2 and 32.2.2.3. Therefore some of the necessary tools are available to proceed. But we will not pursue the question further in this exercise.

6.7.18. Review Exercise 7.17. The machinery associated with (7.377) through (7.380) and used to prove (7.376) is elegant, but not really necessary. Verify that (7.376) can also be proved directly using (7.382) through (7.385).

6.7.19. Review Exercise 6.7.6. The purpose of this exercise is to show, for an F_2 generating function, that use of the failure condition (7.265) yields the result

$$\det(\partial^2 F_2 / \partial q_k \partial P_\ell) = 0. \quad (6.7.411)$$

Verify that, according to (7.265), (5.13.102), (5.13.103), (6.7.56), and (6.7.57), the relevant matrices in this case are

$$C^{\alpha^{-1}} = -J^{2n}(D^\alpha)^T = \begin{pmatrix} 0 & 0 \\ I^n & 0 \end{pmatrix} \quad (6.7.412)$$

and

$$D^{\alpha^{-1}} = J^{2n}(B^\alpha)^T = \begin{pmatrix} I^n & 0 \\ 0 & 0 \end{pmatrix}. \quad (6.7.413)$$

Write G in the block form

$$G = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \quad (6.7.414)$$

and verify that

$$C^{\alpha^{-1}} G + D^{\alpha^{-1}} = \begin{pmatrix} I^n & 0 \\ A & B \end{pmatrix}, \quad (6.7.415)$$

and therefore in this case (7.265) becomes

$$\det(C^{\alpha^{-1}} G + D^{\alpha^{-1}}) = \det(B) = 0. \quad (6.7.416)$$

Recall that G is the Hessian of g . Verify from the work of Exercise 6.7.6 that

$$B_{k\ell} = \partial^2 F_2 / \partial q_k \partial P_\ell, \quad (6.7.417)$$

and that therefore (7.416) implies (7.411).

6.7.20. Consider the 4×4 symplectic matrix M given by

$$M = \begin{pmatrix} m & 0 & 0 & 0 \\ -1/f & 1/m & 0 & 0 \\ 0 & 0 & m & 0 \\ 0 & 0 & -1/f & 1/m \end{pmatrix}. \quad (6.7.418)$$

With phase-space coordinates listed in the order (q_1, p_1, q_2, p_2) , M is the linear part (Gaussian paraxial approximation) of the geometrical optics transfer map for an imaging system. See Section X.7. What generating function-function types F_1 through F_4 and the Poincaré F_+

are compatible/incompatible with M ? You should find that $F_1(q, Q)$ is incompatible with M , and that $F_2(q, P)$, $F_3(p, Q)$, and $F_4(p, P)$ are compatible. Finally, you should find that F_+ is compatible providing $m \neq -1$. Note that it is quite possible for an imaging system to have $m = -1$.

Let \mathcal{M} be the transfer map for some imaging system. This map sends q and p in the object plane to Q and P in the image plane. Suppose M as given by (7.418) is the linear part of \mathcal{M} . Generally an imaging system consists of a collection of lenses sandwiched between leading and trailing drift spaces. Define the *aperture* plane of the imaging system to be some plane (perpendicular to the z axis) that occurs immediately or shortly after the *last* lens in the system and before the image plane. Let d'_R be the distance (along the z axis) from the aperture plane to the image plane. Also, let Q' and P' be the phase-space variables in the aperture plane. Finally, let \mathcal{M}' be the map between the object and aperture planes. That is, \mathcal{M}' sends q, p to Q', P' . Verify that M' , the linear part of \mathcal{M}' , is given by

$$M' = \begin{pmatrix} m - d'_R/f & d'_R/m & 0 & 0 \\ -1/f & 1/m & 0 & 0 \\ 0 & 0 & m - d'_R/f & d'_R/m \\ 0 & 0 & -1/f & 1/m \end{pmatrix}. \quad (6.7.419)$$

Show that $F_1(q, Q')$ is compatible with M' providing $d'_R > 0$.

6.8 Symplectic Invariants

We have seen that Hamiltonian flows produce symplectic maps. We might hope that, with sufficient ingenuity, we could engineer a Hamiltonian that would produce any desired symplectic map. If so, the remaining fundamental problem is to understand the action of symplectic maps on phase space. Is the action completely general, or are there restrictions? If there are restrictions, what is their nature, and are there any associated invariants? This is a difficult and only partially understood subject. Indeed, it is still a matter of intensive theoretical research in the general setting of symplectic geometry and symplectic topology. It is also of great practical interest. Consider, for example, the field of Accelerator Physics. Suppose some charged-particle source produces a collection of low-energy particles described by some initial distribution function. Imagine these particles are now acted upon over time by some combination of electric and magnetic fields to accelerate them to high energy. What is the final distribution function under the assumption that interactions among the particles, e.g. space-charge effects, are ignored? By a suitable choice of electric and magnetic fields can it be made anything one desires, or are there restrictions? In this section we will partially explore some elementary aspects of this subject.

6.8.1 Liouville's Theorem

Consider some $2n$ -dimensional region R_{2n}^i of phase space that will be referred to as an *initial* region. Suppose some symplectic map \mathcal{M} acts on phase space, and in so doing sends R_{2n}^i to some region R_{2n}^f that will be referred to as a *final* region. Let V^i be the volume of the

initial region,

$$V^i = \int_{R_{2n}^i} dz_1^i \cdots dz_{2n}^i, \quad (6.8.1)$$

and let V^f be the volume of the final region,

$$V^f = \int_{R_{2n}^f} dz_1^f \cdots dz_{2n}^f. \quad (6.8.2)$$

Here the z^i are coordinates for the initial region R_{2n}^i , and the z^f are coordinates for the final region R_{2n}^f . They are related by the map \mathcal{M}

$$z^f = \mathcal{M}z^i, \quad (6.8.3)$$

and correspondingly their differentials are related by M ,

$$dz^f = M dz^i. \quad (6.8.4)$$

It follows from the standard rules for changing variables of integration that the volume V^f is also given by the relation

$$V^f = \int_{R_{2n}^f} dz_1^f \cdots dz_{2n}^f = \int_{R_{2n}^i} |\det(M)| dz_1^i \cdots dz_{2n}^i = \int_{R_{2n}^i} dz_1^i \cdots dz_{2n}^i = V^i. \quad (6.8.5)$$

Here we have used the fact that since \mathcal{M} is a symplectic map, M must be a symplectic matrix and therefore must have determinant +1,

$$\det(M) = 1. \quad (6.8.6)$$

We see that the two volumes V^f and V^i are the *same*. Symplectic maps preserve volume in phase space. This result is called *Liouville's theorem*. Note that in the 2-dimensional case, “volume” is simply area. Therefore, in two dimensions, area (and orientation) preserving maps are symplectic maps, and vice versa.

There is a slightly different phrasing of Liouville's theorem that is also worth mentioning. Consider an ensemble of *noninteracting* systems with each member of the ensemble governed by the same Hamiltonian $H(z, t)$. At some initial instant t^i , let each member of the ensemble be characterized by a point in phase space corresponding to its initial conditions. Suppose, further, that all the points of the ensemble at the initial instant t^i occupy a certain region R_{2n}^i of phase space. Now follow all the trajectories of the members of the ensemble through augmented phase space to some later instant t^f . The members of the ensemble will then occupy some final region R_{2n}^f of phase space. Since the map produced by this Hamiltonian flow is symplectic, we have the relation (8.5). The volume in phase space occupied by the ensemble remains constant. Also, by construction, the number of ensemble points in V^f and V^i is the same. Therefore, since V^f equals V^i , one may also say that the *density* of points in phase space (the number of points divided by the volume they occupy) is preserved by Hamiltonian flows. The collection of ensemble points moves about in phase space (and augmented phase space) like an *incompressible* fluid. In the context of Accelerator Physics, this result means that the density of (assumed noninteracting) beam particles in phase space after acceleration can never exceed (and, in fact, must equal) their initial phase-space density at the source provided their motions are all governed by the same Hamiltonian. That is, particles cannot be concentrated in phase space by solely Hamiltonian means.

6.8.2 Gromov's Nonsqueezing Theorem and the Symplectic Camel

According to Liouville's theorem, if an initial region R_{2n}^i of phase space is sent into a final region R_{2n}^f of phase space under the action of a symplectic map \mathcal{M} , then these two regions must have the same volume. One might wonder about the converse: Given two regions of phase space having the same volume, is there a symplectic map that sends one into the other? The answer is *yes* in the case of two-dimensional phase space ($n = 1$), and, as will be done in Chapter 33, it is fairly easy to show that the answer is *no* in the case of four or more phase-space dimensions ($n > 1$) if one is restricted to *linear* symplectic maps. But what about the far more complicated case where nonlinear symplectic maps are allowed? The answer to this question was unknown until 1985 when *Gromov* announced his famous *nonsqueezing theorem* and its application to the *symplectic camel*.²⁵ The proof of his theorem is beyond the scope of this text and is part of the deep new field of *symplectic topology*. However, it is easy to state and understand its contents.

In the spirit of the theoretical physicist who instructed the farmer to first consider a spherical cow, the mathematician Gromov considered a spherical region in phase space, the symplectic ball $B^{2n}(r)$ of radius r given by the relation

$$B^{2n}(r) = \{z \in R^{2n} \mid \sum_{j=1}^n (p_j^2 + q_j^2) \leq r^2\}. \quad (6.8.7)$$

(This ball is called *symplectic* because its definition involves the p_j as well as the q_j .) It is an easy calculation to show that $B^{2n}(r)$ has a finite volume $V(r)$ given by the relation

$$V(r) = r^{2n} \pi^n / [n \Gamma(n)] = r^{2n} \pi^n / n!. \quad (6.8.8)$$

Gromov also considered a symplectic cylinder $C_1^{2n}(r')$ of radius r' given by the relation

$$C_1^{2n}(r') = B_1^2(r') \times R^{2n-2}. \quad (6.8.9)$$

Here $B_1^2(r')$ is the set

$$(p_1^2 + q_1^2) \leq (r')^2 \quad (6.8.10)$$

and, according to (8.9), the remaining variables q_j and p_j for $j > 1$ are allowed to range from minus to plus infinity,

$$q_j \in (-\infty, +\infty), p_j \in (-\infty, +\infty) \text{ for } j \in [2, n]. \quad (6.8.11)$$

Evidently, because of (8.11), $C_1^{2n}(r')$ has infinite volume. We now ask if there is a symplectic map \mathcal{M} , possibly nonlinear, such that when \mathcal{M} is applied to $B^{2n}(r)$ the resulting region lies within (is *embedded* in) $C_1^{2n}(r')$,

$$\mathcal{M}B^{2n}(r) \subset C_1^{2n}(r')? \quad (6.8.12)$$

²⁵"It is easier for a camel to go through the eye of a needle than for a rich man to enter into the kingdom of God", a saying of Jesus as quoted in Matthew 19, Mark 10, and Luke 18. The term *symplectic camel* and the allusion to the Biblical texts is due to

Put another way, if we regard $B^{2n}(r)$ as a “symplectic camel”, can this camel be squeezed into the cylinder $C_1^{2n}(r')$ under the action of some symplectic map? Liouville would not object because the volume of the cylinder, being infinite, would certainly exceed the volume of the camel. However, Gromov showed that there was no symplectic \mathcal{M} that would map the camel to a region lying within the cylinder unless the radius of the cylinder equaled or exceeded that of the camel (ball),

$$r' \geq r. \quad (6.8.13)$$

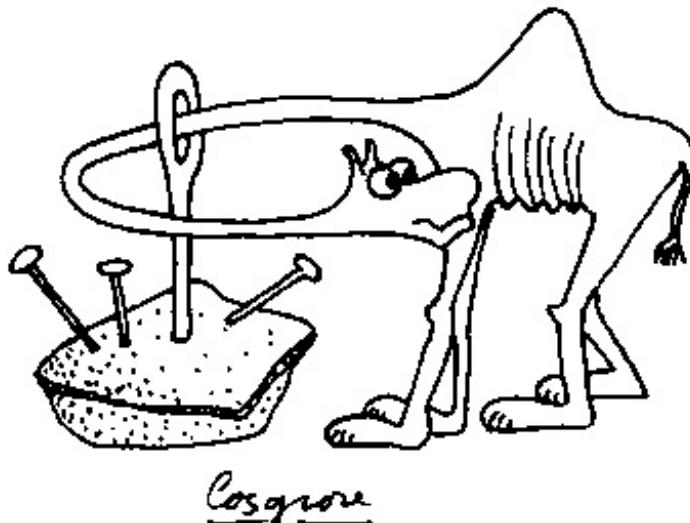


Figure 6.8.1: An *ordinary* camel and needle in R^3 .

A related question, more akin to passing a camel through the eye of a needle, is this: Suppose there is a camel on one side of a wall, and this wall has a hole in it. Is there a continuous family of symplectic maps $\mathcal{M}(\tau)$ such that $\mathcal{M}(0)$ is the identity map \mathcal{I} and the map $\mathcal{M}(1)$ has the property that when it acts on the camel the result is a camel on the other side of the wall? Moreover, is it the case that all points obtained by letting $\mathcal{M}(\tau)$ act on the camel (for $\tau \in [0, 1]$) lie either outside the wall or within the hole in the wall?

Because we are working in dimension four or higher where our intuition may easily fail, let us phrase the question more precisely in mathematical terms. We define the wall W to be the hyperplane $q_1 = 0$,

$$W = \{z \in R^{2n} \mid q_1 = 0\}. \quad (6.8.14)$$

We define the hole in the wall, $H(r')$, to be the set

$$H(r') = \{z \in R^{2n} \mid q_1 = 0 \text{ and } \sum_{j=1}^n (p_j^2 + q_j^2) \leq (r')^2\}. \quad (6.8.15)$$

As for the symplectic camel, we define the two sets $B_+^{2n}(r, a)$ and $B_-^{2n}(r, a)$, with $a > r$, by the rules

$$B_+^{2n}(r, a) = \{z \in R^{2n} \mid (q_1 - a)^2 + p_1^2 + \sum_{j=2}^n (p_j^2 + q_j^2) \leq r^2\}, \quad (6.8.16)$$

$$B_+^{2n}(r, a) = \{z \in R^{2n} \mid (q_1 + a)^2 + p_1^2 + \sum_{j=2}^n (p_j^2 + q_j^2) \leq r^2\}. \quad (6.8.17)$$

Evidently $B_+^{2n}(r, a)$ is a camel centered around the point given by $q_1 = a$ with all remaining coordinates being zero, and $B_-^{2n}(r, a)$ is a camel centered around the point given by $q_1 = -a$ with all remaining coordinates being zero. And since we have assumed $r < a$, no part of either camel is in contact with the wall. Therefore $B_+^{2n}(r, a)$ is a camel located on the side of the wall W with $q_1 > 0$, and $B_-^{2n}(r, a)$ is a camel located on the side of the wall W with $q_1 < 0$.

Now suppose the camel is smaller than the hole in the wall, $r < r'$. Then it is easy to see that the camel can be moved through the hole from one side of the wall to the other by a simple translation along the q_1 axis of the form (6.2.9). Employing notation to be introduced in Section 7.7, we may then write $\mathcal{M}(\tau)$ in the form

$$\mathcal{M}(\tau) = \exp(2\tau a : p_1 :). \quad (6.8.18)$$

It easily verified that there are the relations

$$\mathcal{M}(0)B_+^{2n}(r, a) = B_+^{2n}(r, a), \quad (6.8.19)$$

$$\mathcal{M}(1)B_+^{2n}(r, a) = B_-^{2n}(r, a). \quad (6.8.20)$$

Moreover all the points given by

$$\mathcal{M}(\tau)B_+^{2n}(r, a) = B_+^{2n}(r, a(1 - 2\tau)) \quad (6.8.21)$$

with $q_1 = 0$ satisfy

$$(a(1 - 2\tau))^2 + p_1^2 + \sum_{j=2}^n (p_j^2 + q_j^2) \leq r^2. \quad (6.8.22)$$

From (8.22) we conclude that either $q_1 \neq 0$ or

$$q_1 = 0 \text{ and } \sum_{j=1}^n (p_j^2 + q_j^2) \leq r^2 - (a(1 - 2\tau))^2 \leq (r')^2, \quad (6.8.23)$$

and therefore all points of the camel are either off the wall ($q_1 \neq 0$) or are within the hole $H(r')$ as the camel passes through the wall under the action of $\mathcal{M}(\tau)$. We have moved the camel from the side with $q_1 > 0$ to the side with $q_1 < 0$.

What happens in the more interesting case where the camel is larger than the hole, $r > r'$? In that case Gromov has shown that there is *no* continuous family of symplectic maps $\mathcal{M}(\tau)$ satisfying (8.19) and (8.20) without some points of $\mathcal{M}(\tau)B_+^{2n}(r, a)$ lying in the wall and outside the hole for some intermediate τ values. Thus for a symplectic camel to pass through the eye of a needle under the action of a continuous family of symplectic maps, the eye of the needle must be larger than the camel. By contrast, if one is allowed to use maps that are simply volume preserving but not symplectic, it is easy to see that one can pass the camel through the eye of any needle no matter how large the camel is or how small the eye of the needle is. For example, one may first stretch and thin the camel by

pulling along her tail in the $+q_1$ direction while holding her nose fixed. (We assume the camel is eyeing the eye with some trepidation, and we plan to pass her through head first.) While increasing her length in the q_1 direction, we appropriately compress her in all other directions so that her volume remains unchanged. Then this thinned camel may be safely passed through the eye of the needle. Finally, the camel can be brought back to her original shape by holding her hind quarters fixed, pushing on her nose thereby compressing her q_1 dimension, and letting her other dimensions expand to their original values.

The discussion so far has been concerned with the ‘spherical’ camel $B^{2n}(r)$ given by (8.7). It can be extended to the case of a *general elliptic* camel $E^{2n}(r)$. By a general elliptic camel we mean the set defined by the rule

$$E^{2n}(r) = \{z \in R^{2n} \mid (z, Sz) \leq r^2\} \quad (6.8.24)$$

where S is a positive-definite matrix. Suppose we make the symplectic change of variables

$$z = AZ \text{ or } Z = A^{-1}z \quad (6.8.25)$$

where A is a symplectic matrix. Then there is the relation

$$(z, Sz) = (AZ, SAZ) = (Z, A^T SAZ). \quad (6.8.26)$$

As will be seen in Chapter 33, if S is positive definite, there is always a symplectic A such that

$$A^T SA = S_\lambda \quad (6.8.27)$$

where S_λ is a diagonal matrix with pair-wise degenerate positive entries. (S_λ is called the Williamson diagonal or normal form of S .) In the 4×4 case, for example, S_λ has the form

$$S_\lambda = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 \\ 0 & 0 & 0 & \lambda_2 \end{pmatrix}. \quad (6.8.28)$$

Correspondingly, and in the case of general dimension, there is the relation

$$(z, Sz) = (Z, S_\lambda Z) = \sum_{j=1}^n \lambda_j (P_j^2 + Q_j^2). \quad (6.8.29)$$

Here, without loss of generality, we may select A such the λ_j are ordered in the fashion

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n > 0. \quad (6.8.30)$$

Motivated by this result, we will define a *normal-form elliptic* camel $E_{\text{nf}}^{2n}(r, \lambda)$ by the relation

$$E_{\text{nf}}^{2n}(r, \lambda) = \{z \in R^{2n} \mid \sum_{j=1}^n \lambda_j (p_j^2 + q_j^2) \leq r^2\}. \quad (6.8.31)$$

We conclude that the general elliptic camel can be transformed into a normal-form elliptic camel by a linear symplectic map. But now there is a generalization of the nonsqueezing

result for the spherical camel to the case of a normal-form elliptic camel. It states that a normal-form elliptic camel cannot be imbedded in the cylinder $C_1^{2n}(r')$ by a symplectic map unless

$$r' \geq r/\sqrt{\lambda_1}. \quad (6.8.32)$$

Similarly, the normal-form elliptic camel cannot be passed through the hole $H(r')$ by a family of symplectic maps unless (8.32) holds.

Evidently the nonsqueezing theorem and the symplectic camel results, which are examples of the general subject of symplectic *capacities*, have important applications to Accelerator Physics. The nonsqueezing theorem has implications for the feasibility of *emittance trading* (the hope that one might be able to concentrate particles in some phase-space plane at the expense of possible dilution in other planes), and the symplectic camel results bear on problems of linear and nonlinear beam transport. Of course one would like to have analogous results for camels and needle eyes with more general shapes than the simple spherical and elliptical and cylindrical shapes assumed in this section. Also, even if all of a camel cannot be squeezed into, say, some cylinder or some other volume, what fraction of the camel can be so squeezed, and how? Some important results have been found in these directions. See the references to Symplectic Geometry and Topology given at the end of this chapter.²⁶ The study of such matters is still in its infancy. And even when such results have been obtained and should possibly useful nonlinear symplectic maps be found, there will still be the problem of designing beamline elements and sequences of beamline elements to realize the desired symplectic maps.²⁷ Clearly, in this area as in so many others, there is still much to be learned about the effects of nonlinear maps and how to achieve, exploit, or mitigate them.

Finally, we close this section with a related consideration. Suppose $f(z)$ and $g(z)$ are two phase-space distributions. It would be nice to know whether or not two different phase-space distributions could be sent into each other by a symplectic map. For example, as already asked earlier, given the phase-space distribution coming out of some ion source or electron gun, is there any possible collection of beamline elements that would transform this distribution into some desired distribution at the end of some beamline or accelerator complex? Mathematically stated, one would like to decompose phase-space distributions into equivalence classes. This too is a deep question about which little is known.

²⁶There are many surprises. For example, when $r' < r$ so that according to Gromov 100% of the spherical camel cannot be embedded in the cylinder, nevertheless any fraction less than 100% can be embedded. Moreover, the construction of such embeddings is very complicated, and in some cases only an existence proof is available. Apparently there will always be some points whose images are outside the cylinder, and perhaps quite far outside the cylinder, whose measure can be made as small as one might desire.

²⁷In the case of Accelerator Physics the maps will arise from Hamiltonian flows and will be analytic. With heroic effort it could probably be possible to achieve maps of the form

$\mathcal{M} = \exp(: f_2^c :) \exp(: f_2^a :) \exp(: f_3 :) \exp(: f_4 :) \exp(: f_5 :) \exp(: f_6 :) \cdots \exp(: g_1 :)$
with the $g_1, f_2^a, f_2^c, f_3, f_4, f_5, f_6$ being any desired homogeneous polynomials and the f_m with $m > 6$ being small. (See Section 7.7 for the meaning of this notation.) In these considerations questions of differentiability might be important and the distinction between $ISpM(2n, \mathbb{R})$ [= $Symp(n)$] and $Ham(n)$ might be relevant.

6.8.3 Poincaré Integral Invariants

The volume invariant of Liouville's theorem is actually the last in a hierarchy of invariants called the *Poincaré integral invariants*. The first invariant in the series consists of a certain 2-dimensional integral over a 2-dimensional submanifold in phase space. The next consists of a 4-dimensional integral over a 4-dimensional submanifold, etc. The last consists of a $2n$ dimensional integral, which is just the volume of Liouville's theorem.

A complete and proper discussion of all the Poincaré invariants requires the use of the *exterior calculus of differential forms*. However, the first in the series of invariants is easily discussed using ordinary calculus and the fundamental symplectic 2-form $(\delta z, Jdz)$ introduced earlier, and we will do so shortly. At this point it is worth noting that, in constructing the general higher-order Poincaré invariants, the exterior calculus of differential forms is used to fabricate general $2m$ -forms for $m = 2, 3, \dots, n$ using as the *only* building block the fundamental symplectic 2-form. The invariance of all these forms, including the last of the hierarchy ($m = n$), which is simply the volume element, follows from the invariance (1.19) of the fundamental symplectic 2-form. This invariance is in turn equivalent to the symplectic condition (1.12). Thus, the symplectic condition is really the fundamental condition from which everything else follows. To the author's knowledge, the utility of the $2m$ -forms for the intermediate m values $2 \leq m < n$ is an open question. Finally it is worth remarking that it is the symplectic structure at the classical level of mechanics that makes possible the uncertainty principle at the quantum level.

Let R_2^i be some *initial* 2-dimensional submanifold in phase space. To be more precise, we construct it as follows. We imagine a 2-dimensional Euclidean space with coordinates α, β , and consider a domain Γ_2 in this space. See Figure 8.1 below. We map Γ_2 into R_2^i with the aid of $2n$ relations of the form

$$z_a^i = g_a(\alpha, \beta). \quad (6.8.33)$$

That is, the functions $g_1 \cdots g_{2n}$ specify the mapping of Γ_2 into R_2^i .

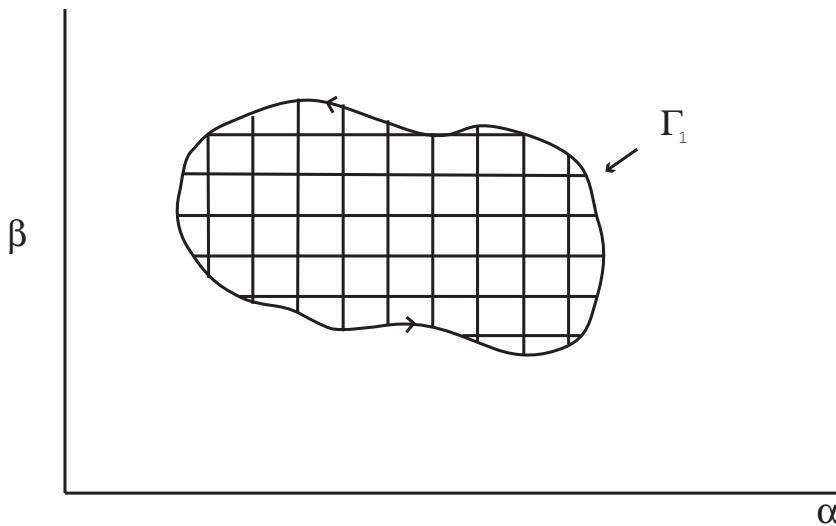


Figure 6.8.2: The domain Γ_2 in α, β space. Also shown is its subdivision into rectangles of sides $d\alpha, d\beta$ and its boundary Γ_1 .

Next we will define a certain integral I_2^i over R_2^i . Subdivide Γ_2 into N rectangles with each rectangle having sides $d\alpha$ and $d\beta$. This subdivision of Γ_2 will produce a corresponding subdivision of R_2^i into “parallelograms” with sides dz^i and δz^i . Here dz^i is the vector formed using (8.33) when only α is allowed to vary,

$$dz_a^i = (\partial g_a / \partial \alpha) d\alpha \quad (6.8.34)$$

or, in vector notation,

$$dz^i = \partial_\alpha g d\alpha. \quad (6.8.35)$$

Similarly, δz^i is the vector formed when only β is allowed to vary,

$$\delta z^i = \partial_\beta g d\beta. \quad (6.8.36)$$

Now, for each parallelogram in R_2^i , compute the quantity $(\delta z^i, Jdz^i)$. By using the relations (8.35) and (8.36) we find the result

$$(\delta z^i, Jdz^i) = (\partial_\beta g, J\partial_\alpha g) d\alpha d\beta. \quad (6.8.37)$$

The left side of (8.37) is the result of evaluating a 2-form in phase space for a small parallelogram in R_2^i . The right side of (8.37) is a 2-form in α, β space. It is called the *pullback* (into α, β space) of the form in phase space.²⁸ We may summarize the situation as follows: The functions g_a provide a mapping of Γ_2 in α, β space into R_2^i in phase space, with small rectangles in Γ_2 mapped into small parallelograms in R_2^i . On phase space there is a 2-form, namely $(\delta z^i, Jdz^i)$, which induces the 2-form $\{(\partial_\beta g, J\partial_\alpha g) d\alpha d\beta\}$ back in the original α, β space. See Exercise 8.3. We remark that in this terminology the integrand in the right side of (4.47) is the pullback to the τ parameter space of the 1-form (4.41) in z space.

As a last step, form a Riemann sum over all parallelograms in R_2^i and a corresponding Riemann sum over all rectangles in Γ_2 . Upon continually refining the subdivision of Γ_2 by letting N go to infinity, we obtain the integrals and the relation

$$I_2^i = \int_{R_2^i} (\delta z^i, Jdz^i) = \int_{\Gamma_2} (\partial_\beta g, J\partial_\alpha g) d\alpha d\beta. \quad (6.8.38)$$

Put another way, the integral over Γ_2 on the right side of (8.38) is well defined; and, based on (8.37), defines what is meant by the integral I_2^i of the 2-form $(\delta z^i, Jdz^i)$ over R_2^i . [We remark that it can be shown, as desired, that the value of the right side of (8.38) is independent of the choice of parameterization.]

Now suppose some symplectic map \mathcal{M} sends the points in R_2^i to some other 2-dimensional submanifold R_2^f according to the rule (6.3). For this submanifold we can compute the associated integral

$$I_2^f = \int_{R_2^f} (\delta z^f, Jdz^f). \quad (6.8.39)$$

²⁸Why is the 2-form on the right side of (8.37) called a pullback? The relations (8.33) provide a mapping from points in Γ_2 to points in phase space. Suppose we regard this as a mapping in the *forward* direction. It is sometimes described by saying that points in Γ_2 are *pushed forward* into points in phase space. By contrast, the relation (8.37) begins with a 2-form in phase space and yields a 2-form *back* in Γ_2 space. If points may be regarded as being pushed forward, then associated forms may be regarded as being related by pulling back.

With the aid of (8.4) and its counterpart for δz , this integral can be pulled back from the final phase space to the initial phase space, and then pulled back further to α, β space to give the result

$$I_2^f = \int_{R_2^f} (\delta z^f, J dz^f) = \int_{R_2^i} (M \delta z^i, JM dz^i) = \int_{\Gamma_2} (M \partial_\beta g, JM \partial_\alpha g) d\alpha d\beta. \quad (6.8.40)$$

Here we have used the relations (8.4) and (8.34) to write

$$dz_a^f = (M dz^i)_a = \sum_c M_{ac} dz_c^i = \sum_c M_{ac} (\partial g_e / \partial \alpha) d\alpha = (M \partial_\alpha g)_a d\alpha, \quad (6.8.41)$$

and similarly for δz^f . But, from the symplectic condition, we have the result

$$(M \partial_\beta g, JM \partial_\alpha g) = (\partial_\beta g, M^T JM \partial_\alpha g) = (\partial_\beta g, J \partial_\alpha g). \quad (6.8.42)$$

See also (1.19). It follows that

$$I_2^f = \int_{\gamma_2} (M \partial_\beta g, JM \partial_\alpha g) d\alpha d\beta = \int (\partial_\beta g, J \partial_\alpha g) d\alpha d\beta = I_2^i. \quad (6.8.43)$$

Under the action of a symplectic map, the 2-dimensional integral based on the fundamental symplectic 2-form is conserved,

$$I_2^f = I_2^i. \quad (6.8.44)$$

Finally, if we wish, we may associate the points on R_2^i with the members of some ensemble at some initial time t^i . Assuming that the members at the ensemble are governed by some Hamiltonian $H(z, t)$, we may follow, as before, the trajectories of the members of the ensemble through augmented phase space to some later instant t^f when they terminate on R_2^f . Since H generates a symplectic map, we have the relation (8.44). Integrals over sums of projected signed areas are conserved. Recall Exercise 1.2.

6.8.4 Connection between Surface and Line Integrals

There is an intimate connection between the 2-form $(\delta z, J dz)$ and the differential form (1-form)

$$(z, J dz). \quad (6.8.45)$$

Moreover, we will learn that this connection and the relation (8.44) are the inspiration for the differential form (6.3).

Let Γ_1 be the *boundary* of Γ_2 as illustrated in Figure 8.1. View Γ_1 as a closed path in α, β space, and parameterize it using the parameter $\tau \in [0, 1]$ by introducing functions $\alpha(\tau)$ and $\beta(\tau)$. Under the mapping (8.33) there is an associated closed phase-space path in R_2^i , call it R_1^i , given by the relations

$$z_a^i(\tau) = g_a(\alpha(\tau), \beta(\tau)). \quad (6.8.46)$$

By construction, R_1^i is the boundary of R_2^i . Now form the closed phase-space path integral

$$I_1^i = \int_{R_1^i} (z^i, J dz^i). \quad (6.8.47)$$

To be more explicit, take the differential of (8.46) to find the result

$$dz_a^i(\tau) = (\partial g_a / \partial \alpha) d\alpha + (\partial g_a / \partial \beta) d\beta \quad (6.8.48)$$

or, in vector notation,

$$dz^i = \partial_\alpha g d\alpha + \partial_\beta g d\beta. \quad (6.8.49)$$

These results enable us to write the relations

$$(z^i, J dz^i) = (z^i, J \partial_\alpha g) d\alpha + (z^i, J \partial_\beta g) d\beta, \quad (6.8.50)$$

$$\begin{aligned} I_1^i = \int_{R_1^i} (z^i, J dz^i) &= \int_{\Gamma_1} [(z^i, J \partial_\alpha g) d\alpha + (z^i, J \partial_\beta g) d\beta] \\ &= \int_0^1 d\tau [(z^i, J \partial_\alpha g)(d\alpha/d\tau) + (z^i, J \partial_\beta g)(d\beta/d\tau)]. \end{aligned} \quad (6.8.51)$$

Observe that the left side of (8.50) is a differential 1-form in phase space, and the right side is a differential 1-form in α, β space. In analogy to our earlier discussion, the differential form in α, β space is the pullback of the 1-form in phase space. And the integrand on the far right side of (8.51) is a differential 1-form in τ space that is a pullback from α, β space, and the pullback of the pullback from phase space.

Introduce the functions C_α and C_β by the rules

$$C_\alpha(\alpha, \beta) = (z^i, J \partial_\alpha g) = (g, J \partial_\alpha g), \quad (6.8.52)$$

$$C_\beta(\alpha, \beta) = (z^i, J \partial_\beta g) = (g, J \partial_\beta g). \quad (6.8.53)$$

They allow us to write the integral over the closed path Γ_1 in the more compact form

$$I_1^i = \int_{\Gamma_1} [(z^i, J \partial_\alpha g) d\alpha + (z^i, J \partial_\beta g) d\beta] = \int_{\Gamma_1} C_\alpha d\alpha + C_\beta d\beta. \quad (6.8.54)$$

Now apply Green's (or Stokes') theorem to convert the path integral over Γ_1 to a surface integral over Γ_2 . Doing so gives the result

$$I_1^i = \int_{\Gamma_1} C_\alpha d\alpha + C_\beta d\beta = \int_{\Gamma_2} (\partial C_\beta / \partial \alpha - \partial C_\alpha / \partial \beta) d\alpha d\beta. \quad (6.8.55)$$

However, from (8.52) and (8.53) we find the results

$$\partial C_\beta / \partial \alpha = (\partial_\alpha g, J \partial_\beta g) + (g, J \partial_\alpha \partial_\beta g), \quad (6.8.56)$$

$$\partial C_\alpha / \partial \beta = (\partial_\beta g, J \partial_\alpha g) + (g, J \partial_\beta \partial_\alpha g). \quad (6.8.57)$$

It follows from the symmetry of mixed partials and the antisymmetry of J that there is the relation

$$\partial C_\beta / \partial \alpha - \partial C_\alpha / \partial \beta = 2(\partial_\alpha g, J \partial_\beta g). \quad (6.8.58)$$

Consequently, (8.55) can be rewritten in the form

$$I_1^i = 2 \int_{\Gamma_2} (\partial_\alpha g, J\partial_\beta g) d\alpha d\beta. \quad (6.8.59)$$

Finally, upon comparing (8.38) and (8.59), we find the key result

$$I_1^i = -2I_2^i. \quad (6.8.60)$$

Note that this result is completely general in that it holds for any surface R_2^i and its boundary R_1^i . We remark that one of the features of the exterior calculus of differential forms, see the beginning of Subsection 6.8.3, is that it incorporates the Poincaré lemma of Exercise 1.1 and its generalizations in a systematic way so that relations like (8.60) become routinely obvious.

Now suppose, as in Subsection 6.8.2, that the symplectic map \mathcal{M} sends R_2^i to R_2^f according to the rule (8.3). It will then send R_1^i to R_1^f where R_1^f is the boundary of R_2^f . Let I_1^f be the result of integrating the differential (z^f, Jdz^f) over the path R_1^f ,

$$I_1^f = \int_{R_1^f} (z^f, Jdz^f). \quad (6.8.61)$$

Based on the result just found, there is the relation

$$I_1^f = -2I_2^f, \quad (6.8.62)$$

no matter what the nature of \mathcal{M} is save that it be differentiable and invertible. (Given R_2^i and R_1^i , R_2^f and R_1^f must be well defined.) But if \mathcal{M} is symplectic, then (8.44) must hold, and we conclude from (8.60) and (8.62) that there must also be the relation

$$I_1^f = I_1^i. \quad (6.8.63)$$

When written out in full, the relation (8.63) reads

$$\int_{R_1^f} (z^f, Jdz^f) = \int_{R_1^i} (z^i, Jdz^i). \quad (6.8.64)$$

By using (8.4) to change variables, the integral on the left side of (8.64) can be rewritten as

$$\int_{R_1^f} (z^f, Jdz^f) = \int_{R_1^i} (z^f, JMdz^i). \quad (6.8.65)$$

Now combine (8.64) and (8.65) to find the result

$$\int_{R_1^i} [(z^f, JMdz^i) - (z^i, Jdz^i)] = 0. \quad (6.8.66)$$

We know that a necessary and sufficient condition for this result to hold for any closed path R_1^i in phase space is that the differential form

$$(z^f, JMdz^i) - (z^i, Jdz^i) \quad (6.8.67)$$

be exact. But, in slightly different notation (identify z^i with z and z^f with Z), this is the differential form (6.4). And we know that a necessary and sufficient condition for this form to be exact is that \mathcal{M} be a symplectic map. What we have found is that (8.64) or (8.66) holding for all closed paths is a necessary and sufficient condition for \mathcal{M} to be a symplectic map.

We close this subsection with some comments. Recall the relation (6.30). There is also the simple result

$$d\left(\sum_j p_j q_j\right) = \sum_j (p_j dq_j + q_j dp_j). \quad (6.8.68)$$

Combining (6.30) and (8.68) gives the relation

$$\sum_j p_j dq_j = -[(z, Jdz) - d(\sum_j p_j q_j)]/2. \quad (6.8.69)$$

Since by definition the quantity $d(\sum_j p_j q_j)$ is an exact differential, there must be the result

$$\int_{R_1} d\left(\sum_j p_j q_j\right) = 0 \quad (6.8.70)$$

for any closed phase-space path R_1 . Therefore from (8.69) and (8.70) we have the general relation

$$\int_{R_1} \sum_j p_j dq_j = -(1/2) \int_{R_1} (z, Jdz) = -(1/2)I_1 \quad (6.8.71)$$

for any closed phase-space path R_1 . It follows that (8.64) can also be written as

$$\int_{R_1^f} \sum_j P_j dQ_j = \int_{R_1^i} \sum_j p_j dq_j. \quad (6.8.72)$$

This relation, although it does not appear to treat the coordinates and momenta on an equal footing, is still true whenever the Q, P and the q, p are related by a symplectic map \mathcal{M} , and frequently occurs in the literature. [An integral quantity of the form appearing on the left (or right) side of (8.72) is sometimes called a *circulation* because, if the q_i are regarded as the coordinates of a “position” vector \mathbf{r} and the p_i are regarded as being proportional to the coordinates of a “velocity” vector \mathbf{v} , the integrand is of the form $\mathbf{v} \cdot d\mathbf{r}$.] Evidently (8.72) holding for all closed paths is also a necessary and sufficient condition for \mathcal{M} to be symplectic. Finally, we note that combining (8.60) and (8.71) gives the relation

$$\int_{R_1} \sum_j p_j dq_j = I_2 = \int_{R_2} (\delta z, Jdz) \quad (6.8.73)$$

for any phase-space surface R_2 whose boundary is the closed phase-space path R_1 .²⁹

²⁹We remark that sometimes the differential form $\sum_j p_j dq_j$ is called the *Liouville form*. By the same token, the differential form (8.45) could be called the *Poincaré form*. There does not seem to be any name for the differential form $\sum_j q_j dp_j$. Like the Liouville form in (8.72) and the Poincaré form in (8.64), it too is “invariant” under the action of symplectic maps. See Exercise 8.7.

6.8.5 Poincaré-Cartan Integral Invariant

Suppose we are given a family of symplectic maps $\mathcal{N}(t)$. Then we know there is an associated generating Hamiltonian $H(z, t)$. Recall Theorem 4.2. Alternatively, given a Hamiltonian, we know from Theorem 4.1 that it generates a family of symplectic maps. In this context, let C^i be a closed path in the associated $(2n + 1)$ dimensional augmented phase space. See Figure 8.2. Specifically, for a parameter $\tau \in [0, 1]$, we describe C^i by $(2n + 1)$ relations of the form

$$z_a^i(\tau) = g_a(\tau), \quad (6.8.74)$$

$$t^i(\tau) = g_{2n+1}(\tau). \quad (6.8.75)$$

Note that different points of C^i may have different values of t .

View each point of C^i as an initial condition. For each point on C^i launch a trajectory governed by the Hamiltonian $H(z, t)$, and follow this trajectory to some final time t^f . Allow this time to vary from trajectory to trajectory by specifying yet one more relation of the form

$$t^f(\tau) = g_{2n+2}(\tau). \quad (6.8.76)$$

So doing produces a set of final conditions that constitutes another closed path C^f in augmented phase space. [Note that all the functions appearing on the right sides of (8.74) through (8.76) are assumed to be periodic in τ with period 1.] Put another way, C^i and C^f are any two augmented phase-space paths that surround a common bundle of phase-space trajectories produced by H .

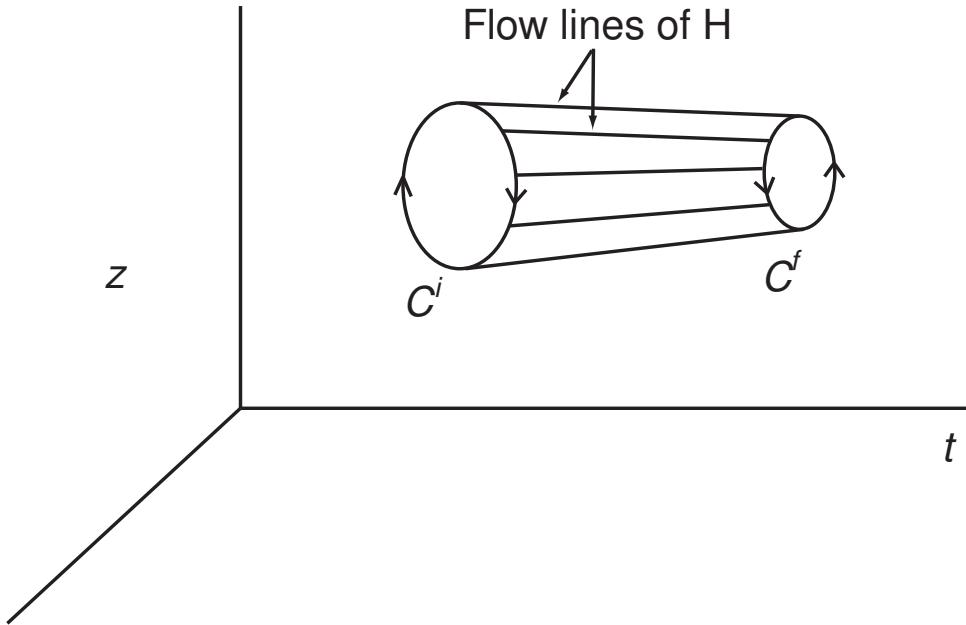


Figure 6.8.3: The closed paths C^i and C^f in augmented phase space and the trajectories that join them.

For augmented phase space consider the differential form

$$\left(\sum_j p_j dq_j \right) - H dt. \quad (6.8.77)$$

Then, according to Poincaré and Cartan, there is the path-integral relation

$$\int_{C^f} [(\sum_j p_j dq_j) - H dt] = \int_{C^i} [(\sum_j p_j dq_j) - H dt]. \quad (6.8.78)$$

Note that in the special case that t is constant on both C^i and C^f , (8.78) reduces to (8.72).

There are several ways to prove the Poincaré-Cartan relation. Our proof will use variational calculus. The trajectories originating on C^i and terminating on C^f form a two-dimensional surface in augmented phase space that is topologically equivalent to a cylinder. Indeed, points on this surface can be viewed as the image of a two-dimensional parameter space region described by τ and t with

$$\tau \in [0, 1], \quad (6.8.79)$$

$$t \in [g_{2n+1}(\tau), g_{2n+2}(\tau)], \quad (6.8.80)$$

and the understanding that the lines $\tau = 0$ and $\tau = 1$ are to be identified. Introduce for the integral on the right side of (8.78) the short-hand notation

$$\int_{C^i} **, \quad (6.8.81)$$

and similarly for the integral on the left side. Also, let

$$\int_{-C^f} ** \quad (6.8.82)$$

denote the integral on the left side of (8.78) with the path traversed in the opposite sense. With these understandings, (8.78) can be rewritten in the form

$$\int_{C^i} ** + \int_{-C^f} ** = 0. \quad (6.8.83)$$

The paths C^i and $-C^f$ are the images of the left and right boundaries of the parameter-space region. See Figure 8.3.

Divide the τ interval (8.79) into N equal pieces of size $\epsilon = 1/N$. For each subdivision consider pairs of parameter-space paths of constant τ traversed in opposite directions. See Figure 8.3. By construction, their images in augmented phase-space are trajectories for the Hamiltonian H traversed forward and backward in time. Imagine integrating the differential form (8.77) over these pairs of augmented phased-space trajectories. So doing will give a null net result because, by construction, the integrals so produced cancel in pairs. Add these self-canceling path integrals to those occurring in (8.83). Evidently the sum of integrals thus obtained can be reorganized into a sum of integrations over N thin loops ℓ_j ,

$$\int_{C^i} ** + \int_{-C^f} ** + \text{canceling integral pairs} = \sum_{j=1}^N \int_{\ell_j} **. \quad (6.8.84)$$

See Figures 8.4 and 8.5.

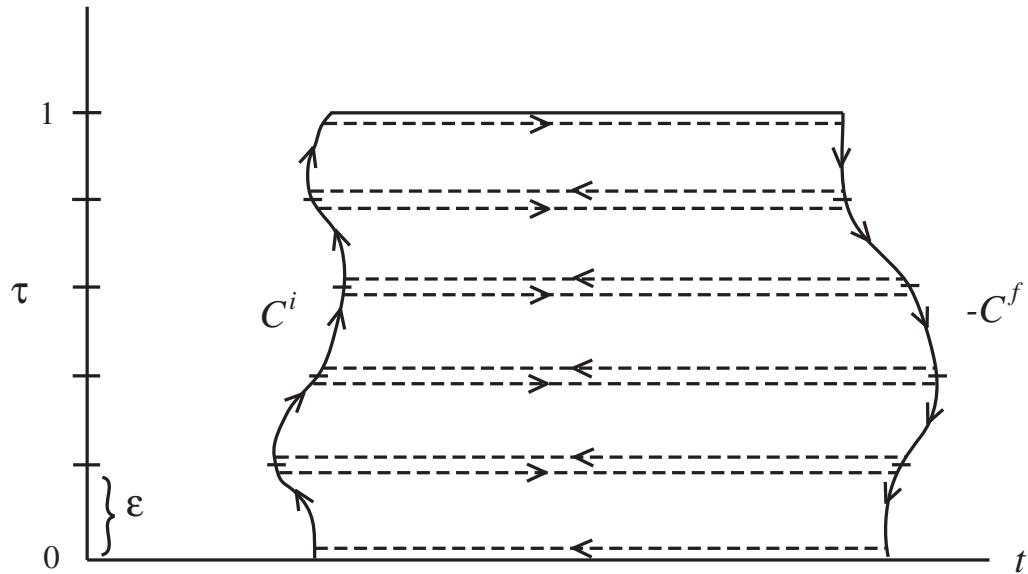


Figure 6.8.4: The t, τ parameter space. The left and right boundaries are the curves $t^i(\tau)$ and $t^f(\tau)$, and their augmented phase-space images are the paths C^i and C^f . Also shown as dashed lines are pairs of parameter-space paths traversed in opposite directions whose images are augmented phase-space trajectories traversed in opposite directions. Note that the lines $\tau = 0$ and $\tau = 1$ have the same image in augmented phase space.

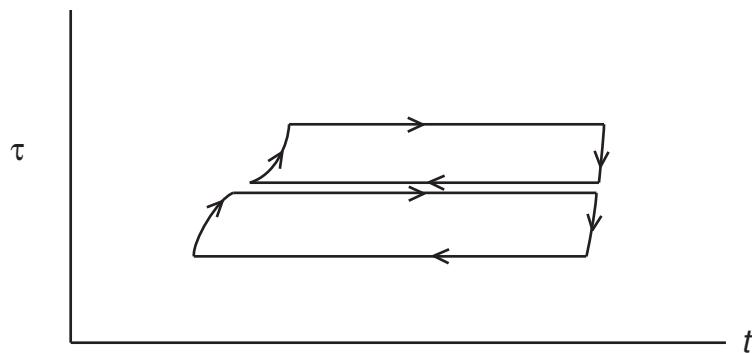


Figure 6.8.5: Two adjacent loops in parameter space.

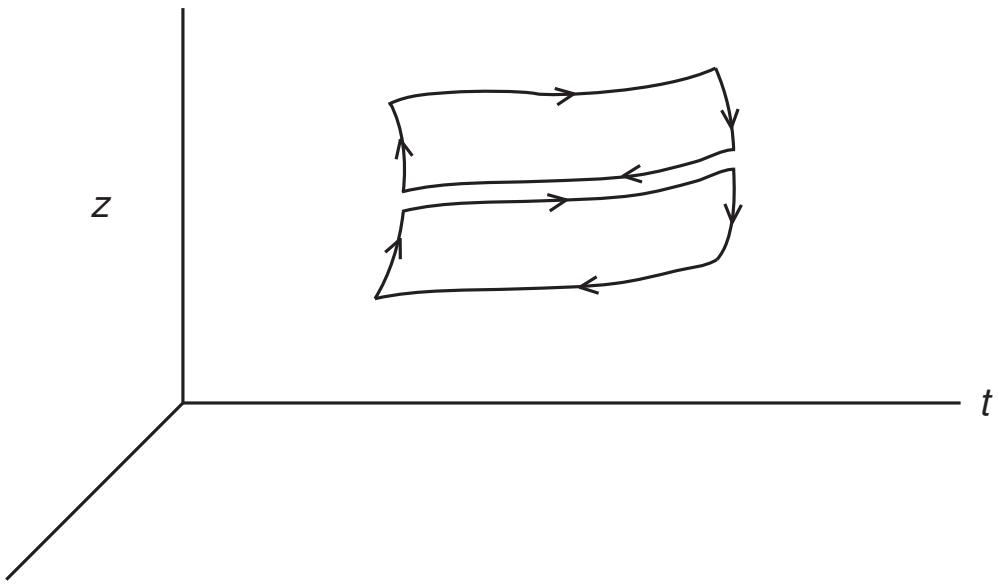


Figure 6.8.6: The loops in augmented phase space corresponding to the two parameter-space loops of Figure 8.4. Note that the long sides of the loops are trajectories for the Hamiltonian H , and the short sides are pieces of C^i and C^f .

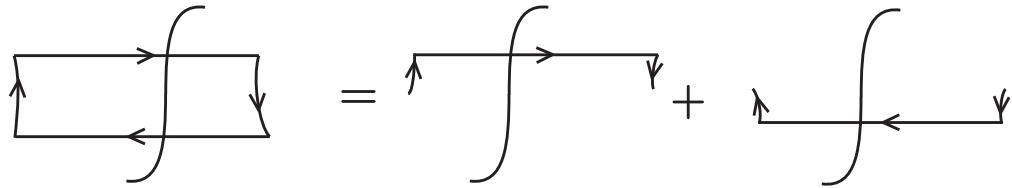


Figure 6.8.7: The integral over a loop is the sum of integrals over top and bottom halves.

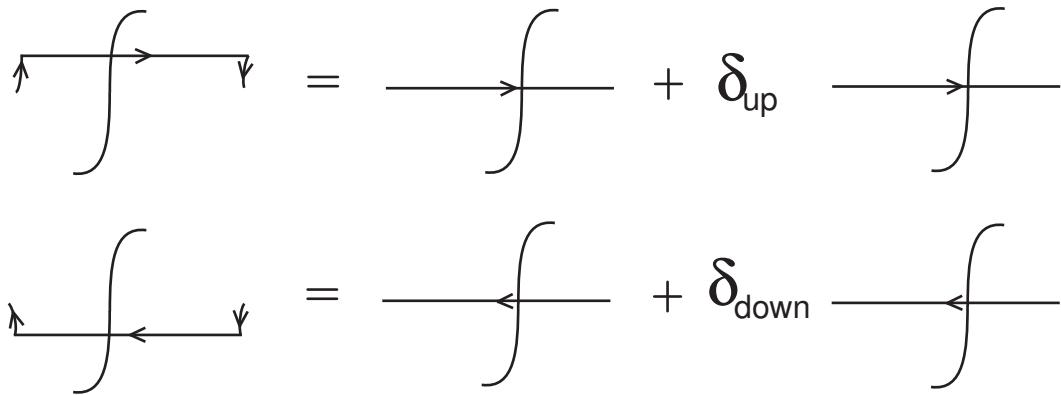


Figure 6.8.8: The integral over a half loop is the integral over a trajectory of H or its reverse plus the change in the integral resulting from deforming this path.

Now consider an individual loop. It can be viewed as a sum of top and bottom halves. See Figure 8.6. Each half can in turn be viewed as the result of deforming (in parameter space) a line of constant τ . See Figure 8.7. Note that the image of a line of constant τ in parameter space is a trajectory for the Hamiltonian H in augmented phase space.

Observe that by definition the sum of a path integral over a trajectory of H and its reverse, see Figure 8.7, cancel. It follows that the integral over any loop ℓ_j is the sum of the “up” and “down” variations about a trajectory of H . See Figures 8.6 and 8.7.

$$\int_{\ell_j} ** = \delta_{\text{up}} \int ** + \delta_{\text{down}} \int **. \quad (6.8.85)$$

But, from Hamilton’s (modified) principle, we know that the functional formed by integrating (8.77) over paths in augmented phase space has an extremum on paths that are trajectories of H . See Exercise 8.8 for details. Therefore each term on the right side of (8.85) vanishes through terms of order ϵ , and we have the result

$$\int_{\ell_j} ** = 0 + O(\epsilon^2). \quad (6.8.86)$$

Insert this result into (8.84) to find the relation

$$\int_{C^i} ** + \int_{-C^f} ** = 0 + O(N\epsilon^2). \quad (6.8.87)$$

Now let $N \rightarrow \infty$ and, correspondingly, $\epsilon \rightarrow 0$. Then $N\epsilon^2 \rightarrow 0$, and (8.87) becomes the desired relation (8.83) or (8.78).

Exercises

6.8.1. Suppose a “burst” of protons is injected into a uniform electric field $\mathbf{E} = E_0 \mathbf{e}_z$. Assume the burst is initially concentrated at x and $y = 0$ and v_x and $v_y = 0$, but is uniformly spread in z and v_z about the values $z = 0$ and $v_z = v_z^0$ within intervals $\pm\Delta z$ and $\pm\Delta v_z$. Thus the problem is essentially that of one-dimensional motion along the z axis. The initial distribution is shown schematically in Figure 8.8. Find the distribution at later times, and verify Liouville’s theorem. Do not assume that Δz and Δv_z are infinitesimal. Neglect Coulomb interactions between particles.

6.8.2. Problem about Liouville’s theorem and divergence theorem and how density transforms.

6.8.3. Exercise showing that, in the case of electromagnetic fields, Liouville’s theorem also holds in terms of spatial coordinates and mechanical momenta.

6.8.4. Verify (8.8).

6.8.5. Construct a nonsymplectic but volume preserving family of maps $\mathcal{N}(\tau)$ that will send any symplectic camel through the eye of any needle.

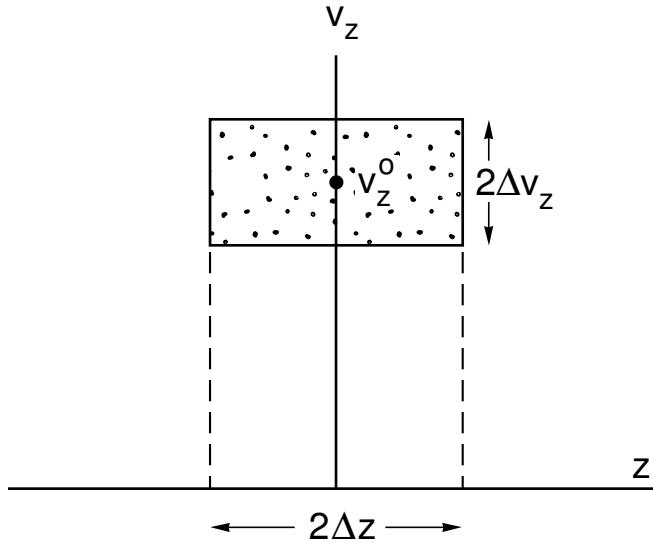


Figure 6.8.9: Initial phase-space distribution for Exercise 8.1.

6.8.6. Show that for elliptic camels there is the result

$$\text{Volume of } E^{2n}(r) = \text{Volume of } E_{\text{nf}}^{2n}(r, \lambda) = r^{2n} \pi^n / [(n!)(\lambda_1 \lambda_2 \lambda_3 \cdots \lambda_n)]. \quad (6.8.88)$$

Show the impossibility of sending a symplectic cigar into a symplectic ball of the same volume using a symplectic map.

6.8.7. Show that (8.72) can also be written as

$$\int_{R_1^f} \sum_j Q_j dP_j = \int_{R_1^i} \sum_j q_j dp_j. \quad (6.8.89)$$

Consider the differential form

$$- [(z, Jdz) - d(\lambda \sum_j p_j q_j)]/2 \quad (6.8.90)$$

where λ is a parameter. Evaluate this form for the cases $\lambda = -1, 0, 1$. Show that it is invariant for all λ .

6.8.8. Refer to Exercise 6.2. Show that the Poincaré-Cartan relation (8.78) can also be written in the more democratic form

$$\int_{C^f} [(z, Jdz)/2 + H(z, t)dt] = \int_{C^i} [(z, Jdz)/2 + H(z, t)dt]. \quad (6.8.91)$$

6.8.9. The observant reader may object that, in deriving the Poincaré-Cartan invariant of Section 6.8.5, we invoked Hamilton's modified principle (1.6.11) and (1.6.12) in an unusual way because we employed paths in augmented phase space along which the time t may possibly both increase and decrease. So, some special explanation is required. To take into

account the possibility of this more general case, suppose the path in (1.6.11) is parameterized by considering the q_i , the p_i , and t itself to be functions of some parameter σ where $\sigma \in [0, 1]$. Introduce the notation

$$q'_i = dq_i/d\sigma, \quad p'_i = dp_i/d\sigma, \quad t' = dt/d\sigma. \quad (6.8.92)$$

Verify that (1.6.11) can be rewritten in the form

$$\mathcal{A} = \int_0^1 d\sigma A \quad (6.8.93)$$

where

$$A(q, q', p, p', t, t') = \sum_i p_i q'_i - Ht'. \quad (6.8.94)$$

Show, employing the usual variational calculus machinery, that the variation in \mathcal{A} for fixed end points q, p, t is given by the relation

$$\begin{aligned} \delta\mathcal{A} &= \int_0^1 d\sigma \left\{ \sum_i [(d/d\tau)(\partial A/\partial q'_i) - \partial A/\partial q_i] \delta q_i + \sum_i [(d/d\tau)(\partial A/\partial p'_i) - \partial A/\partial p_i] \delta p_i \right. \\ &\quad \left. + [(d/d\tau)(\partial A/\partial t') - \partial A/\partial t] \delta t \right\} + O(\epsilon^2). \end{aligned} \quad (6.8.95)$$

Next show that the various partial derivatives in (8.95) are given by the relations

$$\frac{\partial A}{\partial q'_i} = p_i, \quad \frac{\partial A}{\partial q_i} = -t'(\partial H/\partial q_i), \quad (6.8.96)$$

$$\frac{\partial A}{\partial p'_i} = 0, \quad \frac{\partial A}{\partial p_i} = q'_i - t'(\partial H/\partial p_i), \quad (6.8.97)$$

$$\frac{\partial A}{\partial t'} = -H, \quad \frac{\partial A}{\partial t} = -t'(\partial H/\partial t). \quad (6.8.98)$$

From these results, and Hamilton's equations of motion (1.5.11) augmented by (1.5.14), verify the following conclusions about the terms appearing in the integrand of (8.95):

$$[(d/d\tau)(\partial A/\partial q'_i) - \partial A/\partial q_i] = dp_i/d\tau + t'(\partial H/\partial q_i) = p'_i - t'\dot{p}_i = 0, \quad (6.8.99)$$

$$[(d/d\tau)(\partial A/\partial p'_i) - \partial A/\partial p_i] = t'(\partial H/\partial p_i) - q'_i = t'\dot{q}_i - q'_i = 0, \quad (6.8.100)$$

$$[(d/d\tau)(\partial A/\partial t') - \partial A/\partial t] = -dH/d\tau + t'(\partial H/\partial t) = -dH/d\tau + t'(dH/dt) = 0. \quad (6.8.101)$$

We see that in the general case, as claimed, the variation in \mathcal{A} about a trajectory is given by the relation

$$\delta\mathcal{A} = 0 + O(\epsilon^2). \quad (6.8.102)$$

6.9 Poincaré Surface of Section and Poincaré Return Maps

In Section 6.4.1 we saw that Hamiltonian flows between two times t^i and t^f generated symplectic maps. In this section we will study two generalizations of this result. The first, the *Poincaré surface of section map*, is related to the concept of using a coordinate as an

independent variable. For an application of Poincaré surface of section maps see Section 21.7.2.

The second is related to long-term behavior. Recall that Section 1.4.3 illustrated how the determination of the long-term behavior of a periodically driven system could be reduced to the study of the behavior of a certain map, the stroboscopic map, under repeated iteration. For some Hamiltonian problems a similar simplification can be obtained by the use of a *Poincaré return map*. How this can be done will be a second generalization.

Poincaré maps may have other uses as well.

6.9.1 Poincaré Surface of Section Maps

Consider the case of conservative Hamiltonian flows in $2n$ dimensional phase space. That is, we assume $\partial H/\partial t = 0$. In this case we know that H is an integral of motion. Let g and h be two phase-space functions and let S^g and S^h be two $(2n - 2)$ dimensional submanifolds in phase space defined by the equations

$$S^g : H(z) = \mathcal{E} \text{ and } g(z) = 0, \quad (6.9.1)$$

$$S^h : H(z) = \mathcal{E} \text{ and } h(z) = 0. \quad (6.9.2)$$

Note that each of the equations in (9.1) defines a $(2n - 1)$ dimensional submanifold. For their intersection to define a $(2n - 2)$ dimensional submanifold S^g , the gradients $\partial_z H$ and $\partial_z g$ must not be colinear. The analogous condition must also hold for S^h .

Next assume that S^g is *transverse* to the flow generated by H . What does this mean? Suppose z is some point in S^g . Then, we want z to leave S^g under both the forward and backward time evolution generated by H . Under time evolution the change in z is given by

$$dz = (J\partial_z H)dt. \quad (6.9.3)$$

In order for z to leave S^g , the quantity dz must have some component in at least one of the directions $\partial_z H$ and $\partial_z g$. Suppose we require that dz have some component in the direction of $\partial_z g$,

$$(\partial_z g, dz) \neq 0. \quad (6.9.4)$$

In view of (9.3), this requirement is equivalent to the condition

$$(\partial_z g, J\partial_z H) = [g, H] \neq 0. \quad (6.9.5)$$

We also observe that

$$(\partial_z H, dz) = (\partial_z H, J\partial_z H)dt = 0 \quad (6.9.6)$$

due to the antisymmetry of J . Therefore, (9.5) is a necessary and sufficient condition for z to leave S^g . Finally, we note that (9.5) guarantees that $\partial_z H$ and $\partial_z g$ cannot be colinear (proportional). Thus (9.5) is a necessary and sufficient condition both for S^g to be defined by (9.1) and for the flow to cross S^g . The surface S^g is said to be a *surface of section* for the flow generated by H .

Suppose S^h is also a surface of section. Suppose further that for some region R_{2n-2}^g of S^g the points $z \in R_{2n-2}^g$ have the property that their phase-space trajectories generated by

H , when followed sufficiently forward in time, arrive at some region R_{2n-2}^h in S^h . Note that the interval of time required for this to occur may vary from trajectory to trajectory. Also, since H does not depend on the time, without loss of generality we may assume that all trajectories are launched at some common initial time $t = t^i$. See Figure 9.1. Then, by this operation, we have produced a mapping \mathcal{M} , called a *Poincaré surface of section map*, that sends R_{2n-2}^g to R_{2n-2}^h . Moreover, \mathcal{M} is invertible since, given any point in R_{2n-2}^h , we can always follow trajectories backward in time until they reach R_{2n-2}^g .

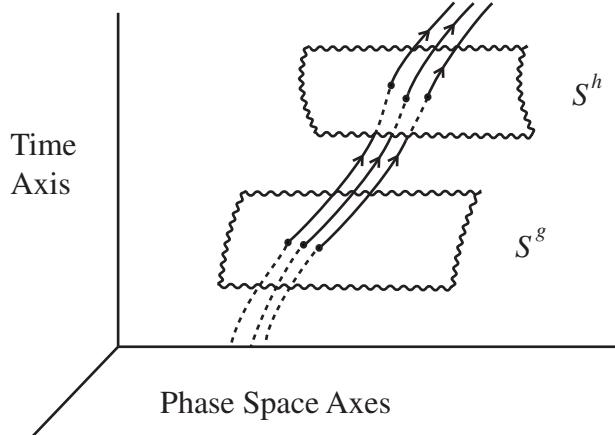


Figure 6.9.1: Two surfaces of section in augmented phase space. Trajectories leaving S^g are assumed to eventually enter and cross S^h , perhaps at different times.

Let R_1^g be any closed path in R_{2n-2}^g , and let R_1^h be its image in R_{2n-2}^h under the action of \mathcal{M} . Then, since R_1^g and R_1^h are related by following trajectories generated by H , the Poincaré-Cartan relation (8.78) takes form

$$\int_{R_1^g} [(\sum_j p_j dq_j) - H dt] = \int_{R_1^h} [(\sum_j p_j dq_j) - H dt]. \quad (6.9.7)$$

Since we assumed all trajectories on R_1^g were launched at $t = t^i$, we have $dt = 0$ for the integral on the left side of (9.7). Therefore, we have the result

$$\int_{R_1^g} [(\sum_j p_j dq_j) - H dt] = \int_{R_1^g} \sum_j p_j dq_j. \quad (6.9.8)$$

Also, since all trajectories lie on the surface $H = \mathcal{E}$, see (9.1) and (9.2), we have the result

$$\int_{R_1^h} (-H) dt = -\mathcal{E} \int_{R_1^h} dt = 0 \quad (6.9.9)$$

because R_1^h is a closed curve.³⁰ Therefore we have the result

$$\int_{R_1^h} [(\sum_j p_j dq_j) - H dt] = \int_{R_1^h} \sum_j p_j dq_j. \quad (6.9.10)$$

³⁰Note that we could have used the same argument to deduce (9.8) without the assumption that all points on S^g are launched with the same times t^i .

It follows, for a Poincaré map, that

$$\int_{R_1^g} \sum_j p_j dq_j = \int_{R_1^h} \sum_j p_j dq_j. \quad (6.9.11)$$

In some cases still more can be said. Suppose it can be arranged, perhaps by a suitable choice of variables, that $g(z)$ and $h(z)$ take the form

$$g(z) = 0 \rightarrow q_1 = \alpha, \quad (6.9.12)$$

$$h(z) = 0 \rightarrow q_1 = \beta, \quad (6.9.13)$$

when α and β are certain constants. Then we have $dq_1 = 0$ on R_1^g and R_1^h , and (9.11) becomes the relation

$$\int_{R_1^g} \sum_{j=2}^n p_j dq_j = \int_{R_1^h} \sum_{j=2}^n p_j dq_j. \quad (6.9.14)$$

Let z be an initial condition in R_{2n-2}^g . We know the value of q_1 from (9.12). Suppose $q_2, p_2 \dots q_n, p_n$ are selected to lie in R_{2n-2}^g . Then p_1 can be determined, perhaps up to a sign, from the condition $H = \mathcal{E}$. The sign ambiguity can be resolved by requiring that the trajectory launched from R_{2n-2}^g reach R_{2n-2}^h when traced forward in time. Thus, we may assume that points in R_{2n-2}^g (and R_{2n-2}^h) are described by the $(2n - 2)$ coordinates $q_2, p_2 \dots q_n, p_n$; and the Poincaré map \mathcal{M} acts on this $(2n - 2)$ dimensional space. Finally, from the results of Section 6.8.4, the relation (9.14) implies that the Poincaré map \mathcal{M} is a symplectic map on this $(2n - 2)$ dimensional space.

6.9.2 Poincaré Return Maps

Many Hamiltonian flows of physical interest have the property that they repeatedly re-enter some region of phase space. For example, in a *Penning* trap or a mirror machine, particles repeatedly return to some midplane region. In a circular accelerator or storage ring, particles repeatedly pass through any given beam-line element. In celestial and galactic dynamics, trajectories sufficiently close to a periodic trajectory nearly repeat themselves.

For such systems there are surfaces of section that are crossed repeatedly by a bundle of trajectories, and such a surface can be used to define a *Poincaré return* map. Let S^g be such a surface of section, and let R_{2n-2}^g be some region in S^g . For any point $z \in R_{2n-2}^g$ suppose the trajectory launched with these initial conditions returns to S^g . Then, by following these trajectories, we obtain a mapping of S^g onto itself,

$$\mathcal{M} : S^g \rightarrow S^g. \quad (6.9.15)$$

Moreover, like the case of a stroboscopic map, the long-term behavior of such a system can be found by studying the repeated action of \mathcal{M} . See (1.4.34). Finally, if coordinates can be selected so that (9.12) holds, the map \mathcal{M} is symplectic.

Consider, for example, a circular accelerator or storage ring as shown schematically in Figure 1.2.9. At the point O we may introduce Cartesian coordinates as in Figure 1.6.1 so that particle trajectories repeatedly cross the plane $z = 0$ as they go around the ring. For

the Hamiltonian we will use H^{eff} as given by (1.6.34). In particular near O we will employ the conjugate coordinate pairs (z, p_z) , (x, p_x) , (y, p_y) , and (t, p_t) , and the independent time-like variable τ . By construction H^{eff} is conserved, and we may restrict our attention to trajectories for which $H^{\text{eff}} = 0$, in which case (1.6.5) holds. Given the values of (z, p_z) , (x, p_x) , (y, p_y) , and (t, p_t) in the plane $z = 0$, we can find p_z as in (1.6.6). Starting with these initial conditions, we follow a trajectory until it again crosses $z = 0$. In this way we find a mapping \mathcal{M} of the surface of section into itself,

$$\mathcal{M} : (x, p_x), (y, p_y), (t, p_t) \rightarrow (\bar{x}, \bar{p}_x), (\bar{y}, \bar{p}_y), (\bar{t}, \bar{p}_t). \quad (6.9.16)$$

Moreover, we have the relation

$$\int_{R_1} (p_x dx + p_y dy + p_t dt) = \int_{\bar{R}_1} (p_x dx + p_y dy + p_t dt) \quad (6.9.17)$$

for any closed path R_1 in the phase-space surface $z = 0$ and its image \bar{R}_1 under the action of \mathcal{M} . Therefore, \mathcal{M} is a symplectic map. Finally, determining the long-term behavior of trajectories in the ring is equivalent to determining the effect of the repeated action of \mathcal{M} on points in the surface of section.

6.10 Overview and Preview

We have studied symplectic maps and have seen their intimate connection with Hamiltonian dynamics. Thus, a key goal is to be able to produce, manipulate, and apply symplectic maps.

We have also learned that symplectic maps can be produced using mixed-variable generating functions. However, while often useful, this method has some disadvantages. As we have seen, the relations between old and new variables are initially implicit, and must be made explicit. This fact makes it difficult to apply, multiply, and invert symplectic maps specified in terms of generating functions.

In subsequent chapters we will learn that symplectic maps can also be produced using Lie transformations. This approach has the advantage of being explicit. Moreover, we will develop tools for inverting, multiplying, and otherwise manipulating symplectic maps in Lie form. Finally, the use of Lie methods yields physical insight and facilitates high-order perturbation theory.

Bibliography

Legendre Transformations

(See also the results of Googling “Legendre transformation”.)

- [1] V.I. Arnold, *Mathematical Methods of Classical Mechanics*, Second Edition, page 61, Springer Verlag (1989).
- [2] R. K. P. Zia, E. F. Redish, and S. R. McKay, “Making Sense of the Legendre Transform”, *American Journal of Physics* **77**, 614-622 (2009).

Transformation (Generating) Functions

(See also the Variational Calculus section of the Bibliography for Chapter 1.)

- [3] A. Wintner, *the Analytical Foundations of Celestial Mechanics*, Princeton University Press (1947).
- [4] C. Carathéodory, *Calculus of Variations and Partial Differential Equations of the First Order, Parts I and II*, Holden-Day (1965).
- [5] H. Goldstein, *Classical Mechanics*, Addison-Wesley (1980).
- [6] J.V. Jose and E.J. Salatan, *Classical Dynamics: A Contemporary Approach*, Cambridge University Press (1998).
- [7] H. Poincaré, *New Methods of Celestial Mechanics* (American Institute of Physics), New York, (1893/1993), Vol. 3, chap. 28, section 319.
- [8] V.I. Arnold, *Mathematical Methods of Classical Mechanics*, Second Edition, Springer Verlag (1989).
- [9] R. Abraham and J. Marsden, *Foundations of Mechanics*, American Mathematical Society (2008).
- [10] S. H. Benton, *The Hamilton-Jacobi Equation: A Global Approach*, Academic Press (1977).
- [11] Feng Kang, Wu Hua-mo, Qin Meng-shao, and Wang Dao-liu, “Construction of Canonical Difference Schemes for Hamiltonian Formalism via Generating Functions”, *Journal of Computational Mathematics* **11**, p. 71 (1989).

- [12] Feng Kang, “The Calculus of Generating Functions and the Formal Energy for Hamiltonian Algorithms”, *Journal of Computational Mathematics* **16**, p. 481 (1998).
- [13] Feng Kang and Mengzhao Qin, *Symplectic Geometric Algorithms for Hamiltonian Systems*, Zhejiang Publishing and Springer-Verlag (2010).
- [14] A. Weinstein, “Symplectic Manifolds and their Lagrangian Submanifolds”, *Advances in Math.* **6**, 329 (1971).
- [15] A. Weinstein, “The Invariance of Poincaré’s Generating Function for Canonical Transformations”, *Invent. Math.* **16**, 202 (1972).
- [16] A. Weinstein, “Lagrangian Submanifolds and Hamiltonian Systems”, *Ann. of Math.* **98**, 377 (1973).
- [17] A. Weinstein, “Normal Modes for Nonlinear Hamiltonian Systems”, *Invent. Math.* **20**, 47 (1973).
- [18] A. Weinstein, *Lectures on symplectic manifolds*, CBMS Reg. Conf. Series in Math. 29, American Mathematical Society, Providence, RI (1977).
- [19] A. Weinstein, “Symplectic Geometry”, *Bulletin Amer. Math. Soc. (N.S.)* **5**, 1-13 (1981).
- [20] J. Amiet and P. Huguenin, “Generating functions of canonical maps”, *Helv. Phys. Acta* **53**, 377 (1980).
- [21] A. Ozorio de Almeida, “On the Symplectically Invariant Variational Principle and Generating Functions”, *Proc. R. Soc. Lond. A* **431**, 403 (1990).
- [22] M. Sewell and I. Roulstone, “Anatomy of the Canonical Transformation”, *Phil. Trans. R. Soc. Lond. A* **345**, 577 (1993).
- [23] H. Gzyl, *Hamiltonian Flows and Evolution Semigroups*, Longman Group (1990).
- [24] B. Erdélyi, “Symplectic Approximation of Hamiltonian Flows and Accurate Simulation of Fringe Field Effects”, Michigan State University Physics and Astronomy Department Ph.D. Thesis (2001).
- [25] B. Erdélyi and M. Berz, “Optimal Symplectic Approximation of Hamiltonian Flows”, *Physical Review Letters* **87**, 114302 (2001).
- [26] B. Erdélyi and M. Berz, “Local Theory and Applications of Extended Generating Functions”, *International Journal of Pure and Applied Mathematics* **11**, 241 (2004).
- [27] Alex Haro, “The Primitive Function of an Exact Symplectomorphism”, *Nonlinearity* **13**, 1483-1500, (2000).
- Symplectic Geometry and Topology
- [28] M. Gromov, “Pseudo-holomorphic curves in symplectic manifolds”, *Invent. Math.* **82**, 307-347, (1985).

- [29] H. Hofer and E. Zehnder, *Symplectic Invariants and Hamiltonian Dynamics*, Birkhäuser (1994).
- [30] D. McDuff and D. Salamon, *Introduction to Symplectic Topology*, Clarendon Press (1995).
- [31] D. Salamon, Edit., *Symplectic Geometry*, London Mathematical Society Lecture Note Series 192, Cambridge University Press (1993).
- [32] A. T. Fomenko, *Symplectic Geometry*, Gordon and Breach (1995).
- [33] A. Crumeyrolle and J. Grifone, Edit., *Symplectic Geometry*, Pitman (1983).
- [34] L. Polterovich, *The Geometry of the Group of Symplectic Diffeomorphisms*, Birkhäuser (2000).
- [35] R. Berndt, *An Introduction to Symplectic Geometry*, American Mathematical Society (2001).
- [36] A. Banyaga, *The Structure of Classical Diffeomorphism Groups*, Kluwer Academic Publishers (1997).
- [37] B. Khesin and R. Wendt, *The Geometry of Infinite-Dimensional Groups*, Springer (2009).
- [38] V.I. Arnold, “Symplectic geometry and topology”, *J. of Math. Phys.* **41**, 6, 3307 (2000).
- [39] V.I. Arnold, *Dynamical Systems IV, Symplectic Geometry and its Applications*, Springer-Verlag (1990).
- [40] M.A. de Gosson, *The principles of Newtonian and quantum mechanics: the need for Planck’s constant, \hbar* , Imperial College Press, London (2001).
- [41] M.A. de Gosson, “The symplectic camel and phase space quantization”, *J. Phys. A: Math. Gen.* **34** p. 10085 (2001).
- [42] M.A. de Gosson, “The ‘symplectic camel principle’ and semiclassical mechanics”, *J. Phys A: Math. Gen.* **35**, p. 6825 (2002).
- [43] M.A. de Gosson, “Symplectically covariant Schrödinger equation in phase space”, *J. Phys A: Math. Gen.* **38**, p. 9263 (2005).
- [44] M.A. de Gosson, “Uncertainty Principle, Phase-Space Ellipsoids, and Weyl Calculus”, *Operator Theory: Advances and Applications*, Vol. 164, p. 121, Birkhäuser Verlag (2006).
- [45] M.A. de Gosson, *Symplectic geometry and quantum mechanics*, Birkhäuser Verlag (2006).

- [46] M.A. de Gosson and F. Luef, “Symplectic capacities and the geometry of uncertainty: The irruption of symplectic topology in classical and quantum mechanics”, *Physics Reports* 484, 131-179, Elsevier (2009).
- [47] F. Schlenk, *Embedding Problems in Symplectic Geometry*, de Gruyter Expositions in Mathematics, vol. 40, Berlin (2005).
- [48] L. Traynor, Book Review of *Embedding Problems in Symplectic Geometry* by F. Schlenk, *Bulletin of the American Mathematical Society* **43**, p. 593 (2006).
- [49] Y. Eliashberg and L. Traynor, Editors, *Symplectic Geometry and Topology*, American Mathematical Society (1999).
- [50] K. Cieliebak, H. Hofer, J. Latschev, and F. Schlenk, “Quantitative symplectic geometry”, 22 May 2006, arXiv:math.SG/0506191 v1 10 June 2005.
- [51] A. Cannas da Silva, *Lectures on Symplectic Geometry*, Corrected second printing, Springer Verlag (2008).
- [52] Y. Long, *Index Theory for Symplectic Paths with Applications*, Birkhäuser (2002).
- [53] V. Guillmen and S. Sternberg, *Symplectic Techniques in Physics*, Cambridge (1984).
- [54] V. Guillmen and S. Sternberg, *Geometric Asymptotics*, American Mathematical Society (1977).
- [55] P. Libermann and C-M. Marle, *Symplectic Geometry and Analytical Mechanics*, D. Reidel (1987).

Differential Manifolds and Forms

- [56] H.M. Edwards, *Advanced Calculus: A Differential Forms Approach*, Birkhäuser (1994).
- [57] H. Flanders, *Differential Forms with Applications to the Physical Sciences*, Academic Press (1963).
- [58] B. Schultz, *Geometrical Methods of Mathematical Physics*, Cambridge University Press (1980).
- [59] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, Vols. 1-5, Publish or Perish (1999).
- [60] M. Schreiber, *Differential Forms, A Heuristic Introduction*, Springer-Verlag (1977).
- [61] R. Courant and F. John, *Introduction to Calculus and Analysis*, Vol. I, Vol. II/1, Vol. II/2, Springer-Verlag (1998, 1999, 2000).
- [62] R. Abraham, J. Marsden, and T. Ratiu, *Manifolds, Tensor Analysis, and Applications*, Springer-Verlag (1988).

- [63] J. Marsden and T. Ratiu, *Introduction to Mechanics and Symmetry*, Second Edition, Springer (1999).
- [64] J. Marsden, R. Montgomery, and T. Ratiu, *Reduction, Symmetry , and Phases in Mechanics*, American Mathematical Society (1990).
- [65] C. J. Isham, *Modern Differential Geometry for Physicists*, Second Edition, World Scientific (1999).
- [66] T. Frankel, *The Geometry of Physics*, Second Edition, Cambridge University Press (2004).
- [67] S. Lang, *Fundamentals of Differential Geometry*, Springer-Verlag (1999).
- [68] W. Rudin, *Principles of Mathematical Analysis*, Third Edition, Mc-Graw-Hill (1976).
- [69] J. Dieudonné, *Treatise on Analysis*, Volumes 10-III and 10-IV in the series Pure and Applied Mathematics, Academic Press (1972 and 1974).
- [70] P. Iglesias-Zemmour, *Diffeology*, (2010), <http://math.huji.ac.il/~piz/documents/Diffeology.pdf>.

Poincaré Maps

- [71] 3-body problem
- [72] Galactic Dynamics
- [73] A.J. Dragt, “Trapped Orbits in a Magnetic Dipole Field”, *Rev. Geophys.* **3**, p. 255-298 (1965).
- [74] A.J. Dragt and J.M. Finn, “Insolubility of Trapped Particle Motion in a Magnetic Dipole Field”, *J. Geophys. Res.* **81** p. 2327-2340 (1976).
- [75] A.J. Dragt and J.M. Finn, “Normal Form for Mirror Machine Hamiltonians”, *J. Math. Physics* **20**, p. 2649-2660 (1979).

The Galilean Group and Group Contraction/Deformation

- [76] Theodore Jacobson provided the ideas for Exercise 6.2.5.
- [77] E. Inönü and E.P. Wigner, “On the contraction of groups and their representations”, *Proc. Nat. Acad. Sci. U.S.A.* **39**, 510-524 (1953).
- [78] J. Löhmus, E. Paal, and L. Sorgsepp, *Nonassociative Algebras in Physics*, Hadronic Press (1994).

Chapter 7

Lie Transformations and Symplectic Maps

Chapter 6 showed that there is an intimate connection between symplectic maps and Hamiltonian flows, and showed how symplectic maps could be produced (in implicit form) with the aid of mixed-variable generating functions. This chapter explores how Lie transformations can be used for the same purpose, and how their use produces symplectic maps in explicit form. It also displays how the group of all symplectic maps is a Lie group whose Lie algebra is the Poisson bracket Lie algebra of all phase-space functions.

7.1 Production of Symplectic Maps

Let $f(z, t)$ be any dynamical variable, and let $\exp(: f(z, t) :)$ be the Lie transformation associated with f . (Here, as in Section 6.1, the time t simply plays the role of a parameter.) This Lie transformation can be used to define a map \mathcal{M} that produces new variables $\bar{z}(z, t)$ by the rule

$$\bar{z}_a(z, t) = \exp(: f(z, t) :) z_a , \quad a = 1, 2, \dots, 2n. \quad (7.1.1)$$

The relations (1.1) can also be expressed more compactly by writing

$$\bar{z} = \mathcal{M}z, \quad (7.1.2)$$

$$\mathcal{M} = \exp(: f :). \quad (7.1.3)$$

Note that in writing (1.1) we have indicated explicitly the arguments of f . Generally these arguments will be omitted for simplicity of notation. However, it is always important to keep in mind what these arguments are, and they should and will always be stated explicitly whenever there is any possibility for confusion.

Consider the Poisson brackets of the various \bar{z} 's with each other. Using the definition (1.1), the isomorphism condition (5.4.14), and (5.4.22), we find the result

$$\begin{aligned} [\bar{z}_a, \bar{z}_b]_z &= [\exp(: f :) z_a, \exp(: f :) z_b]_z \\ &= \exp(: f :)[z_a, z_b]_z \\ &= \exp(: f :) J_{ab} = J_{ab}. \end{aligned} \quad (7.1.4)$$

It follows from (1.4) that \mathcal{M} is a symplectic map! What has been shown is that every Lie transformation may be viewed as a symplectic map. Consequently, Lie transformations produce an endless supply of symplectic maps. And, unlike the case for mixed-variable generating functions (see Section 6.5.1), these maps are immediately in explicit form. Finally, we see from (5.4.15) and its generalization that products of Lie transformations also produce symplectic maps.

Consider the map $\mathcal{M}(\lambda)$ depending on the parameter λ and defined by the relation

$$\mathcal{M}(\lambda) = \exp(\lambda : f :). \quad (7.1.5)$$

This map produces the transformation

$$\begin{aligned} \bar{z}(z, t; \lambda) &= \exp(\lambda : f :) z \\ &= z + \lambda : f : z + (\lambda^2/2!) : f :^2 z + \dots \end{aligned} \quad (7.1.6)$$

Evidently we have the relations

$$\mathcal{M}(0) = \mathcal{I} = \text{identity map}, \quad (7.1.7)$$

$$\mathcal{M}(1) = \mathcal{M}. \quad (7.1.8)$$

Next let $\mathcal{M}(\lambda_1)$ and $\mathcal{M}(\lambda_2)$ be two maps of the form (1.5) corresponding to the λ values λ_1 and λ_2 , respectively. Consider the product map given by the relation

$$\mathcal{M}(\lambda_1)\mathcal{M}(\lambda_2) = \exp(\lambda_1 : f :) \exp(\lambda_2 : f :). \quad (7.1.9)$$

Because Lie operators are linear operators, their behavior is in many ways analogous to that of matrices. In particular, observe that we may attempt to combine the exponents appearing on the right side of (1.9) into a single exponent using the Baker-Campbell-Hausdorff (BCH) series. See (3.7.33) and (3.7.34). According to (5.3.14), the Lie operators $\lambda_1 : f :$ and $\lambda_2 : f :$ commute. Consequently, the exponents in (1.9) simply add to give the result

$$\begin{aligned} \mathcal{M}(\lambda_1)\mathcal{M}(\lambda_2) &= \exp(\lambda_1 : f :) \exp(\lambda_2 : f :) \\ &= \exp(\lambda_1 : f : + \lambda_2 : f :) \\ &= \exp((\lambda_1 + \lambda_2) : f :) \\ &= \mathcal{M}(\lambda_1 + \lambda_2). \end{aligned} \quad (7.1.10)$$

Section 6.2 showed that the set of all symplectic maps forms a group. The relation (1.10) shows that the subset of symplectic maps given by (1.5) forms a one-parameter subgroup of symplectic maps. Moreover (1.10), together with (1.7) and (1.8), shows that any Lie transformation lies on a one-parameter subgroup of symplectic maps. That is, any Lie transformation is continuously connected to the identity map \mathcal{I} by a path whose points are all elements of some common subgroup of symplectic maps.

We have seen that Lie transformations produce symplectic maps that act on the phase-space variables z . According to (5.4.11), Lie transformations also act on general functions. Let $g^{\text{old}}(z, t)$ be any function of the phase-space variables z and perhaps the time t . Then

the Lie transformation (1.3), using (5.4.11), produces a *new* function $g^{\text{new}}(z, t)$ according to the rule

$$\begin{aligned} g^{\text{new}}(z, t) &= \mathcal{M}g^{\text{old}}(z, t) = \exp(:f:)g^{\text{old}}(z, t) \\ &= g^{\text{old}}(\exp(:f:)z, t) = g^{\text{old}}(\bar{z}(z, t), t) \\ &= g^{\text{old}}(\mathcal{M}z, t). \end{aligned} \quad (7.1.11)$$

Note that the relation (1.11) is analogous to (6.3.6).

The symplectic map defined by (1.3) has the particular property that f is an *invariant function* for the map. That is, there is the relation

$$f(\bar{z}, t) = f(z, t), \quad \text{or} \quad f^{\text{new}}(z, t) = f^{\text{old}}(z, t). \quad (7.1.12)$$

To see the truth of this assertion, apply (5.4.11) to the case where $g = f$. We find, using the notation of (1.1), the result

$$\exp(:f:)f(z, t) = f(\bar{z}, t). \quad (7.1.13)$$

However, using the expression (5.4.2), we also obtain the result

$$\begin{aligned} \exp(:f:)f(z, t) &= f + [f, f] + [f[f, f]]/2! + \dots \\ &= f(z, t) \end{aligned} \quad (7.1.14)$$

since the Poisson bracket $[f, f]$ is zero by the antisymmetry condition. Comparison of (1.13) and (1.14) shows that (1.12) is indeed correct. Note again that in all these calculations, the time t plays no essential role and may be regarded simply as a parameter.

Suppose the symplectic map $\exp(-:f(z, t):)$ is applied to both sides of (1.1). We find the result

$$\exp(-:f:)\bar{z}_a = \exp(-:f:)\exp(:f:)z_a. \quad (7.1.15)$$

Consider first the problem of evaluating the right side of (1.15). Observe that the Lie operators $:f:$ and $-:f:$ commute. Consequently, the exponents on the right side of (1.15) can be added to give the result

$$\exp(-:f:)\exp(:f:) = \exp(:0:) = \mathcal{I}. \quad (7.1.16)$$

Correspondingly, when read from right to left, (1.15) may be rewritten in the form

$$z_a = \exp(-:f:)\bar{z}_a. \quad (7.1.17)$$

The right side of (1.17), which is the left side of (1.15), can be viewed in two ways. First, both f and the \bar{z}_a can be regarded as functions of z (and perhaps the time t), and all indicated Poisson brackets are to be taken with respect to the variables z . The result of these operations is simply to produce the functions z_a as indicated by the left side of (1.17). Alternatively, f may be viewed as a function of \bar{z} by writing the relation

$$f(z, t) = f^*(\bar{z}, t) = f(z(\bar{z}, t), t). \quad (7.1.18)$$

See (6.3.3) and (6.3.5). Then, thanks to the preservation of Poisson brackets under symplectic maps as expressed by (6.3.11) and (6.3.21), the relation (1.17) can be written in the form

$$z_a(\bar{z}, t) = \exp(- : f^*(\bar{z}, t) :) \bar{z}_a \quad (7.1.19)$$

where now all Poisson brackets are to be taken with respect to the variables \bar{z} . However, the invariance condition (1.12) can be written in the form

$$f^*(\bar{z}, t) = f(z, t) = f(\bar{z}, t). \quad (7.1.20)$$

Consequently, (1.17) can also be written in the final form

$$z_a(\bar{z}, t) = \exp(- : f(\bar{z}, t) :) \bar{z}_a \quad (7.1.21)$$

where all Poisson brackets are to be taken with respect to the variables \bar{z} . What has been shown is that if \mathcal{M} is given by the relations (1.1) through (1.3), then, when due regard is taken for the variables involved, the inverse relation (1.21) can be written in the compact form

$$z = \mathcal{M}^{-1}\bar{z} \quad (7.1.22)$$

with

$$\mathcal{M}^{-1} = \exp(- : f :). \quad (7.1.23)$$

Exercises

7.1.1. Verify in detail the steps leading from (1.15) to (1.23).

7.1.2. Suppose f and g are two phase-space functions in involution. That is,

$$[f, g] = 0. \quad (7.1.24)$$

Show from the power series definition (5.4.1) that in this case there is the relation

$$\exp(: f :) \exp(: g :) = \exp(: f + g :). \quad (7.1.25)$$

See Exercise 3.7.11.

7.1.3. Consider the map of Exercise 5.4.6 written in the form

$$\bar{q}(q, p) = \exp(: f :)q = q(1 - \lambda p)^2, \quad (7.1.26)$$

$$\bar{p}(q, p) = \exp(: f :)p = p/(1 - \lambda p). \quad (7.1.27)$$

Verify by direct computation that $[\bar{q}, \bar{p}]_z = 1$. Verify that $f = \lambda qp^2$ is an invariant function, that is $f(\bar{q}, \bar{p}) = f(q, p)$. Solve (1.26) and (1.27) for $q(\bar{q}, \bar{p}), p(\bar{q}, \bar{p})$. Verify by direct computation that $[q, p]_{\bar{z}} = 1$. Verify directly that \mathcal{M}^{-1} is given by (1.23).

7.1.4. Repeat Exercise 1.3 for the f of Exercise 5.4.5.

7.2 Realization of the Group $Sp(2n)$ and Its Subgroups

7.2.1 Realization of General Group Element

Let \mathcal{M} be the map given by the relation

$$\mathcal{M} : z \rightarrow \bar{z} = Mz, \quad (7.2.1)$$

where M is a symplectic matrix. Then, according to Exercise 6.2.1, \mathcal{M} is a symplectic map. Since M is a symplectic matrix, it can be written in the form

$$M = PO = \exp(JS^a) \exp(JS^c). \quad (7.2.2)$$

See (3.8.1) and (3.8.24).

Define a quadratic polynomial f_2^a in terms of the matrix S^a appearing in the decomposition (2.2) by the relation

$$f_2^a = -(1/2) \sum_{de} S_{de}^a z_d z_e. \quad (7.2.3)$$

Now consider the Lie operator $: f_2^a :$. Suppose this Lie operator acts on the various z 's. We find the result

$$\begin{aligned} : f_2^a : z_b &= -(1/2) \sum_{de} S_{de}^a [z_d z_e, z_b] \\ &= -(1/2) \sum_{de} S_{de}^a \{[z_d, z_b] z_e + [z_e, z_b] z_d\} \\ &= -(1/2) \sum_{de} S_{de}^a \{J_{db} z_e + J_{eb} z_d\} \\ &= \sum_d (JS^a)_{bd} z_d. \end{aligned} \quad (7.2.4)$$

Here use has been made of the antisymmetry of J and the symmetry of S^a . Using matrix and vector notation, (2.4) can also be written in the more compact form

$$: f_2^a : z = (JS^a)z. \quad (7.2.5)$$

From this form it is easy to see that there is the general relation

$$: f_2^a :^m z = (JS^a)^m z. \quad (7.2.6)$$

Finally, it follows from (2.6) that we also have the relation

$$\exp(: f_2^a :) z = \exp(JS^a) z = Pz. \quad (7.2.7)$$

In a similar way, define a quadratic polynomial f_2^c in terms of the matrix S^c appearing in (2.2),

$$f_2^c = -(1/2) \sum_{de} S_{de}^c z_d z_e. \quad (7.2.8)$$

Correspondingly, we have the relation

$$\exp(: f_2^c :)z = \exp(JS^c)z = Oz. \quad (7.2.9)$$

Now define a symplectic map \mathcal{M} by the relation

$$\mathcal{M} = \exp(: f_2^c :) \exp(: f_2^a :). \quad (7.2.10)$$

Here \mathcal{M} is intended to act on the phase-space variables z , and both f_2^a and f_2^c are functions of z . We find, using (2.7) and (2.9), the result

$$\begin{aligned} \mathcal{M}z_b &= \exp(: f_2^c :) \exp(: f_2^a :)z_b \\ &= \exp(: f_2^c :) \sum_d P_{bd} z_d \\ &= \sum_d P_{bd} \exp(: f_2^c :)z_d \\ &= \sum_{de} P_{bd} O_{de} z_e = (Mz)_b. \end{aligned} \quad (7.2.11)$$

This result may be written in the more compact form

$$\bar{z} = \mathcal{M}z = Mz. \quad (7.2.12)$$

Notice two things. First, we have shown that any linear symplectic transformation of the form (2.1) can be realized as the product of two Lie transformations. Second, comparison of (2.2) and (2.10) shows that the corresponding factors appear in opposite order. That is, when Lie transformations all involve the *same* phase-space variables, they act from *left to right*. This particular feature of Lie transformations will be explored in greater detail in Section 8.3. There it will also be explained why the difference in sign between relations such as (5.5.1) and (2.3) is not arbitrary. The reader will soon come to realize that Lie transformations lead lives of their own, and possess many unexpected properties.

7.2.2 Realization of Various Subgroups

We next employ Lie transformations to study various aspects of subgroups of $Sp(2n)$. We begin with symplectic matrices of the form (3.3.9). As shown in Section (3.10), such matrices are related to matrices S of the form (3.10.2). Correspondingly, let f_2^B be the quadratic polynomial given by the relation

$$f_2^B = -(1/2) \sum_{de} S_{de} z_d z_e = -(1/2) \sum_{jk} B_{jk} p_j p_k. \quad (7.2.13)$$

Then we have the relation

$$\bar{z} = \exp(: f_2^B :)z = Mz, \quad (7.2.14)$$

with M given by (3.3.9) or (3.10.5). We have learned that the subgroup of symplectic matrices of the form (3.3.9) is produced by Lie transformations whose Lie operators arise from

monomials of the form $p_j p_k$. Evidently, monomials of this form are mutually in involution. See (5.5.14). Correspondingly, the subgroup is Abelian, as has already been seen earlier.

In a similar fashion, it is easily checked that the subgroup of symplectic matrices of the form (3.3.10) is generated by the Lie operator $:f_2^C:$ given by the relation

$$f_2^C = -(1/2) \sum_{de} S_{de} z_d z_e = (1/2) \sum_{jk} C_{jk} q_j q_k \quad (7.2.15)$$

with S given by (3.10.7). That is, we have the relation

$$\bar{z} = \exp(:f_2^C:)z = Mz \quad (7.2.16)$$

with M given by (3.3.10). We have learned that Lie transformations arising from monomials of the form $q_j q_k$ produce the subgroup of symplectic matrices of the form (3.3.10). Monomials of the form $q_j q_k$ are also mutually in involution, and the corresponding subgroup is again Abelian. See (5.5.15).

Next consider the subgroup of matrices of the form (3.3.11). Let f_2 be the quadratic polynomial defined by the relation

$$\begin{aligned} f_2 &= -(1/2) \sum_{de} S_{de} z_d z_e = -(1/2)(z, Sz) \\ &= -(1/2)[(q, a^T p) + (p, aq)] = -(q, a^T p) \\ &= -\sum_{jk} a_{jk}^T q_j p_k. \end{aligned} \quad (7.2.17)$$

Here S is given by (3.10.13). Then, for matrices M of the form (3.3.11) and sufficiently near the identity, we have the relation

$$\bar{z} = \exp(:f_2:)z = Mz \quad (7.2.18)$$

with f_2 given by (2.17). We have learned that Lie transformations arising from monomials of the form $q_j p_k$ produce symplectic matrices of the form (3.3.11). It is easily verified that the set of monomials of the form $q_j p_k$ forms a Lie algebra under the Poisson bracket operation. See (5.5.16). Correspondingly, matrices of the form (3.3.11) constitute a group. As we saw in Section 3.10, this group is $GL(n, \mathbb{R})$.

As a special case of (2.18), consider the Lie transformation $\exp(:-\lambda q_\ell p_\ell:)$ where ℓ is some integer satisfying $0 \leq \ell \leq n$, and λ is a parameter. Then for $j \neq \ell$ we have the relations

$$\begin{aligned} \bar{q}_j &= \exp(:-\lambda q_\ell p_\ell:)q_j = q_j, \\ \bar{p}_j &= \exp(:-\lambda q_\ell p_\ell:)p_j = p_j. \end{aligned} \quad (7.2.19)$$

And for $j = \ell$ we find the result

$$\begin{aligned} \bar{q}_\ell &= \exp(:-\lambda q_\ell p_\ell:)q_\ell = (e^\lambda)q_\ell, \\ \bar{p}_\ell &= \exp(:-\lambda q_\ell p_\ell:)p_\ell = (e^{-\lambda})p_\ell. \end{aligned} \quad (7.2.20)$$

See Exercise 5.4.4. We conclude that $\exp(: -\lambda q_\ell p_\ell :)$ scales q_ℓ and p_ℓ by the (positive) factors e^λ and $e^{-\lambda}$, respectively, and leaves the remaining q_j and p_j untouched.

Consider next Lie transformations corresponding to the quadratic polynomials f_2^ℓ given by the definition

$$f_2^\ell = -(1/2)\theta_\ell(q_\ell^2 + p_\ell^2). \quad (7.2.21)$$

Then for $j \neq \ell$ we have the relations

$$\bar{q}_j = \exp(: f_2^\ell :)q_j = q_j,$$

$$\bar{p}_j = \exp(: f_2^\ell :)p_j = p_j. \quad (7.2.22)$$

And for $j = \ell$ we find the results

$$\begin{aligned} \bar{q}_\ell &= \exp(: f_2^\ell :)q_\ell = q_\ell \cos \theta_\ell + p_\ell \sin \theta_\ell, \\ \bar{p}_\ell &= \exp(: f_2^\ell :)p_\ell = -q_\ell \sin \theta_\ell + p_\ell \cos \theta_\ell. \end{aligned} \quad (7.2.23)$$

See Exercise 5.4.5. We conclude that in this case $\exp(: f_2^\ell :)$ produces a *rotation* by angle θ_ℓ in the q_ℓ, p_ℓ plane. Because these two variables are conjugate, such a rotation is sometimes referred to as a *phase advance*.

At this point we remark that there is a correspondence between phase advances and the maximal $Sp(2n, \mathbb{R})$ torus described at the end of Section 3.9. From (2.22) and (2.23) we find the result

$$\exp(: f_2^1 + f_2^2 + \cdots + f_2^n :)z_a = \sum_b [N(\theta_1, \theta_2, \dots, \theta_n)]_{ab} z_b \quad (7.2.24)$$

where N is given by (3.8.85), or (3.5.60) and (3.5.61). Here we have used the ordering (3.2.4).

Finally, consider Lie transformations corresponding to the quadratic polynomials f_2^{jk} given by the definition

$$f_2^{jk} = \theta_{jk}(q_j p_k - q_k p_j). \quad (7.2.25)$$

It is easily verified that the set of such polynomials is closed under the Poisson bracket operation, and thus constitutes a Lie algebra. Furthermore, for $\ell \neq j, k$ we have the evident result

$$\begin{aligned} \bar{q}_\ell &= \exp(: f_2^{jk} :)q_\ell = q_\ell, \\ \bar{p}_\ell &= \exp(: f_2^{jk} :)p_\ell = p_\ell. \end{aligned} \quad (7.2.26)$$

Also, explicit calculation gives the results

$$\begin{aligned} \bar{q}_j &= \exp(: f_2^{jk} :)q_j = q_j \cos \theta_{jk} + q_k \sin \theta_{jk}, \\ \bar{q}_k &= \exp(: f_2^{jk} :)q_k = -q_j \sin \theta_{jk} + q_k \cos \theta_{jk}, \\ \bar{p}_j &= \exp(: f_2^{jk} :)p_j = p_j \cos \theta_{jk} + p_k \sin \theta_{jk}, \\ \bar{p}_k &= \exp(: f_2^{jk} :)p_k = -p_j \sin \theta_{jk} + p_k \cos \theta_{jk} \end{aligned} \quad (7.2.27)$$

We conclude that in this case $\exp(: f_2^{jk} :)$ produces a rotation by angle θ_{jk} in the q_j, q_k plane, and simultaneously, the same rotation in the p_j, p_k plane. These rotations provide a realization of the special orthogonal group $SO(n, \mathbb{R})$. See Exercise 2.5.

Finally, it can be verified that the f_2^ℓ given by (2.21) and the f_2^{jk} given by (2.25) all correspond to matrices S^c that commute with J . Consequently, the transformations given by (2.22), (2.23) and (2.26), (2.27) are all in the $U(n)$ subgroup of $Sp(2n)$.

7.2.3 Another Proof of Transitive Action of $Sp(2n)$ on Phase Space

Near the end of Section 3.6 we showed in effect that if \tilde{z}^i and \tilde{z}^f are *any* two points in phase space *distinct* from the origin, then there is a *linear* symplectic map of the form (2.1) such that

$$\tilde{z}^f = M\tilde{z}^i. \quad (7.2.28)$$

See (3.6.115). To recapitulate, with the exception of the origin, any point in phase space can be sent into any other point by a linear symplectic transformation. (The origin is obviously sent into itself.) Following the terminology of Section 5.12, we say that, with the exception of the origin, $Sp(2n)$ acts *transitively* on phase space.

We will now provide another proof of this result using a series of constructive steps that have some instructive merit. First, suppose that \tilde{z}^i is not the origin. Perform successive phase advances of the form (2.22), (2.23) to remove all “ p ” type components from \tilde{z}^i . Next perform a rotation of the form (2.26), (2.27) in the $(n-1), n$ plane to remove any q_n component. In so doing, no p_n component is produced. Thus both p_n and q_n components have been removed. Next perform a rotation in the $(n-2), (n-1)$ plane to remove any q_{n-1} component, etc. The net result of a sequence of such rotations is that all components have been transformed to zero save for the q_1 component. Also, this component cannot be zero, because transformations of the form (2.22), (2.23) and (2.26), (2.27) evidently preserve the inner product (z, z) , and this quantity cannot vanish if \tilde{z}^i is not the origin. Moreover, the q_1 component can be taken to be positive. [If it is not, simply increase θ_{12} by π . See (2.27).] Finally, apply a scaling transformation of the form (2.19), (2.20) with $\ell = 1$ to transform the q_1 component so that it has the numerical value 1. Since all the transformations just described are linear symplectic maps, and linear symplectic maps form a group, it follows that there is a symplectic matrix M^i such that

$$M^i \tilde{z}^i = z^1. \quad (7.2.29)$$

Here z^1 is a vector (phase-space point) whose q_1 component is 1, and all others are zero,

$$z_a^1 = \delta_{a1}. \quad (7.2.30)$$

By an analogous argument, there is also a symplectic matrix M^f such that

$$M^f \tilde{z}^f = z^1. \quad (7.2.31)$$

Upon combining (2.30) and (2.28), we get the result

$$\tilde{z}^f = (M^f)^{-1} z^1 = (M^f)^{-1} M^i \tilde{z}^i. \quad (7.2.32)$$

That is, the advertised result (2.27) is correct with M given by the relation

$$M = (M^f)^{-1} M^i. \quad (7.2.33)$$

Note that M as given by (2.33) is again symplectic as a consequence of the group property.

Introduce the term *punctured phase space* to refer to the set of all points in phase space with the exception of the origin. We have learned that $Sp(2n, \mathbb{R})$ acts transitively on punctured phase space. Consequently, according to the discussion in Section 5.12, punctured phase space is a homogeneous space with respect to $Sp(2n, \mathbb{R})$, and therefore must be a coset space of $Sp(2n, \mathbb{R})$ with respect to one of its subgroups. What is this subgroup? See Exercise 7.4 in Section 7.7.

Exercises

7.2.1. Verify the relations (2.4) through (2.7).

7.2.2. Verify (2.17) and (2.18).

7.2.3. Verify (2.19) and (2.20).

7.2.4. Verify (2.22) and (2.23).

7.2.5. The orthogonal group $O(n, \mathbb{R})$ is defined by the set of real $n \times n$ matrices satisfying (5.10.13). Show that such matrices do indeed form a group. See Exercise (3.7.24). Show that (5.10.13) implies the relation

$$\det O = \pm 1.$$

Orthogonal matrices with determinant $+1$ are called *proper*. Show that proper $O(n, \mathbb{R})$ matrices form a subgroup of $O(n, \mathbb{R})$. Recall that this subgroup is $SO(n, \mathbb{R})$, the special orthogonal group. Show that the set of $O(n, \mathbb{R})$ matrices with determinant -1 (called *improper* orthogonal) does not form a subgroup, and is disconnected from $SO(n, \mathbb{R})$. Show that any matrix of the form (3.3.11) with

$$A = D = O, \quad (7.2.34)$$

and O orthogonal, is symplectic. Recall Exercise 6.5.2.

If O is special (proper) real orthogonal, then it can be written in the form

$$O = \exp(F) \quad (7.2.35)$$

where F is $n \times n$, real, and antisymmetric,

$$F^T = -F. \quad (7.2.36)$$

Show that M as given by (3.3.11) and (2.34) has the form

$$M = \exp(JS^c) \quad (7.2.37)$$

where JS^c is the matrix

$$JS^c = \begin{pmatrix} F & 0 \\ 0 & F \end{pmatrix}. \quad (7.2.38)$$

Show that S^c is given by the relation

$$S^c = \begin{pmatrix} 0 & -F \\ F & 0 \end{pmatrix}. \quad (7.2.39)$$

Show that S^c is symmetric and, as the notation suggests, commutes with J . See (3.9.6). Use (2.8) and (2.39) to derive the result

$$\begin{aligned} f_2^c &= -(1/2)(z, S^c z) = (q, Fp) \\ &= \sum_{jk} F_{jk} q_j p_k \\ &= (1/2) \sum_{jk} F_{jk} (q_j p_k - q_k p_j). \end{aligned} \quad (7.2.40)$$

7.2.6. Show that the set of polynomials of the form (2.25) or (2.40) constitutes a Lie algebra under the Poisson bracket operation. This Lie algebra is $so(n, \mathbb{R})$, the Lie algebra of $SO(n, \mathbb{R})$.

7.2.7. Verify (2.26) and (2.27).

7.2.8. Verify that the transformations (2.22), (2.23) and (2.26), (2.27) preserve the inner product (w, z) .

7.2.9. Show that for any (square) matrix G there is the identity

$$\exp(G) = \cosh G + \sinh G. \quad (7.2.41)$$

Using (3.1.3) and the series expansion for cosh and sinh, verify the relation

$$\exp(\lambda J) = I \cos \lambda + J \sin \lambda. \quad (7.2.42)$$

Find quadratic polynomials f_2 such that \mathcal{M} given by $\mathcal{M} = \exp(: f_2 :)$ satisfies (2.1) with $M = \pm J$ and $M = \pm I$.

7.2.10. Review Exercises 6.2.6 and 6.2.7. There we learned that Lorentz transformations are symplectic maps, and in fact are linear symplectic maps. Therefore, based on the work of this section, we suspect that they can be written as Lie transformations. Lorentz transformations consist of rotations about the x, y, z axes and velocity transformations (sometimes called *boosts*) along these axes. (We remark that the factorization of Lorentz transformations into rotations and boosts arises naturally in a polar decomposition of the Lorentz group.) The relations (2.27) show that rotations can be written as Lie transformations. We want to show that the same is true of boosts. For simplicity we will consider boosts along the z axis. Boosts along the other axes, and in arbitrary directions, can be treated analogously.

Verify that the quantities β, γ defined by (6.2.54) and (6.2.55) satisfy the relation

$$\gamma^2 - \gamma^2 \beta^2 = 1. \quad (7.2.43)$$

Therefore, we can define a quantity χ called the *rapidity* such that

$$\sinh \chi = \beta \gamma, \quad (7.2.44)$$

$$\cosh \chi = \gamma. \quad (7.2.45)$$

With this notation, show that (6.2.58) and (6.2.59), and their momentum counterparts, can be written on the form

$$\tilde{x}^3 = x^3 \cosh \chi + x^4 \sinh \chi, \quad (7.2.46)$$

$$\tilde{x}^4 = x^3 \sinh \chi + x^4 \cosh \chi, \quad (7.2.47)$$

$$\tilde{p}^3 = p^3 \cosh \chi + p^4 \sinh \chi, \quad (7.2.48)$$

$$\tilde{p}^4 = p^3 \sinh \chi + p^4 \cosh \chi. \quad (7.2.49)$$

Using the metric tensor \bar{g} given by (1.6.75), show that (2.48) and (2.49) can be rewritten as

$$\tilde{p}_3 = p_3 \cosh \chi - p_4 \sinh \chi, \quad (7.2.50)$$

$$\tilde{p}_4 = -p_3 \sinh \chi + p_4 \cosh \chi. \quad (7.2.51)$$

Let f_2 be the quadratic polynomial defined by the relation

$$f_2 = -\chi(x^3 p_4 + x^4 p_3). \quad (7.2.52)$$

Verify the relations

$$: f_2 : x^3 = \chi x^4, \quad (7.2.53)$$

$$: f_2 : x^4 = \chi x^3, \quad (7.2.54)$$

$$: f_2 : p_3 = -\chi p_4, \quad (7.2.55)$$

$$: f_2 : p_4 = -\chi p_3. \quad (7.2.56)$$

Finally, by summing the relevant infinite series, show that

$$\tilde{x}^3 = \exp(: f_2 :) x^3, \quad (7.2.57)$$

$$\tilde{x}^4 = \exp(: f_2 :) x^4, \quad (7.2.58)$$

$$\tilde{p}_3 = \exp(: f_2 :) p_3, \quad (7.2.59)$$

$$\tilde{p}_4 = \exp(: f_2 :) p_4. \quad (7.2.60)$$

7.2.11. We have seen that, apart from the origin, $Sp(2n)$ acts transitively on the $2n$ -dimensional Euclidean space E^{2n} . Does $O(2n, \mathbb{R})$ act transitively on E^{2n} ?

7.2.12. Consider the 3-dimensional *isotropic* harmonic oscillator described by the Hamiltonian

$$H = \sum_1^3 p_j^2/(2m) + (k/2)q_j^2. \quad (7.2.61)$$

Show that there is a linear canonical transformation that brings H to the form

$$H = (\omega/2) \sum_1^3 (p_j^2 + q_j^2). \quad (7.2.62)$$

See Exercises (5.4.4) and (6.4.3). Consider the set of all linear canonical transformations that leaves H invariant. Show that these transformations form a group isomorphic to $U(3)$. See Section (5.8). Show that there is an even larger group of linear and nonlinear canonical transformations that leaves H invariant.

7.2.13. Equation (5.11.39) provides the partial Iwasawa decomposition for any element in the group $Sp(2n, \mathbb{R})$. The purpose of this exercise is to find the corresponding decomposition of the Lie algebra $sp(2n, \mathbb{R})$. From (5.11.39) we see that we must study the elements in the Lie algebra associated with $M(Z)$ given by (5.11.18), and the elements in the Lie algebra associated with $M(m)$ given by (3.9.19). The case of $M(m)$ has already been discussed, and is realized in terms of Lie transformations by symplectic maps of the form $\exp(: f_2^c :)$. Thus, the associated elements in the Lie algebra $sp(2n, \mathbb{R})$ are the polynomials f_2^c when the Poisson bracket realization is used, the Lie operators $: f_2^c :$ when the Lie operator realization is used, and the matrices of the form JS^c when the matrix realization is used. We now turn to the case of $M(Z)$. According to (5.11.18) it can be written as the product of two factors. Consider first the second factor. It can be written in the form (5.11.41). Show that matrices of this form are equivalent to those given by (3.3.9). That is, show that any real symmetric B can be written in the form

$$B = Y^{-1/2}XY^{-1/2} \quad (7.2.63)$$

with X real symmetric, and Y real symmetric and positive definite. Thus, this case has already been treated in the discussion surrounding (2.13) and (2.14). Finally, consider the first factor in (4.11.18). It can be written in the form given by (5.11.43) and (5.11.44). Show that this case is a special case of (3.3.11) with A symmetric. Do symplectic matrices of the form (3.3.11) with A symmetric form a subgroup? According to Exercise 5.11.9, $\log(Y)$ is real and symmetric. Thus, show that this case is a special case of that treated in the discussion surrounding (2.17). Specifically, show in this case that the matrix a appearing in (2.17) is real and symmetric. Consider the Lie algebra $sp(2, \mathbb{R})$. According to Section 5.6, it has a Poisson bracket realization in terms of the basis polynomials b^0 , f , and g . See (5.6.6), (5.6.11), and (5.6.12). Show that the partial Iwasawa basis for $sp(2, \mathbb{R})$ is given by the polynomials b^0 , $p^2 = b^0 - f$, and $qp = g$. Show that the partial Iwasawa basis for the quadratic polynomial realization of $sp(4, \mathbb{R})$ is given by the polynomials b^0, b^1, b^2, b^3 ; $p_1^2 = (1/2)(b^0 + b^3 + f^1 - g^2)$, $p_1 p_2 = (1/2)(b^1 - f^3)$, $p_2^2 = (1/2)(b^0 - b^3 - f^1 - g^2)$; $q_1 p_1 = -(1/2)(f^2 + g^1)$, $q_2 p_2 = (1/2)(g^1 - f^2)$, $q_1 p_2 + q_2 p_1 = g^3$. See Section 5.7.

7.2.14. The *center* of a group G consists of those elements of G that commute with all elements of G . Show that the center of a group forms a subgroup of G . Show that the

center of $Sp(2n, \mathbb{R})$ consists of the elements $\pm I$. What is the center of $Sp(2n, \mathbb{C})$? What is the center of $U(n)$? What is the center of $SU(n)$? What is the center of $O(n, \mathbb{R})$?

7.2.15. Refer to Exercise 4.5.4. Show that static symplectic matrices form a group. Show that this group is generated by quadratic polynomials f_2 that obey $\partial f_2 / \partial t = 0$.

7.3 Invariant Scalar Product

In Section 5.8, in the context of describing representations of $su(3)$, the need for a suitable scalar product was mentioned. In this section we will introduce a particularly convenient scalar product that has special properties with regard to the symplectic group.

7.3.1 Definition of Scalar Product

For simplicity, we will treat the case of $sp(6)$. From this treatment it will be easy to read off the results for the general case $sp(2n)$. Let $G(\mu; \nu)$ denote the general monomial defined by the relation

$$G(\mu; \nu) = (\mu_1! \nu_1! \mu_2! \nu_2! \mu_3! \nu_3!)^{-1/2} p_1^{\mu_1} q_1^{\nu_1} p_2^{\mu_2} q_2^{\nu_2} p_3^{\mu_3} q_3^{\nu_3}. \quad (7.3.1)$$

It is evident that the $G(\mu; \nu)$ form a basis for the set of all phase-space functions.

For reasons that will become clear shortly, let us pause to consider the Lie operators associated with the quadratic polynomials $q_1^2, p_1^2, q_1 q_2, p_1 p_2, q_1 p_1$, and $q_1 p_2$. Explicit calculation gives the relations

$$: q_1^2 : G(\mu_1 \mu_2 \mu_3; \nu_1 \nu_2 \nu_3) = 2\sqrt{\mu_1(\nu_1 + 1)} G(\mu_1 - 1, \mu_2, \mu_3; \nu_1 + 1, \nu_2, \nu_3), \quad (7.3.2)$$

$$: p_1^2 : G(\mu_1 \mu_2 \mu_3; \nu_1 \nu_2 \nu_3) = -2\sqrt{\nu_1(\mu_1 + 1)} G(\mu_1 + 1, \mu_2, \mu_3; \nu_1 - 1, \nu_2, \nu_3), \quad (7.3.3)$$

$$\begin{aligned} : q_1 q_2 : G(\mu_1 \mu_2 \mu_3; \nu_1 \nu_2 \nu_3) &= \sqrt{\mu_1(\nu_2 + 1)} G(\mu_1 - 1, \mu_2, \mu_3; \nu_1, \nu_2 + 1, \nu_3) \\ &\quad + \sqrt{\mu_2(\nu_1 + 1)} G(\mu_1, \mu_2 - 1, \mu_3; \nu_1 + 1, \nu_2, \nu_3), \end{aligned} \quad (7.3.4)$$

$$\begin{aligned} : p_1 p_2 : G(\mu_1 \mu_2 \mu_3; \nu_1 \nu_2 \nu_3) &= -\sqrt{\nu_1(\mu_2 + 1)} G(\mu_1, \mu_2 + 1, \mu_3; \nu_1 - 1, \nu_2, \nu_3) \\ &\quad - \sqrt{\nu_2(\mu_1 + 1)} G(\mu_1 + 1, \mu_2, \mu_3; \nu_1, \nu_2 - 1, \nu_3), \end{aligned} \quad (7.3.5)$$

$$\begin{aligned} : q_1 p_2 : G(\mu_1 \mu_2 \mu_3; \nu_1 \nu_2 \nu_3) &= \sqrt{\mu_1(\mu_2 + 1)} G(\mu_1 - 1, \mu_2 + 1, \mu_3; \nu_1 \nu_2 \nu_3) \\ &\quad - \sqrt{\nu_2(\nu_1 + 1)} G(\mu_1 \mu_2 \mu_3; \nu_1 + 1, \nu_2 - 1, \nu_3). \end{aligned} \quad (7.3.6)$$

$$: q_1 p_1 : G(\mu_1 \mu_2 \mu_3; \nu_1 \nu_2 \nu_3) = (\mu_1 - \nu_1) G(\mu_1 \mu_2 \mu_3; \nu_1 \nu_2 \nu_3). \quad (7.3.7)$$

With this detour behind us, define a scalar product among the basis elements $G(\mu; \nu)$ by the rule

$$\langle G(\mu'; \nu'), G(\mu; \nu) \rangle = \delta_{\mu' \mu} \delta_{\nu' \nu}. \quad (7.3.8)$$

Here we use the short-hand notation

$$\delta_{\mu' \mu} = \delta_{\mu'_1 \mu_1} \delta_{\mu'_2 \mu_2} \delta_{\mu'_3 \mu_3}, \text{ etc.} \quad (7.3.9)$$

That is, the basis elements are defined to be an orthonormal set. Note that although the notation is the same, this scalar product is not to be confused with that introduced for phase-space vectors in Section 3.5.

It is easily verified that the rule (3.8) induces a positive-definite scalar product among the set of all phase-space functions. Let f and g be any two (possibly complex) functions. Make the expansions

$$f = \sum_{\mu\nu} f_{\mu\nu} p_1^{\mu_1} q_1^{\nu_1} p_2^{\mu_2} q_2^{\nu_2} p_3^{\mu_3} q_3^{\nu_3}, \quad (7.3.10)$$

$$g = \sum_{\mu\nu} g_{\mu\nu} p_1^{\mu_1} q_1^{\nu_1} p_2^{\mu_2} q_2^{\nu_2} p_3^{\mu_3} q_3^{\nu_3}. \quad (7.3.11)$$

Then we have the relation

$$\langle f, g \rangle = \sum_{\mu\nu} \bar{f}_{\mu\nu} g_{\mu\nu} \mu_1! \nu_1! \mu_2! \nu_2! \mu_3! \nu_3!. \quad (7.3.12)$$

Examination of (3.12) shows that there is another equivalent way of defining the scalar product. Let ∂_z denote the set of partial differentiation operators, $\partial_z = (\partial/\partial z_1, \partial/\partial z_2, \dots, \partial/\partial z_6)$. Then we also have the result

$$\langle f, g \rangle = \bar{f}(\partial_z) g(z)|_{z=0} = g(\partial_z) \bar{f}(z)|_{z=0}. \quad (7.3.13)$$

There is a corollary that will be of later use. Let h be any phase-space function. Then, from (3.13), we find the result

$$\langle h f, g \rangle = \langle f, \bar{h}(\partial_z) g \rangle. \quad (7.3.14)$$

We close this subsection with the remark that in the definition of the scalar product given by (3.1) and (3.8) it was convenient to treat the q 's and p 's separately. For a somewhat different notation that treats them on the same footing, see Exercise 3.23.

7.3.2 Definition of Hermitian Conjugate

Given a scalar product and any linear operator O , the Hermitian conjugate O^\dagger is defined by the relation

$$\langle f, O^\dagger g \rangle = \langle O f, g \rangle. \quad (7.3.15)$$

The virtue of the scalar product (3.8) is that the Lie operators associated with quadratic polynomials have particularly simple Hermitian conjugates. From the relations (3.2) through (3.7) and their generalizations to all $z_a z_b$ pairs, and the definition (3.15), we find the pleasing results

$$: q_j q_k :^\dagger = - : p_j p_k :, \quad (7.3.16)$$

$$: p_j p_k :^\dagger = - : q_j q_k :, \quad (7.3.17)$$

$$: q_j p_k :^\dagger = : q_k p_j :. \quad (7.3.18)$$

Indeed, let \mathcal{L}_{ab} be any vector field of the form

$$\mathcal{L}_{ab} = z_a (\partial/\partial z_b). \quad (7.3.19)$$

See Section 5.3. Then we find the result

$$(\mathcal{L}_{ab})^\dagger = \mathcal{L}_{ba}. \quad (7.3.20)$$

Consider the quadratic polynomials b^0 through b^8 defined by (5.8.5). It is easily verified from their definitions, and the relations (3.16) through (3.18), that their associated Lie operators are anti-Hermitian,

$$: b^j :^\dagger = - : b^j :. \quad (7.3.21)$$

Also, from (5.8.50), (5.8.51), and (3.21), we have the relations

$$: c^j :^\dagger =: c^j :, \quad (7.3.22)$$

$$: r(\boldsymbol{\mu}) :^\dagger =: r(-\boldsymbol{\mu}) :. \quad (7.3.23)$$

Thus the $: c^j :$ are Hermitian as desired for the construction of a representation theory. Finally, the Lie operators associated with the quadratic polynomials f^j and g^j given by (5.8.63) are Hermitian,

$$: f^j :^\dagger =: f^j :, \quad (7.3.24)$$

$$: g^j :^\dagger =: g^j :. \quad (7.3.25)$$

Suppose f_2^c is a real quadratic polynomial defined in terms of a real matrix S^c as in (2.8). Then, in the case of a 6-dimensional phase space, such an f_2^c can be written as a linear combination of the polynomials b^0 through b^8 , with real coefficients. Correspondingly, the Lie operator associated with f_2^c is anti-Hermitian,

$$: f_2^c :^\dagger = - : f_2^c :. \quad (7.3.26)$$

Let \mathcal{M} be the symplectic map associated with f_2^c ,

$$\mathcal{M} = \exp(: f_2^c :). \quad (7.3.27)$$

Then, from (3.26), we find the result

$$\mathcal{M}^\dagger = \exp(: f_2^c :^\dagger) = \exp(- : f_2^c :) = \mathcal{M}^{-1}. \quad (7.3.28)$$

It follows that \mathcal{M} is *unitary* with respect to the scalar product (3.12). That is, we have the relation

$$\langle \mathcal{M}f, \mathcal{M}g \rangle = \langle f, \mathcal{M}^\dagger \mathcal{M}g \rangle = \langle f, g \rangle. \quad (7.3.29)$$

We already know that Lie transformations of the form (3.27) are a realization of the group $U(3)$. From this perspective, the relation (3.28) indicates that the scalar product defined by (3.12) is *invariant* under $U(3)$.

Remarkably, the scalar product (3.12) is in fact invariant under the full group $USp(6)$. Suppose f_2^a is a real quadratic polynomial defined in terms of a real matrix S^a as in (2.3). Then it is easily verified from (5.8.43), (3.24), and (3.25) that the Lie operator $: f_2^a :$ is Hermitian,

$$: f_2^a :^\dagger =: f_2^a :. \quad (7.3.30)$$

Let \mathcal{M} be any (complex) symplectic map of the form

$$\mathcal{M} = \exp(: f_2^c :)\exp(i : f_2^a :). \quad (7.3.31)$$

Then, from (3.26) and (3.30), we find the result

$$\begin{aligned} \mathcal{M}^\dagger &= \exp(-i : f_2^a :^\dagger)\exp(: f_2^c :^\dagger) \\ &= \exp(-i : f_2^a :)\exp(- : f_2^c :) \\ &= \mathcal{M}^{-1}. \end{aligned} \quad (7.3.32)$$

It follows as before that \mathcal{M} is unitary with respect to the scalar product (3.12). Moreover, we know from Section 5.10 and (2.10) that symplectic maps of the form (3.31) are a realization of the group $USp(6)$. Thus, the scalar product (3.12) is invariant under $USp(6)$.

One might wonder if it is possible to define a scalar product that would be invariant under all of the *real* symplectic group $Sp(6, \mathbb{R})$. The answer is no. Let \mathcal{M} be the symplectic map associated with the monomial $\lambda q_1 p_1$ with λ real,

$$\mathcal{M} = \exp(\lambda : q_1 p_1 :). \quad (7.3.33)$$

From (3.7) we have the result

$$\mathcal{M}G(\mu; \nu) = \exp[\lambda(\mu_1 - \nu_1)]G(\mu; \nu). \quad (7.3.34)$$

It follows that for any definition of the scalar product, there is the relation

$$\langle \mathcal{M}G(\mu; \nu), \mathcal{M}G(\mu; \nu) \rangle = \exp[2\lambda(\mu_1 - \nu_1)]\langle G(\mu; \nu), G(\mu; \nu) \rangle. \quad (7.3.35)$$

We conclude that if the scalar product is such that the elements $G(\mu; \nu)$ are normalizable (have *finite* norm), then this scalar product cannot be invariant under all of $Sp(6, \mathbb{R})$. What we are observing here is a consequence of the fact that the group $Sp(6, \mathbb{R})$ is not compact. It can be shown that a noncompact group cannot have finite-dimensional *unitary* representations. Note that any \mathcal{M} of the form (2.10), when acting on a homogeneous polynomial, preserves the degree of that polynomial. See Lemma 7.6.3. Consequently, any realization of $Sp(6, \mathbb{R})$ associated with a polynomial basis must be finite dimensional, and thus, by the comment above, cannot be unitary. Finally, we remark that $U(3)$ is the largest compact subgroup of $Sp(6, \mathbb{R})$, and therefore is the largest subgroup for which we can hope to obtain finite-dimensional unitary representations.

While the relations (3.2) through (3.7) are fresh in the mind, we take this opportunity to observe that any Lie operator of the form $: f_2 :$ is traceless. Correspondingly, according to (3.7.56), any map \mathcal{M} of the form

$$\mathcal{M} = \exp(: f_2 :) \quad (7.3.36)$$

has determinant +1. These statements may seem somewhat surprising since both $: f_2 :$ and \mathcal{M} given by (3.36) may be viewed as infinite dimensional matrices in the sense that they are both linear operators that act on infinite dimensional vector spaces. We therefore have to be more precise.

We have already noted that any \mathcal{M} of the form (2.10) and therefore also of the form (3.36), when acting on a homogeneous polynomial, preserves the degree of that polynomial. We capitalize on this fact by slightly changing the notation for the monomials $G(\mu, \nu)$ defined by (3.1). Specifically, we denote the same monomials by the symbols G_r^m where m now denotes the degree of the monomial, and r is some index that labels the various possibilities for the exponents μ_i and ν_j subject to their sum being equal to m ,

$$\sum (\mu_i + \nu_i) = m, \quad (7.3.37)$$

$$r = \{\mu_i, \nu_j\}. \quad (7.3.38)$$

With this notation, the scalar product (3.8) takes the form

$$\langle G_{r'}^{m'}, G_r^m \rangle = \delta_{m'm} \delta_{r'r}. \quad (7.3.39)$$

Let us pause for a moment to make the selection of the index r more specific. Consider all monomials of degree m in d variables. Let $N(m, d)$ be the number of such monomials. Combinatorial considerations (see Exercises 3.9 through 3.13) show that N is given by the relation

$$N(m, d) = \binom{m+d-1}{m} = \frac{(m+d-1)!}{m!(d-1)!}. \quad (7.3.40)$$

Table 3.1 below shows values of $N(m, d)$ for various values of m for the case of 6 dimensional phase space ($d = 6$), and for other values of d that may be of interest later. Consequently, for $d = 6$ and each value of m , we may take for the index r the integers running from $r = 1$ through $r = N(m, 6)$. More sophisticated indexing schemes are described in Section 27.2.

7.3.3 Matrices Associated with Quadratic Lie Generators

We now return to our main discussion. As is evident from (3.2) through (3.7), any quantity of the form $: f_2 : G_r^m$ must be a homogeneous polynomial of degree m . See also Lemma 7.6.3. Thus, we must have a result of the form

$$: f_2 : G_r^m = \sum_{r'} F_{r'r}^m G_{r'}^m \quad (7.3.41)$$

where the $F_{r'r}^m$ are coefficients yet to be determined. In fact, from (3.39) and (3.41) we have the result

$$F_{r'r}^m = \langle G_{r'}^m, : f_2 : G_r^m \rangle. \quad (7.3.42)$$

Let \mathcal{P}_m denote the space of all homogeneous polynomials of degree m . We know its dimension is $N(m, 6)$. What we have made explicit is that the general $: f_2 :$ sends \mathcal{P}_m into itself. Indeed, the action of $: f_2 :$ on \mathcal{P}_m for each value of m is described by the $N(m, 6) \times N(m, 6)$ matrix F^m given by (3.41) and (3.42). Let us compute the matrices corresponding to powers of $: f_2 :$. From (3.41) we find the result

$$\begin{aligned} : f_2 :^2 G_r^m &= \sum_{r'} F_{r'r}^m : f_2 : G_{r'}^m = \sum_{r'r''} F_{r'r}^m F_{r''r'}^m G_{r''}^m \\ &= \sum_{r''} \left(\sum_{r'} F_{r''r'}^m F_{r'r}^m \right) G_{r''}^m = \sum_{r''} [(F^m)^2]_{r''r'} G_{r''}^m. \end{aligned} \quad (7.3.43)$$

Table 7.3.1: Number of monomials of degree m in various numbers of variables.

m	$N(m, 4)$	$N(m, 5)$	$N(m, 6)$	$N(m, 7)$	$N(m, 8)$	$N(m, 9)$	$N(m, 10)$	$N(m, 11)$
0	1	1	1	1	1	1	1	1
1	4	5	6	7	8	9	10	11
2	10	15	21	28	36	45	55	66
3	20	35	56	84	120	165	220	286
4	35	70	126	210	330	495	715	1001
5	56	126	252	462	792	1287	2002	3003
6	84	210	462	924	1716	3003	5005	8008
7	120	330	792	1716	3432	6435	11440	19448
8	165	495	1287	3003	6435	12870	24310	43758
9	220	715	2002	5005	11440	24310	48620	92378
10	286	1001	3003	8008	19448	43758	92378	184756
11	364	1365	4368	12376	31824	75582	167960	352716
12	455	1820	6188	18564	50388	125970	293930	646646

It follows that the matrix corresponding to the action of $: f_2 :^2$ on \mathcal{P}_m is $(F^m)^2$. Similarly, the matrix corresponding to the action of $: f_2 :^\ell$ on \mathcal{P}_m is $(F^m)^\ell$. Correspondingly, it follows that the action of $\mathcal{M} = \exp(: f_2 :)$ on \mathcal{P}_m is given by a relation of the form

$$\mathcal{M}G_r^m = \sum_{r'} M_{r'r}^m G_{r'}^m, \quad (7.3.44)$$

and the matrices M^m are related to the F^m by the equations

$$M^m = \exp(F^m). \quad (7.3.45)$$

Let us now examine the form of F^m using (3.2) through (3.7) and (3.42). We see from (3.2) through (3.6) and (3.42) that the matrices F^m associated with any $: f_2 :$ made from monomials of the form $q_i^2, p_i^2, q_i q_j, p_i p_j, q_i p_j$ have no diagonal entries. Thus, all such F^m must be traceless. The only monomials that can produce diagonal entries in the F^m are of the form $q_i p_i$. But, from (3.7), we see that these entries are either zero or occur in positive and negative pairs. For example, let α and β be any two positive integers. Then, referring

to (3.7), for the case $\mu_1 = \alpha$ and $\nu_1 = \beta$ there must also be the case $\mu_1 = \beta$ and $\nu_1 = \alpha$. We conclude that all the F^m must be traceless,

$$\text{tr } F^m = 0. \quad (7.3.46)$$

Correspondingly, the M^m given by (3.44) and (3.45) must have unit determinant,

$$\det M^m = 1. \quad (7.3.47)$$

We close this subsection with a final remark. The relations (3.26) and (3.30) show that $:f_2 :^\dagger$ is a Lie operator for any f_2 . This need not be the case for $:f_m :^\dagger$ with $m > 2$. See Exercise 3.21.

Exercises

7.3.1. Verify the relations (3.2) through (3.7).

7.3.2. Verify the relations (3.13) and (3.14), and (3.16) through (3.20).

7.3.3. Verify the relations (3.21) through (3.23).

7.3.4. Verify the relations (3.24) and (3.25).

7.3.5. Verify the relations (3.26) and (3.30).

7.3.6. Suppose that the scalar product (3.8) is generalized to the case of a $2n$ dimensional phase space in the obvious way. See Exercise 3.23. Show that the relations (3.26) and (3.30) still hold. It follows that this scalar product is invariant under $U(n)$ and $USp(2n)$. Suggestion: Let f_2 and f_2^* be quadratic polynomials defined by the equations

$$f_2 = -(1/2)(z, Sz), \quad (7.3.48)$$

$$f_2^*(z) = f_2(Jz) = -(1/2)(Jz, SJz). \quad (7.3.49)$$

Here S is a real symmetric matrix. Prove the relation

$$:f_2 :^\dagger = - :f_2^* :. \quad (7.3.50)$$

7.3.7. Show that f_2 and f_2^* as defined by (3.48) and (3.49) are connected by the relation

$$\exp[-(\pi/2) : b^0 :] f_2 = f_2^*. \quad (7.3.51)$$

7.3.8. Verify the scalar product relations

$$\langle z_a z_b, z_c z_d \rangle = \delta_{ac} \delta_{bd} + \delta_{ad} \delta_{bc}. \quad (7.3.52)$$

For two *quadratic* polynomials f and g written in the forms (5.5.1) and (5.5.2), verify that

$$\langle f, g \rangle = (1/2) \text{tr} (S^f S^g). \quad (7.3.53)$$

7.3.9. Review Section 5.9.3 and Exercise 3.8 above. Let

$$\mathcal{R} = \exp(: f_2^c :) \quad (7.3.54)$$

be the map corresponding to the matrix R given by (5.9.28). Show that all of the $U(n)$ subgroup is covered by elements of the form (3.54) when

$$\langle f_2^c, f_2^c \rangle = (1/2)\text{tr}[(S^c)^2] \leq n\pi^2. \quad (7.3.55)$$

7.3.10. Consider the Lie algebra $sp(2)$. Show that b^0 as given by (5.6.6) has a squared norm of 1. That is, $\langle b^0, b^0 \rangle = 1$. Show that f and g given by (5.6.11) and (5.6.12) also have a squared norm of 1. Show that all these $sp(2)$ elements are mutually orthogonal. Consider the Lie algebra $sp(4)$. Show that b^0 through b^3 as given by (5.7.4) have a squared norm of 2. Show that the f^j and the g^j given by (5.7.30) and (5.7.31) also have a squared norm of 2. Show that all these $sp(4)$ elements are mutually orthogonal. Consider the Lie algebra $sp(6)$. Show that b^0 as given by (5.8.5) has a squared norm of 3. Show that b^1 through b^8 as given by (5.8.5) have a squared norm of 2. Show that h^1, h^3 , and h^5 as given by (5.8.37) have a squared norm of 2. Show that h^2, h^4 , and h^6 have a squared norm of 4. Carry out an analogous computation for the $sp(6)$ elements \bar{h}^j . Show that all these $sp(6)$ elements are mutually orthogonal. The group-theoretical reason for this orthogonality is that these $sp(6)$ elements either have different $su(3)$ weights or belong to different $su(3)$ representations. Show that all f_2^a are orthogonal to all f_2^c . For some purposes we may wish to renormalize the $sp(6)$ elements so that they all have the same norm. As shown in Chapter 27, in a suitable Cartan basis all the basis elements are orthonormal.

7.3.11. Prove (3.40). Hint: First show that the binomial coefficients obey the recursion relations

$$\binom{n+1}{m} = \binom{n}{m} + \binom{n}{m-1}, \quad (7.3.56)$$

and hence

$$\binom{n+1}{m} = \binom{n}{m} + \binom{n-1}{m-1} + \cdots + \binom{n-m}{0}, \quad (7.3.57)$$

and the identity

$$\binom{n}{m} = \binom{n}{n-m}. \quad (7.3.58)$$

Next, by definition, $N(m, d)$ is the number of monomials of degree m in d variables. Verify the relations

$$N(m, 1) = 1, \quad (7.3.59)$$

$$N(m, 2) = m + 1. \quad (7.3.60)$$

In fact, show that $N(m, d)$ satisfies the recursion relation

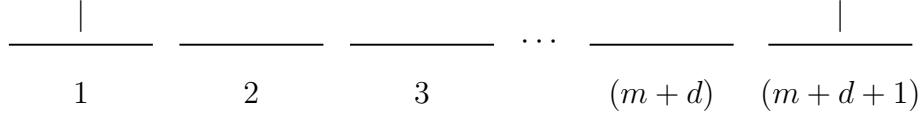
$$N(m, d+1) = N(m, d) + N(m-1, d) + N(m-2, d) + \cdots + N(0, d) = \sum_{j=0}^m N(j, d). \quad (7.3.61)$$

(Let z_{d+1} be the extra variable that is added to pass from d to $d + 1$ variables. Then the monomials of degree m in $d + 1$ variables may be partitioned into subsets that contain z_{d+1} to the zero power, z_{d+1} to the first power, z_{d+1} to the second power, etc.) Finally, show that $N(m, d)$ as given by (3.40) satisfies the recursion relation (3.59) and the initial condition (3.57), and verify that these facts specify $N(m, d)$ uniquely.

7.3.12. The fact that, according to (3.40), $N(m, d)$ is given simply by a binomial coefficient suggests that there should be a simple proof of this result. There is: Given n things, the number of combinations of these n things taken ℓ at a time is specified by the binomial coefficient

$${}_nC_\ell = C[n, \ell] = n \text{ choose } \ell = \binom{n}{\ell} = \frac{n!}{\ell!(n-\ell)!}. \quad (7.3.62)$$

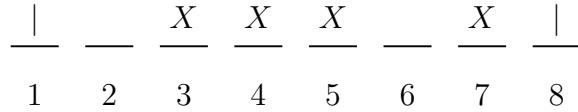
On a sheet of paper lay out $(m + d + 1)$ spaces as shown below, and number them from 1 to $(m + d + 1)$. Place a vertical bar “|” in the first and last spaces.



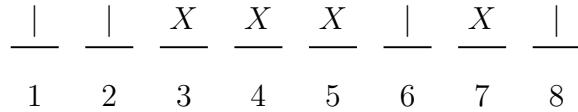
After this construction $(m+d-1)$ empty spaces remain. Select m of these spaces, and place an “ X ” in each. The number of ways of doing this is given by the binomial coefficient

$$C[(m+d-1), m] = \binom{m+d-1}{m}. \quad (7.3.63)$$

For example, suppose $m = 4$ and $d = 3$. Then one way of placing the X ’s is shown below.



Next, put vertical bars in the remaining empty spaces. Doing so for the $m = 4$ and $d = 3$ example just cited yields this picture.



Evidently these are $(d + 1)$ vertical bars after this is done. Now regard the $(d + 1)$ vertical bars as representing the walls of d cells, and count the number of X ’s in each cell. In the case shown above, and proceeding from left to right, these counts are 0,3,1, respectively. Consider the monomial $z_1^{j_1} z_2^{j_2} \cdots z_d^{j_d}$ subject to the homogeneity condition

$$j_1 + j_2 + \cdots + j_d = m. \quad (7.3.64)$$

We may regard the numbers j_1, j_2, \dots, j_d as possible cell counts for the cells 1, 2, \dots , d since our construction automatically satisfies (3.62). We now see that (3.58) is the number of ways of selecting d non-negative integers j_1, j_2, \dots, j_d such that (3.62) is satisfied. It follows that (3.40) is correct. Verify the argument just given for several examples.

7.3.13. There is another way to derive (3.40). Define a *composition* of m into d parts to be a representation of the form

$$m = j_1 + j_2 + \cdots + j_d, \quad (7.3.65)$$

where j_1, j_2, \dots, j_d are d non-negative integers, and the order of the summands is significant. Evidently $N(m, d)$ is the number of compositions for a given m and d . Suppose we have several power series, and we want to find their product. How can we calculate the coefficients of the product series? Consider, for example, the product

$$(\sum a_i x^i)(\sum b_j x^j)(\sum c_k x^k) = \sum d_\ell x^\ell. \quad (7.3.66)$$

Then we have the relation

$$d_\ell = \sum_{i+j+k=\ell} a_i b_j c_k. \quad (7.3.67)$$

On the right side of (3.65) there will be a term for each composition of ℓ into 3 parts. Consider, by inspiration, the power series

$$f(x) = 1 + x + x^2 + \cdots. \quad (7.3.68)$$

Verify the relation

$$[f(x)]^d = \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} \cdots \sum_{j_d=0}^{\infty} x^{j_1+j_2+\cdots+j_d}. \quad (7.3.69)$$

Collect terms with the same exponent, and show that the exponent x^m occurs $N(m, d)$ times. Thus, verify the relation

$$[f(x)]^d = \sum_{\ell=0}^{\infty} N(\ell, d) x^\ell. \quad (7.3.70)$$

You have shown that the functions $g(x; d)$ defined by the relations

$$g(x; d) = [f(x)]^d \quad (7.3.71)$$

are the *generating functions* (for each value of d) for the quantities $N(\ell, d)$. Next verify the relations

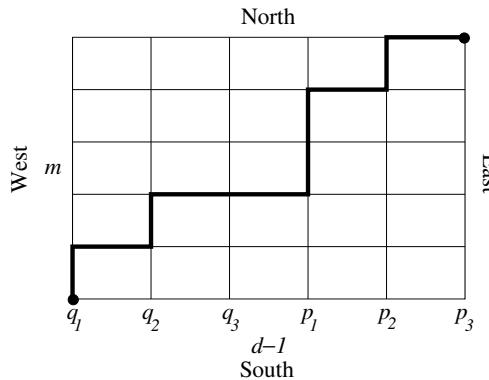
$$[f(x)]^d = (1 - x)^{-d}, \quad (7.3.72)$$

and, by the binomial theorem,

$$(1 - x)^{-d} = \sum_{\ell=0}^{\infty} \binom{\ell + d - 1}{\ell} x^\ell. \quad (7.3.73)$$

Now compare (3.68) through (3.71) to prove (3.40). We remark that apparently *de Moivre* was the first person to view collections of numbers or functions as coefficients in the power series of some master function. Laplace subsequently championed the use of such master functions, for which he coined the term *generating functions*.

7.3.14. Here is yet another way to derive (3.40). Suppose a walker in Manhattan wants to go from **A** to **B** (see the picture below where **A** is taken to be at the lower left corner and **B** at the upper right corner). While walking, he is thinking of the problem of finding the number of all the distinct monomials $N(m, d)$ of degree m in d variables. He soon realizes that $N(m, d)$ equals the number of different paths he can walk through from **A** to **B**¹, provided that the number of blocks in the East-West direction is $d - 1$ and the number of blocks in the South-North direction is m (in the picture $m = 5$, $d = 6$). He can easily associate a path with a monomial. First he labels each street going North with a variable name. Before leaving **A** he sets to zero the exponents of all the variables of a monomial. Then he increases by one the exponent of any variable in the monomial each time he goes North by one block along the corresponding street. The monomial he is left with when he reaches **B** is the one associated with the particular path he has gone through. For example, the path shown in the picture represents the monomial $q_1 q_2 p_1^2 p_2$.



How many different paths can he walk through? In his way from **A** to **B** he has to decide if he wants to go East or North at the corner of each block. He has to take $m + (d - 1)$ decisions. The only constraint is that, overall, he needs to choose to go North (N) m times, and go East (E) $d - 1$ times. With each path we can associate a sequence of N's and E's. The following table represents the path shown in the picture.

N	E	N	E	E	N	N	E	N	E
---	---	---	---	---	---	---	---	---	---

The problem is equivalent to finding the number of all possible rearrangements of such sequences. Each sequence has $m + (d - 1)$ slots; the symbol N has to appear m times, while the symbol E appears $d - 1$ times. Therefore the number of all the possible rearrangements is given by the relation

$$N(m, d) = \frac{[m + (d - 1)]!}{m!(d - 1)!}.$$

7.3.15. Show from (3.40) or (3.59) that $N(m, d)$ can be generated by the recursion relation

$$N(m, d) = N(m, d - 1) + N(m - 1, d) \quad (7.3.74)$$

with the boundary conditions

$$N(m, 1) = 1, \quad (7.3.75)$$

¹Only those paths that minimize the walking distance count. He never walks south or west.

$$N(0, d) = 1 \text{ or } N(1, d) = d. \quad (7.3.76)$$

7.3.16. From (3.37) with $d = 6$ we find the results $N(0, 6) = 1$, $N(1, 6) = 6$, $N(2, 6) = 21$, $N(3, 6) = 56$, etc. See Table 3.1. Equations (5.8.24) and (5.8.28) describe how homogeneous polynomials f_ℓ of degree ℓ in 6 variables can be decomposed into irreducible representations of $su(3)$. Equation (5.8.18) gives the dimension of these representations. Show by explicit calculation that the dimension counts for both sides of (5.8.24) match for the cases $\ell = 0, 2, 4$. Do the same for (5.8.28) for the cases $\ell = 1$ and 3. Can you show that the dimension counts agree for general ℓ ?

7.3.17. Verify that the matrix corresponding to the action of $:f_2:\ell$ on \mathcal{P}_m is $(F^m)^\ell$. Derive (3.45) from (3.36), (3.41), and (3.44).

7.3.18. Show that the matrices JS^a , JS^c and P, O given by (2.4), (2.7) etc. are special cases of the matrices F^m and M^m , respectively. What is m in this case?

7.3.19. Strictly speaking, what has been shown in the text is that all matrices M^m arising from the \mathcal{M} given by (3.36) must satisfy (3.47). Show that (3.47) also holds for matrices M^m arising from \mathcal{M} of the form (2.10).

7.3.20. Show that the matrices F^m associated with Lie operators of the form $:f_2^a:$ are real and symmetric. Show that the matrices F^m associated with Lie operators of the form $:f_2^c:$ are real and antisymmetric.

7.3.21. The dimension of $sp(2n)$ is given by (3.7.35). Compare this dimension with $N(2, 2n)$ as given by (3.40), and explain why these two numbers must agree.

7.3.22. Let q, p be coordinates in a two-dimensional phase space. Show that $:q^3:\dagger$ is not a derivation, and therefore not a vector field. Hint: Evaluate its action on q and q^2 .

7.3.23. The purpose of this exercise is to introduce a somewhat different notation for the scalar product of Subsection 3.1 with the aim of treating the q 's and p 's more democratically. For monomials introduce the notation

$$z^k = z_1^{k_1} z_2^{k_2} \cdots z_{2n}^{k_{2n}}, \quad (7.3.77)$$

$$\delta_{kk'} = \delta_{k_1 k'_1} \delta_{k_2 k'_2} \cdots \delta_{k_{2n} k'_{2n}}, \quad (7.3.78)$$

$$k! = k_1! k_2! \cdots k_{2n}!. \quad (7.3.79)$$

In terms of this notation, show that the scalar product of Subsection 3.1 is given by the relation

$$\langle z^k, z^{k'} \rangle = \delta_{kk'} k!. \quad (7.3.80)$$

Show that the scalar product based on (3.80) is positive definite. That is, verify that

$$\langle f, f \rangle = 0 \quad (7.3.81)$$

when $f = 0$, and

$$\langle f, f \rangle > 0 \quad (7.3.82)$$

otherwise.

7.4 Symplectic Map for Flow of Time-Independent Hamiltonian

Suppose the Hamiltonian of Theorem 6.4.1 does not explicitly depend on the time. [In this case, we say that the differential equations (1.5.6) generated by the Hamiltonian H are *autonomous*.] Then the symplectic map (6.4.1) obtained by following the Hamiltonian flow specified by H can be written immediately in the form

$$\mathcal{M} = \exp\{-(t^f - t^i) : H :\}. \quad (7.4.1)$$

That is, we have the relation

$$z^f = \mathcal{M}z^i. \quad (7.4.2)$$

To verify (4.1) and (4.2), let \mathcal{M} act on z^i to give the result

$$z^f = \mathcal{M}z^i = \sum_{m=0}^{\infty} (1/m!)(t^f - t^i)^m : -H :^m z^i. \quad (7.4.3)$$

However, Taylor's theorem gives the result

$$z^f = z(t^f) = z(t^i) + \sum_{m=1}^{\infty} (1/m!)(t^f - t^i)^m (d/dt)^m z(t)|_{t^i}. \quad (7.4.4)$$

Also, Hamilton's equations of motion for the z 's can be written in the form

$$\begin{aligned} \dot{z} &= [z, H] = [-H, z] =: -H : z, \\ \ddot{z} &= [-H, \dot{z}] =: -H : \dot{z} =: -H :^2 z, \\ (d^3 z)/(dt)^3 &=: -H :^3 z, \text{ etc.} \end{aligned} \quad (7.4.5)$$

Upon inserting the results of (4.5) into (4.4), we obtain the desired result (4.3).

At this point several observations are possible and in order. Suppose we replace the final time t^f by a general time t . Then (4.1) and (4.2) can be written in the form

$$z(t) = \mathcal{M}(t)z^i, \quad (7.4.6)$$

with

$$\mathcal{M}(t) = \exp\{(t - t^i) : -H :\}. \quad (7.4.7)$$

Note that here the Hamiltonian H depends on the initial conditions z^i , and does not depend explicitly on the time,

$$H = H(z^i). \quad (7.4.8)$$

Suppose that (4.7) is differentiated with respect to the time. Doing so gives, in accord with Appendix C, the result

$$\begin{aligned} \dot{\mathcal{M}} &= \exp\{(t - t^i) : -H :\} : -H : \\ &= \mathcal{M} : -H :. \end{aligned} \quad (7.4.9)$$

The relation (4.9) provides an equation of motion for \mathcal{M} that, although only derived so far for the case of a time independent Hamiltonian, will eventually be shown to hold in general. Also, \mathcal{M} evidently satisfies the initial condition

$$\mathcal{M}(t^i) = \mathcal{I}. \quad (7.4.10)$$

It will eventually be shown that the equation of motion (4.9) and the initial condition (4.10) specify $\mathcal{M}(t)$ completely. Conversely, if $\mathcal{M}(t)$ is known, the Hamiltonian H can be found, up to an immaterial constant, from the relation

$$: -H := \mathcal{M}^{-1} \dot{\mathcal{M}}. \quad (7.4.11)$$

See (5.3.15) and (5.3.16).

Next, suppose we differentiate (4.6) with respect to the time. Doing so and making use of (4.9) gives the result

$$\begin{aligned} \dot{z}(t) &= \dot{\mathcal{M}}(t) z^i = \mathcal{M} : -H : z^i \\ &= \mathcal{M}[-H, z^i]_{z^i} = \mathcal{M}[z^i, H]_{z^i}. \end{aligned} \quad (7.4.12)$$

The right side of (4.12) can be manipulated further using the relations (5.4.15) and (5.4.11) to give the result

$$\begin{aligned} \mathcal{M}[z^i, H]_{z^i} &= [\mathcal{M}z^i, \mathcal{M}H]_{z^i} = [\mathcal{M}z^i, H(\mathcal{M}z^i)]_{z^i} \\ &= [z(t), H(z(t))]_{z^i}. \end{aligned} \quad (7.4.13)$$

Also, we should really write $z(t)$ in the more explicit form

$$z = z(z^i, t) \quad (7.4.14)$$

to indicate that $z(t)$ depends on the initial conditions z^i . Finally, because the mapping between z^i and z is symplectic, the Poisson bracket on the right side of (4.13) can be rewritten to give the result

$$\begin{aligned} [z(t), H(z(t))]_{z^i} &= [z(z^i, t), H(z(z^i, t))]_{z^i} \\ &= [z, H(z)]_z. \end{aligned} \quad (7.4.15)$$

See (6.3.3) and (6.3.10) and let z^i play the role of \bar{z} . Putting all these results together, we find the final expected relation

$$\dot{z} = [z, H(z)]_z. \quad (7.4.16)$$

As a generalization of (4.1), suppose that the Hamiltonian H is not necessarily time independent, but does have the property that the Lie operators $: H(z, t) :$ for various times all commute. That is, one has the relation

$$\{ : H(z, t) :, : H(z, t') : \} = 0 \text{ for all } t, t'. \quad (7.4.17)$$

Alternatively, because of the homomorphism (5.3.14), we may require that $H(z, t)$ and $H(z, t')$ be in involution for all t, t' . It can be shown that in either case the symplectic map obtained by following the Hamiltonian flow specified by H can be written in the form

$$\mathcal{M} = \exp\left(- \int_{t^i}^{t^f} : H : dt\right). \quad (7.4.18)$$

See Section 10.3.

Exercises

7.4.1. Verify in detail the steps leading from (4.12) to (4.16).

7.4.2. Prove (4.18) given the assumption (4.17).

7.4.3. Consider the map $\mathcal{M}(t)$ given by (1.4.13). Find H using (4.11).

7.4.4. Consider the map

$$q(q^i, p^i, t) = q^i(1 - tp^i)^2, \quad (7.4.19)$$

$$p(q^i, p^i, t) = p^i/(1 - tp^i). \quad (7.4.20)$$

Verify that this map is symplectic. Find H using (4.11). Sketch the flow generated by H .

7.4.5. Use the result (1.4.13) to carry out Exercise 5.4.5 and to derive (2.23). Use the results (1.4.21), (1.4.22), (1.4.23), and (1.4.24) to carry out Exercises 5.4.1 through 5.4.4 and Exercise 5.4.6, and to derive (2.20).

7.4.6. Consider the Hamiltonian H given by

$$H = (p^2 + \sigma q^2)/2 \quad (7.4.21)$$

and the linear symplectic map \mathcal{M} generated by H ,

$$\mathcal{M}(\sigma, t) = \exp(-t : H :). \quad (7.4.22)$$

Let $M(\sigma, t)$ be the symplectic matrix associated with \mathcal{M} as in (7.2.1). Find M explicitly and show that, in accord with Poincaré's Theorem 3.3 given in Section 1.3, M is analytic in the variables σ and t . Write M in the form

$$M = \exp(JS) \quad (7.4.23)$$

and show that the Hamiltonian matrix (JS) is analytic in σ and t . Find the eigenvalues of M and (JS) and plot them in the complex plane as a function of σ . See Section 3.4 and Exercise 3.7.12. Show that the eigenvalues of M and (JS) have square-root branch points (singularities) at $\sigma = 0$.

7.4.7. Let H be a quadratic and time independent Hamiltonian, and write it in the form

$$H = (1/2)(z, Sz) \quad (7.4.24)$$

where S is a time independent symmetric matrix. It generates the linear symplectic map

$$\mathcal{M} = \exp(-t : H :) \quad (7.4.25)$$

Show that the symplectic matrix M associated with \mathcal{M} is given by the relation

$$M = \exp(tJS). \quad (7.4.26)$$

7.5 Taylor Maps and Jets

Let \mathcal{N} be a symplectic map, and suppose \mathcal{N} sends the particular point \tilde{z}^i to the point \tilde{z}^f . Consider points z near \tilde{z}^i by writing the relation

$$z = \tilde{z}^i + \zeta, \quad (7.5.1)$$

and define points \bar{z} near \tilde{z}^f by writing the relation

$$\bar{z} = \tilde{z}^f + \bar{\zeta}. \quad (7.5.2)$$

Then, by construction, we have the relation

$$\bar{\zeta} = 0 \text{ if } \zeta = 0. \quad (7.5.3)$$

Also, the mappings (5.1) and (5.2) are symplectic. See Exercise 6.2.2. It follows from the group property for symplectic maps that the mapping between ζ and $\bar{\zeta}$, call it \mathcal{M} , is also symplectic. We write the relation

$$\bar{\zeta} = \mathcal{M}\zeta, \quad (7.5.4)$$

and observe that according to (5.3), \mathcal{M} sends the origin into itself.

Suppose the map \mathcal{N} is *analytic* in z around the point \tilde{z}^i . Then \mathcal{M} will be analytic in ζ around the origin. Correspondingly, we may write a *Taylor* expansion of the form

$$\bar{\zeta}_a = \sum_b R_{ab}\zeta_b + \sum_{bc} T_{abc}\zeta_b\zeta_c + \sum_{bcd} U_{abcd}\zeta_b\zeta_c\zeta_d + \dots. \quad (7.5.5)$$

Note that the expansion has no constant terms due to (5.3). Expansions of the form (5.5) often are used both in magnetic particle optics and in light ray optics. In this context the coefficients R describe paraxial optics, and the remaining coefficients T , U , \dots describe aberration effects. For this reason we will refer to expansions of the form (5.5) either as *Taylor maps* or *aberration expansions*.

We have already seen in Section 6.4 that Hamiltonian flows produce symplectic maps. Also, according to Theorem 3.3.3, if the Hamiltonian has suitable analytic properties, which is often the case, then the symplectic map it produces will also be analytic. See Chapter 26 and Appendix F. Thus, *analytic* symplectic maps are of great interest. Without loss of generality, such maps may be taken to be of the form (5.5).

Finally, suppose $f(\zeta)$, a function of the phase-space variables ζ , is analytic at the origin. Suppose further that the Taylor expansion of f begins with quadratic terms. Then evidently the symplectic map given by the Lie transformation $\exp(: f :)$ is of the form (5.5).

By combining (5.1) through (5.5) we may write the relation

$$\bar{z} = \mathcal{N}z \quad (7.5.6)$$

in the form

$$\begin{aligned} \bar{z}_a &= \tilde{z}_a^f + \sum_b R_{ab}(z - \tilde{z}^i)_b \\ &\quad + \sum_{bc} T_{abc}(z - \tilde{z}^i)_b(z - \tilde{z}^i)_c \\ &\quad + \sum_{bcd} U_{abcd}(z - \tilde{z}^i)_b(z - \tilde{z}^i)_c(z - \tilde{z}^i)_d + \dots. \end{aligned} \quad (7.5.7)$$

Let $g_a(m; z)$ denote a homogeneous polynomial of degree m in the components of z . With this notation (5.7) can also be written in the form

$$\bar{z}_a = \sum_{m=0}^{\infty} g_a[m; (z - \tilde{z}^i)]. \quad (7.5.8)$$

Suppose \mathcal{N}' is some other symplectic map that sends \tilde{z}^i to \tilde{z}^f , and suppose \mathcal{N}' has an expansion of the form

$$\bar{z}_a = \sum_{m=0}^{\infty} g'_a[m; (z - \tilde{z}^i)]. \quad (7.5.9)$$

Since both \mathcal{N} and \mathcal{N}' send \tilde{z}^i to \tilde{z}^f , we have the relation

$$g_a(0; z) = g'_a(0; z). \quad (7.5.10)$$

Now suppose that in fact g_a and g'_a agree for $m \leq k$:

$$g_a(m; z) = g'_a(m; z) \text{ for } m \leq k. \quad (7.5.11)$$

We express this condition symbolically by writing

$$\mathcal{N}' \overset{k}{\sim} \mathcal{N} \quad (7.5.12)$$

and say that \mathcal{N}' and \mathcal{N} are *equivalent* through terms of degree k . It can be shown, as the notation and terminology are meant to suggest, that (5.12) defines an equivalence relation among maps that send \tilde{z}^i to \tilde{z}^f . See Exercise 5.2. With the aid of this equivalence relation, we may define equivalence classes of maps. For any given k , an equivalence class is called a *k -jet*. Evidently, an equivalence class (a k -jet) is determined by specifying the polynomials $h_a(0; z), h_a(1; z), \dots, h_a(k; z)$. Put another way, we may say that two maps \mathcal{N} and \mathcal{N}' represent the same k -jet if their derivatives agree at \tilde{z}_i through order k . Or, what amounts to the same thing, we may view a k -jet as being represented by a point \tilde{z}^i and a Taylor series map (about this point) truncated beyond terms of degree k .

Finally, suppose a Taylor map is a Taylor expansion of a symplectic map. We will refer to the jet obtained by truncating such an expansion as a *symplectic jet*. It is important to note that a symplectic k -jet is generally *not* a symplectic map, but rather is a k -jet that satisfies the symplectic condition through terms of degree $(k - 1)$. See Exercise 5.3.

Exercises

7.5.1. Suppose that $f(\zeta)$ is analytic at the origin and begins with quadratic terms. Show that the symplectic map given by $\exp(: f :)$ is of the form (5.5).

7.5.2. Show that (5.11) and (5.12) produce an equivalence relation on the set of differentiable maps. See Exercise 5.12.7.

7.5.3. Exercise about symplectic jets.

7.6 Factorization Theorem

Note what has been accomplished so far. Section 3.7 showed that matrices of the form JS with S symmetric produce a Lie algebra. It also showed that any symplectic matrix sufficiently near the identity can be written in the form $\exp(JS)$. Section 3.8 showed that any symplectic matrix can be written as the product of two exponentials. Similarly, Section 5.3 showed that the set of Lie operators $:f:$ forms a Lie algebra. And Section 7.1 showed that Lie transformations $\exp(:f:)$ are symplectic maps. Finally, we have just seen that if f is analytic at the origin and begins with quadratic terms, then the Lie transformation $\exp(:f:)$ produces a map of the form (5.5). What remains to be studied is the question of whether any symplectic map \mathcal{M} can be written in exponential form. The answer to this question is given by the *factorization theorem*.

Theorem 6.1 Let \mathcal{M} be an *analytic* symplectic map that sends the origin into itself. That is, the relation between z and \bar{z} is assumed to be expressible in a Taylor series of the form

$$\bar{z}_a = \sum_b R_{ab} z_b + \sum_{bc} T_{abc} z_b z_c + \sum_{bcd} U_{abcd} z_b z_c z_d + \dots \quad (7.6.1)$$

In the terminology of the last section, truncating this series beyond terms of degree k yields, for any k , a symplectic k -jet. (Here, to avoid proliferation of notation, we again use the symbols \bar{z} and z to denote general phase-space variables.) Then, remarkably, there are *unique* functions $f_2^c(z), f_2^a(z), f_3(z), f_4(z), \dots$ such that the relation (6.1) can be written in the form

$$\bar{z} = \mathcal{M}z, \quad (7.6.2)$$

with \mathcal{M} expressed as a product of Lie transformations in the form

$$\mathcal{M} = \exp(:f_2^c:) \exp(:f_2^a:) \exp(:f_3:) \exp(:f_4:) \dots \quad (7.6.3)$$

Furthermore, each of the functions $f_m(z)$ is a *homogeneous polynomial* of degree m in the variables z .

The proof of this theorem is best achieved in stages by verifying a series of lemmas.

Lemma 6.1 The matrix R of (6.1) is symplectic. To see this, compute the Jacobian matrix $M(z)$ corresponding to the transformation (6.1) using (6.1.2). We find the result

$$M(0) = R. \quad (7.6.4)$$

Then, since \mathcal{M} is assumed to be symplectic, $M(z)$ must be symplectic for all z , and hence R must be symplectic.

Lemma 6.2 Let $g_1(z) \dots g_{2n}(z)$ be a set of $2n$ functions. (Here, as usual, $2n$ is the dimensionality of the phase space in question.) Suppose these functions satisfy the relations

$$[z_a, g_b] = [z_b, g_a]. \quad (7.6.5)$$

Then such functions exist if and only if there is a function h such that

$$g_a = [h, z_a] =: h : z_a. \quad (7.6.6)$$

The function h is unique up to an additive constant.

To verify this assertion, first suppose that each g_a is given by (6.6). Then we quickly demonstrate (6.5). We find the result

$$\begin{aligned} [z_a, g_b] - [z_b, g_a] &= [z_a, [h, z_b]] - [z_b, [h, z_a]] \\ &= [h, [z_a, z_b]] = [h, J_{ab}] = 0. \end{aligned} \quad (7.6.7)$$

Here we have used the Jacobi identity (5.1.7).

Next, suppose (6.5) is true. We are now in a situation analogous to that of Section 6.4. Compare (6.4.30) and (6.5). As before, define functions η_c using (6.4.39),

$$\eta_c = \sum_d J_{dc} g_d, \quad (7.6.8)$$

and define an associated function h by the integral

$$h = - \int^z \sum_a \eta_a dz'_a. \quad (7.6.9)$$

As we have seen, the integral is path independent, and has the property

$$\partial h / \partial z_b = -\eta_b. \quad (7.6.10)$$

Following (6.4.31), we find that the Poisson bracket $[h, z_a]$ has the value

$$\begin{aligned} [h, z_a] &= -[z_a, h] = - \sum_b J_{ab} (\partial h / \partial z_b) \\ &= \sum_b J_{ab} \eta_b = \sum_{bc} J_{ab} J_{cb} g_c \\ &= \sum_{bc} J_{ab} (J^T)_{bc} g_c = \sum_c (JJ^T)_{ac} g_c = g_a, \end{aligned} \quad (7.6.11)$$

as desired. Here we have also used (6.8).

Lemma 6.3 Let f_m be a homogeneous polynomial in z of degree m . Also, let \mathcal{P}_r denote the set of all homogeneous polynomials of degree r . Then, for any two homogeneous polynomials f_m and f_n , we have the relation

$$[f_m, f_n] \in \mathcal{P}_{m+n-2}. \quad (7.6.12)$$

To put it another way, define a *degree* functional by the rule

$$\deg(f_m) = m. \quad (7.6.13)$$

Then we have the relation

$$\deg([f_m, f_n]) = m + n - 2. \quad (7.6.14)$$

This lemma is obviously true because the Poisson bracket operation simply involves two differentiations and multiplication.

We now have the necessary tools to prove Theorem 6.1. First consider the linear part of the transformation (6.1) that is described by the matrix R . Polar decomposition can be used to write R in the standard form

$$R = PO. \quad (7.6.15)$$

And we know from our earlier work that P is real symmetric, positive definite, and symplectic; and O is real orthogonal and symplectic. See (2.2) etc. where M plays the role of R . Let $f_2^a(z)$ and $f_2^c(z)$ be the polynomials associated with R using (2.3) and (2.8). Then, according to (2.11), (2.12), and (1.23), we have the result

$$\exp(- : f_2^a :) \exp(- : f_2^c :) z = R^{-1}z. \quad (7.6.16)$$

Suppose both sides of (6.1) are acted on by $\exp(- : f_2^a :) \exp(- : f_2^c :)$. Doing so, and using (6.16), gives the result

$$\exp(- : f_2^a :) \exp(- : f_2^c :) \bar{z}_b = z_b + r_b(> 1). \quad (7.6.17)$$

Here the notation $r_b(> m)$ denotes *any* “remainder” series consisting of terms of degree higher than m .

To proceed further, suppose the remainder terms $r_b(> 1)$ are decomposed into second degree terms $g_b(2; z)$ and higher degree terms by writing the relations

$$r_b(> 1) = g_b(2; z) + r_b(> 2). \quad (7.6.18)$$

With this notation, we may rewrite (6.17) in the form

$$\exp(- : f_2^a :) \exp(- : f_2^c :) \bar{z}_b = z_b + g_b(2; z) + r_b(> 2). \quad (7.6.19)$$

Take the Poisson bracket of both sides of (6.19) with themselves for different values of the index b . Doing so, and making use of (5.4.15), (5.4.16), and (6.12), gives the result

$$J_{bc} = J_{bc} + [z_b, g_c(2)] + [g_b(2), z_c] + r_{bc}(> 1). \quad (7.6.20)$$

Finally, upon equating terms of like degree in (6.20), we find the relation

$$[z_b, g_c(2)] + [g_b(2), z_c] = 0. \quad (7.6.21)$$

At this point the results of Lemma 6.2 come into play. According to this lemma, there is a function h such that g_a is given by (6.6). Indeed, we may use (6.9) to compute h explicitly. Inserting the definition (6.8) for the functions η_a into (6.9) gives the result

$$h = - \int^z \sum_{ab} g_a J_{ab} dz'_b. \quad (7.6.22)$$

Suppose we consider the case where each g_a is a homogeneous polynomial of degree m , call it $g_a(m, z)$, and suppose we denote by f_{m+1} the result of computing h in this case. Then the path integral is conveniently evaluated along the path

$$z'_b = \tau z_b, \quad (7.6.23)$$

where the parameter τ ranges from 0 to 1. Use of (6.22) and (6.23) gives the result

$$f_{m+1}(z) = -[1/(m+1)] \sum_{ab} g_a(m; z) J_{ab} z_b. \quad (7.6.24)$$

As the notation suggests, $f_{m+1}(z)$ is a homogeneous polynomial of degree $(m+1)$. In particular, use of (6.24) with the functions $g_b(2; z)$ produces the third-degree polynomial $f_3(z)$.

As a warm-up exercise for the next step, consider the effect of applying the Lie transformation $\exp(- : f_3 :)$ to z . We find the result

$$\exp(- : f_3 :) z_b = z_b + \underbrace{ : -f_3 : z_b}_{\text{quadratic terms}} + (1/2!) \underbrace{-f_3 :^2 z_b}_{\text{cubic terms}} + \dots \quad (7.6.25)$$

Here, in accord with (6.13), the degrees of the various terms have been indicated.

Now apply $\exp(: -f_3 :)$ to both sides of (6.19). Doing so, and again making use of (6.14), gives the result

$$\begin{aligned} & \exp(- : f_3 :) \exp(- : f_2^a :) \exp(- : f_2^c :) \bar{z}_b = \\ & z_b + : -f_3 : z_b + g_b(2; z) + r_b (> 2). \end{aligned} \quad (7.6.26)$$

However, according to Lemma 6.2, f_3 has the property

$$: -f_3 : z_b + g_b(2; z) = 0. \quad (7.6.27)$$

Consequently, (6.26) can be rewritten in the form

$$\exp(- : f_3 :) \exp(- : f_2^a :) \exp(- : f_2^c :) \bar{z}_b = z_b + r_b (> 2). \quad (7.6.28)$$

Comparison of the right sides of (6.17) and (6.28) shows that the degree of the remainder term has been raised by 1. At this stage it should also be clear that the degree of the remainder term can be increased indefinitely by finding f_4, f_5, \dots and applying the Lie transformations $\exp(: -f_4 :), \exp(- : f_5 :), \dots$. That is, for any s we have the general result

$$\exp(- : f_s :) \cdots \exp(: -f_3 :) \exp(- : f_2^a :) \exp(- : f_2^c :) \bar{z}_b = z_b + r_b [> (s-1)]. \quad (7.6.29)$$

We are ready for the final step. Rewrite (6.29) in the form

$$\bar{z}_b = \exp(: f_2^c :) \exp(: f_2^a :) \exp(: f_3 :) \cdots \exp(: f_s :) z_b + r_b [> (s-1)], \quad (7.6.30)$$

and let $s \rightarrow \infty$. Then, if the remainder term tends to zero, we obtain the advertised result (6.2) and (6.3). Otherwise the result is true only formally. In this latter case the infinite product (6.3) is also not convergent.

We have proved a key result. Recall that in Section 6.4 it was shown that Hamiltonian flows produce symplectic maps. Also, Theorem 1.3.3 shows that for many systems of physical interest such maps are analytic. Now, thanks to Theorem 6.1, it is possible to describe the most general analytic symplectic map (which sends the origin into itself) simply in terms of various homogeneous polynomials. Finally, it will be shown in the next section that the

restriction of preserving the origin can be removed by including Lie transformations of the form $\exp(: f_1 :)$ where f_1 is a suitably chosen polynomial linear in the z 's. Consequently, any analytic symplectic map can be represented uniquely as a product of Lie transformations generated by homogeneous polynomials. Conversely, any product of Lie transformations generated by homogeneous polynomials is a symplectic map.

At this point, two comments are appropriate. First, suppose the factored product representation (6.3) is truncated at any point. Then the resulting expression is still a symplectic map because each term in the product is a symplectic map. Also, if the truncation consists of dropping all terms in the product (6.3) beyond $\exp(: f_m :)$ for some m , then according to (6.31) the power-series expansion for the truncated map agrees with that of the original Taylor map (6.1) through terms of degree $(m - 1)$. Consequently, a truncated product map provides a symplectic approximation to the exact map. By contrast, as we know from Exercise 5.3, simply truncating a Taylor map generally *violates* the symplectic condition.

Second, suppose (6.3) is decomposed, as shown below, into those factors involving only quadratic polynomials, and the remaining factors involving cubic and higher degree polynomials,

$$\mathcal{M} = \overbrace{\exp(: f_2^c :)}^{\text{"Gaussian optics"}} \exp(: f_2^a :)^a \times \overbrace{\exp(: f_3 :)}^{\text{Aberrations or nonlinear corrections}} \exp(: f_4 :)^a \cdots . \quad (7.6.31)$$

It will be demonstrated in subsequent sections that dropping all terms beyond those involving the quadratic polynomials leads to a lowest-order approximation for \mathcal{M} that is equivalent to the paraxial Gaussian optics approximation in the case of light optics, and the usual linear matrix approximation in the case of charged-particle beam optics. Moreover, the remaining factors $\exp(: f_3 :)^a \exp(: f_4 :)^a \cdots$ represent aberrations or nonlinear corrections to the lowest-order approximation. In particular, in the case of charged-particle beam optics, the factor $\exp(: f_3 :)$ describes various chromatic effects and the effects due to sextupole magnets. Similarly, the factor $\exp(: f_4 :)$ describes higher-order chromatic effects, the effects due to iterated sextupoles, and the effects due to octupoles. Finally in some cases f_3, f_4 , etc. also describe what may be called “kinematic” nonlinearities in the equations of motion. They arise, for example, from the fact that the equations of motion generated by the Hamiltonians (1.6.16) and (1.6.17) are intrinsically nonlinear even in the absence of electric and magnetic fields. Let D be an integer with $D \geq 3$. In general, as will be shown later, retaining in the product (6.31) only those terms with f_m satisfying $m \leq D$ amounts to neglecting aberrations of degree D and higher.

At this point it is appropriate to issue a cautionary note: the quantities f_2^c and f_2^a that occur in the factorization (2.10), or more generally in the factorization (6.3), can have what may seem to be surprising properties. Suppose, for example, that \mathcal{M} is a *linear* symplectic map. Employ the notation $z = (q_1, q_2, \dots; p_1, p_2, \dots)$ and write $\bar{z} = \mathcal{M}z = (\bar{q}_1, \bar{q}_2, \dots; \bar{p}_1, \bar{p}_2, \dots)$. Suppose moreover that \mathcal{M} is known to have the property

$$\bar{p}_1 = \mathcal{M}p_1 = p_1. \quad (7.6.32)$$

Such maps obviously form a group, and any map of the form

$$\mathcal{M}_g = \exp(: g_2 :)^a \quad (7.6.33)$$

where g_2 does not depend on the variable q_1 ,

$$\partial g_2 / \partial q_1 = 0, \quad (7.6.34)$$

will have this property. Suppose that h_2 is another function that does not depend on q_1 . Then it is clear that all linear combinations of g_2 and h_2 and their single and multiple Poisson brackets will also be independent of q_1 . Thus, the set of all such functions forms a Lie subalgebra. Now let \mathcal{M} be any product of maps which individually are exponentials of Lie operators associated with q_1 -independent quadratic polynomials, and let f_2^c and f_2^a be the quadratic polynomials associated with a factorization of \mathcal{M} in the form (2.10) or (6.3),

$$\mathcal{M} = \exp(: g_2 :) \exp(: h_2 :) \cdots = \exp(: f_2^c :) \exp(: f_2^a :). \quad (7.6.35)$$

Then, it is tempting to assume that f_2^c and f_2^a will also be independent of q_1 . This would be the case if f_2^c and f_2^a were in the Lie subalgebra generated by g_2, h_2, \dots . However, a simple counter-example shows that this need not be true. See Exercise 6.14. Although f_2^c and f_2^a are in the Lie algebra of quadratic polynomials, they need not be in the subalgebra generated by g_2, h_2, \dots . This is because, by their construction, f_2^c and f_2^a are required to have specific properties with respect to J , and it may happen that these properties can only be achieved by including Lie elements outside the subalgebra. For example, in the case of $sp(2, \mathbb{R})$, f_2^c is proportional to $q^2 + p^2$; and f_2^a is a linear combination of $-q^2 + p^2$ and qp . See Section 5.6. Note that both f_2^c and f_2^a , when considered individually, depend on *both* the variables q and p .

In the context of Accelerator Physics a particularly confusing/irritating example of this phenomena occurs in the case of *static/time-independent* maps where the differential transit time t plays the role of q_1 in the present discussion and its conjugate momentum (energy) p_t plays the role of p_1 . With the exception of maps for radio-frequency cavities and maps for magnets with time-dependent fields, most maps for accelerator beam-line elements are static. Yet when the f_2^c and f_2^a are computed for a *static* map, they may turn out to be *time dependent*.² But, of course, when the effects of these f_2^c and f_2^a are combined to compute a net map, this net map will leave p_t unchanged even though this fact is not readily apparent simply by looking at f_2^c and f_2^a .

We close this section with the observation that the factorization (6.3) can also be written simply in the form

$$\mathcal{M} = \mathcal{R} \exp(: f_3 :) \exp(: f_4 :) \cdots \quad (7.6.36)$$

where \mathcal{R} is the linear symplectic map

$$\mathcal{R} = \exp(: f_2^c :) \exp(: f_2^a :). \quad (7.6.37)$$

Evidently \mathcal{R} has the property

$$\mathcal{R}z = Rz \quad (7.6.38)$$

²For this and other reasons, the charged-particle beam transport code MaryLie does not work directly with the polynomials f_2^c and f_2^a . It works with 6×6 symplectic matrices R to represent the linear part of the map \mathcal{M} , and only computes f_2^c and f_2^a from R when requested. The static/dynamic nature of R is readily apparent upon inspection of its matrix elements.

with R being the matrix appearing in (6.1). Indeed, (6.38) may be taken to be the definition of \mathcal{R} . In this way, as we will often do, we can bypass polar decomposition and the use of f_2^c and f_2^a unless needed for some specific purpose.³

Exercises

7.6.1. Verify (6.7).

7.6.2. Verify (6.8) through (6.10).

7.6.3. Show that any h that satisfies (6.6) is unique up to an additive constant.

7.6.4. Verify (6.12) and (6.14).

7.6.5. Verify (6.20).

7.6.6. Verify (6.11) and (6.22). Carry out the path integral described to get (6.24).

7.6.7. Suppose that $f(m, z)$ is a homogeneous polynomial of degree m . Then it must satisfy *Euler's* relation

$$\sum_a z_a (\partial f / \partial z_a) = mf. \quad (7.6.39)$$

See Exercise 1.5.11. Given (6.5), verify by direct calculation that f_{m+1} as given by (6.24) satisfies the relation

$$: f_{m+1} : z_b = g_b(m; z). \quad (7.6.40)$$

7.6.8. Verify (6.26).

7.6.9. Justify the passage from (6.29) to (6.30).

7.6.10. Consider the two-variable map, a variant of the *Hénon* map, given by the relations

$$\begin{aligned} \bar{q} &= \lambda[q + (q - p)^2], \\ \bar{p} &= (1/\lambda)[p + (q - p)^2], \end{aligned} \quad (7.6.41)$$

where λ is a parameter (positive or negative). Show that this map is symplectic. Find the factorization (6.3). That is, determine the polynomials f_m .

7.6.11. Consider the two-variable map, a variant of the *Hénon* map, given by the relations

$$\begin{aligned} \bar{q} &= q \cos \alpha + p \sin \alpha + p^2 \cos \alpha, \\ \bar{p} &= -q \sin \alpha + p \cos \alpha - p^2 \sin \alpha, \end{aligned} \quad (7.6.42)$$

where α is a parameter. Show that this map is symplectic. Find the factorization (6.3). That is, determine the polynomials f_m .

³Recall that, in addition to proof of the factorization theorem, polar decomposition is useful, for example, for writing Lorentz transformations as products of boosts and rotations.

7.6.12. Given the factorization (6.31), show how to compute the R , T , and U of (6.1).

7.6.13. Find the restrictions on the coefficients R , T , and U in (6.1) that are entailed by the symplectic condition.

7.6.14. This exercise is a sequel to Exercise 5.6.7, which you should review. Its purpose is to examine two linear two-dimensional symplectic maps, find their single exponential forms where applicable, find their polar decompositions and associated polynomials f_2^a and f_2^c , and examine the q and p content of these polynomials.

As the first case to be studied, let \mathcal{M} be a linear symplectic map acting on the two-dimensional phase space q_1, p_1 and described by the matrix L given by the relation

$$L = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \quad (7.6.43)$$

Evidently L is symplectic and \mathcal{M} satisfies (6.32). Verify that, in fact, \mathcal{M} can be written in the form

$$\mathcal{M} = \exp(: g_2 :) \quad (7.6.44)$$

with

$$g_2 = -p_1^2/2. \quad (7.6.45)$$

Observe that g_2 does not depend on q_1 .

Let us continue on to compute the quadratic polynomials f_2^c and f_2^a appearing in (2.10) to examine their q_1 behavior and then find them explicitly. Begin by polar decomposing L as in (6.15). From Exercise 5.6.7 we know that P is given by (5.6.34) and O is given by (5.6.40). If we write O and P in the exponential forms (3.8.17) and (3.8.25), we see that neither S^c nor S^a can vanish since neither O nor P are equal to the identity. Use the parameterization of (5.9.10) to write

$$JS^c = \beta_0 B^0, \quad (7.6.46)$$

$$JS^a = \phi F + \gamma G. \quad (7.6.47)$$

See (5.6.7), (5.6.13), and (5.6.14). Show that f_2^c and f_2^a are given by the relations

$$f_2^c = -(\beta_0/2)(p_1^2 + q_1^2), \quad (7.6.48)$$

$$f_2^a = -(\phi/2)(p_1^2 - q_1^2) - \gamma q_1 p_1. \quad (7.6.49)$$

Because neither S^c nor S^a can vanish, verify that consequently both f_2^c and f_2^a , unlike g_2 , must depend on q_1 .

Let us now compute f_2^c and f_2^a explicitly. With regard to O , use (5.9.12) to deduce the relation

$$O = \exp(\beta_0 B^0) = I \cos \beta_0 + B^0 \sin \beta_0, \quad (7.6.50)$$

and thereby obtain the results

$$\cos \beta_0 = 2/\sqrt{5}, \quad (7.6.51)$$

$$\sin \beta_0 = 1/\sqrt{5}, \quad (7.6.52)$$

$$\tan(\beta_0) = 1/2, \quad (7.6.53)$$

$$\beta_0 = \tan^{-1}(1/2) = .463 \dots \quad (7.6.54)$$

Here, in view of the 2π periodicity of the right side of (6.50), we have restricted β_0 to the interval $\beta_0 \in [-\pi, \pi]$.

With regard to P , deduce from (5.9.11) the relation

$$\begin{aligned} P = \exp(\phi F + \gamma G) &= I \cosh[(\phi^2 + \gamma^2)^{1/2}] \\ &+ [(\phi F + \gamma G)]/(\phi^2 + \gamma^2)^{1/2} \sinh[(\phi^2 + \gamma^2)^{1/2}]. \end{aligned} \quad (7.6.55)$$

Using the explicit form (5.6.27) for P , take the trace of both sides of (6.55) to find the result

$$\cosh[(\phi^2 + \gamma^2)^{1/2}] = (1/2)\sqrt{5}. \quad (7.6.56)$$

Next multiply both sides of (6.55) by F and again take traces to find the result

$$[\phi/(\phi^2 + \gamma^2)^{1/2}] \sinh[(\phi^2 + \gamma^2)^{1/2}] = -1/\sqrt{5}. \quad (7.6.57)$$

Finally, multiply both sides of (6.55) by G and take traces to find the result

$$[\gamma/(\phi^2 + \gamma^2)^{1/2}] \sinh[(\phi^2 + \gamma^2)^{1/2}] = 1/(2\sqrt{5}). \quad (7.6.58)$$

Show that (6.57) and (6.58) are consistent with (6.56), and from them deduce the relation

$$\phi = -2\gamma. \quad (7.6.59)$$

Solve (6.56) through (6.58) to obtain the results

$$\phi = -.430 \dots, \quad (7.6.60)$$

$$\gamma = .215 \dots \quad (7.6.61)$$

Taken together, the relations (6.48) and (6.49), with β_0 and ϕ and γ given by (6.54) and (6.60) and (6.61), specify f_2^c and f_2^a explicitly. Note that both f_2^c and f_2^a depend on both q_1 and p_1 even though (6.32) holds and g_2 is independent of q_1 .

As the second case to be studied, let \mathcal{M} be the linear symplectic map described by the matrix M given by the relation

$$M = -L = \begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix}. \quad (7.6.62)$$

Evidently M is symplectic. Moreover, we know from Exercise 3.7.12 that this M cannot be written in single exponential form.

Let us continue on to compute and find explicitly the quadratic polynomials f_2^c and f_2^a for \mathcal{M} . From Exercise 5.6.7 we know that M has the polar decomposition (5.6.41). It follows from (5.6.42) that f_2^a is the same as the f_2^a found in the first part of this exercise. What remains is to find f_2^c . Evidently the Ansatz (6.48) continues to hold, but with a different value of β_0 . Show that in this case there is the relation

$$O' = \exp(\beta_0 B^0) = I \cos \beta_0 + B^0 \sin \beta_0 \quad (7.6.63)$$

with O' given by (5.6.43). Show that now there are the results

$$\cos \beta_0 = -2/\sqrt{5}, \quad (7.6.64)$$

$$\sin \beta_0 = -1/\sqrt{5}, \quad (7.6.65)$$

$$\tan(\beta_0) = 1/2, \quad (7.6.66)$$

$$\beta_0 = -\pi + \tan^{-1}(1/2) = -\pi + .463 \dots = -2.677 \dots \quad (7.6.67)$$

7.7 Inclusion of Translations

Consider transformations of the form

$$\bar{z}_b = z_b + \delta_b, \quad (7.7.1)$$

where the quantities $\delta_1, \dots, \delta_{2n}$ are parameters. It is easy to verify that (7.1) is a symplectic map. See Exercise 6.2.2. Define a related set of parameters δ_a^* by the rule

$$\delta_a^* = \sum_b J_{ab} \delta_b, \text{ or } \delta^* = J\delta. \quad (7.7.2)$$

Also, define a first-degree polynomial $g_1(z)$ by the rule

$$\begin{aligned} g_1(z) &= \sum_{ab} J_{ab} z_a \delta_b = (z, \delta^*) = (z, J\delta) \\ &= (J^T z, \delta) = (\delta, J^T z) = -(\delta, Jz) \\ &= -(\delta, z^*). \end{aligned} \quad (7.7.3)$$

Then, use of (6.10) shows that g_1 obeys the relations

$$\begin{aligned} :g_1: z_b &= [g_1, z_b] = -[z_b, g_1] \\ &= -\partial g_1 / \partial z_b^* = \delta_b, \end{aligned} \quad (7.7.4)$$

$$:g_1:^m z_b = 0 \text{ for } m > 1. \quad (7.7.5)$$

Consequently, we have the relation

$$\exp(:g_1:) z_b = z_b + \delta_b. \quad (7.7.6)$$

That is, Lie transformations of the form $\exp(:g_1:)$ produce translations in phase space, and any translation can be written in this form.

Suppose the Taylor map (6.1) is generalized by the addition of constant terms to give a transformation of the form

$$\bar{z}_a = \delta_a + \sum_b R_{ab} z_b + \sum_{bc} T_{abc} z_b z_c + \sum_{bcd} U_{abcd} z_b z_c z_d + \dots \quad (7.7.7)$$

Then, a slight modification of Theorem (6.1) shows that (6.31) is generalized to become the relation

$$\begin{aligned} \exp(- : f_s :) \cdots \exp(- : f_3 :) \exp(- : f_2^a :) \exp(- : f_2^c :) \bar{z}_b \\ = z_b + \delta_b + r_b[> (s-1)]. \end{aligned} \quad (7.7.8)$$

Here the homogeneous polynomials f_m are the same as before. Next use (7.6) in (7.8) to get the relation

$$\begin{aligned} \exp(- : f_s :) \cdots \exp(- : f_3 :) \exp(- : f_2^a :) \exp(- : f_2^c :) \bar{z}_b \\ = \exp(: g_1 :) z_b + r_b[> (s-1)]. \end{aligned} \quad (7.7.9)$$

Finally, rewrite (7.9) in the form

$$\bar{z}_b = \exp(: f_2^c :) \exp(: f_2^a :) \exp(: f_3 :) \cdots \exp(: f_s :) \exp(: g_1 :) z_b + r_b[> (s-1)], \quad (7.7.10)$$

and let $s \rightarrow \infty$. We see that the generalized transformation (7.7) can be written in the form

$$\bar{z} = \mathcal{M}z \quad (7.7.11)$$

where \mathcal{M} has the factorization

$$\mathcal{M} = [\exp(: f_2^c :) \exp(: f_2^a :) \exp(: f_3 :) \exp(: f_4 :) \cdots] \exp(: g_1 :). \quad (7.7.12)$$

As in Section 7.5, let \mathcal{N} be a symplectic map, and suppose \mathcal{N} sends the particular point \tilde{z}^i to the point \tilde{z}^f . Also, again suppose \mathcal{N} is analytic in z around the point \tilde{z}^i . Then, the relations (5.1), (5.2), and (5.5) can also be written in the form

$$\begin{aligned} \bar{z}_a &= \tilde{z}_a^f + \sum_b R_{ab}(z - \tilde{z}^i)_b + \sum_{bc} T_{abc}(z - \tilde{z}^i)_b(z - \tilde{z}^i)_c \\ &+ \sum_{bcd} U_{abcd}(z - \tilde{z}^i)_b(z - \tilde{z}^i)_c(z - \tilde{z}^i)_d + \cdots. \end{aligned} \quad (7.7.13)$$

Let $h_1(z)$ be a first-degree polynomial defined by the relation

$$h_1(z) = (z, J\tilde{z}^i). \quad (7.7.14)$$

Then, by construction and the discussion at the beginning of this section, $h_1(z)$ has the property

$$\exp(: h_1 :) z = z + \tilde{z}^i, \quad (7.7.15)$$

or

$$\exp(: h_1 :)(z - \tilde{z}^i) = z. \quad (7.7.16)$$

Apply $\exp(: h_1 :)$ to both sides of (7.13). Doing so, and making use of (7.16) and (5.4.11), gives the result

$$\exp(: h_1 :) \bar{z}_a = \tilde{z}_a^f + \sum_b R_{ab} z_b + \sum_{bc} T_{abc} z_b z_c + \sum_{bcd} U_{abcd} z_b z_c z_d + \cdots. \quad (7.7.17)$$

We are now again in the situation described by (7.7) with \tilde{z}^f playing the role of k . Consequently, we may write the relation

$$\begin{aligned} \exp(- : f_s :) & \cdots \exp(- : f_3 :) \exp(- : f_2^a :) \exp(- : f_2^c :) \exp(: h_1 :) \bar{z}_b \\ & = \exp(: g_1 :) z_b + r_b[> (s-1)]. \end{aligned} \quad (7.7.18)$$

Here the homogeneous polynomials f_m are again the same as before, and g_1 is given by the relation

$$g_1(z) = (z, J\tilde{z}^f). \quad (7.7.19)$$

Finally, rewrite (7.18) in the form

$$\begin{aligned} \bar{z}_b & = \exp(- : h_1 :) \exp(: f_2^c :) \exp(: f_2^a :) \exp(: f_3 :) \cdots \exp(: f_s :) \exp(: g_1 :) z_b \\ & + \exp(- : h_1 :) r_b[> (s-1)]. \end{aligned} \quad (7.7.20)$$

Again let $s \rightarrow \infty$. Then, providing the remainder term tends to zero,

$$\lim_{s \rightarrow \infty} \exp(- : h_1 :) r_b[> (s-1)] = 0, \quad (7.7.21)$$

we have the result

$$\bar{z} = \mathcal{N}z \quad (7.7.22)$$

where \mathcal{N} has the factorization

$$\mathcal{N} = \exp(- : h_1 :)[\exp(: f_2^c :) \exp(: f_2^a :) \exp(: f_3 :) \exp(: f_4 :) \cdots] \exp(: g_1 :). \quad (7.7.23)$$

We conclude that the general analytic symplectic map \mathcal{N} given by (7.13) can be written in the factored product form (7.23).

Note that Sections 5.1 and 5.3 showed that the set of Lie operators forms an infinite-dimensional Lie algebra, and Section 6.2 showed that symplectic maps form a group. Theorem 6.1 and (7.23) [see also (8.1) and (8.2)] show that Lie operators form the Lie algebra of the group of symplectic maps. *Thus, the group of symplectic maps is an infinite-dimensional Lie group, and its Lie algebra is the Lie algebra of Lie operators.* From (2.10) we see that the subgroup of all symplectic maps that preserve the origin and are linear, namely $Sp(2n, \mathbb{R})$, has as its Lie algebra $sp(2n, \mathbb{R})$ all Lie operators of the form $: f_2 :$. Next consider $SpM(2n, \mathbb{R})$, the group of all symplectic maps that preserve the origin. From (6.3) we see that its Lie algebra, $spm(2n, \mathbb{R})$, consists of all Lie operators of the form $: f_m :$ with $m = 2, 3, \dots$. Finally, consider $ISpM(2n, \mathbb{R})$, the group of all symplectic maps. We see from (7.23) [see also (8.1) and (8.2)] that its Lie algebra, $ispm(2n, \mathbb{R})$, consists of all Lie operators of the form $: f_m :$ with $m = 1, 2, 3, \dots$.

We close this section with the remark that if one considers the set of all invertible analytic maps, and not just the subset of analytic symplectic maps, then this set of all invertible analytic maps also forms a group. This group, sometimes called the group of *analytic diffeomorphisms*, is also an infinite-dimensional Lie group, and has as its Lie algebra the set of all general Lie operators of the form (5.3.17) with the g_b being analytic functions. This group is sometimes called $Diff(\mathbb{R}^m)$ where m is the dimension of the space under consideration. If one is more careful, which we generally are not because we assume analyticity, one should

make distinctions between maps that are merely continuous (C^0), or have some specified number of derivatives (C^k), or have an infinite number of derivatives (C^∞), or are analytic (C^ω).⁴ With more careful notation, the group of *analytic* diffeomorphisms should be called $Diff^\omega(R^m)$. We also remark that often C^∞ functions are called *smooth*.

Exercises

7.7.1. Verify (7.8).

7.7.2. Find the factorization of the form (7.12) for the map (6.2.10). Show that Lie operators of the form : f_1 : and : f_2 : generate a Lie algebra under commutation. This is the Lie algebra $isp(2n, \mathbb{R})$, the Lie algebra of the inhomogeneous symplectic group $ISp(2n, \mathbb{R})$. What is the dimension of this Lie algebra?

Show that the polynomials f_0 (where $f_0 = \text{any constant}$) and f_1 generate a Lie algebra under the Poisson bracket operation. This is the Lie algebra of the *Heisenberg* group. What is its dimension?

Show that the polynomials f_0 , f_1 , and f_2 generate a Lie algebra under the Poisson bracket operation. This is the Lie algebra of the *Jacobi* group. For lack of a standard notation, we will denote the Jacobi group by the symbols $J(2n, \mathbb{R})$, and its Lie algebra by $j(2n, \mathbb{R})$. What is the dimension of this Lie algebra? Show that this algebra is homomorphic to that of the inhomogeneous symplectic group. See (5.3.14), (5.3.15), and (5.3.16).

7.7.3. Consider the matrices Q , P , and E defined by the rules

$$Q = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (7.7.24)$$

$$P = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad (7.7.25)$$

$$E = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (7.7.26)$$

Show that these matrices form a Lie algebra with the commutator as a Lie product. Consider a two-dimensional phase space with coordinates q and p . Show that the functions q , p , and 1 form a Lie algebra with the Poisson bracket as a Lie product. According to Exercise 7.2 above, this Lie algebra is the Heisenberg Lie algebra in the case of a two-dimensional phase space. Show that the commutator Lie algebra associated with the matrices Q , P , and E has the same structure constants as the Heisenberg Lie algebra, and therefore provides a matrix representation of the Heisenberg Lie algebra. Show that this representation is *not* the adjoint representation.

⁴In the context of Accelerator Physics where one is often concerned with charged-particle motion in electromagnetic fields in vacuum, this analyticity can be proved. See Appendix F.

7.7.4. As shown in Section 7.2, the symplectic group $Sp(2n, \mathbb{R})$ acts transitively on punctured phase space. Consequently, according to the discussion in Section 5.12, it must be possible to view punctured phase space as a coset space of $Sp(2n, \mathbb{R})$ with respect to one of its subgroups. The purpose of this exercise is to find this subgroup. Following the general procedure of Section 5.12, we must look for all $Sp(2n, \mathbb{R})$ transformations of punctured phase space that leave some point fixed. Without loss of generality, this point can be taken to be the point z^1 given by (2.30).

Let M be a symplectic matrix that preserves the vector (phase-space point) z^1 ,

$$Mz^1 = z^1. \quad (7.7.27)$$

Show from (2.29) and (7.27) that the M_{a1} matrix elements obey the relations

$$M_{a1} = \delta_{a1}. \quad (7.7.28)$$

Let $\mathcal{M}(M)$ be a symplectic map associated with M by the relation

$$\mathcal{M}(M)z_a = \sum_b (M^T)_{ab}z_b. \quad (7.7.29)$$

Show from (7.29) that \mathcal{M} is symplectic, and from (7.28) and (7.29) that \mathcal{M} satisfies the relation

$$\mathcal{M}q_1 = q_1. \quad (7.7.30)$$

Conversely, show that if \mathcal{M} is a symplectic map of the form (7.29) that also satisfies (7.30), then (7.28) and (7.27) are satisfied. Consequently, we can concentrate on finding the general solution to (7.30).

We begin by working near the identity, and write \mathcal{M} in the form

$$\mathcal{M} = \exp(: f_2 :). \quad (7.7.31)$$

Show that requiring (7.30) for \mathcal{M} near the identity is equivalent to requiring that f_2 satisfy the relation

$$: f_2 : q_1 = 0. \quad (7.7.32)$$

Show that any f_2 that satisfies (7.32) cannot depend on p_1 , and is therefore a linear combination of the polynomials q_1^2 , $q_1\tilde{f}_1$, and \tilde{f}_2 . Here \tilde{f}_1 and \tilde{f}_2 denote homogeneous polynomials in the remaining variables $q_2 \cdots q_n$ and $p_2 \cdots p_n$ of degrees 1 and 2, respectively. Show that the polynomials \tilde{f}_2 give a representation of the Lie algebra $sp[2(n-1), R]$. Review Exercise 7.2, and consider the Poisson bracket Lie algebra generated by f_0 , f_1 , and f_2 , the Jacobi Lie algebra $j(2n, \mathbb{R}R)$. You should have found that it has dimension $n(2n+3)+1$. Show that the polynomials q_1^2 , $q_1\tilde{f}_1$, and \tilde{f}_2 have the Lie algebra $j[2(n-1), R]$ under the Poisson bracket operation, and that the Lie operators $: q_1^2 :$, $: q_1\tilde{f}_1 :$, and $: \tilde{f}_2 :$ have that same Lie algebra under commutation. Show that the dimensions of $sp(2n, \mathbb{R})$ and $j[2(n-1), \mathbb{R}]$ are related by the equation

$$\dim\{sp(2n, \mathbb{R})\} - \dim\{j[2(n-1), \mathbb{R}]\} = 2n. \quad (7.7.33)$$

Let $J[2(n-1), \mathbb{R}]$ be the Lie group generated by the Lie operators $:q_1^2: :q_1\tilde{f}_1: :$ and $:\tilde{f}_2: :$. Show that the general \mathcal{M} in $J[2(n-1), \mathbb{R}]$ can be written in the form

$$\mathcal{M} = \exp(\alpha : q_1^2 :) \exp(: \tilde{f}_2^c :) \exp(: \tilde{f}_2^a :) \exp(: q_1 \tilde{f}_1 :), \quad (7.7.34)$$

where α is an arbitrary parameter. Show that (7.34) is the most general $Sp(2n, \mathbb{R})$ transfer map that satisfies the relation

$$\mathcal{M}q_1 = q_1. \quad (7.7.35)$$

Hint: If you are having difficulty, see the beginning of Section 9.4.

Let H be the subgroup of $Sp(2n, \mathbb{R})$ consisting of all matrices M that satisfy (7.27). Suppose M^1 and M^2 are in H . Then we have the relation

$$\begin{aligned} \mathcal{M}(M^1)\mathcal{M}(M^2)z_a &= \mathcal{M}(M^1) \sum_b [(M^2)^T]_{ab} z_b \\ &= \sum_{b,c} [(M^2)^T]_{ab} [(M^1)^T]_{bc} z_c \\ &= \sum_c [(M^2)^T (M^1)^T]_{ac} z_c = \sum_c [(M^1 M^2)^T]_{ac} z_c \\ &= \mathcal{M}(M^1 M^2)z_a, \end{aligned} \quad (7.7.36)$$

or, more compactly put,

$$\mathcal{M}(M^1)\mathcal{M}(M^2) = \mathcal{M}(M^1 M^2). \quad (7.7.37)$$

From (7.37) we see that the subgroup H is a matrix realization of $J[2(n-1), \mathbb{R}]$. Let G be the group $Sp(2n, \mathbb{R})$. Show that the coset space G/H has dimension $2n$, the expected dimension for the phase space under consideration.

7.7.5. Section 6.1 defined a symplectic map to be a map (of a $2n$ -dimensional space into itself) whose Jacobian matrix M is symplectic. Suppose that M is instead required to be *orthogonal*. Show that such maps also form a group, which might be called the group of orthogonal maps. Consider the Taylor expansion of such a map in the 2-dimensional case. Show that, unlike the expansion for symplectic maps, *only* constant and linear terms can occur in the Taylor expansion! You have verified a special case of the fact that, unlike symplectic maps, orthogonal maps are trivial in the sense that they consist only of translations and *linear* orthogonal transformations.

7.8 Other Factorizations

In addition to the factorization (7.23), there are other factorizations that are often useful. First, as will be shown in Chapter 9, it is possible to bring all first degree polynomials over to the left. In this case, \mathcal{N} has the factorization

$$\mathcal{N} = \exp(: f_1 :) \exp(: f_2^c :) \exp(: f_2^a :) \exp(: f_3 :) \exp(: f_4 :) \cdots. \quad (7.8.1)$$

Here the polynomials f_m are generally different from those in (7.23).⁵ The factorization (8.1) will be called *forward* or *ascending* factorization. Second, it is often useful to have a factorization of the form

$$\mathcal{N} = \cdots \exp(: f_4 :) \exp(: f_3 :) \exp(: f_2^a :) \exp(: f_2^c :) \exp(: f_1 :). \quad (7.8.2)$$

Here again the polynomials f_m are generally different from those in (8.1) or (7.23). The factorization (8.2) will be called *reverse* or *descending* factorization. Finally, it is often useful to have *mixed* factorizations where the f_m terms with $m > 1$ are ascending or descending, and the $\exp(: f_1 :)$ term is at the beginning, or at the end.

Exercises

7.8.1. Find other factorizations for the inhomogeneous symplectic group. See Exercises (3.9.2) and (7.2).

7.9 Coordinates and Connectivity

Suppose G is a finite-dimensional Lie group, and let B_1, B_2, \dots, B_n be a basis for the associated Lie algebra L . To be more specific, suppose G is realized as a group of matrices, and suppose that some element g in G is sufficiently near the identity so that it can be written in the form

$$g = \exp\left(\sum_{j=1}^n \xi_j B_j\right). \quad (7.9.1)$$

The parameters ξ_j are called *canonical coordinates* (for g) of the *first kind*. Another possibility is to write g in the form

$$g = \exp(\eta_1 B_1) \exp(\eta_2 B_2) \cdots \exp(\eta_n B_n). \quad (7.9.2)$$

The parameters η_j are called *canonical coordinates* of the *second kind*.⁶ For the case of a finite-dimensional Lie group, at least in some neighborhood of the identity, one may pass in principle from one kind of coordinates to the other with the aid of the BCH formula. (See Section 3.7. See also Exercise 10.4.2.) This may not be possible in the infinite-dimensional case because then the BCH series may not have any domain of convergence. See Section 33.7.

Reference to (7.23) shows that if the factorization process succeeds, the general analytic symplectic map \mathcal{N} has been given coordinates that are a hybrid of canonical coordinates of the first and second kinds. The map is written as a product of exponentials as in (9.2), and each exponential is a sum of terms as in (9.1).

⁵In Section 6.6 we saw that the $2n$ functions required to specify a map in $2n$ variables can be related to a single function in the symplectic case. This fact is also evident from (8.1) if we formally regard the f_j as the homogeneous parts of a single function f .

⁶Note that there are in principle as many as $n!$ different canonical coordinates of the second kind because, since the B_j may not commute, the order of the factors in (9.2) may be important.

Suppose \mathcal{N} can be factored as in (7.23). Then it is easy to see that \mathcal{N} is *connected* to the identity map \mathcal{I} by a continuous family of symplectic maps. Indeed, let λ be a parameter and let $\mathcal{N}(\lambda)$ be the map

$$\mathcal{N}(\lambda) = \exp(-\lambda : h_1 :) [\exp(\lambda : f_2^c :) \exp(\lambda : f_2^a :) \exp(\lambda : f_3 :) \cdots] \exp(\lambda : g_1 :). \quad (7.9.3)$$

It is evident that $\mathcal{N}(\lambda)$ is a symplectic map for all λ with the properties

$$\mathcal{N}(0) = \mathcal{I}, \quad \mathcal{N}(1) = \mathcal{N}. \quad (7.9.4)$$

The argument just given lacks generality because we had to assume that \mathcal{N} is analytic and that the factorization process converges. However, we can do better. Suppose we assume only that \mathcal{N} has at least a first few derivatives so that (7.13) can be written in the form

$$\bar{z}_a = \tilde{z}_a^f + \sum_b R_{ab}(z - \tilde{z}^i)_b + O[(z - \tilde{z}^i)^2]. \quad (7.9.5)$$

Then, by arguments similar to those of Section 7.7, the map \mathcal{N} can be rewritten in the form

$$\mathcal{N} = \exp(- : h_1 :) \exp(: f_2^c :) \exp(: f_2^a :) \mathcal{P} \exp(: g_1 :) \quad (7.9.6)$$

where \mathcal{P} is a symplectic map that sends the origin into itself and has an expansion of the form

$$\bar{z}_a = \mathcal{P}z_a = z_a + W_a(z) \quad (7.9.7)$$

with

$$W_a(z) = O[(z)^2]. \quad (7.9.8)$$

Define a map $\mathcal{P}(\lambda)$ by the relation

$$\bar{z}_a = \mathcal{P}(\lambda)z_a = (1/\lambda)[\lambda z_a + W_a(\lambda z)] = z_a + (1/\lambda)W_a(\lambda z). \quad (7.9.9)$$

Here λz denotes the collection of quantities λz_b . Evidently, in view of (9.8), we have the relations

$$\mathcal{P}(0) = \mathcal{I}, \quad \mathcal{P}(1) = \mathcal{P}. \quad (7.9.10)$$

Also, $\mathcal{P}(\lambda)$ is a symplectic map for all values of λ . To see this, observe that $\mathcal{P}(\lambda)$ can be written as a product of three maps in the form

$$\mathcal{P}(\lambda) = [(1/\lambda)\mathcal{I}][\mathcal{P}][\lambda\mathcal{I}]. \quad (7.9.11)$$

Although the maps $[(1/\lambda)\mathcal{I}]$ and $[\lambda\mathcal{I}]$ are not symplectic (if $\lambda \neq 1$), the product (9.11) is. Indeed, denoting by $P(\lambda, z)$ and $P(z)$ the Jacobian matrices of $\mathcal{P}(\lambda)$ and \mathcal{P} , respectively, use of the chain rule gives the relation

$$P(\lambda, z) = [(1/\lambda)I][P(\lambda z)][\lambda I] = P(\lambda z). \quad (7.9.12)$$

Since we know that \mathcal{P} is a symplectic map, $P(z)$ will be a symplectic matrix. According to (9.12) $P(\lambda, z)$, being equal to $P(\lambda z)$, is also a symplectic matrix because $P(z)$ is a

symplectic matrix for *all* values of z . Finally, because $P(\lambda, z)$ is a symplectic matrix, $\mathcal{P}(\lambda)$ is a symplectic map.⁷ In analogy with (9.6), we now define $\mathcal{N}(\lambda)$ by writing

$$\mathcal{N}(\lambda) = \exp(-\lambda : h_1 :) \exp(\lambda : f_2^c :) \exp(\lambda : f_2^a :) \mathcal{P}(\lambda) \exp(\lambda : g_1 :). \quad (7.9.13)$$

Evidently $\mathcal{N}(\lambda)$ is a symplectic map for all λ and has the desired properties (9.4). We have shown that if a symplectic map \mathcal{N} has at least a first few derivatives, then it is connected to the identity map by a continuous family of symplectic maps.

Of course, the family just constructed will generally differ from that given by (9.3). There are many families of symplectic maps that connect a given symplectic map to the identity map.⁸ We note that, according to Section 6.4, for each family there is a corresponding Hamiltonian that generates it.

7.10 Storage Requirements

How much computer memory is required to store a symplectic map in the Taylor form (7.7), and how much memory is required to store the corresponding Lie form (8.1)? Suppose the Taylor map (7.7) is truncated by discarding all terms having degree $(D+1)$ and higher. We denote this truncated Taylor map by \mathcal{T}_{D+1} . Then (7.7) has the truncated form

$$\tilde{z}_a = \mathcal{T}_{D+1} z_a = \sum_{m=0}^D g_a(m; z) \quad (7.10.1)$$

where, as in Section 7.6, each $g_a(m; z)$ is a homogeneous polynomial of degree m . According to (8.1) and our discussion of the relation between Taylor and Lie maps, there is a map \mathcal{M} in Lie form that corresponds to \mathcal{T}_{D+1} (they both have the same Taylor expansions through terms of degree D), and this map has a truncated factored product representation of the form

$$\mathcal{M} = \exp(: f_1 :) \exp(: f_2^c :) \exp(: f_2^a :) \exp(: f_3 :) \exp(: f_4 :) \cdots \exp(: f_{D+1} :). \quad (7.10.2)$$

Let $S(m, d)$ be the total number of monomials in d variables having degrees 1 through m . We know from (3.40) that the total number of monomials in d variables having degree m is $N(m, d)$. Consequently, $S(m, d)$ is given by the relation

$$S(m, d) = \sum_{k=1}^m N(k, d). \quad (7.10.3)$$

The sum (10.3) can be evaluated to give the result

$$S(m, d) = \binom{m+d}{m} - 1 = \frac{(m+d)!}{m!d!} - 1. \quad (7.10.4)$$

⁷This scaling by some factor λ that we have used is sometimes called *Alexander's Trick*.

⁸For example, in (9.13) one might replace λ as the coefficient of f_2^c by λ^2 . The reader should be able to construct other examples.

See Exercises 10.1 and 10.2. Table 10.1 below lists values of $S(m, d)$ for various values of m and d .

Now let $S_L(D, d)$ be the number of storage locations required to specify the Lie map (10.2). We know that the specification of an f_m requires $N(m, d)$ numbers (with $d = 6$ for the case of a 6-dimensional phase space). Consequently, comparison of (10.2) and (10.3) gives the result

$$S_L(D, d) = S(D + 1, d) = \binom{D + d + 1}{D + 1} - 1 = \frac{(D + d + 1)!}{(D + 1)!d!} - 1. \quad (7.10.5)$$

Correspondingly, let S_T be the number of locations required to specify the truncated Taylor map (10.1). We know that the specification of a particular $g_a(m, z)$ requires $N(m, d)$ numbers. Consequently, $S_T(D, d)$ must be given by the relation

$$S_T(D, d) = d \sum_{k=0}^D N(k, d) = d[S(D, d) + 1] = \frac{d(D + d)!}{D!d!}. \quad (7.10.6)$$

Note that d must be even in both (10.5) and (10.6) because phase space is even dimensional.

Finally, let us compare $S_T(D, d)$ and $S_L(D, d)$ for various values of d and D . Table 10.2 below lists values of S_T , S_L , and the ratio S_T/S_L , for $d = 4$ and $d = 6$ and various values of D . We conclude that (for modest values of D) storing a 6-dimensional phase-space map in Taylor form requires about 3 times more storage locations than the equivalent Lie form. For large D values this ratio approaches 6. This difference in storage requirements for the Taylor and Lie forms of a symplectic map arises from the fact that the Taylor form makes no use of the symplectic condition. Indeed, the Lie form contains exactly the minimal information required to specify a symplectic map while the Taylor form has all the coefficients required to specify the most general (analytic diffeomorphic) map.⁹

⁹Observe that $S_L(3, 6)$, the number of storage locations required to store a 6-variable symplectic map through third order in Lie form, has the value $S_L(3, 6) = 209$. Curiously, according to the Los Angles Times, in 1987 (when MaryLie 3.0 was being written) the Chinese Communist Party Central Committee, China's highest governing body, had 209 members.

Table 7.10.1: Number of monomials of degree 1 through m in various numbers of variables.

m	$S(m, 4)$	$S(m, 5)$	$S(m, 6)$	$S(m, 7)$	$S(m, 8)$	$S(m, 9)$	$S(m, 10)$	$S(m, 11)$
1	4	5	6	7	8	9	10	11
2	14	20	27	35	44	54	65	77
3	34	55	83	119	164	219	285	363
4	69	125	209	329	494	714	1000	1364
5	125	251	461	791	1286	2001	3002	4367
6	209	461	923	1715	3002	5004	8007	12375
7	329	791	1715	3431	6434	11439	19447	31823
8	494	1286	3002	6434	12869	24309	43757	75581
9	714	2001	5004	11439	24309	48619	92377	167959
10	1000	3002	8007	19447	43757	92377	184755	352715
11	1364	4367	12375	31823	75581	167959	352715	705431
12	1819	6187	18563	50387	125969	293929	646645	1352077

Table 7.10.2: Storage Requirements for Taylor and Lie Maps.

D	$S_T(D, 4)$	$S_L(D, 4)$	$S_T(D, 4)/S_L(D, 4)$	$S_T(D, 6)$	$S_L(D, 6)$	$S_T(D, 6)/S_L(D, 6)$
2	60	34	1.8	168	83	2.0
3	140	69	2.0	504	209	2.4
4	280	125	2.2	1260	461	2.7
5	504	209	2.4	2772	923	3.0
6	840	329	2.6	5544	1715	3.2
7	1320	494	2.7	10,296	3002	3.4
8	1980	714	2.8	18,018	5004	3.6
9	2860	1000	2.9	30,030	8007	3.8
10	4004	1364	2.9	48,048	12,375	3.9
11	5460	1819	3.0	74,256	18,563	4.0
12	7280	2379	3.1	111,384	27,131	4.1

Exercises

7.10.1. Verify (10.3) through (10.6). [Hint: Use the relations (3.52) through (3.54).] Show that S can be generated using the recursion relation

$$S(m, d) = S(m, d - 1) + S(m - 1, d) + 1 \quad (7.10.7)$$

with the initial conditions

$$S(m, 1) = m, \quad (7.10.8)$$

$$S(1, d) = d. \quad (7.10.9)$$

Show that S also satisfies the relation

$$S(m, d) = S(m - 1, d) + N(m, d). \quad (7.10.10)$$

7.10.2. The relation (10.4) can be derived directly from the definition of $S(m, d)$ as the total number of monomials in d variables having degrees 1 through m . Let $S_0(m, d)$ be the

total number of monomials in d variables having degrees 0 through m . Then we evidently have the relation

$$S_0(m, d) = S(m, d) + 1. \quad (7.10.11)$$

Show that from its definition $S_0(m, d)$ obeys the relations

$$S_0(1, 1) = 2, \quad (7.10.12)$$

$$S_0(m, 1) = m + 1, \quad (7.10.13)$$

$$S_0(2, 2) = 6. \quad (7.10.14)$$

Next show that S_0 obeys the recursion relation

$$S_0(m, d) = S_0(m - 1, d) + S_0(m, d - 1). \quad (7.10.15)$$

Hint: The number of monomials in d variables having degrees 0 through m is the number of monomials having degree 0 through $m - 1$, which is $S_0(m - 1, d)$, plus the number of monomials having degree m . We have already agreed to let $N(m, d)$ be the number of monomials having degree m . Homogeneous monomials of degree m in the d variables $z_1 \cdots z_d$ can be viewed as monomials in the variables $z_1 \cdots z_{d-1}$ of degree 0 through m multiplied by the appropriate power of z_d to make the total degree exactly m . Thus, we have the relation

$$N(m, d) = S_0(m, d - 1). \quad (7.10.16)$$

Finally, use the recursion relation (10.15) with the initial conditions (10.12) through (10.14) to show that S_0 is given by the relation

$$S_0(m, d) = \binom{m+d}{m} = \frac{(m+d)!}{m!d!}. \quad (7.10.17)$$

Hint: Recall that the binomial coefficients satisfy the recursion relation (3.52).

7.10.3. Evaluate the ratio S_T/S_L for various values of d (say $d = 4$ and $d = 6$) and various values of D . Show that this ratio approaches d in the limit of large D .

7.10.4. Compute the quantity $[S_L(D+1, d) - S_L(D, d)]/S_L(D, d)$ for large D . This quantity is the limiting fractional incremental increase in storage required to include one order higher aberration effects.

Bibliography

Combinatorics

- [1] A. Nijenhuis and H.S. Wilf, *Combinatorial Algorithms for Computers and Calculators*, Academic Press (1978).
- [2] J. Riordan, *An Introduction to Combinatorial Analysis*, John Wiley (1958).
- [3] The method of Exercise 3.12 is due to M. Venturini.

Taylor Maps and Jets

- [4] E. Hille and R. Phillips, *Functional Analysis and Semi-groups*, American Mathematical Society Colloquium Publications, Volume 31 (1957).
- [5] P.W. Michor, *Manifolds of Differentiable Mappings*, Shiva Publishing Limited (1980).

Factorization

- [6] A. Dragt and J. Finn, “Lie Series and Invariant Functions for Analytic Symplectic Maps”, *J. Math. Phys.* **17**, 2215 (1976).

Wigner Rotation

- [7] See the Web site https://en.wikipedia.org/wiki/Wigner_rotation.

Jacobi Group

- [8] R. Berndt and R. Schmidt, *Elements of the Representation Theory of the Jacobi Group*, (Progress in Mathematics; Vol. 163), Birkhäuser Verlag (1998).

Connectivity

- [9] P.J. Chanell, “Hamiltonian suspensions of symplectomorphisms: an alternative approach to design problems”, *Physica D* **127**, pp. 117-130 (1999).

Group Theory and Analyticity

- [10] B. Beers and A. Dragt, “New Theorems about Spherical Harmonic Expansions and $SU(2)$ ”, *Journal of Mathematical Physics*, Volume 11, pp. 2313-2328 (1970).

Computation of Charged-Particle Beam Transport

- [11] A. Dragt et al., *MaryLie 3.0 Users’ Manual* (2003). See www.physics.umd.edu/dsat/.

Chapter 8

A Calculus for Lie Transformations and Noncommuting Operators

Section 6.4 showed that Hamiltonian flows produce symplectic maps, and Sections 7.2 and 7.7 showed that the general analytic symplectic map (7.7.13) can be written in the factored product form (7.7.23). In addition, (7.4.1) gives an explicit representation for the symplectic map in the case of a time-independent Hamiltonian. See also (7.4.18). In subsequent sections these results will be applied to charged particle beam transport, light optics, and orbits in circular machines. The purpose of this chapter is to provide a collection of formulas for the manipulation of Lie transformations and noncommuting operators in general. Some formulas will be used to compute the product of two symplectic maps when each is written in factored product form. Others will be used to combine various exponents in a factored product decomposition into a single exponent. Still others will be used to produce factored product decompositions. Where necessary, discussion will be restricted to symplectic maps that send the origin into itself. (See Section 7.6.) This restriction will subsequently be removed in Chapter 9.

8.1 Adjoint Lie Operators and the Adjoint Lie Algebra

Work with noncommuting quantities is often facilitated by the concept of an *adjoint* Lie operator. Let $:f:$ be some Lie operator, and let $:g:$ be any other Lie operator. The *adjoint* of the Lie operator $:f:$, which will be denoted by the symbol $\# :f:\#$, is a kind of super operator that acts on other Lie operators according to the rule

$$\# :f:\# :g := \{ :f:, :g :\}. \quad (8.1.1)$$

Here, the right side of (1.1) denotes the commutator as in (5.3.10). Thanks to (5.3.14), the relation (1.1) can also be written in the form

$$\# :f:\# :g := \{ :f:, :g :\} =: [f, g] :. \quad (8.1.2)$$

We see that adjoint Lie operators act on Lie operators, and send them to other Lie operators. Furthermore, this action is linear. That is, we have the relation

$$\# : f : \#(a : g : + b : h :) = a\# : f : \# : g : + b\# : f : \# : h : . \quad (8.1.3)$$

We remark that the word *adjoint* is much overused in mathematics, and is not to be confused in this context with the Hermitian conjugate (also sometimes referred to as a Hermitian adjoint) defined in (7.3.15). To simplify notation in some cases where no confusion can arise, the set of colons in the symbol $\# : f : \#$ for the adjoint of the Lie operator $: f :$ will often be omitted. That is, the abbreviated symbol $\#f\#$ will often be used to serve for the complete symbol $\# : f : \#$.

Powers of $\# : f : \#$ or $\#f\#$ can be defined by repeated application of (1.1). For example, $\#f\#^2$ is defined by the relation

$$\#f\#^2 : g := \{ : f : , \{ : f : , : g : \} \}. \quad (8.1.4)$$

Also, $\#f\#$ to the zero power is defined to be the identity operator,

$$\#f\#^0 : g := : g : . \quad (8.1.5)$$

The set of adjoint Lie operators $\# : f : \#$ can also be made into a Lie algebra in its own right. This Lie algebra is called the *adjoint Lie algebra*. First, there is obviously the relation

$$a\#f\# + b\#g\# = \#(af + bg)\#. \quad (8.1.6)$$

That is, the set of adjoint Lie operators forms a linear vector space. Next, we define the Lie product of any two adjoint Lie operators to be their commutator,

$$\{\#f\#, \#g\# \} = \#f\#\#g\# - \#g\#\#f\#. \quad (8.1.7)$$

We note that this definition of a Lie product obviously satisfies the requirements 1 through 4 listed in Section (3.7) for a Lie algebra. It also satisfies requirement 5 since commutators satisfy the Jacobi condition. Of course, we must also show that the Lie product (commutator) of two adjoint Lie operators is again an adjoint Lie operator. Let $: h :$ be an arbitrary Lie operator. Then we have the results

$$\begin{aligned} \#f\#\#g\# : h &:= \#f\#\{ : g : , : h : \} = \{ : f : , \{ : g : , : h : \} \}, \\ \#g\#\#f\# : h &:= \#g\#\{ : f : , : h : \} = \{ : g : , \{ : f : , : h : \} \}, \\ \{\#f\#, \#g\# \} : h &:= \{ : f : , \{ : g : , : h : \} \} - \{ : g : , \{ : f : , : h : \} \} \\ &= \{ : f : , \{ : g : , : h : \} \} + \{ : g : , \{ : h : , : f : \} \}. \end{aligned} \quad (8.1.8)$$

Now use the Jacobi condition for the Lie algebra of Lie operators to find the relation

$$\{ : f : , \{ : g : , : h : \} \} + \{ : g : , \{ : h : , : f : \} \} = -\{ : h : , \{ : f : , : g : \} \} = \{ \{ : f : , : g : \} , : h : \}. \quad (8.1.9)$$

We see that (1.8) can be rewritten in the form

$$\{\#f\#, \#g\#} : h := \{\{ : f : , : g : \}, : h : \} = \#\{ : f : , : g : \} \# : h : . \quad (8.1.10)$$

Since the Lie operator $: h :$ is arbitrary, it follows that we have the result

$$\{\#f\#, \#g\#} = \#\{ : f : , : g : \} \# = \# : [f, g] : \# . \quad (8.1.11)$$

Thus the Lie product of two adjoint Lie operators is indeed an adjoint Lie operator.

Our discussion should have a familiar ring. It parallels, in fact, the material at the end of Section 3.7 and the treatment of Lie operators given in Section 5.3. Reviewing these sections, we see that the commutator Lie algebra of Lie operators is actually the adjoint Lie algebra of the underlying Poisson bracket Lie algebra. And, consequently, the “adjoint” we have been discussing is really the “adjoint-adjoint” of the basic Poisson bracket Lie algebra.

Exercises

8.1.1. Prove (1.3).

8.1.2. Prove (1.6). Verify that requirements 1 through 5 listed in Section 3.7 for a Lie algebra are satisfied by the adjoint Lie algebra.

8.1.3. Describe in detail how the adjoint Lie algebra is the adjoint-adjoint of the basic Poisson bracket Lie algebra. What would the adjoint-adjoint-adjoint Lie algebra be?

8.2 Formulas Involving Adjoint Lie Operators

There are several useful formulas involving adjoint Lie operators. First, we have the relations

$$\begin{aligned} \#f\#^0 : g &:=: g :=: (: f :^0 g) :, \\ \#f\# : g &:= \{ : f : , : g : \} =: [f, g] :=: (: f : g) : . \end{aligned} \quad (8.2.1)$$

Here use has been made of (5.3.14). From these relations we have by induction the general result

$$\#f\#^n : g :=: (: f :^n g) : . \quad (8.2.2)$$

Second, the definition of $\#f\#$ can be extended to let $\#f\#$ act on any sum or product, or sum of products, or even power series, of Lie operators. Suppose $F(: g : , : h : , \dots)$ is any function of a collection of Lie operators $: g : , : h : , \dots$. Then we define the action of $\#f\#$ on F in analogy to (1.1) by the rule

$$\#f\#F = \{ : f : , F \} . \quad (8.2.3)$$

As a special case of (2.3) we have the relation

$$\#f\#(: g :: h :) = \{ : f : , : g :: h : \}$$

$$\begin{aligned}
&= \{ : f : , : g : \} : h : + : g : \{ : f : , : h : \} \\
&= (\#f\# : g :) : h : + : g : (\#f\# : h :).
\end{aligned} \tag{8.2.4}$$

We see that the adjoint Lie operator $\#f\#$ is a *derivation* with respect to the multiplication of Lie operators.

Now suppose that $: f :$ and $: g :$ are any two Lie operators. We then find that

$$\exp(: f :) : g : \exp(- : f :) = \exp(\#f\#) : g : . \tag{8.2.5}$$

Here, as the notation suggests,

$$\exp(\#f\#) = \sum_{m=0}^{\infty} \#f\#^m / m!. \tag{8.2.6}$$

This result is sometimes called *Hadamard's lemma*.¹ To see that (2.5) is correct, consider the operator function $O(\tau)$ defined by the equation

$$O(\tau) = \exp(\tau : f :) : g : \exp(-\tau : f :) , \tag{8.2.7}$$

where τ is a parameter. Then we have the relation

$$O(0) =: g : . \tag{8.2.8}$$

Further, we find by differentiation of (2.7) the relation

$$dO/d\tau =: f : O - O : f := \{ : f :, O \} = \#f\#O. \tag{8.2.9}$$

The solution to this differential equation with the initial condition (2.8) is given by the relation

$$O(\tau) = \exp(\tau \#f\#) : g : . \tag{8.2.10}$$

Now set $\tau = 1$ in (2.10) to obtain the desired result.

From (2.2) it follows that we also have the relation

$$\exp(\#f\#) : g := \exp(: f :)g : . \tag{8.2.11}$$

Consequently, (2.5) can also be written in the form

$$\exp(: f :) : g : \exp(- : f :) =: \exp(: f :)g : . \tag{8.2.12}$$

Because $\#f\#$ is a derivation, see (2.4), there is an even more general result. Let $F(: g :, : h :, \dots)$ be a function of a collection of Lie operators of the type described above. Then we have the relations

$$\exp(: f :)F(: g :, : h :, \dots) \exp(- : f :) = \exp(\#f\#)F(: g :, : h :, \dots), \tag{8.2.13}$$

$$\begin{aligned}
\exp(\#f\#)F(: g :, : h :, \dots) &= F(\exp(\#f\#) : g :, \exp(\#f\#) : h :, \dots) \\
&= F(: \exp(: f :)g :, : \exp(: f :)h :, \dots).
\end{aligned} \tag{8.2.14}$$

¹A Web search reveals that there are also other Hadamard lemmas.

As a special case of (2.13) and (2.14) we find the results

$$\exp(:f:) : g :^m \exp(- :f:) = [\exp(\#f\#) : g :]^m, \quad (8.2.15)$$

$$\exp(:f:) \exp(:g:) \exp(- :f:) = \exp[\exp(\#f\#) : g :]. \quad (8.2.16)$$

This discussion should also have a familiar ring. See Section 5.4. Here we are exploiting the fact that $\exp(\#f\#)$ is an *isomorphism* with respect to Lie operator multiplications.

The relations (2.13) and (2.14) can also be derived directly. Consider, for example, the simple case

$$\begin{aligned} & \exp(:f:) : g :: h : \exp(- :f:) = \exp(\#f\#) : g :: h : \\ & = (\exp(\#f\#) : g :)(\exp(\#f\#) : h :) =: \exp(:f:)g :: \exp(:f:)h : . \end{aligned} \quad (8.2.17)$$

The relation (2.17) can also be found by using the fact that the expression $\exp(- :f:) \exp(:f:)$ is the identity operator and employing (2.11) and its analog for $:h:$,

$$\begin{aligned} \exp(:f:) : g :: h : \exp(- :f:) &= \exp(:f:) : g : \exp(- :f:) \exp(:f:) : h : \exp(- :f:) \\ &= : \exp(:f:)g :: \exp(:f:)h : . \end{aligned} \quad (8.2.18)$$

We can carry (2.15) a step further using (2.11) to find the relation

$$\exp(:f:) : g :^m \exp(- :f:) =: \exp(:f:)g :^m, \quad (8.2.19)$$

which is a generalization of (2.12). Moreover, (2.19) in turn, or direct use of (2.16) and (2.11), yields the relation

$$\begin{aligned} \exp(:f:) \exp(:g:) \exp(- :f:) &= \exp[: \exp(:f:)g :] \\ &= \exp[: g(\exp :f: z) :]. \end{aligned} \quad (8.2.20)$$

This relation gives a result for the multiplication of a particular combination of Lie transformations.

The relations (2.2), (2.13), and (2.14) have obvious generalizations to the case of several Lie operators. Consider, for example, the case of two Lie operators $:e:$ and $:f::$. Then, (2.2) has the generalization

$$\#e\#^m \#f\#^n : g :=: (:e :^m :f :^n g) : . \quad (8.2.21)$$

Indeed, suppose E is any function consisting of sums, products, sums of products, or even power series in two arguments. Then (2.2) has the generalization

$$E(\#e\#, \#f\#) : g :=: E(:e :, :f :)g : . \quad (8.2.22)$$

Analogous results hold for any number of Lie operators and functions of any number of arguments.

As for the relations (2.13) and (2.14), they can be generalized to any number of factors. For example, for the case of two factors, we have the results

$$\begin{aligned} & \exp(:e:) \exp(:f:) F(:g :, :h :, \dots) \exp(- :f:) \exp(- :e:) \\ & = \exp(\#e\#)(\exp \#f\#) F(:g :, :h :, \dots), \end{aligned} \quad (8.2.23)$$

$$\begin{aligned} & \exp(\#e\#) \exp(\#f\#) F(:g :, :h :, \dots) \\ & = F(: \exp(:e:) \exp(:f:)g :, : \exp(:e:) \exp(:f:)h :, \dots). \end{aligned} \quad (8.2.24)$$

Analogous results hold for any number of factors. Consider, for example, the factors that compose (7.6.3). In this case we have the result

$$\begin{aligned}
& \mathcal{M}F(:g,:,:h:\cdots)\mathcal{M}^{-1} \\
&= \exp(:f_2^c:) \exp(:f_2^a:) \exp(:f_3:) \exp(:f_4:) \cdots \times \\
& F(:g,:,:h:\cdots) \cdots \exp(-:f_4:) \exp(-:f_3:) \exp(-:f_2^a:) \exp(-:f_2^c:) \\
&= \exp(\#f_2^c\#) \exp(\#f_2^a\#) \exp(\#f_3\#) \exp(\#f_4\#) \cdots F(:g,:,:h:\cdots) \\
&= F(\exp(\#f_2^c\#) \exp(\#f_2^a\#) \exp(\#f_3\#) \exp(\#f_4\#) \cdots :g:\cdots) \\
&= F(:\exp(:f_2^c:)\exp(:f_2^a:)\exp(:f_3:)\exp(:f_4:)\cdots:g:\cdots) \\
&= F(:\mathcal{M}g,:,:h(\mathcal{M}z):\cdots) \\
&= F(:g(\mathcal{M}z):,:h(\mathcal{M}z):\cdots).
\end{aligned} \tag{8.2.25}$$

As special cases of (2.25) we have the result

$$\mathcal{M}:g(z):\mathcal{M}^{-1} =: \mathcal{M}g(z) := g(\mathcal{M}z):, \tag{8.2.26}$$

which is an extension of (2.12), and the result

$$\mathcal{M}[\exp :g(z):]\mathcal{M}^{-1} = \exp : \mathcal{M}g(z) := \exp :g(\mathcal{M}z):, \tag{8.2.27}$$

which is an extension of (2.20).

We close this section with another useful result for the multiplication of Lie transformations. It is the analog of formulas (3.7.33) and (3.7.34) for Lie operators. Suppose $:f:$ and $:g:$ are any two Lie operators. Then one has the BCH formula

$$\begin{aligned}
& \exp(s:f:) \exp(t:g:) = \exp(s:f:+t:g:) \\
& + (st/2)\{ :f:, :g: \} + (s^2t/12)\{ :f:, \{ :f:, :g: \} \} \\
& + (st^2/12)\{ :g:, \{ :g:, :f: \} \} + \cdots.
\end{aligned} \tag{8.2.28}$$

Moreover, using (5.3.14) and (2.2), (2.28) can also be written in the form

$$\exp(s:f:) \exp(t:g:) = \exp(:h:) \tag{8.2.29}$$

with

$$\begin{aligned}
h &= sf + tg + (st/2)[f,g] \\
& + (s^2t/12):f:^2g + (st^2/12):g:^2f + \cdots.
\end{aligned} \tag{8.2.30}$$

Exercises

8.2.1. Prove (2.2).

8.2.2. Prove (2.4).

8.2.3. Carry out the steps that lead to (2.5), (2.11), and (2.12). Also verify (2.5) term by term for at least the first few terms by comparing power series expansions.

8.2.4. Prove (2.13) and (2.14). Hint: Imitate the proof of (2.5) and (2.12). Also verify (2.13) term by term for at least the first few terms by comparing power series expansions.

8.2.5. Construct a general proof of (2.13) and (2.14) by employing the method used to prove (2.17).

8.2.6. Prove (2.20).

8.2.7. Prove (2.21) and (2.22).

8.2.8. Prove (2.23), (2.24), and (2.25). Show that (2.25) also holds for \mathcal{M} of the form (7.7.23).

8.2.9. Prove (2.30).

8.3 Questions of Order and other Miscellaneous Mysteries

Lie operators and Lie transformations have remarkable properties, and in many ways seem to lead lives of their own. The purpose of this section is to discuss various questions of operator ordering that are often confusing to the uninitiated, and sometimes puzzling to even the enlightened. We will also extend some previous results, and resolve some mysteries of sign that arose in previous sections.

8.3.1 Questions of Order and Map Multiplication

Suppose \mathcal{M}_f is a symplectic map that sends the general point z in phase space to the point \bar{z} , and suppose \mathcal{M}_g is another symplectic map that sends \bar{z} to the point $\bar{\bar{z}}$. The reason for our naming convention using the subscripts f and g will become apparent shortly. See Figure 3.1

Equivalently, in the context of charged particle beam transport, we may think of a beam that first passes through beam line element f , the action of which is described by the map \mathcal{M}_f , and then through beam line element g whose action is described by the map \mathcal{M}_g . See Figure 3.2.

Now consider the composite mapping \mathcal{M} which sends z to $\bar{\bar{z}}$ and which, following usual mathematical notation, would be written in the form

$$\mathcal{M} = \mathcal{M}_g \mathcal{M}_f, \quad (8.3.1)$$

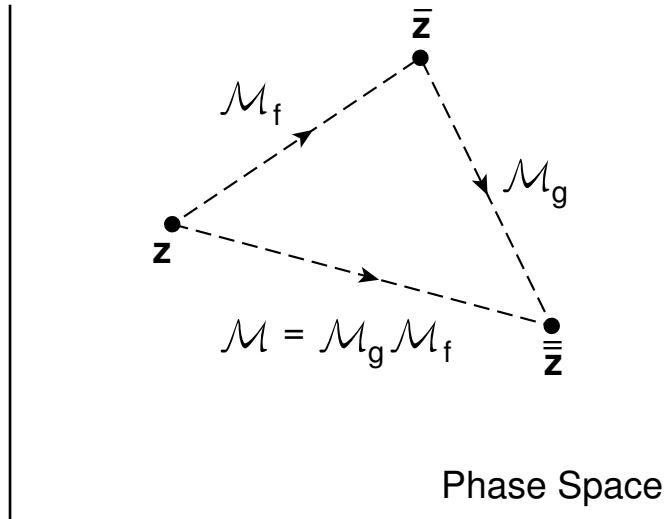


Figure 8.3.1: The composite action of two maps \mathcal{M}_f and \mathcal{M}_g .

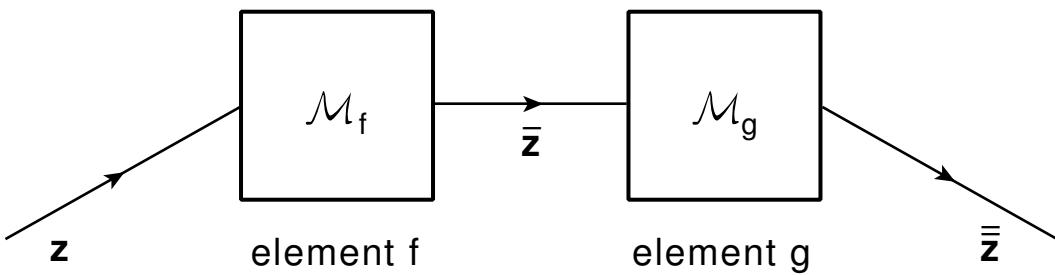


Figure 8.3.2: Successive passage of a trajectory with initial condition z through beam line elements f and g resulting in the intermediate condition \bar{z} and final condition $\bar{\bar{z}}$.

$$\mathcal{M}_f : z \rightarrow \bar{z}, \quad (8.3.2)$$

$$\mathcal{M}_g : \bar{z} \rightarrow \bar{\bar{z}}, \quad (8.3.3)$$

$$\mathcal{M}_g \mathcal{M}_f : z \rightarrow \bar{\bar{z}}. \quad (8.3.4)$$

Note that, when reading (3.1) from left to right, the maps \mathcal{M}_f and \mathcal{M}_g occur in the *opposite* order from which they actually act. See Fig. 3.2. That is, \mathcal{M}_f acts first, but appears last in (3.1); and \mathcal{M}_g acts last, but appears first in (3.1). This order follows the standard mathematical convention for maps (including matrices as a special case), and is in accord with the ordering used earlier in Section 6.2 and Equations (6.4.5) and (6.4.7).

Suppose, for purposes of discussion, that both \mathcal{M}_f and \mathcal{M}_g can be written in exponential form using single exponents,

$$\mathcal{M}_f = \exp(: f :), \quad (8.3.5)$$

$$\mathcal{M}_g = \exp(: g :). \quad (8.3.6)$$

In this case (3.2) and (3.3) can be written in the more explicit form

$$\bar{z}(z) = \mathcal{M}_f z = \exp(: f(z) :) z, \quad (8.3.7)$$

$$\bar{\bar{z}}(\bar{z}) = \mathcal{M}_g \bar{z} = \exp(: g(\bar{z}) :) \bar{z}. \quad (8.3.8)$$

Also, if we regard \bar{z} as a function of z , as done in (3.7), then (3.8) can also be written in the form

$$\bar{\bar{z}}(z) = \bar{\bar{z}}(\bar{z}(z)) = \exp(: g(\bar{z}(z)) :) \bar{z}(z). \quad (8.3.9)$$

[Note that the Poisson brackets implied in (3.9) can be evaluated using either the variables z or \bar{z} with the same result. See (6.3.10), (6.3.11), and (6.3.20).] Finally, suppose we substitute (3.7) into (3.9). Doing so gives the result

$$\bar{\bar{z}}(z) = \exp(: g(\bar{z}(z)) :) \exp(: f(z) :) z. \quad (8.3.10)$$

This result is simply (3.4) written in explicit form.

Next suppose the identity operator \mathcal{I} , written in the form

$$\mathcal{I} = \exp(: f(z) :) \exp(- : f(z) :), \quad (8.3.11)$$

is inserted right after the equal sign in (3.10). This insertion brings (3.10) to the form

$$\bar{\bar{z}}(z) = \exp(: f(z) :) \exp(- : f(z) :) \exp(: g(\bar{z}(z)) :) \exp(: f(z) :) z. \quad (8.3.12)$$

Consider the quantity $\exp(- : f(z) :) \exp(: g(\bar{z}(z)) :) \exp(: f(z) :)$ that appears in (3.12). According to (2.20) we have the result

$$\exp(- : f(z) :) \exp(: g(\bar{z}(z)) :) \exp(: f(z) :) = \exp(: \exp(- : f(z) :) g(\bar{z}(z)) :). \quad (8.3.13)$$

According to (3.7) and (5.4.11) we have the result

$$g(\bar{z}(z)) = g(\exp(: f(z) :) z) = \exp(: f(z) :) g(z). \quad (8.3.14)$$

With this information in hand, we may rewrite (3.13) in the form

$$\begin{aligned} & \exp(-: f(z) :) \exp(: g(\bar{z}(z)) :) \exp(: f(z) :) \\ &= \exp(: \exp(-: f(z) :) g(\bar{z}(z)) :) \\ &= \exp(: \exp(-: f(z) :) \exp(: f(z) :) g(z) :) \\ &= \exp(: g(z) :). \end{aligned} \quad (8.3.15)$$

Finally, use of (3.15) in (3.12) gives the remarkable result

$$\bar{\bar{z}}(z) = \exp(: f(z) :) \exp(: g(z) :) z. \quad (8.3.16)$$

Observe that (3.4) can be written in the form

$$\bar{\bar{z}}(z) = \mathcal{M}z = \mathcal{M}_g\mathcal{M}_fz, \quad (8.3.17)$$

while (3.16) can be written in the form

$$\bar{\bar{z}}(z) = \mathcal{M}z = \mathcal{M}_f\mathcal{M}_gz. \quad (8.3.18)$$

What is going on here to produce two seemingly contradictory results? The difference between (3.17) and (3.18) is as follows: In (3.17), as examination of (3.7), (3.8), and (3.10) shows, f is a function of the *initial* variable z while g is a function of the *intermediate* variable \bar{z} . By contrast in (3.18), as (3.16) shows, *both* f and g are functions of the *initial* variable z . What we have learned is that if a beam passes successively through beam line elements f and g , and in that order, then the map for the composite system is $\mathcal{M}_f\mathcal{M}_g$ where both f and g are taken to be functions of the initial variable z . We see that when the factors in a map (which is expressed as a product of factors all of the initial variable z) are read from left to right, they are encountered in the *same* order as they are encountered by the beam.

Strictly speaking, the last two sentences in the previous paragraph have been shown to be true for two maps with both maps assumed to be expressible in exponential form using single exponents as in (3.5) and (3.6). However, by similar arguments, analogous results can be shown to hold in general. For example, suppose \mathcal{M}_f , \mathcal{M}_g , and \mathcal{M}_h are any three (analytic) maps. Then, from (7.8.1), \mathcal{M}_f has a factorization of the form

$$\mathcal{M}_f = \exp(: f_1 :) \exp(: f_2^c :) \exp(: f_2^a :) \exp(: f_3 :) \exp(: f_4 :) \cdots ; \quad (8.3.19)$$

and \mathcal{M}_g and \mathcal{M}_h have similar factorizations. Suppose a trajectory with *initial* condition z^i passes successively through the beam line elements f , g , and h described by the maps \mathcal{M}_f , \mathcal{M}_g , and \mathcal{M}_h , respectively. Then the *final* condition z^f as a result of this passage is given by the relation

$$\begin{aligned} z^f(z^i) &= \mathcal{M}_f\mathcal{M}_g\mathcal{M}_hz^i \\ &= \exp(: f_1 :) \exp(: f_2^c :) \exp(: f_2^a :) \exp(: f_3 :) \exp(: f_4 :) \cdots \times \\ &\quad \exp(: g_1 :) \exp(: g_2^c :) \exp(: g_2^a :) \exp(: g_3 :) \exp(: g_4 :) \cdots \times \\ &\quad \exp(: h_1 :) \exp(: h_2^c :) \exp(: h_2^a :) \exp(: h_3 :) \exp(: h_4 :) \cdots z^i \end{aligned} \quad (8.3.20)$$

when all the homogeneous polynomials f_j , g_j , and h_j are taken to be functions of the initial variable z^i .

8.3.2 Questions of Order in the Linear Case

We are now prepared to revisit some questions of order that were passed over in earlier sections. In (7.2.3) and (7.2.8) we constructed polynomials f_2^a and f_2^c in such a way that

$$\exp(: f_2^a :)z_b = \sum_d P_{bd} z_d, \quad (8.3.21)$$

$$\exp(: f_2^c :)z_d = \sum_e O_{de} z_e. \quad (8.3.22)$$

In so doing, we made the z 's transform as the components of a vector under the actions of the matrices P and O . Then we found the result

$$\exp(: f_2^c :) \exp(: f_2^a :) z_b = \sum_e (PO)_{be} z_e. \quad (8.3.23)$$

See (7.2.10) and (7.2.11). Here both f_2^c and f_2^a were quadratic polynomials in the variable z . Since P and O were defined in such a way that the z 's transformed under their actions as the components of a vector, we expect to have the matrix product PO as the result of O acting first and then followed by P . On the other hand, since both $\exp(: f_2^c :)$ and $\exp(: f_2^a :)$ are functions of the initial variable z , we expect to have the operator product $\exp(: f_2^c :) \exp(: f_2^a :)$ when $\exp(: f_2^c :)$ acts first and is then followed by the action of $\exp(: f_2^a :)$. But, as we see from (3.21) and (3.22), $\exp(: f_2^a :)$ corresponds to P and $\exp(: f_2^c :)$ corresponds to O . Thus, the orders on both sides of (3.23) are just as they should be.

Next consider the operator and matrix orders in (7.7.34). Here the operator and matrix orders are the *same* rather than reversed! But look at (7.7.26). We see that in this case the z 's transformed under the action of the *transposed* matrix M^T . Thus the definition of M as given in (7.7.26) is different from the definitions of P and O as given by (3.21) and (3.22). One definition involved matrix transposition and the other did not. What (7.7.26) and (7.7.34) teach us is that the inclusion of a transpose in the definition of M makes it possible to have both operator and matrix orders the same.

To gain further insight into what is going on, it is useful to consider the general subject of linear operators and matrices. We begin our discussion by examining simple transformation properties of vectors under the action of linear operators. Consider a vector space V spanned by basis vectors e_α that are orthonormal under some scalar product $(,)$. Let \mathcal{L} be a *linear* operator that sends V into itself. Suppose \mathcal{L} sends e_α into f_α . Then we have a relation of the form

$$f_\alpha = \mathcal{L}e_\alpha = \sum_\beta L_{\beta\alpha} e_\beta. \quad (8.3.24)$$

The coefficients $L_{\beta\alpha}$ are given in terms of the scalar product by the matrix elements

$$L_{\beta\alpha} = (e_\beta, f_\alpha) = (e_\beta, \mathcal{L}e_\alpha). \quad (8.3.25)$$

Let \mathbf{A} be some vector in V . Since the e_α form a basis, \mathbf{A} must have an expansion of the form

$$\mathbf{A} = \sum_\alpha a_\alpha e_\alpha. \quad (8.3.26)$$

Consider a vector \mathbf{B} defined by

$$\mathbf{B} = \mathcal{L}\mathbf{A}. \quad (8.3.27)$$

It must have an expansion of the form

$$\mathbf{B} = \sum_{\beta} b_{\beta} \mathbf{e}_{\beta}. \quad (8.3.28)$$

From (3.25) through (3.28) and the orthonormality condition we find that the components b_{β} are given by the relation

$$\begin{aligned} b_{\beta} &= (\mathbf{e}_{\beta}, \mathbf{B}) = (\mathbf{e}_{\beta}, \mathcal{L}\mathbf{A}) = \sum_{\alpha} a_{\alpha} (\mathbf{e}_{\beta}, \mathcal{L}\mathbf{e}_{\alpha}) \\ &= \sum_{\alpha} L_{\beta\alpha} a_{\alpha}. \end{aligned} \quad (8.3.29)$$

We note that the summation in (3.24) is over the first index in L , and that in (10.6) is over the second. Let us rewrite (3.24) in the form

$$\mathbf{f}_{\beta} = \mathcal{L}\mathbf{e}_{\beta} = \sum_{\alpha} L_{\alpha\beta} \mathbf{e}_{\alpha} = \sum_{\alpha} (L^T)_{\beta\alpha} \mathbf{e}_{\alpha}. \quad (8.3.30)$$

(Note that the indices α and β are dummy indices, and can be changed at will.) Upon comparing (3.29) and (3.30), we see that if *components* are transformed by the matrix L , then *basis vectors* are transformed by the transpose matrix L^T , and vice versa.

8.3.3 Application to General Operators and General Monomials to Construct Matrix Representations

Let us apply what we have just learned to operators \mathcal{M} of the general form

$$\mathcal{M} = \exp(: f_1 :) \exp(: f_2^c :) \exp(: f_2^a :) \exp(: f_3 :) \exp(: f_4 :) \cdots, \quad (8.3.31)$$

and monomials $G(\mu; \nu)$ of the form (7.3.1). [Here the f_j are homogeneous polynomials, and are not to be confused with the \mathbf{f}_{α} of (3.24), which are abstract vectors.] We will view these monomials as basis vectors for the vector space of all polynomials. For this purpose, it is convenient to subsume the indices μ, ν into a single index, which we will again call r . (For a concrete way of making a one-to-one correspondence between the indices μ, ν and the integers r , see Section 27.2.) We will thus work with a set of basis monomials G_r , and in accord with (7.3.8) we will assign them a scalar product $\langle \cdot, \cdot \rangle$ by the rule

$$\langle G_{r'}, G_r \rangle = \delta_{r'r}. \quad (8.3.32)$$

Now let \mathcal{M} act on a general basis vector G_r . Then we find a result of the form

$$\mathcal{M}G_r = \sum_s M_{sr} G_s \quad (8.3.33)$$

where the coefficients M_{sr} are given in terms of the scalar product by the matrix elements

$$M_{sr} = \langle G_s, \mathcal{M}G_r \rangle. \quad (8.3.34)$$

Suppose \mathcal{N} is another operator of the form (3.31). Then we also have the relation

$$\mathcal{N}G_r = \sum_s N_{sr}G_s \quad (8.3.35)$$

with the matrix N_{sr} given by

$$N_{sr} = \langle G_s, \mathcal{N}G_r \rangle. \quad (8.3.36)$$

Now consider the effect of the operator product \mathcal{MN} on G_r . Here all the Lie generators f_j appearing in \mathcal{M} and \mathcal{N} , as well as all the functions G_r , are assumed to depend on the *same* set of variables z . Then, we find the result

$$\begin{aligned} \mathcal{MN}G_r &= \mathcal{M} \sum_s N_{sr}G_s = \sum_s N_{sr}\mathcal{M}G_s \\ &= \sum_s N_{sr} \sum_t M_{ts}G_t = \sum_{st} M_{ts}N_{sr}G_t \\ &= \sum_t (MN)_{tr}G_t. \end{aligned} \quad (8.3.37)$$

We note that the ordering of the subscripts in (3.33) through (3.37) is analogous to that used in (7.3.37) through (7.3.40). Indeed, let \mathcal{A} and \mathcal{B} be any two *linear operators*. They could, for example, be Lie operators, or products of Lie operators, or sums of products of Lie operators, or infinite sums of products, etc., including Lie transformations and their products such as occur in (3.31). Define associated matrices $O(\mathcal{A})$ and $O(\mathcal{B})$ by rules of the form

$$O_{sr}(\mathcal{B}) = \langle G_s, \mathcal{B}G_r \rangle. \quad (8.3.38)$$

Then, since the G 's form a basis (are a complete set), we have the results

$$\mathcal{B}G_r = \sum_s O_{sr}(\mathcal{B})G_s, \quad (8.3.39)$$

$$\begin{aligned} \mathcal{AB}G_r &= \mathcal{A} \sum_s O_{sr}(\mathcal{B})G_s = \sum_s O_{sr}(\mathcal{B})\mathcal{A}G_s \\ &= \sum_s O_{sr}(\mathcal{B}) \sum_t O_{ts}(\mathcal{A})G_t = \sum_{st} O_{ts}(\mathcal{A})O_{sr}(\mathcal{B})G_t \\ &= \sum_t (O(\mathcal{A})O(\mathcal{B}))_{tr}G_t. \end{aligned} \quad (8.3.40)$$

From (3.38) and (3.40) we obtain the general relation

$$O(\mathcal{AB}) = O(\mathcal{A})O(\mathcal{B}). \quad (8.3.41)$$

We have found a matrix representation of the algebra of linear operators acting on function space. It can be shown, in the case that these operators are Lie algebra or Lie group elements, that these matrices are related to the adjoint representation. See Section 8.9.²

Note that, for a 6-dimensional phase space and for polynomials of degree 0 through m , the matrices O are $[S(m, 6) + 1] \times [S(m, 6) + 1]$. See Section 7.10. For example, in the case $m = 4$, 210×210 matrices are required. And, in the case $m = 8$, 3003×3003 matrices are required.

8.3.4 Application to Linear Transformations of Phase Space

Let us also apply what we have learned to the subject of linear transformations of *phase space* into itself. Set up a Euclidean coordinate system in phase space with unit vectors e_a along the coordinate and momentum axes. Then the general point in phase space may be identified with the vector \mathbf{z} from the origin to that point, and \mathbf{z} may be written in the form

$$\mathbf{z} = \sum_a z_a e_a \quad (8.3.42)$$

where the z_a are the usual coordinate variables. Suppose \mathcal{L} is a linear transformation of phase space into itself, and suppose \mathcal{L} sends the vector \mathbf{z} to the vector $\bar{\mathbf{z}}$,

$$\bar{\mathbf{z}} = \mathcal{L}\mathbf{z}. \quad (8.3.43)$$

The vector $\bar{\mathbf{z}}$ must have an expansion of the form

$$\bar{\mathbf{z}} = \sum_b \bar{z}_b e_b. \quad (8.3.44)$$

Correspondingly, in analogy to (3.25) through (3.29), the quantities \bar{z}_b and z_a are related by the equation

$$\bar{z}_b = \sum_a L_{ba} z_a. \quad (8.3.45)$$

That is, the z 's transform as the *components* of a vector, as is consistent with relations of the form (7.1.1) through (7.1.3), (7.6.1), (3.21), and (3.22).

8.3.5 Dual role of the Phase-Space Coordinates z_a

So far we have regarded the z_a as the components of a vector as in (3.42). However, the z_a are also *functions* on phase space. Indeed the z_a are special cases of the functions G_r , and, according to (7.3.1) and (7.3.8), satisfy the orthonormality conditions

$$\langle z_a, z_b \rangle = \delta_{ab}. \quad (8.3.46)$$

²The fact that linear operators acting on function space can be represented by matrices is familiar to any student of Quantum Mechanics. In the context of differential equations and maps, the matrix representation can be realized by *Carleman linearization*, a construction suggested by Poincaré. See the references at the end of this chapter.

Suppose \mathcal{M} is of the form (7.7.28). Then, following (3.39), we find the result

$$\bar{z}_b(z) = \mathcal{M}z_b = \sum_a M_{ab}^1 z_a \quad (8.3.47)$$

where the matrix M^1 is given by the relation

$$M_{ab}^1 = \langle z_a, \mathcal{M}z_b \rangle. \quad (8.3.48)$$

In addition, according to (7.3.41) through (7.3.45) or relations of the form (3.41), we have the result

$$M^1 = \exp(F^1), \quad (8.3.49)$$

where

$$F_{ab}^1 = \langle z_a, : f_2 : z_b \rangle. \quad (8.3.50)$$

Now observe that (3.47) can also be written in the form

$$\bar{z}_b(z) = \sum_a [(M^1)^T]_{ba} z_a. \quad (8.3.51)$$

Thus in view of our earlier discussion, see (3.30) for example, and noting that a and b are dummy indices, we conclude the convention used in (7.7.26) is equivalent to viewing the z_a as *basis vectors* (which is consistent with treating the G_r as basis vectors) rather than as components of a vector.

We have learned that the z_a play a dual role. If they are viewed as the components of a displacement vector as in (3.42), then it is appropriate to write their transformation law in the form (3.45) or (7.6.1). If they are viewed as functions, and therefore as special cases of the basis vectors (functions) G_r , then it may be more convenient to write their transformation law in the form (3.47) or, more generally, (3.33).

8.3.6 Extensions

We now turn to extensions of two results found previously. In our initial discussion of symplectic maps in Section 6.1, a symplectic map was defined as a mapping of phase space into itself that obeyed certain equivalent conditions such as (6.1.3) or (6.1.6) or (6.1.10). In Chapter 7 we learned that symplectic maps can be written in terms of Lie transformations, and obtained the factorizations (7.7.23) and (7.8.1). We also know from (5.4.13) and its generalizations that Lie transformations act on functions, and that the action of a Lie transformation on a function is determined once its action is known on phase space. Indeed, if a and b are any functions, then from (7.7.23), or (7.8.1), and (5.4.10) we have the results

$$\mathcal{M}a(z) = a(\mathcal{M}z), \quad \mathcal{M}a(z)b(z) = \mathcal{M}a(z)\mathcal{M}b(z) = a(\mathcal{M}z)b(\mathcal{M}z). \quad (8.3.52)$$

Thus, we may also view symplectic maps as entities that act on functions. [Note that we have already encountered this idea in (6.3.6) and (7.1.11) when use is made of (7.7.11)]. Conversely, if we view symplectic maps as entities that act on functions, then, since the z_a are functions, we get the action of symplectic maps on phase space from (7.7.11).

The property (3.52) is a consequence of the fact that Lie transformations are isomorphisms with respect to (ordinary) multiplication. See Section 5.4. We also know from Section 5.4 that Lie transformations are isomorphisms with respect to Poisson bracket multiplication. See (5.4.14). It follows from (5.4.15) and its generalizations, and from (7.7.23) or (7.8.1), that we also have the result

$$\mathcal{M}[a, b] = [\mathcal{M}a, \mathcal{M}b]. \quad (8.3.53)$$

Again, this result should already be familiar. When combined with (3.48), it yields the result (6.3.20).

8.3.7 Sign Differences

The last question to be discussed in this section is the difference in sign between relations such as (5.5.1) and (7.2.3). The simple answer is that the sign in (5.5.1) was selected to achieve the correspondence (5.5.13), and that in (7.2.3) was selected to make (7.2.7) hold. But why should the signs turn out to be different? Our discussion of this topic may seem somewhat discursive. However, we shall learn some interesting concepts and facts along the way.

Exercise 3.7.33 studied how, given some matrix representation of a Lie algebra, one might find other similar or possibly different representations. Now let us carry out the analogous discussion for the corresponding Lie group. Suppose some set of matrices gives a representation of some group. To every representation matrix M we associate another matrix M' by the rule

$$M' = \bar{M} \quad (8.3.54)$$

where a bar denotes complex conjugation. Then these matrices satisfy the relation

$$M'_1 M'_2 = (M_1 M_2)', \quad (8.3.55)$$

and therefore also provide a representation of some group. If the matrices M are real, then nothing new has been found. However, if the matrices M are complex and the structure constants of the underlying Lie algebra cannot be made real by some appropriate basis choice, then one must determine whether the resulting group is the same as the original group. If the structure constants are real, then the group will be the same and the representation given by the matrices M' may be different than that given by the matrices M . For example, in the case of the group $SU(3)$, if the matrices M provide the representation $\Gamma(m, n)$, then the matrices M' provide the representation $\Gamma(n, m)$. See Section 5.8.

Suppose, instead of using (3.54) to define M' , we use the rule

$$M' = (M^T)^{-1} = (M^{-1})^T. \quad (8.3.56)$$

Then the M' matrices defined in this way also satisfy (3.55), and also provide a representation of the group in question. As an example, consider the case where the matrices M are symplectic. The symplectic condition (3.1.2) can be written in the form

$$M' = (M^T)^{-1} = J M J^{-1}. \quad (8.3.57)$$

From (3.57) we see that in this case the matrices M' and M are *similar*, and hence the representations of the group in question carried by M' and M are the *same*. As a second example, suppose the matrices M are unitary,

$$M^\dagger = (\bar{M})^T = M^{-1}. \quad (8.3.58)$$

Then we find the result

$$M' = (M^{-1})^T = \bar{M}. \quad (8.3.59)$$

In this case the definitions (3.54) and (3.56) coincide. What happens if the matrices M belong to $SU(2)$? Since these matrices are complex, we might hope that the “priming” operation (3.59) would give something new. However, since matrices in $SU(2)$ are 2×2 and have determinant +1, they must also be symplectic. See the comment after Exercise 3.1.3. Consequently, (3.57) must also hold, and we in fact find that both M' and M carry the *same* representation.

Let M^f be a symplectic matrix. Associate with M^f a symplectic map $\mathcal{M}(M^f)$ by the rule

$$\mathcal{M}(M^f)z_a = \sum_b [(M^f)^{-1}]_{ab} z_b. \quad (8.3.60)$$

Note that (3.60) is analogous to (7.7.26) except that M^T has been replaced by M^{-1} . Let M^g be another symplectic matrix, and make the definitions

$$\mathcal{M}_f = \mathcal{M}(M^f), \quad (8.3.61)$$

$$\mathcal{M}_g = \mathcal{M}(M^g). \quad (8.3.62)$$

Then if we regard \mathcal{M}_f and \mathcal{M}_g as composed of Lie transformations all involving the same variable z we find, in analogy to (7.7.33), the result

$$\begin{aligned} \mathcal{M}_f \mathcal{M}_g z_a &= \mathcal{M}(M^f) \mathcal{M}(M^g) z_a = \mathcal{M}(M^f) \sum_b [(M^g)^{-1}]_{ab} z_b \\ &= \sum_{b,c} [(M^g)^{-1}]_{ab} [(M^f)^{-1}]_{bc} z_c \\ &= \sum_c [(M^g)^{-1} (M^f)^{-1}]_{ac} z_c = \sum_c [(M^f M^g)^{-1}]_{ac} z_c \\ &= \mathcal{M}(M^f M^g) z_a. \end{aligned} \quad (8.3.63)$$

We see that the definition (3.60) makes it possible to have both operator and matrix orders the same just as the definition (7.7.26) did. This result is not surprising in view of (3.56) and (3.55).

From a group theory perspective, the advantage of (3.60) compared to (7.7.26) is that the computation of M^{-1} is a *group* operation whereas the computation of M^T is not.

Now that we have a relation that has both operator and matrix orders the same, we can compare their respective Lie algebras. Let S^f and S^g be two symmetric matrices and use them to define functions f_2 and g_2 as in (5.5.1) and (5.5.2),

$$f_2 = (1/2) \sum_{a,b} S^f_{ab} z_a z_b, \quad (8.3.64)$$

$$g_2 = (1/2) \sum_{a,b} S_{ab}^g z_a z_b. \quad (8.3.65)$$

Then we have results of the form

$$: f_2 : z = (-JS^f)z, \quad (8.3.66)$$

and hence

$$\exp(: f_2 :)z = \exp(-JS^f)z = [\exp(JS^f)]^{-1}z. \quad (8.3.67)$$

Upon comparing (3.60) and (3.67), we find the results

$$\mathcal{M}_f = \exp(: f_2 :) = \mathcal{M}[\exp(JS^f)], \quad (8.3.68)$$

$$\mathcal{M}_g = \exp(: g_2 :) = \mathcal{M}[\exp(JS^g)]. \quad (8.3.69)$$

Now consider the product $\exp(\epsilon : f_2 :) \exp(\epsilon : g_2 :) \exp(-\epsilon : f_2 :) \exp(-\epsilon : g_2 :)$ where ϵ is a small quantity. Then, as a consequence of (3.63), we find the relation

$$\begin{aligned} & \exp(\epsilon : f_2 :) \exp(\epsilon : g_2 :) \exp(-\epsilon : f_2 :) \exp(-\epsilon : g_2 :) \\ &= \mathcal{M}[\exp(\epsilon JS^f) \exp(\epsilon JS^g) \exp(-\epsilon JS^f) \exp(-\epsilon JS^g)]. \end{aligned} \quad (8.3.70)$$

The products occurring in (3.70) may be viewed as the group analog of what would be a commutator at the Lie algebraic level. Indeed, from (2.27) and (2.28) we find through terms of order ϵ^2 the result

$$\exp(\epsilon : f_2 :) \exp(\epsilon : g_2 :) \exp(-\epsilon : f_2 :) \exp(-\epsilon : g_2 :) = \exp(\epsilon^2 : [f_2, g_2] :). \quad (8.3.71)$$

Similarly, from (3.7.34) we find through terms of order ϵ^2 the result

$$\exp(\epsilon JS^f) \exp(\epsilon JS^g) \exp(-\epsilon JS^f) \exp(-\epsilon JS^g) = \exp(\epsilon^2 \{JS^f, JS^g\}). \quad (8.3.72)$$

Now compare (3.70) through (3.72). Doing so gives through terms of order ϵ^2 the result

$$\exp(\epsilon^2 : [f_2, g_2] :) = \mathcal{M}[\exp(\epsilon^2 \{JS^f, JS^g\})]. \quad (8.3.73)$$

We are ready for the final step. Let h_2 be the second-degree polynomial defined by the relation

$$h_2 = [f_2, g_2]. \quad (8.3.74)$$

Then, with S^h defined by

$$h_2 = (1/2) \sum_{a,b} S_{ab}^h z_a z_b, \quad (8.3.75)$$

we have the result

$$\mathcal{M}_h = \exp(: h_2 :) = \mathcal{M}[\exp(JS^h)]. \quad (8.3.76)$$

Now compare (3.73) and (3.76) to get the result

$$\mathcal{M}[\exp(\epsilon^2 JS^h)] = \mathcal{M}[\exp(\epsilon^2 \{JS^f, JS^g\})]. \quad (8.3.77)$$

We see that for consistency we must have the relation

$$JS^h = \{JS^f, JS^g\}. \quad (8.3.78)$$

This relation is identical to that in (5.5.13). The moral of this rather long tale is that the sign in relations such as (5.5.1) was chosen so that, when (3.60) is used, it is possible to have relations of the form (3.68) and to have both operator and matrix orders the same in (3.63); and when the orders are the same it is easy to compare Lie algebras (exponents) as was done in (3.68) through (3.73). On the other hand, the sign in relations such as (7.2.3) was chosen to achieve relations such as (7.2.7) and (7.2.9), which are to be compared to (3.67).

Exercises

8.3.1. Verify (3.20).

8.3.2. From (3.38) show that $O(\mathcal{A}) + O(\mathcal{B}) = O(\mathcal{A} + \mathcal{B})$.

8.3.3. Verify (3.52) using (5.4.13) and (7.7.23) or (7.8.1).

8.3.4. Verify (3.53) using (5.4.15) and (7.7.23) or (7.8.1).

8.3.5. Given (3.56), verify (3.55).

8.3.6. Show that the map (3.60) is indeed symplectic.

8.3.7. Verify (3.66) and (3.67).

8.3.8. Verify (3.70).

8.3.9. Verify (3.71) and (3.72).

8.4 Lie Concatenation Formulas

As an application of the formulas and ideas developed so far, consider the problem of computing the product of two symplectic maps when each is expressed in factored product form. This problem arises in accelerator physics, for example, in the case that one knows the effect of each of two beam elements separately, and one wants to know the net effect when one beam element is followed by another. For simplicity, in this section we will consider maps that send the origin into itself, i.e. maps of the form (7.6.3). The most general case of maps that include leading and trailing translations, i.e. maps of the form (7.7.23), will be treated in the next chapter.

Let \mathcal{M}_f and \mathcal{M}_g denote the symplectic maps given by the expressions

$$\mathcal{M}_f = \exp(: f_2^c :) \exp(: f_2^a :) \exp(: f_3 :) \exp(: f_4 :) \cdots, \quad (8.4.1)$$

$$\mathcal{M}_g = \exp(: g_2^c :) \exp(: g_2^a :) \exp(: g_3 :) \exp(: g_4 :) \cdots. \quad (8.4.2)$$

Also, let \mathcal{M}_h be the product of \mathcal{M}_f and \mathcal{M}_g ,

$$\mathcal{M}_h = \mathcal{M}_f \mathcal{M}_g. \quad (8.4.3)$$

The problem is to find polynomials h_2^c, h_2^a, h_3, \dots such that

$$\mathcal{M}_h = \exp(: h_2^c :) \exp(: h_2^a :) \exp(: h_3 :) \exp(: h_4 :) \dots. \quad (8.4.4)$$

That is, the problem is to express \mathcal{M}_h as given by (4.3) in the factored product form (4.4). For simplicity, only expressions for $h_2^c, h_2^a, h_3, \dots, h_8$ will be found explicitly. Here, as described in Section 8.3, all polynomials f_j, g_j , and h_j are taken to be functions of the same variable z .

Before proceeding further, it is necessary to establish a few simple facts. Suppose g_2 is a quadratic polynomial written in the form

$$g_2 = -(1/2) \sum_{de} S_{de} z_d z_e = -(1/2)(z, S z), \quad (8.4.5)$$

where S is some symmetric matrix. Suppose further that f_m is some homogeneous polynomial of degree m . Then $\exp(: g_2 :) f_m$ is also a homogeneous polynomial of degree m . Indeed, we have the result

$$\exp(: g_2 :) f_m(z) = f_m[\exp(: g_2 :) z] = f_m(M^g z), \quad (8.4.6)$$

where M^g is the linear transformation defined by the equation

$$M^g = \exp(JS). \quad (8.4.7)$$

See (5.4.11) and Section 7.2. Suppose further that g_2^c and g_2^a are quadratic polynomials written in the forms

$$g_2^a = -(1/2)(z, S^a z), \quad (8.4.8)$$

$$g_2^c = -(1/2)(z, S^c z). \quad (8.4.9)$$

Then we have the result

$$\exp(: g_2^c :) \exp(: g_2^a :) f_m(z) = f_m[\exp(: g_2^c :) \exp(: g_2^a :) z] = f_m(R^g z), \quad (8.4.10)$$

where R^g is the linear transformation defined by the equations

$$R^g = P^g O^g, \quad (8.4.11)$$

$$P^g = \exp(JS^a), \quad (8.4.12)$$

$$O^g = \exp(JS^c). \quad (8.4.13)$$

See Section 7.2. Let us introduce the symplectic map \mathcal{R}_g defined by the equation

$$\mathcal{R}_g = \exp(: g_2^c :) \exp(: g_2^a :). \quad (8.4.14)$$

What we have established is the relation

$$\mathcal{R}_g f_m(z) = f_m(R^g z). \quad (8.4.15)$$

Actually, (4.15) is not quite what will be needed. What we will need is the relation

$$(\mathcal{R}_g)^{-1} f_m(z) = f_m[(R^g)^{-1} z]. \quad (8.4.16)$$

This relation can be established in a similar fashion. Note that, thanks to the symplectic condition, the matrix $(R^g)^{-1}$ required in (4.16) is easily calculated using (3.1.9).

We are now ready to continue. Simply from its definition (4.3), \mathcal{M}_h can be written in the form

$$\mathcal{M}_h = \mathcal{R}_f \exp(: f_3 :) \exp(: f_4 :) \cdots \mathcal{R}_g \exp(: g_3 :) \exp(: g_4 :) \cdots. \quad (8.4.17)$$

Here we have used (4.14) and an analogous definition for \mathcal{R}_f . Next, by insertion of a factor of $\mathcal{R}_g(\mathcal{R}_g)^{-1}$, (4.17) can be rewritten in the form

$$\mathcal{M}_h = \mathcal{R}_f \mathcal{R}_g (\mathcal{R}_g)^{-1} \exp(: f_3 :) \exp(: f_4 :) \cdots \mathcal{R}_g \exp(: g_3 :) \exp(: g_4 :) \cdots. \quad (8.4.18)$$

Evidently, comparison of (4.4) and (4.18) shows that h_2^c and h_2^a are determined by the equation

$$\mathcal{R}_h = \mathcal{R}_f \mathcal{R}_g. \quad (8.4.19)$$

Indeed, we have the result

$$R^h = R^g R^f, \quad (8.4.20)$$

where R^g is defined by (4.11) through (4.13), and R^f and R^h are defined by analogous relations. See Section 8.3.

Next define a function F of the Lie operators $: f_3 :; : f_4 :; \dots$ by the relation

$$F(: f_3 :; : f_4 :; \dots) = \exp(: f_3 :) \exp(: f_4 :) \cdots. \quad (8.4.21)$$

Evidently (4.18) contains the factor $(\mathcal{R}_g)^{-1} F \mathcal{R}_g$. As a consequence of (2.25) we have the result

$$(\mathcal{R}_g)^{-1} F(: f_3 :; : f_4 :; \dots) \mathcal{R}_g = F(: f_3 [(R^g)^{-1} z] :; : f_4 [(R^g)^{-1} z] :; \dots). \quad (8.4.22)$$

In order to simplify further expressions, introduce the notation

$$f_m^{tr}(z) = f_m[(R^g)^{-1} z], \quad (8.4.23)$$

which indicates that the homogeneous polynomial $f_m(z)$ of degree m has been *transformed* to the new homogeneous polynomial $f_m[(R^g)^{-1} z]$. With this notation, (4.22) can be written in the more compact form

$$(\mathcal{R}_g)^{-1} F(: f_3 :; : f_4 :; \dots) \mathcal{R}_g = F(: f_3^{tr} :; : f_4^{tr} :; \dots) = \exp(: f_3^{tr} :) \exp(: f_4^{tr} :) \cdots. \quad (8.4.24)$$

Putting together the work done so far, one finds that (4.18) can also be written in the form

$$\mathcal{M}_h = \mathcal{R}_h \exp(: f_3^{tr} :) \exp(: f_4^{tr} :) \cdots \exp(: g_3 :) \exp(: g_4 :) \cdots. \quad (8.4.25)$$

Upon comparing (4.4) and (4.25), we find the result

$$\exp(: h_3 :) \exp(: h_4 :) \cdots = \exp(: f_3^{tr} :) \exp(: f_4^{tr} :) \cdots \exp(: g_3 :) \exp(: g_4 :) \cdots. \quad (8.4.26)$$

We are now prepared to compute h_3, h_4, \dots in terms of $f_3^{tr}, f_4^{tr}, \dots$ and g_3, g_4, \dots . The tool for doing so will be the BCH formula as given by (2.27) and (2.28). We will also use the degree function defined in (7.6.13) and the relation (7.6.14). They are reproduced below for ready reference,

$$\deg(f_m) = m, \quad (8.4.27)$$

$$\deg([f_m, f_n]) = m + n - 2. \quad (8.4.28)$$

Consider the result of combining all exponents on the left side of (4.26) into one grand exponent h by repeated use of the BCH formula. Then, thanks to (4.28), it is relatively easy to pick out and collect various terms according to their degree. One finds the result

$$h = h_3 + h_4 + \{(1/2)[h_3, h_4] + h_5\} + \cdots. \quad (8.4.29)$$

Next, consider the result of combining all exponents on the right side of (4.26) into one grand exponent e . One finds the result

$$e = \{f_3^{tr} + g_3\} + \{(1/2)[f_3^{tr}, g_3] + f_4^{tr} + g_4\} + \cdots. \quad (8.4.30)$$

Now compare (4.29) and (4.30). By equating terms of like degree, we immediately obtain the results

$$h_3 = f_3^{tr} + g_3, \quad (8.4.31)$$

$$h_4 = f_4^{tr} + g_4 + [f_3^{tr}, g_3]/2. \quad (8.4.32)$$

With further work, it is possible to find the polynomials h_5, h_6 , etc. Doing so one finds, for example, the results

$$h_5 = f_5^{tr} + g_5 - [g_3, f_4^{tr}] + \frac{1}{3} : g_3 :^2 f_3^{tr} - \frac{1}{6} : f_3^{tr} :^2 g_3, \quad (8.4.33)$$

$$\begin{aligned} h_6 &= f_6^{tr} + g_6 - [g_3, f_5^{tr}] + \frac{1}{2} : g_3 :^2 f_4^{tr} + \frac{1}{2} [f_4^{tr}, g_4] - \frac{1}{4} [f_4^{tr}, [f_3^{tr}, g_3]] \\ &- \frac{1}{4} [g_4, [f_3^{tr}, g_3]] + \frac{1}{24} : f_3^{tr} :^3 g_3 - \frac{1}{8} : g_3 :^3 f_3^{tr} \\ &+ \frac{1}{8} [f_3^{tr}, [g_3, [f_3^{tr}, g_3]]], \end{aligned} \quad (8.4.34)$$

$$\begin{aligned} h_7 &= f_7^{tr} + g_7 - : g_3 : f_6^{tr} - : g_4 : f_5^{tr} + \frac{1}{2} : g_3 :^2 f_5^{tr} \\ &+ \frac{1}{2} : f_4^{tr} :: g_3 : f_4^{tr} + : g_4 :: g_3 : f_4^{tr} + \frac{1}{3} : g_3 :: f_4^{tr} :: f_3^{tr} : g_3 \\ &- \frac{1}{6} : g_3 :^3 f_4^{tr} - \frac{1}{6} : g_4 :: f_3^{tr} :: g_3 : f_3^{tr} - \frac{1}{6} : f_4^{tr} :: f_3^{tr} :: g_3 : f_3^{tr} \\ &- \frac{1}{3} : g_4 :: g_3 :^2 f_3^{tr} - \frac{1}{3} [: g_3 : f_3^{tr}, : g_3 : f_4^{tr}] \\ &- \frac{1}{120} : f_3^{tr} :^4 g_3 - \frac{1}{30} : g_3 :: f_3^{tr} :^3 g_3 - \frac{1}{20} : g_3 :^2 f_3^{tr} :^2 g_3 \\ &+ \frac{1}{30} : g_3 :^4 f_3^{tr} + \frac{1}{30} [: f_3^{tr} : g_3, : f_3^{tr} :^2 g_3] + \frac{1}{15} [: g_3 : f_3^{tr}, : g_3 :^2 f_3^{tr}], \end{aligned} \quad (8.4.35)$$

$$\begin{aligned}
h_8 = & f_8^{tr} + g_8 - :g_3:f_7^{tr} - :g_4:f_6^{tr} - \frac{1}{2} :g_5:f_5^{tr} \\
& + \frac{1}{2} :g_3:^2 f_6^{tr} + \frac{1}{2} :f_5^{tr} ::g_3:f_4^{tr} + :g_4 ::g_3:f_5^{tr} + \frac{1}{2} :g_5 ::g_3:f_4^{tr} + \frac{1}{3} :g_4:^2 f_4^{tr} \\
& + \frac{1}{6} :f_4^{tr} ::g_4:f_4^{tr} - \frac{1}{6} :g_3:^3 f_5^{tr} - \frac{1}{12} :f_5^{tr} ::f_3^{tr} ::g_3:f_3^{tr} - \frac{1}{6} :g_3 ::f_5^{tr} ::g_3:f_3^{tr} \\
& - \frac{1}{12} :g_5 ::f_3^{tr} ::g_3:f_3^{tr} - \frac{1}{6} :g_5 ::g_3:^2 f_3^{tr} - \frac{1}{6} [:g_3:f_3^{tr},:g_3:f_5^{tr}] \\
& - \frac{1}{6} :f_4:^2:g_3:f_3^{tr} - \frac{1}{4} :g_3 ::f_4^{tr} ::g_3:f_4^{tr} - \frac{1}{3} :g_4 ::f_4^{tr} ::g_3:f_3^{tr} \\
& - \frac{1}{2} :g_4 ::g_3:^2 f_4^{tr} - \frac{1}{6} [:g_3:f_3^{tr},:g_4:f_4^{tr}] - \frac{1}{6} :g_4:^2:g_3:f_3^{tr} \\
& + \frac{1}{24} :f_4^{tr} ::f_3:^2:g_3:f_3^{tr} + \frac{1}{8} :g_3 ::f_4^{tr} ::f_3^{tr} ::g_3:f_3^{tr} + \frac{1}{8} :g_3:^2:f_4^{tr} ::g_3:f_3^{tr} \\
& + \frac{1}{24} :g_3:^4 f_4^{tr} - \frac{1}{24} [:g_3:f_4^{tr},:f_3^{tr} ::g_3:f_3^{tr}] - \frac{1}{12} [:g_3:f_4^{tr},:g_3:^2 f_3^{tr}] \\
& + \frac{1}{24} [:g_3:f_3^{tr},:f_4^{tr} ::g_3:f_3^{tr}] + \frac{1}{8} [:g_3 ::f_3^{tr},:g_3:^2 f_4^{tr}] + \frac{1}{24} :g_4 ::f_3^{tr}:^2:g_3:f_3^{tr} \\
& + \frac{1}{8} :g_4 ::g_3 ::f_3^{tr} ::g_3:f_3^{tr} + \frac{1}{8} :g_4 ::g_3:^3 f_3^{tr} : + \frac{1}{24} [:g_3:f_3^{tr},:g_4 ::g_3:f_3^{tr}] \\
& - \frac{1}{720} :f_3:^4:g_3:f_3^{tr} - \frac{1}{144} :g_3 ::f_3:^3:g_3:f_3^{tr} - \frac{1}{144} [:g_3:f_3^{tr},:f_3:^2:g_3:f_3^{tr}] \\
& - \frac{1}{72} :g_3:^2:f_3:^2:g_3:f_3^{tr} - \frac{1}{48} [:g_3:f_3^{tr},:g_3 ::f_3^{tr} ::g_3:f_3^{tr}] - \frac{1}{72} :g_3:^3:f_3^{tr} ::g_3:f_3^{tr} \\
& - \frac{1}{48} [:g_3:f_3^{tr},:g_3:^3 f_3^{tr}] - \frac{1}{144} :g_3:^5 f_3^{tr}.
\end{aligned} \tag{8.4.36}$$

Upon examining the expressions for h_4 , h_5 , h_6 , etc. we see that they contain both what we will call *direct* terms and what we will call *feed-up* terms. For example, consider h_4 as given by (4.32). It contains the direct terms f_4^{tr} and g_4 which come from like terms in \mathcal{M}_f and \mathcal{M}_g . It also contains the feed-up term $[f_3^{tr}, g_3]$ which comes from lower-order terms in \mathcal{M}_f and \mathcal{M}_g . We see that low-order nonlinearities, when combined, can lead to higher-order nonlinearities.

There is also a way of getting relations such as (4.31) through (4.36) directly without use of the BCH formula. Suppose we expand all the exponentials appearing in (4.26). Doing so gives a relation of the form

$$\begin{aligned}
& (1 + :h_3:+:h_3:^2/2!+\cdots)(1 + :h_4:+\cdots)\cdots \\
& = (1 + :f_3^{tr}:+:f_3^{tr}:^2/2!+\cdots)(1 + :f_4^{tr}:+\cdots)\cdots \times \\
& \quad (1 + :g_3:+:g_3:^2/2!+\cdots)(1 + :g_4:+\cdots)\cdots.
\end{aligned} \tag{8.4.37}$$

Next carry out the indicated multiplications and group terms according to the degree of the polynomial that would be produced if these terms were to act on z . We find the result

$$\begin{aligned}
& 1 + :h_3:+(:h_3:^2/2!+ :h_4:) + \cdots = 1 + (:f_3^{tr}:+:g_3:) \\
& \quad + (:f_3^{tr} ::g_3:+:f_3^{tr}:^2/2!+ :g_3:^2/2!+ :f_4^{tr}:+:g_4:) + \cdots.
\end{aligned} \tag{8.4.38}$$

Now equate terms of like degree to find results of the form

$$: h_3 :=: f_3^{tr} : + : g_3 :, \quad (8.4.39)$$

$$: h_3 :^2 / 2! + : h_4 : = : f_3^{tr} :: g_3 : + : f_3^{tr} :^2 / 2! + : g_3 :^2 / 2! + : f_4^{tr} : + : g_4 :. \quad (8.4.40)$$

Evidently (4.39) is equivalent to (4.31). Also, if we substitute (4.39) into (4.40) and rearrange terms, we find the result

$$\begin{aligned} : h_4 : &= : f_4^{tr} : + : g_4 : + : f_3^{tr} :: g_3 : - (1/2)(: f_3^{tr} :: g_3 : + : g_3 :: f_3^{tr} :) \\ &= : f_4^{tr} : + : g_4 : + (1/2)\{ : f_3^{tr} :: g_3 : \}. \end{aligned} \quad (8.4.41)$$

We see, with the aid of (4.3.14), that (4.41) corresponds to (4.32). Similarly, if we retain more terms in (4.38), we can derive the relations (4.33) through (4.36), etc.

There is an equivalent but somewhat more elegant way of carrying out the same calculation. As before we expand all the exponentials appearing on the right side of (4.26); however, we retain the left side in factored product form. As a result we find the relation

$$\begin{aligned} \exp(: h_3 :) \exp(: h_4 :) \cdots &= (1 + : f_3^{tr} : + : f_3^{tr} :^2 / 2! + \cdots) \times \\ (1 + : f_4^{tr} : + \cdots) \cdots &\times (1 + : g_3 : + : g_3 :^2 / 2! + \cdots)(1 + : g_4 : + \cdots) \cdots \\ &= 1 + (: f_3^{tr} : + : g_3 :) + (: f_3^{tr} :: g_3 : + : f_3^{tr} :^2 / 2! + : g_3 :^2 / 2! + : f_4^{tr} : + : g_4 :) + \cdots. \end{aligned} \quad (8.4.42)$$

From (4.42) we infer, as before, the relation (4.39). But now we multiply both sides of (4.42) on the left by $\exp(- : h_3 :)$ and make use (4.39) to find the result

$$\begin{aligned} \exp(: h_4 :) \cdots &= \exp[-(: f_3^{tr} : + : g_3 :)] \times [1 + (: f_3^{tr} : + : g_3 :) \\ &+ (: f_3^{tr} :: g_3 : + : f_3^{tr} :^2 / 2! + : g_3 :^2 / 2! + : f_4^{tr} : + : g_4 :) + \cdots]. \end{aligned} \quad (8.4.43)$$

Next we expand the exponential on the right side of (4.43) and carry out the indicated multiplications to get the relation

$$\exp(: h_4 :) \cdots = 1 + [(: f_3^{tr} :: g_3 : - : g_3 :: f_3^{tr} :) / 2 + : f_4^{tr} : + : g_4 :] + \cdots. \quad (8.4.44)$$

At this point we are ready to repeat the process: From (4.44) we infer, again as before, the relation (4.41). Next we multiply both sides of (4.44) on the left by $\exp(- : h_4 :)$, make use of (4.41), etc. This process is reminiscent of the factorization algorithm employed in Section 7.6. It is clear that it can be repeated indefinitely to find expressions for the $: h_m :$ for ever larger values of m . The only major problem (which also occurred before) is to write these expressions in commutator form so that the outer colons can be removed [using (5.3.14)] to obtain final results in the form (4.31) through (4.36), etc.

The problem of writing expressions in commutator form is solved by a result of *Dynkin*. Let x_1, x_2, \dots, x_n be a collection of noncommuting variables. Suppose P is a polynomial in these variables of the form

$$P = \sum a_{i_1 i_2 \dots i_k} x_{i_1} x_{i_2} \cdots x_{i_k}. \quad (8.4.45)$$

Note that each term contains k factors, not necessarily distinct, and therefore P is homogeneous of degree k . For each monomial $x_{i_1} x_{i_2} \cdots x_{i_k}$ form a related multiple commutator $(x_{i_1} x_{i_2} \cdots x_{i_k})^0$ by the rule

$$(x_{i_1} x_{i_2} \cdots x_{i_k})^0 = (1/k) \{ \cdots \{ x_{i_1}, x_{i_2} \}, x_{i_3} \}, \cdots x_{i_k} \}. \quad (8.4.46)$$

Also suppose it is known in principle that P can be written in terms of commutators (as is our situation thanks to the BCH formula). Then, Dynkin proved, *one* such commutator form for P is

$$P = \sum a_{i_1 i_2 \dots i_k} (x_{i_1} x_{i_2} \cdots x_{i_k})^0. \quad (8.4.47)$$

Here it is helpful to work out an example. In the process of calculating h_5 based on (4.37) one finds the intermediate result

$$: h_5 :=: f_5^{tr} : + : g_5 : + : f_4^{tr} :: g_3 : - : g_3 :: f_4^{tr} : + P \quad (8.4.48)$$

with P given by the relation

$$\begin{aligned} P(: f_3^{tr} :, : g_3 :) = & - : f_3^{tr} :^2 : g_3 : / 6 + : f_3^{tr} :: g_3 :: f_3^{tr} : / 3 \\ & + : f_3^{tr} :: g_3 :^2 / 3 - : g_3 :: f_3^{tr} :^2 / 6 - (2/3) : g_3 :: f_3^{tr} :: g_3 : \\ & + : g_3 :^2 : f_3^{tr} : / 3. \end{aligned} \quad (8.4.49)$$

The problem is to put P in commutator form. Following (4.46) gives the results

$$\begin{aligned} (: f_3^{tr} :^2 : g_3 :)^0 &= (1/3) \{ \{ : f_3^{tr} :, : f_3^{tr} : \}, : g_3 : \} = 0, \\ (: f_3^{tr} :: g_3 :: f_3^{tr} :)^0 &= (1/3) \{ \{ : f_3^{tr} :, : g_3 : \}, : f_3^{tr} : \}, \\ (: f_3^{tr} :: g_3 :^2)^0 &= (1/3) \{ \{ : f_3^{tr} :, : g_3 : \}, : g_3 : \}, \\ (: g_3 :: f_3^{tr} :^2)^0 &= (1/3) \{ \{ : g_3 :, : f_3^{tr} : \}, : f_3^{tr} : \}, \\ (: g_3 :: f_3^{tr} :: g_3 :)^0 &= (1/3) \{ \{ : g_3 :, : f_3^{tr} : \}, : g_3 : \}, \\ (: g_3 :^2 : f_3^{tr} :)^0 &= (1/3) \{ \{ : g_3 :, : g_3 : \}, : f_3^{tr} : \} = 0. \end{aligned} \quad (8.4.50)$$

Consequently, according to Dynkin, P can be written in the commutator form

$$\begin{aligned} P &= (1/9) \{ \{ : f_3^{tr} :, : g_3 : \}, : f_3^{tr} : \} + (1/9) \{ \{ : f_3^{tr} :, : g_3 : \}, : g_3 : \} \\ &\quad - (1/18) \{ \{ : g_3 :, : f_3^{tr} : \}, : f_3^{tr} : \} - (2/9) \{ \{ : g_3 :, : f_3^{tr} : \}, : g_3 : \}. \end{aligned} \quad (8.4.51)$$

Note that h_5 as given by (4.48) with P given by (4.51) bears only some resemblance to the h_5 in (4.33). However, repeated use of the antisymmetry condition (3.7.41) gives the results

$$\begin{aligned} \{ \{ : f_3^{tr} :, : g_3 : \}, : f_3^{tr} : \} &= - \{ \{ : f_3^{tr} :, : f_3^{tr} : \}, : g_3 : \}, \\ \{ \{ : f_3^{tr} :, : g_3 : \}, : g_3 : \} &= \{ \{ : g_3 :, : f_3^{tr} : \}, : g_3 : \}, \\ \{ \{ : g_3 :, : f_3^{tr} : \}, : f_3^{tr} : \} &= \{ \{ : f_3^{tr} :, : f_3^{tr} : \}, : g_3 : \}, \\ \{ \{ : g_3 :, : f_3^{tr} : \}, : g_3 : \} &= - \{ \{ : g_3 :, : g_3 : \}, : f_3^{tr} : \}. \end{aligned} \quad (8.4.52)$$

Inserting these results into P as given by (4.51) brings it to the form

$$P = (1/3) \{ \{ : g_3 :, : g_3 : \}, : f_3^{tr} : \} - (1/6) \{ \{ : f_3^{tr} :, : f_3^{tr} : \}, : g_3 : \}, \quad (8.4.53)$$

and h_5 as given by (4.48) with this P is the “colonized” version of (4.33).

This example illustrates both the use of Dynkin's theorem and a further complication. The complication is to determine when two expressions involving multiple Lie products are in fact equivalent when due account is taken of the antisymmetry condition (3.7.41) and/or the Jacobi condition (3.7.42). One method to compare expressions is to realize the Lie products in terms of commutators and then expand out all commutators to obtain a sum of monomials. If the two monomial sums agree term by term, then the two multiple Lie product expressions are equivalent. For example, expanding out P as given by (4.51) or (4.53) both produce (4.49). However, this expansion method is awkward since the expanded version may contain a very large number of terms. Another method is to employ some *basis* in standard form in which only certain multiple Lie product terms occur (with all other possible terms being brought to standard form by use of the antisymmetry and Jacobi conditions). Two multiple Lie product expressions are then equivalent if they agree term by term when re-expressed in some standard form basis. (Obviously their expanded commutator realizations will then also agree.) Two possible such bases are the *Hall* and *Chen-Fox-Lyndon-Shirshov* bases. See Appendix C.

There are yet another concerns when one is interested in numerical implementations. Because Lie multiplications (Poisson bracketing) are time consuming, it is desirable to minimize their number. For example, if only h_3 through h_5 were needed, the relations (4.32) and (4.33) could be rewritten and utilized in the form

$$h_4 = f_4^{tr} + g_4 - (1/2)[g_3, f_3^{tr}], \quad (8.4.54)$$

$$h_5 = f_5^{tr} + g_5 + [g_3, -f_4^{tr} + (1/3)[g_3, f_3^{tr}]] + (1/6)[f_3^{tr}, [g_3, f_3^{tr}]]. \quad (8.4.55)$$

In this form a total of only three Poisson brackets is required. Also there is the sometimes conflicting desire to rearrange terms so that quantities already calculated can be reused to maximum benefit. Thus, the strategy might change if one wished to compute h_6, h_7, \dots as well. Finally, in actual practice, the quantities g_3, g_4, \dots may be *sparse* (most possible monomials in them having vanishing coefficients). In this case it is desirable to arrange the Poisson bracket terms in such a way that Poisson bracket routines designed to exploit sparseness can be employed. [The expansions (4.54) and (4.55) above have been arranged to exploit possible sparseness in g_3 .] See Section 27.8.

We close this section by noting that there is yet another way of finding Lie concatenation formulas without needing the BCH coefficients, and it has the added advantage of immediately yielding results in Lie form. At present we do not have at our disposal all the tools required for its presentation. They will be developed in Chapter 10. See Section 10.6 where the subject of Lie concatenation is again discussed.

Exercises

8.4.1. Verify (4.16).

8.4.2. Verify (4.29) and (4.30).

8.4.3. Verify (4.31) through (4.33).

8.4.4. Starting with (4.37), verify (4.38) through (4.41).

8.4.5. Derive (4.48) and (4.49) from (4.37).

8.4.6. Expand out (4.51) and (4.53), and verify that both expansions produce (4.49).

8.4.7. Let us refer to a multiple commutator of the kind that appears in (4.46) as a *left nest*. Show that every left nest can be re-expressed as a *right nest*. That is show, by repeated use of the antisymmetry condition, that there is the relation

$$\begin{aligned} \{\cdots\{x_{i_1}, x_{i_2}\}, x_{i_3}\}, \cdots x_{i_k}\} &= (-1)^{k-1}\{x_{i_k}, \{x_{i_{k-1}}, \{x_{i_{k-2}}, \cdots \{x_{i_2}, x_{i_1}\} \cdots \} \} \\ &= (-1)^{k-1}\#x_{i_k}\#\#x_{i_{k-1}}\#\#x_{i_{k-2}}\#\cdots\#\#x_{i_2}\#\#x_{i_1}. \end{aligned} \quad (8.4.56)$$

8.5 Map Inversion and Reverse Factorization

Suppose the map \mathcal{M}_f is written in the factored product form

$$\mathcal{M}_f = \mathcal{R}_f \exp(: f_3 :) \exp(: f_4 :) \cdots. \quad (8.5.1)$$

Here, as in the previous section, \mathcal{R}_f denotes the map

$$\mathcal{R}_f = \exp(: f_2^c :) \exp(: f_2^a :) \quad (8.5.2)$$

that is associated with the linear transformation R^f given by the matrix relation

$$R^f = \exp(JS^a) \exp(JS^c). \quad (8.5.3)$$

It follows immediately from (5.1) that the *inverse* of \mathcal{M}_f has the representation

$$(\mathcal{M}_f)^{-1} = \cdots \exp(-: f_4 :) \exp(-: f_3 :)(\mathcal{R}_f)^{-1}. \quad (8.5.4)$$

Although (5.4) gives a possible representation for the inverse of \mathcal{M}_f , it is in the form of a *reverse* factorization. We would also like to have a representation in the standard *forward* factorization. That is, we wish also to have a representation for the inverse of \mathcal{M}_f in the form

$$(\mathcal{M}_f)^{-1} = \mathcal{R}_h \exp(: h_3 :) \exp(: h_4 :) \cdots. \quad (8.5.5)$$

See Section 7.8. This is easily accomplished with the aid of the concatenation formulas of the previous section. We simply write (5.4) and (5.5) in the form

$$\cdots [\exp(-: f_4 :)][\exp(-: f_3 :)][(\mathcal{R}_f)^{-1}] = \mathcal{R}_h \exp(: h_3 :) \exp(: h_4 :) \cdots \quad (8.5.6)$$

where we have used square brackets to indicate that the various maps are to be concatenated together. See Exercise 5.1. In particular, as needed in the next paragraph, we have the results

$$\mathcal{R}_h = (\mathcal{R}_f)^{-1}, \quad (8.5.7)$$

$$R^h = (R^f)^{-1}. \quad (8.5.8)$$

Note again, as a result of the symplectic condition, that the matrix $(R^f)^{-1}$ is easily calculated using (3.1.9).

The relation (5.5) also provides a procedure for reverse factorizing a map. Suppose we wish to represent \mathcal{M}_f in reverse factorized form. That is, we wish to find generators g_m such that

$$\mathcal{M}_f = \mathcal{R}_f \exp(: f_3 :) \exp(: f_4 :) \cdots = \cdots \exp(: g_4 :) \exp(: g_3 :) \mathcal{R}_g. \quad (8.5.9)$$

Simply take the inverse of both sides of (5.9) and use (5.5) to get the relation

$$(\mathcal{R}_g)^{-1} \exp(- : g_3 :) \exp(- : g_4 :) \cdots = \mathcal{R}_h \exp(: h_3 :) \exp(: h_4 :) \cdots. \quad (8.5.10)$$

From (5.10) and (5.7) we find the desired results

$$\mathcal{R}_g = (\mathcal{R}_h)^{-1} = \mathcal{R}_f, \quad (8.5.11)$$

$$g_m = -h_m. \quad (8.5.12)$$

Exercises

8.5.1. Verify (5.7) and (5.8). Show that h_3 , h_4 , and h_5 are given by the formulas

$$h_3 =, \quad (8.5.13)$$

$$h_4 =, \quad (8.5.14)$$

$$h_5 =. \quad (8.5.15)$$

8.5.2. Verify (5.11) and (5.12).

8.6 Taylor and Hybrid Taylor-Lie Concatenation and Inversion

Section 8.4 treated the problem of concatenating two maps, both of which were in factored-product Lie form, to obtain their product, again in factored-product Lie form. We also know that maps can be written in Taylor form. See Section 7.5. For some applications it is useful to have concatenation procedures for which one or more of the maps is in Taylor form. Several possibilities arise, as illustrated in Figure 6.1. Of course, we can always pass back and forth between the Taylor and factored-product Lie forms (see Section 7.6 and Exercise 7.6.12) so that in principle we already have all needed results. However, it is also desirable to have procedures that work directly with Taylor maps. Four cases of particular interest are discussed below.

Let us begin with the case where both \mathcal{M}_1 and \mathcal{M}_2 are in Taylor form, and we desire as well to represent the product $\mathcal{M}_3 = \mathcal{M}_1 \mathcal{M}_2$ in Taylor form. Suppose that \mathcal{M}_1 sends z to \bar{z} , and we express this fact in the form of a Taylor series that is truncated beyond terms of degree D ,

$$\mathcal{M}_1 : z \rightarrow \bar{z} \quad (8.6.1)$$

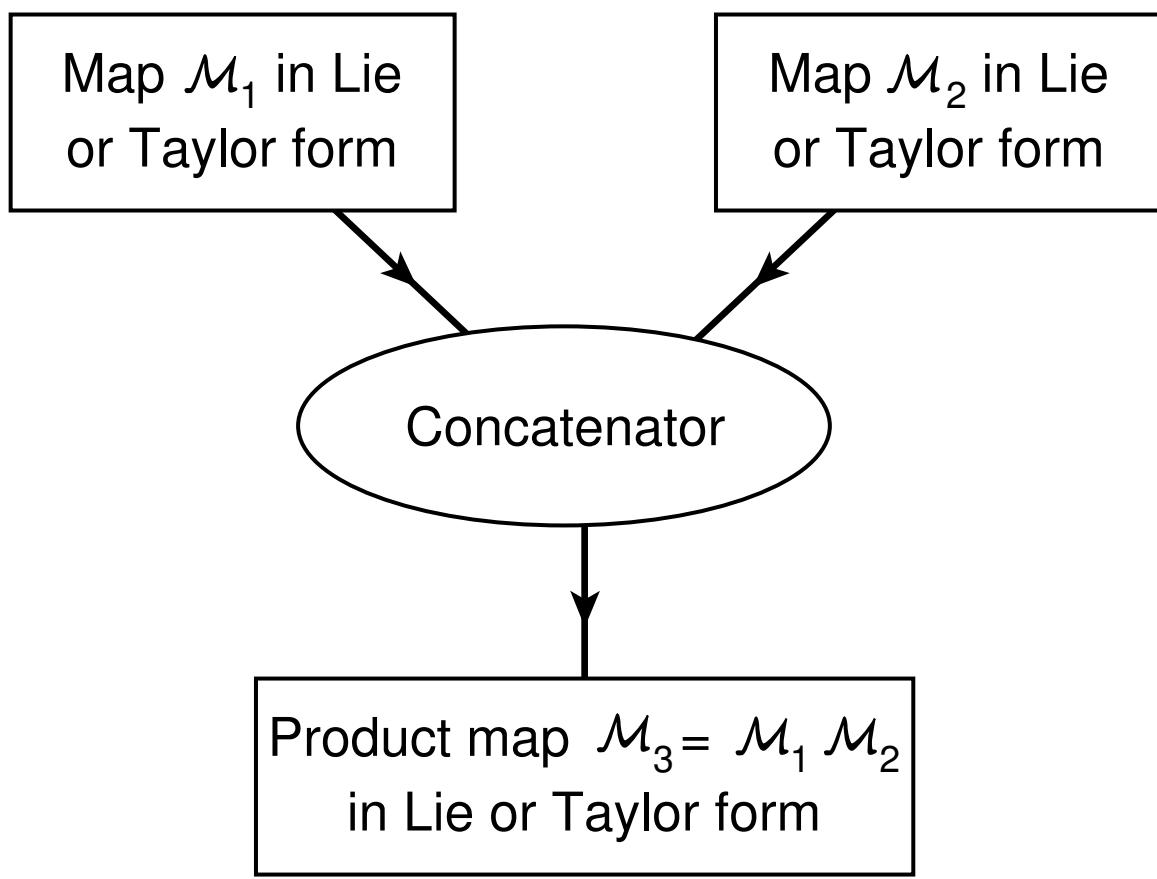


Figure 8.6.1: Various possibilities for the representation of maps in the operation of concatenation.

with

$$\bar{z}_a = \bar{z}_a(z) = \sum_{m=1}^D g_a^1(m; z). \quad (8.6.2)$$

Here the $g_a^1(m; z)$ denote homogeneous polynomials of degree m in the variables z . Similarly, \mathcal{M}_2 sends \bar{z} to $\bar{\bar{z}}$,

$$\mathcal{M}_2 : \bar{z} \rightarrow \bar{\bar{z}} \quad (8.6.3)$$

with

$$\bar{\bar{z}}_a = \bar{\bar{z}}_a(\bar{z}) = \sum_{m'=1}^D g_a^2(m'; \bar{z}). \quad (8.6.4)$$

What we desire is a representation for \mathcal{M}_3 of the form

$$\mathcal{M}_3 : z \rightarrow \bar{\bar{z}} \quad (8.6.5)$$

with

$$\bar{\bar{z}}_a = \bar{\bar{z}}_a(z) = \sum_{m''=1}^D g_a^3(m''; z). \quad (8.6.6)$$

Upon comparing (6.4) and (6.6) we see that the polynomials g_a^3 are given by the relations

$$g_a^3(m''; z) = P_{m''} \sum_{m'=1}^D g_a^2(m'; \bar{z}(z)). \quad (8.6.7)$$

Here $P_{m''}$ denotes a *projection* operator that retains only terms of degree m'' in the variables z .

To verify the truth of (6.7), we observe that the quantities $g_a^2(m'; \bar{z})$ in (6.4) are linear combinations of monomials in the \bar{z} 's of degree m' . When these monomials are computed using (6.2), the results are linear combinations of monomials in the z 's of degree as high as $m'D$. For example, second-order monomials in the \bar{z} 's are given by the relation

$$\bar{z}_c \bar{z}_d = \sum_{m'=1}^D g_c^1(m'; z) \sum_{m''=1}^D g_d^1(m''; z). \quad (8.6.8)$$

From these monomials we need to extract the terms of degree m in the z 's in order to find their contribution to the $g_a^3(m; z)$,

$$P_m(\bar{z}_c \bar{z}_d) = \sum_{m'+m''=m} g_c^1(m'; z) g_d^1(m''; z). \quad (8.6.9)$$

We conclude that the operation of concatenating maps in Taylor form involves the multiplication of truncated Taylor series, the extraction of terms of various degrees from the resulting products, and the assembly of linear combinations of these terms to form the truncated Taylor expansion (6.6) for the resulting map \mathcal{M}_3 . All these operations can in principle be carried out in a straight-forward manner to arbitrary order on a computer using various algorithms for *Truncated Power Series Algebra* (TPSA).

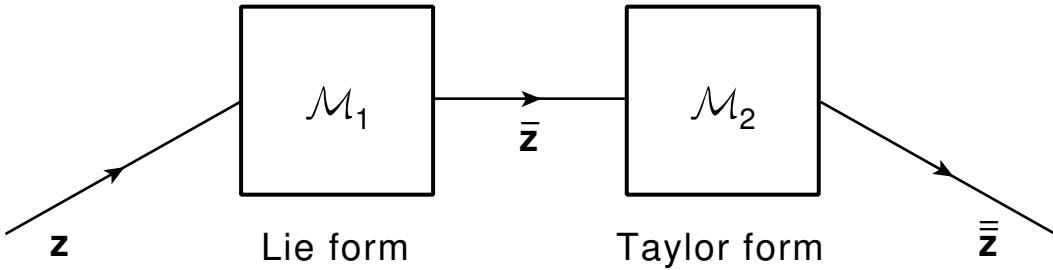


Figure 8.6.2: Product of a map in Lie form with a map in Taylor form.

In a second important case \mathcal{M}_1 is in Lie form, \mathcal{M}_2 is in Taylor form, and we desire or are content to know their product in Taylor form. See Figure 6.2. According to (6.4) we can express the action of \mathcal{M}_2 in Taylor form by writing the relations

$$\bar{\bar{z}}_a(\bar{z}) = T_a^D(\bar{z}) \quad (8.6.10)$$

where

$$T_a^D(\bar{z}) = \sum_{m'=1}^D g_a^2(m'; \bar{z}). \quad (8.6.11)$$

We also have the relations

$$\bar{z}_a(z) = \mathcal{M}_1 z, \quad (8.6.12)$$

$$g_a^2(m'; \bar{z}(z)) = g_a^2(m'; \mathcal{M}_1 z) = \mathcal{M}_1 g_a^2(m'; z). \quad (8.6.13)$$

Here we have used (5.4.13). Consequently, we have the result

$$\bar{\bar{z}}_a(z) = \mathcal{M}_1 T_a^D(z). \quad (8.6.14)$$

At this point we recognize that there are three common ways that \mathcal{M}_1 may be specified in Lie form. First, suppose that \mathcal{M}_1 is given in terms of a single exponent,

$$\mathcal{M}_1 = \exp(: h :), \quad (8.6.15)$$

where h has a homogeneous polynomial expansion of the form

$$h = h_2 + h_3 + \cdots + h_{D+1}. \quad (8.6.16)$$

[Note that, consistent with truncating maps beyond terms of degree D , we have truncated h beyond terms of degree $(D+1)$.] Maps of this kind arise from autonomous systems. See Sections 7.4 and 10.5. In this case we may expand $\exp(: h :)$ to get the result

$$\bar{\bar{z}}_a(z) = \sum_{\ell=0}^{\infty} (1/\ell!) : h :^\ell T_a^D(z). \quad (8.6.17)$$

Correspondingly, we have from (6.6) the result

$$g_a^3(m, z) = P_m \sum_{m'=1}^D \sum_{\ell=0}^{\infty} (1/\ell!) : h :^\ell g_a^2(m'; z). \quad (8.6.18)$$

In the circumstance that $h_2 = 0$, each sum over ℓ (for a given m and m') reduces to a finite sum because of (7.6.16). In the case that h_2 does not vanish, an infinite sum (for each value of m') is generally required. It can be shown, by an argument similar to that given in Section 10.5, that these sums always converge thanks to the $(1/\ell!)$ factor. However, all the caveats described in Section 4.1 concerning the use of Taylor series to evaluate the exponential function also apply here.

Suppose, as a second possibility, that \mathcal{M}_1 is given in the factored product form

$$\mathcal{M}_1 = \mathcal{R}_f \exp(: f_3 :) \exp(: f_4 :) \cdots \exp(: f_{D+1} :). \quad (8.6.19)$$

Let \mathcal{N}_f be the nonlinear part of \mathcal{M}_1 ,

$$\mathcal{N}_f = \exp(: f_3 :) \exp(: f_4 :) \cdots \exp(: f_{D+1} :). \quad (8.6.20)$$

According to (6.14) we need to find the quantities

$$\mathcal{M}_1 T_a^D(z) = \mathcal{R}_f \mathcal{N}_f T_a^D(z). \quad (8.6.21)$$

Introduce the intermediate results \tilde{T}_a^D and $\tilde{g}_a^3(m; z)$ defined by the equations

$$\tilde{T}_a^D(z) = \mathcal{N}_f T_a^D(z), \quad (8.6.22)$$

$$\tilde{g}_a^3(m; z) = P_m \sum_{m'=1}^D \mathcal{N}_f g_a^2(m'; z). \quad (8.6.23)$$

Then, by construction, we have the relation

$$\tilde{T}_a^D(z) = \sum_{m=1}^D \tilde{g}_a^3(m; z). \quad (8.6.24)$$

Next expand the exponentials appearing in (6.20). Doing so brings (6.23) to the form

$$\tilde{g}_a^3(m; z) = P_m \sum_{m'=1}^D \sum_{\ell_3=0}^{\infty} (1/\ell_3!) : f_3 :^{\ell_3} \cdots \sum_{\ell_{D+1}=0}^{\infty} (1/\ell_{D+1}!) : f_{D+1} :^{\ell_{D+1}} g_a^2(m'; z). \quad (8.6.25)$$

Again because of (7.6.16), each of the sums over $\ell_3 \cdots \ell_{D+1}$ (for a given m and m') reduces to a finite sum. The remaining task is to take \mathcal{R}_f into account. This is easily done. Again by construction we have the relation

$$g_a^3(m; z) = \mathcal{R}_f \tilde{g}_a^3(m; z). \quad (8.6.26)$$

Now use the analog of (8.4.15) to get the final result

$$g_a^3(m; z) = \tilde{g}_a^3(m; R^f z). \quad (8.6.27)$$

A third common possibility is that \mathcal{M}_1 arises in Lie form as a result of the use of some kind of Zassenhaus (symplectic integration) approximation. In this case \mathcal{M}_1 is typically a product of Lie transformations of the form

$$\mathcal{M}_1 = \exp(w_1 h : A :) \exp(w_2 h : B :) \cdots \exp(w_m h : A :) \quad (8.6.28)$$

where the w_j are various weights and h is the integration step size. See Section 10.8. Here the function A is typically a second-degree polynomial, and the function B has a homogeneous polynomial expansion consisting of terms of degree three and higher. If A is a second-degree polynomial, then $\exp(w_j h : A :)$ is a linear transformation that can be represented by some matrix R , and we can use methods analogous to (6.26) and (6.27). If B consists only of terms of degree three and higher, then $\exp(w_j h : B :)$ can be expanded in a Taylor series to give results analogous to (6.18) and for which only a finite number of ℓ values contribute.

We have seen how to find, in Taylor form, the product of a map in Lie form with a second map in Taylor form. Consider, as case three, the situation in which the two maps to be multiplied are both in Lie form, and we want their product in Taylor form. An obvious approach is to convert the second map from Lie to Taylor form, and then proceed as just described. This conversion is easily carried out as in Exercise 7.6.12. Equivalently, we may use the machinery just developed. The map \mathcal{M}_2 can always be written as

$$\mathcal{M}_2 = \mathcal{M}_2 \mathcal{I} \quad (8.6.29)$$

where \mathcal{I} is the identity map. But the identity map has the immediate Taylor expansion

$$\mathcal{I}z_a = z_a. \quad (8.6.30)$$

Therefore, to find \mathcal{M}_2 in Taylor form, we simply concatenate \mathcal{M}_2 in Lie form with the identity map \mathcal{I} in Taylor form.

As a generalization of this approach, suppose we wish to concatenate m maps in Lie form and obtain the net result in Taylor form,

$$\mathcal{M}_{\text{net}} = \mathcal{M}_1 \mathcal{M}_2 \mathcal{M}_3 \cdots \mathcal{M}_m. \quad (8.6.31)$$

Rewrite the desired result in the form

$$\mathcal{M}_{\text{net}} = \mathcal{M}_1 (\mathcal{M}_2 (\mathcal{M}_3 \cdots (\mathcal{M}_m \mathcal{I}) \cdots)), \quad (8.6.32)$$

and observe that each map \mathcal{M}_j in Lie form is now to be concatenated with a map in Taylor form to produce a map again in Taylor form. Thus, after m concatenations [namely, $\mathcal{M}_m \mathcal{I}$, $\mathcal{M}_{m-1}(\mathcal{M}_m \mathcal{I})$, $\mathcal{M}_{m-2}(\mathcal{M}_{m-1}(\mathcal{M}_m \mathcal{I}))$, etc.] we obtain \mathcal{M}_{net} in Taylor form. At this stage we may, if desired, obtain \mathcal{M}_{net} in Lie form from \mathcal{M}_{net} in Taylor form by carrying out the steps of the Factorization Theorem of Section 7.6.

A fourth case of interest is that in which the two maps to be multiplied are both in Lie form, and we also want their product in Lie form. This case has already been discussed in Section 8.4 where the BCH formula was used to find the quantities h_3, h_4, \dots in the relation (4.26). The result was explicit formulas of the form (4.31) through (4.36). These formulas become ever more complicated as the order is increased.

If one is content with numerical results, which is often the case, then the h_m can be computed algorithmically for any order m , without recourse to the BCH series, by Taylor methods as described above. To be explicit, and with reference to (4.26), define variables $\bar{z}_a(z)$ by the relations

$$\bar{z}_a = \exp(: f_3^{tr} :) \exp(: f_4^{tr} :) \cdots \exp(: g_3 :) \exp(: g_4 :) \cdots z_a. \quad (8.6.33)$$

Then we have the result

$$\exp(: h_3 :) \exp(: h_4 :) \cdots z_a = \bar{z}_a(z). \quad (8.6.34)$$

Next let \mathcal{T}^D be a *truncation* operator that acts on Taylor series. It is defined to be a linear operator that retains all terms in a Taylor series of degree less than or equal to D , and discards all terms of degree greater than D . With the aid of this operator we may define truncated Taylor series $T_a^D(z)$ by the relation

$$T_a^D(z) = \mathcal{T}^D \bar{z}_a = \mathcal{T}^D \exp(: f_3^{tr} :) \exp(: f_4^{tr} :) \cdots \exp(: g_3 :) \exp(: g_4 :) \cdots z_a. \quad (8.6.35)$$

Evidently, in view of (7.6.14), to compute the T_a^D it is only necessary to retain the factors containing $f_3^{tr} \cdots f_{D+1}^{tr}$ and $g_3 \cdots g_{D+1}$ in (6.35). Moreover, only a finite number of terms need be retained in each exponential series. Therefore, for a fixed D , only a finite number of operations are required to evaluate (6.35) to find the T_a^D . Finally, examination of the proof of the Factorization Theorem of Section 7.6 shows that the desired quantities $h_3 \cdots h_{D+1}$ can be found from the T_a^D by a finite number of operations. Moreover, unlike the case of the BCH series whose coefficients are very complicated, the only coefficients required are simply the factorials in the exponential series and those that arise in the use of (7.6.24).

In summary the virtue of the Taylor method just described is that, even when the maps to be concatenated are both in Lie factored product form and the desired product is also required in this form, results can be obtained to any desired degree ($D + 1$) in the Lie generators by means of a relatively simple algorithm. The price to be paid for this simplicity is increased computation [compared to that required for the direct formulas of the form (4.31) through (4.36)] and the increased (but temporary) storage associated with the intermediate truncated Taylor series T_a^D (see Section 7.9).

The last topic to be discussed in this section is the inversion of maps in Taylor form. By way of introduction, consider the simple quadratic equation

$$\bar{x} = \alpha x + \beta x^2. \quad (8.6.36)$$

This equation can immediately be solved to find x in terms of \bar{x} ,

$$x = \{-\alpha \pm [\alpha^2 + 4\beta\bar{x}]^{1/2}\}/(2\beta). \quad (8.6.37)$$

The solution that vanishes when $\bar{x} = 0$ has the expansion

$$\begin{aligned} x &= \{-\alpha + [\alpha^2 + 4\beta\bar{x}]^{1/2}\}/(2\beta) \\ &= [\alpha/(2\beta)]\{-1 + [1 + 4\beta\bar{x}/\alpha^2]^{1/2}\} \\ &= [\alpha/(2\beta)]\{(1/2)(4\beta\bar{x}/\alpha^2) - (1/8)(4\beta\bar{x}/\alpha^2)^2 + (1/16)(4\beta\bar{x}/\alpha^2)^3 + \cdots\} \\ &= (1/\alpha)\bar{x} - (\beta/\alpha^3)\bar{x}^2 + (2\beta^2/\alpha^5)\bar{x}^3 + \cdots. \end{aligned} \quad (8.6.38)$$

Equation (6.36) may be viewed as a one-dimensional Taylor map that sends x to \bar{x} , and (6.38) is the Taylor expansion of its inverse.

Suppose we had not been able to solve (6.36) explicitly for x as in (6.37). Is there any other way to obtain the inverse series (6.38)? The answer is yes. The inverse series can also be found by a process of *recursion* or *iteration*: First rewrite (6.36) in the form

$$x = (1/\alpha)\bar{x} - (\beta/\alpha)x^2 = (1/\alpha)\bar{x} + n(x) \quad (8.6.39)$$

where $n(x)$ is the *nonlinear* term

$$n(x) = -(\beta/\alpha)x^2. \quad (8.6.40)$$

Now consider the recursion relation for functions $x^{(m)}(\bar{x})$ specified by the rule

$$x^{(m+1)}(\bar{x}) = (1/\alpha)\bar{x} + n[x^{(m)}(\bar{x})], \quad (8.6.41)$$

with the starting relation

$$x^{(1)}(\bar{x}) = (1/\alpha)\bar{x}. \quad (8.6.42)$$

Upon carrying out the indicated operations, we find the results

$$\begin{aligned} m = 1 : \quad x^{(1)} &= (1/\alpha)\bar{x}, \\ m = 2 : \quad x^{(2)} &= (1/\alpha)\bar{x} - (\beta/\alpha)[(1/\alpha)\bar{x}]^2 = (1/\alpha)\bar{x} - (\beta/\alpha^3)\bar{x}^2, \\ m = 3 : \quad x^{(3)} &= (1/\alpha)\bar{x} - (\beta/\alpha)[(1/\alpha)\bar{x} - (\beta/\alpha^3)\bar{x}^2]^2 \\ &= (1/\alpha)\bar{x} - (\beta/\alpha^3)\bar{x}^2 + (2\beta^2/\alpha^2)\bar{x}^3 + O(\bar{x}^4), \text{ etc.} \end{aligned} \quad (8.6.43)$$

Evidently m applications of the rule (6.41) reproduces the series (6.38) through terms of degree m . We also note the possible appearance, at each stage, of still higher degree terms that may not yet be correct. We may remind ourselves not to bother computing these terms by using the truncation operator \mathcal{T}^m . With the aid of this operator, the recursion relation (6.41) can be modified to take the more convenient form

$$x^{(m+1)}(\bar{x}) = (1/\alpha)\bar{x} + \mathcal{T}^{m+1}n[x^{(m)}(\bar{x})]. \quad (8.6.44)$$

The iteration method we have just used to invert the simple quadratic equation (6.34) can also be used to invert general Taylor maps. Let us rewrite the Taylor representation (6.2) for the map \mathcal{M}_1 in the form

$$\bar{z}_{b'}(z) = \sum_b R_{b'b} z_b + N_{b'}(z) \quad (8.6.45)$$

where the quantities $N_{b'}(z)$ are nonlinear terms of degree 2 and higher. Equation (6.43) can be partially solved to give the result

$$z_a = (R^{-1}\bar{z})_a + \tilde{N}_a(z)$$

where \tilde{N}_a also contains terms only of degree 2 and higher, and is given by the relation

$$\tilde{N}_a = - \sum_{b'} (R^{-1})_{ab'} N_{b'}. \quad (8.6.46)$$

Note that we have assumed R^{-1} exists, as is required by the inverse function theorem for a map to have an inverse, and as will be the case for symplectic matrices. Now form the recursion relation

$$z_a^{(m+1)}(\bar{z}) = (R^{-1}\bar{z})_a + \mathcal{T}^{m+1}\tilde{N}_a[z^{(m)}(\bar{z})] \quad (8.6.47)$$

with the starting relation

$$z_a^{(1)}(\bar{z}) = (R^{-1}\bar{z})_a. \quad (8.6.48)$$

Application of this recursion relation D times produces the Taylor representation, through terms of degree D , for the map \mathcal{M}_1^{-1} .

Finally, we remark that the operations needed to carry out the recursion relation (6.47), as well as the Poisson brackets needed in procedures such as (6.18) and (6.35), can all be performed to arbitrary order on a computer programmed to handle TPSA.

Exercises

8.6.1. According to (4.31) and (4.32) the *direct* determination of h_3 and h_4 requires the computation of *one* Poisson bracket. How many Poisson brackets must be computed to find h_3 and h_4 by Taylor methods? Compare the amounts of work required for the direct and Taylor methods.

8.6.2. Prove that use of the recursion relation (6.47) does indeed produce the Taylor representation of the inverse of (6.45).

8.7 Working with Exponents

8.7.1 Formulas for Combining Exponents

The General Case

Sometimes, as will be shown later, it is useful to be able to write the product of two Lie transformations as a *single* Lie transformation. This is what the BCH formula (2.27) attempts to do. In general, there are no known convenient expressions for all the terms on the right side of (2.29). However, it is possible to sum the series completely with respect to s and the first few powers in t . One such result can be written in the form

$$h = sf + s : f : [1 - \exp(-s : f :)]^{-1}(tg) + O(t^2). \quad (8.7.1)$$

Here the operator expression involving $: f :$ is to be interpreted as the infinite series

$$\begin{aligned} s : f : [1 - \exp(-s : f :)]^{-1} &= s : f : [1 - \sum_{m=0}^{\infty} (-s : f :)^m / m!]^{-1} \\ &= s : f : [- \sum_{m=1}^{\infty} (-s : f :)^m / m!]^{-1} = 1 + (s/2) : f : + (s^2/12) : f :^2 + \dots \end{aligned} \quad (8.7.2)$$

Equations (2.27) and (7.1) may be combined to give the result

$$\exp(s : f :) \exp(t : g :) = \exp[s : f : + : \{s : f : [1 - \exp(-s : f :)]^{-1}(tg)\} : + O(t^2)]. \quad (8.7.3)$$

See Appendix C where the $O(t^2)$ term is also worked out.

Suppose we succeed in writing a product of two or more Lie transformations as a single Lie transformation. Then, as shown in Section 7.1, the map corresponding to the product of Lie transformations has an invariant function. See (7.1.12) through (7.1.14). We will learn later that generically symplectic maps do *not* have invariant functions. Correspondingly, the series (7.1) is generally divergent. We recall from Section 7.7 that the Lie algebra $sp_m(2n, \mathbb{R})$ is *infinite* dimensional. Typically what happens in the infinite dimensional case is that inverses of the form $[1 - \exp(-s : f :)]^{-1}$ may fail to exist. Other difficulties can also arise. Put another way, the BCH series (3.7.34) may have no domain of convergence in the case of an infinite dimensional Lie algebra. See Section 38.7.

The Case of $Sp(2)$

There is one instructive case for which the sum of the BCH series is known exactly. That is the case of $Sp(2)$ or, more generally, $Sp(2, \mathbb{C}) = SL(2, \mathbb{C})$. We will see that the result is quite complicated. Presumably yet more complicated formulas exist in the Platonic realm for the still more interesting cases of $Sp(4)$, $Sp(6)$, etc. But, to the author's knowledge, these formulas have not yet been brought down to Earth.

Given f_2 and g_2 , there are the associated maps

$$\mathcal{M}_f = \exp(: f_2 :), \quad (8.7.4)$$

$$\mathcal{M}_g = \exp(: g_2 :). \quad (8.7.5)$$

Our task is to find h_2 such that

$$\mathcal{M}_h = \exp(: h_2 :) = \mathcal{M}_f \mathcal{M}_g. \quad (8.7.6)$$

According to (3.64), (3.65), and (3.76) there are symmetric matrices S^f , S^g , and S^h associated with f_2 , g_2 , and h_2 respectively. Also, according to Section 8.3, our task is equivalent to that of finding the matrix S^h such that

$$\exp(JS^h) = \exp(JS^f) \exp(JS^g). \quad (8.7.7)$$

In the case of $sp(2)$, we know that the vector space of matrices of the form JS is spanned by the matrices B^0 , F , and G given by (5.6.7), (5.6.13), and (5.6.14). Upon comparison with the Pauli matrices (5.7.3), we find the results

$$B^0 = i\sigma^2, \quad (8.7.8)$$

$$F = -\sigma^1, \quad (8.7.9)$$

$$G = \sigma^3. \quad (8.7.10)$$

Let us define 3-vectors (with possibly complex components) \mathbf{v}^f , \mathbf{v}^g , and \mathbf{v}^h by the rules

$$\mathbf{v}^f \cdot \boldsymbol{\sigma} = v_1^f \sigma^1 + v_2^f \sigma^2 + v_3^f \sigma^3 = JS^f, \text{ etc.} \quad (8.7.11)$$

With these results and definitions, the condition (7.7) in the $Sp(2)$ case is equivalent to the requirement

$$\exp(\mathbf{v}^h \cdot \boldsymbol{\sigma}) = \exp(\mathbf{v}^f \cdot \boldsymbol{\sigma}) \exp(\mathbf{v}^g \cdot \boldsymbol{\sigma}). \quad (8.7.12)$$

The matrix $\exp(\mathbf{v}^h \cdot \boldsymbol{\sigma})$ can be found analytically using (5.7.40). We begin by noting that (5.7.40) can be written in the form

$$(\mathbf{u} \cdot \boldsymbol{\sigma})(\mathbf{v} \cdot \boldsymbol{\sigma}) = \sigma^0 \mathbf{u} \cdot \mathbf{v} + i(\mathbf{u} \times \mathbf{v}) \cdot \boldsymbol{\sigma}, \quad (8.7.13)$$

where \mathbf{u} and \mathbf{v} are any 3-vectors. Next, define the length of \mathbf{v} , denoted by v , by the rule

$$v = (\mathbf{v} \cdot \mathbf{v})^{1/2}. \quad (8.7.14)$$

Note that v may possibly be complex, and is specified only up to a sign. Using (7.13) and (7.14), we find the result

$$\begin{aligned} \exp(\mathbf{v} \cdot \boldsymbol{\sigma}) &= \cosh(\mathbf{v} \cdot \boldsymbol{\sigma}) + \sinh(\mathbf{v} \cdot \boldsymbol{\sigma}) \\ &= \sigma^0 \cosh v + \mathbf{v} \cdot \boldsymbol{\sigma} (\sinh v)/v. \end{aligned} \quad (8.7.15)$$

We observe that both the functions $\cosh v$ and $(\sinh v)/v$ are even in v , and therefore unaffected by the sign ambiguity in (7.14). At this point it is convenient to introduce the 3-vector $\boldsymbol{\tau}(\mathbf{v})$ defined by the equation

$$\boldsymbol{\tau}(\mathbf{v}) = \mathbf{v}(\tanh v)/v. \quad (8.7.16)$$

Equation (7.16) has as its inverse the relation

$$\mathbf{v}(\boldsymbol{\tau}) = \boldsymbol{\tau}(\tanh^{-1} \tau)/\tau \quad (8.7.17)$$

where

$$\tau = (\boldsymbol{\tau} \cdot \boldsymbol{\tau})^{1/2}. \quad (8.7.18)$$

Again observe that $(\tanh v)/v$ and $(\tanh^{-1} \tau)/\tau$ are even functions. With this definition, (7.15) can be written in the equivalent form

$$\exp(\mathbf{v} \cdot \boldsymbol{\sigma}) = [\cosh v][\sigma^0 + \boldsymbol{\tau}(\mathbf{v}) \cdot \boldsymbol{\sigma}]. \quad (8.7.19)$$

Now use (7.19) and (7.13) in (7.12). Doing so gives the result

$$\begin{aligned} \exp(\mathbf{v}^h \cdot \boldsymbol{\sigma}) &= [\cosh v^h][\sigma^0 + \boldsymbol{\tau}(\mathbf{v}^h) \cdot \boldsymbol{\sigma}] \\ &= [\cosh v^f][\sigma^0 + \boldsymbol{\tau}(\mathbf{v}^f) \cdot \boldsymbol{\sigma}][\cosh v^g][\sigma^0 + \boldsymbol{\tau}(\mathbf{v}^g) \cdot \boldsymbol{\sigma}] \\ &= [\cosh v^f][\cosh v^g]\{\sigma^0[1 + \boldsymbol{\tau}(\mathbf{v}^f) \cdot \boldsymbol{\tau}(\mathbf{v}^g)] \\ &\quad + [\boldsymbol{\tau}(\mathbf{v}^f) + \boldsymbol{\tau}(\mathbf{v}^g) + i\boldsymbol{\tau}(\mathbf{v}^f) \times \boldsymbol{\tau}(\mathbf{v}^g)] \cdot \boldsymbol{\sigma}\}. \end{aligned} \quad (8.7.20)$$

Use (5.7.41) to equate like terms on both sides of (7.20), and thereby find the relations

$$\cosh v^h = [\cosh v^f][\cosh v^g][1 + \boldsymbol{\tau}(\mathbf{v}^f) \cdot \boldsymbol{\tau}(\mathbf{v}^g)], \quad (8.7.21)$$

$$(\cosh v^h)\boldsymbol{\tau}(\mathbf{v}^h) = [\cosh v^f][\cosh v^g][\boldsymbol{\tau}(\mathbf{v}^f) + \boldsymbol{\tau}(\mathbf{v}^g) + i\boldsymbol{\tau}(\mathbf{v}^f) \times \boldsymbol{\tau}(\mathbf{v}^g)]. \quad (8.7.22)$$

Upon dividing (7.22) by (7.21) we obtain the final and remarkable result

$$\boldsymbol{\tau}(\mathbf{v}^h) = [\boldsymbol{\tau}(\mathbf{v}^f) + \boldsymbol{\tau}(\mathbf{v}^g) + i\boldsymbol{\tau}(\mathbf{v}^f) \times \boldsymbol{\tau}(\mathbf{v}^g)][1 + \boldsymbol{\tau}(\mathbf{v}^f) \cdot \boldsymbol{\tau}(\mathbf{v}^g)]^{-1}. \quad (8.7.23)$$

Given \mathbf{v}^f and \mathbf{v}^g , (7.23) specifies $\boldsymbol{\tau}(\mathbf{v}^h)$ which, in turn by using (7.17), gives \mathbf{v}^h . Taken together, we will call (7.16), (7.17), and (7.23) the *Sp(2, C) BCH function*.

Let us examine the singularity structure of the *Sp(2, C) BCH function*, namely the relationship between \mathbf{v}^f , \mathbf{v}^g , and \mathbf{v}^h implied by (7.16), (7.17), and (7.23). We see from (7.16) that $\boldsymbol{\tau}(\mathbf{v})$ is analytic in \mathbf{v} for small \mathbf{v} , has poles when $v = i(\pi/2 + n\pi)$, and has an essential singularity at $v = \infty$. We also note that since \mathbf{v} is possibly complex, \mathbf{v} can tend toward infinity in various directions while v remains bounded. Thus the singularity structure at infinity is quite complicated. Near the origin $\boldsymbol{\tau}(\mathbf{v})$ has the convergent expansion

$$\boldsymbol{\tau}(\mathbf{v}) = \mathbf{v}(1 - v^2/3 + 2v^4/15 - 17v^6/315 + \dots). \quad (8.7.24)$$

We see from (7.17) that $\mathbf{v}(\boldsymbol{\tau})$ is analytic in $\boldsymbol{\tau}$ for small $\boldsymbol{\tau}$ and has branch points at $\tau = \pm 1$. It also has a pole in τ at $\tau = 0$ on the Riemann sheets reached by circling these branch points. Moreover, there is a complicated singularity at infinity. Near the origin on the *principal sheet* $\mathbf{v}(\boldsymbol{\tau})$ is analytic and has the convergent expansion

$$\mathbf{v}(\boldsymbol{\tau}) = \boldsymbol{\tau}(1 + \tau^2/3 + \tau^4/5 + \tau^6/7 + \dots). \quad (8.7.25)$$

Finally, we see that (7.23) has the denominator $[1 + \boldsymbol{\tau}(\mathbf{v}^f) \cdot \boldsymbol{\tau}(\mathbf{v}^g)]$ which can possibly vanish, but cannot vanish for small \mathbf{v}^f and \mathbf{v}^g because of (7.24). We conclude that the BCH series for

$Sp(2)$ converges for small \mathbf{v}^f and \mathbf{v}^g , but presumably cannot converge everywhere because of the singularities just described. This result is consistent with our expectations. We know that any symplectic matrix can be written in the form (3.8.24). If it were always possible to combine the two exponents in (3.8.24) into one grand exponent using the BCH series, then (3.7.97) could be written in the form (3.7.36), which we know is false. Indeed, the reader will have the pleasure of showing in Exercise 7.9 that the offending singularity is the pole at $\tau = 0$ on a nonprincipal sheet of $\mathbf{v}(\tau)$.

What is the source of all these singularities? The fault does not lie with the group $Sp(2, \mathbb{C})$ itself. Indeed, if elements of $Sp(2, \mathbb{C})$ are parameterized by 2×2 possibly complex matrices, the operation of group element multiplication is simply matrix multiplication, and entries in the product of two matrices are *entire* functions of the entries in the matrices being multiplied.³ Rather the fault lies in the use of canonical coordinates of the first kind, which is what the Ansatz (7.12) essentially does. See Section 7.9. And the use of canonical coordinates of the first kind depends on the properties of the exponential map. See Section 3.8. Thus, the source of singularities in this case can be traced back to the (not globally possible/successful) use of the exponential map for $Sp(2, \mathbb{C})$. Seeking the impossible results in singularities.

8.7.2 Nature of Single Exponential Form

Let us explore further what elements of $Sp(2, \mathbb{R})$ can be written in single exponential form. In the case of two-dimensional phase space, the most general (real) f_2 can be written in the form

$$f_2 = -(bp^2 + 2aqp + cq^2)/2, \quad (8.7.26)$$

where a, b, c are (real) parameters. We define an associated symplectic matrix $R(a, b, c)$ by the rule

$$\exp(: f_2 :) z_d = \sum_e R_{de} z_e. \quad (8.7.27)$$

Our goal is to find an explicit expression for $R(a, b, c)$ in terms of the quantities a, b, c . So doing amounts to finding the exponential map from $sp(2, \mathbb{R})$ to $Sp(2, \mathbb{R})$.⁴

Direct calculation gives the result

$$: f_2 : z_d = -(1/2) : (bp^2 + 2aqp + cq^2) : z_d = \sum_e F_{de} z_e \quad (8.7.28)$$

where F is the Hamiltonian matrix

$$F = \begin{pmatrix} a & b \\ -c & -a \end{pmatrix}. \quad (8.7.29)$$

We readily verify that F has the property

$$F^2 = \Delta I. \quad (8.7.30)$$

³An entire function is a function that is analytic everywhere except at infinity.

⁴As announced, we will seek results for the case of $Sp(2, \mathbb{R})$. Partial results are also known for the more complicated case of $Sp(4, \mathbb{R})$. In particular, there is an explicit formula for matrices of the form $\exp(JS^a)$. See the references at the end of this chapter.

Here Δ is the *discriminant* of the quadratic form (7.26),

$$\Delta = a^2 - bc. \quad (8.7.31)$$

We know from Section 7.2 or (7.3.41) that R is given by the relation

$$R = \exp(F) = \cosh(F) + \sinh(F). \quad (8.7.32)$$

The term $\cosh(F)$ has the expansion

$$\begin{aligned} \cosh(F) &= F^0 + F^2/2! + F^4/4! + \dots \\ &= I(1 + \Delta/2! + \Delta^2/4! + \dots) \\ &= I \cosh(\Delta^{1/2}). \end{aligned} \quad (8.7.33)$$

Here use has been made of (7.30). For $\sinh(F)$ we find the result

$$\begin{aligned} \sinh(F) &= F + F^3/3! + F^5/5! + \dots \\ &= F(I + F^2/3! + F^4/5! + \dots) \\ &= F(1 + \Delta/3! + \Delta^2/5! + \dots) \\ &= (F/\Delta^{1/2})(\Delta^{1/2} + \Delta^{3/2}/3! + \Delta^{5/2}/5! + \dots) \\ &= F[\sinh(\Delta^{1/2})]/\Delta^{1/2}. \end{aligned} \quad (8.7.34)$$

Note that both $\cosh(\Delta^{1/2})$ and $\{\sinh(\Delta^{1/2})]/\Delta^{1/2}\}$ are even functions of $\Delta^{1/2}$, and thus do not depend on which root we take in computing $\Delta^{1/2}$. In fact, they are *analytic* functions of Δ and hence of a, b, c . Putting everything in (7.32) together gives for R the result

$$R = \begin{pmatrix} \cosh(\Delta^{1/2}) + a[\sinh(\Delta^{1/2})]/\Delta^{1/2} & b[\sinh(\Delta^{1/2})]/\Delta^{1/2} \\ -c[\sinh(\Delta^{1/2})]/\Delta^{1/2} & \cosh(\Delta^{1/2}) - a[\sinh(\Delta^{1/2})]/\Delta^{1/2} \end{pmatrix}. \quad (8.7.35)$$

Let us compute the eigenvalues of R . It has the characteristic polynomial

$$P(\lambda) = \det(R - \lambda I) = \lambda^2 - 2\lambda \cosh(\Delta^{1/2}) + 1. \quad (8.7.36)$$

This polynomial has the roots

$$\lambda = \exp(\Delta^{1/2}), \exp(-\Delta^{1/2}). \quad (8.7.37)$$

Note that if a, b, c are real, then so is Δ . It follows that $\Delta^{1/2}$ is real if $\Delta \geq 0$, and pure imaginary if $\Delta < 0$. Correspondingly, the eigenvalues of R are real if $\Delta > 0$, and have the *hyperbolic* configuration shown in Case 1 of Figure 3.4.1. If $\Delta < 0$, then the eigenvalues of R are on the unit circle corresponding to the *elliptic* configuration shown in Case 3 of Figure 3.4.1. The possibilities $\Delta = 0$ and $\Delta = -\pi^2$ are discussed further in Exercise 7.11, and correspond to the *parabolic* and *inversion parabolic* configurations shown in Cases 4 and 5 of Figure 3.4.1. We note that the *inversion hyperbolic* configuration shown in Case 2 does *not* occur. It follows that such symplectic matrices cannot be written in single exponential form with real exponents. [The use of complex exponents may be possible in some cases. See Exercises 2.16 and 7.12. But we know that even this expedient fails for the matrices M and N given by (3.7.134) and (3.7.135).] We have proved earlier that all (real) symplectic matrices, including the inversion hyperbolic case, can be written in the product form (3.8.26) with real exponents. Also, all the coefficients in the BCH series are real. It follows that the BCH series must *diverge* if we try to combine the exponents in (3.8.26) for the inversion hyperbolic case: if the series converged, the resulting single exponent would be real, and we have seen that a single real exponent never gives an inversion hyperbolic symplectic matrix.

Exercises

- 8.7.1.** Verify the expansion (7.2).
- 8.7.2.** Verify (7.8) through (7.10).
- 8.7.3.** Verify (7.13).
- 8.7.4.** Verify (7.15).
- 8.7.5.** Given (7.16), verify (7.17).
- 8.7.6.** Verify (7.19) and (7.20).
- 8.7.7.** Verify (7.21) through (7.23).
- 8.7.8.** Verify the singularity statements made about $\tau(\mathbf{v})$ and $\mathbf{v}(\tau)$, and verify (7.24) and (7.25).
- 8.7.9.** Review Exercises 3.7.11 and 5.9.3. Consider the matrix $N = -M$. [Note that this N is not that given by (3.7.105).] Show that N can be written in the form (3.7.36) with S given by the equation

$$S = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \quad (8.7.38)$$

Consider the polar decompositions of M and N given by (3.8.1). Show that both M and N have the same P given by (5.9.16). Find the O matrices for M and N . Show that they have the form

$$O = \exp(i\theta\sigma^2), \quad (8.7.39)$$

and find θ in each case. Following (3.8.15) and (3.8.23) find the matrix S^a associated with P and the matrices S^c associated with the matrices O . Compute the corresponding vectors \mathbf{v}^a , \mathbf{v}^c , $\tau^a = \tau(\mathbf{v}^a)$, and $\tau^c = \tau(\mathbf{v}^c)$. Make the identifications $\mathbf{v}^a = \mathbf{v}^f$, $\tau^a = \tau^f$, $\mathbf{v}^c = \mathbf{v}^g$, and $\tau^c = \tau^g$. See (3.8.24) and (7.12). Consider the vector $\mathbf{v}^h(\theta)$ defined by the relation

$$\exp[\mathbf{v}^h(\theta) \cdot \boldsymbol{\sigma}] = \exp(\mathbf{v}^f \cdot \boldsymbol{\sigma}) \exp(i\theta\sigma^2). \quad (8.7.40)$$

Use (7.23) to find $\tau^h(\theta) = \tau^h(\mathbf{v}^h(\theta))$. Starting from $\theta = 0$, follow the quantities θ , $\tau^h(\theta)$, and $\mathbf{v}^h(\theta)$ to the θ value for N . Repeat this same process, again starting at $\theta = 0$, and continuing to the θ value for M . Show that $\mathbf{v}^h(\theta)$ is well defined for the θ value corresponding to N and produces (7.38), and that $\mathbf{v}^h(\theta)$ is singular at the θ value corresponding to M .

- 8.7.10.** Verify (7.28) through (7.37).

- 8.7.11.** Show that $R = -I$ when $\Delta = -\pi^2$. What happens when $\Delta = -4\pi^2$? Show that R takes the form

$$R = \begin{pmatrix} 1+a & b \\ -c & 1-a \end{pmatrix} \quad (8.7.41)$$

in the case $\Delta = 0$. Show that this R can be diagonalized only if $b = c = 0$. Hint: When $\Delta = 0$, $R = I + F$ and, according to (7.30), $F^2 = 0$. Show that such an F can be diagonalized only if $F = 0$.

8.7.12. Consider the inversion hyperbolic symplectic matrix M given by

$$M = \begin{pmatrix} -\lambda & 0 \\ 0 & -1/\lambda \end{pmatrix} \quad (8.7.42)$$

where λ is real and positive. We know that M cannot be written in single exponent form with a real exponent. But can it be written in single exponent form with a complex exponent? Let S be the symmetric matrix given by

$$S = \begin{pmatrix} 0 & a \\ a & 0 \end{pmatrix}. \quad (8.7.43)$$

Verify the results

$$JS = \begin{pmatrix} a & 0 \\ 0 & -a \end{pmatrix} \quad (8.7.44)$$

and

$$N = \exp(JS) = \begin{pmatrix} \exp(a) & 0 \\ 0 & \exp(-a) \end{pmatrix}. \quad (8.7.45)$$

Evidently M can be written in single exponent form if one can satisfy the relation

$$\exp(a) = -\lambda. \quad (8.7.46)$$

Define a quantity α by the rule

$$\alpha = \log(\lambda). \quad (8.7.47)$$

Show that a solution to (7.46) is

$$a = \alpha + \pi i. \quad (8.7.48)$$

You have shown that, although the inversion hyperbolic symplectic matrix M cannot be written in single exponent form with a real exponent, it can be written in single exponent form with a complex exponent.

8.7.13. For some purposes it is useful to have an $SU(2)$ version of the BCH formula (7.23). Recall the 2×2 matrices K^j defined in Exercise 3.7.30 and manufactured from the Pauli matrices by the rules

$$K^j = (-i/2)\sigma^j. \quad (8.7.49)$$

Verify that they satisfy the multiplication rules

$$K^j K^k = (-1/4)\delta_{jk}I + (1/2) \sum_{\ell} \epsilon_{jkl} K^{\ell}, \quad (8.7.50)$$

and recall that these rules can be summarized in the “vector” form

$$(\mathbf{a} \cdot \mathbf{K})(\mathbf{b} \cdot \mathbf{K}) = -(1/4)(\mathbf{a} \cdot \mathbf{b})I + (1/2)(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{K}. \quad (8.7.51)$$

See (3.7.176).

Verify that any matrix $u \in SU(2)$ can be written in the form

$$u(\mathbf{v}) = \exp(\mathbf{v} \cdot \mathbf{K}). \quad (8.7.52)$$

Show that the infinite series implied by (7.52) can be summed to give the explicit result

$$\exp(\mathbf{v} \cdot \mathbf{K}) = I \cos(v/2) + (\mathbf{v} \cdot \mathbf{K})(2/v) \sin(v/2) \quad (8.7.53)$$

where

$$v = (\mathbf{v} \cdot \mathbf{v})^{1/2}. \quad (8.7.54)$$

Define a vector $\tau(\mathbf{v})$ by the rule

$$\tau(\mathbf{v}) = \mathbf{v}(2/v) \tan(v/2). \quad (8.7.55)$$

Show that (7.55) has the inverse

$$\mathbf{v}(\tau) = \tau(2/\tau) \tan^{-1}(\tau/2) \quad (8.7.56)$$

where

$$\tau = (\tau \cdot \tau)^{1/2}. \quad (8.7.57)$$

Show that (7.53) can also be written in the form

$$\exp(\mathbf{v} \cdot \mathbf{K}) = \cos(v/2)[I + (\tau \cdot \mathbf{K})]. \quad (8.7.58)$$

Given two vectors \mathbf{v}^a and \mathbf{v}^b , your task is to find a third vector \mathbf{v}^c such that

$$\exp(\mathbf{v}^a \cdot \mathbf{K}) \exp(\mathbf{v}^b \cdot \mathbf{K}) = \exp(\mathbf{v}^c \cdot \mathbf{K}). \quad (8.7.59)$$

Show that there is the formula

$$\tau(\mathbf{v}^c) = [\tau(\mathbf{v}^a) + \tau(\mathbf{v}^b) + (1/2)\tau(\mathbf{v}^a) \times \tau(\mathbf{v}^b)][1 - (1/4)\tau(\mathbf{v}^a) \cdot \tau(\mathbf{v}^b)]^{-1}. \quad (8.7.60)$$

8.7.14. Review Exercise 8.7.13. Determine the analytic behavior of \mathbf{v}^c as a function of \mathbf{v}^a and \mathbf{v}^b .

8.8 Zassenhaus or Factorization Formulas

The BCH formula (3.7.33) and (3.7.34) attempts to combine two exponents into one. There are related formulas, called *Zassenhaus* formulas, that attempt the reverse: They try to write a single exponent term as a product of several such terms. One simple such formula is the relation

$$\exp(sA + tB) = \exp(sA) \exp(tB) \exp(Z), \quad (8.8.1)$$

where Z has the expansion

$$Z = -(st/2)[A, B] + (s^2t/6)[A, [A, B]] - (st^2/3)[B, [B, A]] + O(s^3t, s^2t^2, st^3). \quad (8.8.2)$$

It will be seen in Sections 10.8 through 10.10 that Zassenhaus formulas are useful in constructing symplectic integrators and computing maps.

Equation (8.1) writes a single exponent term as a product of *three* such terms. It may also be desirable to write a single exponent term as a product of *two* such terms, and to

attempt to sum some of the infinite series that occur. Consider (7.3), which gives a formula for combining two exponentials into one grand exponential. Sometimes, as in for example the construction of a factored product decomposition, it is useful to be able to turn the process around. Define a quantity h by writing

$$s : f : [1 - \exp(-s : f :)]^{-1} g = h. \quad (8.8.3)$$

Observe that (8.3) may be solved for the quantity g to give the relation

$$g = \{[1 - \exp(-s : f :)]/[s : f :]\}h. \quad (8.8.4)$$

Here the operator expression appearing on the right of (8.4) is interpreted to be the series

$$\begin{aligned} [1 - \exp(-s : f :)]/[s : f :] &= - \sum_{m=1}^{\infty} (-s)^m : f :^{m-1} / [m! s : f :] \\ &= \sum_{m=1}^{\infty} (-s)^{m-1} : f :^{m-1} / m!. \end{aligned} \quad (8.8.5)$$

Now insert (8.3) into (7.3). One finds, upon reading right to left, the result

$$\exp[s : f : + t : h : + O(t^2)] = \exp(s : f :) \exp(t : g :). \quad (8.8.6)$$

Finally, the term of $O(t^2)$ can be taken from the left to the right side of (7.6) to produce the relation

$$\exp[s : f : + t : h :] = \exp(s : f :) \exp(t : g :) \exp[: O(t^2) :]. \quad (8.8.7)$$

Equation (8.7) gives a formula for writing the exponential of the sum of two exponents as a product of two exponentials.

It is worth remarking that the operation described by (8.4), which is required for evaluating (8.7), can be written in a more compact form. First, observe the formal integral identity

$$[1 - \exp(-s : f :)]/[s : f :] = \int_0^1 d\tau \exp(-\tau s : f :). \quad (8.8.8)$$

Let us define the function $\text{iex}(w)$, called the *integrated* exponential function, by the rule

$$\text{iex}(w) = \int_0^1 d\tau \exp(\tau w) = (e^w - 1)/w = \sum_{m=0}^{\infty} w^m / (m+1)!.. \quad (8.8.9)$$

Evidently iex is an *entire* analytic function. (Like the exponential function, it has no singularities in the complex plane except at infinity, and its Taylor series has an infinite radius of convergence.) By using the identity (8.8) and the definition (8.9), the relation (8.4) can be written in the forms

$$g = \int_0^1 d\tau \exp(-\tau s : f :) h = \text{iex}(-s : f :) h. \quad (8.8.10)$$

But now (5.4.11) can be employed to give the final integral formula

$$g(z) = \int_0^1 d\tau h[\exp(-\tau s : f :)z]. \quad (8.8.11)$$

In summary, we have the operator identity

$$\begin{aligned} \exp(s : f : + t : h :) &= \exp(s : f :) \exp[\text{iex}(-s \# f \#)(t : h :)] \exp[: O(t^2) :] \\ &= \exp(s : f :) \exp[: \text{iex}(-s : f :)(th :) :] \exp[: O(t^2) :]. \end{aligned} \quad (8.8.12)$$

We also note that in (8.4), unlike in (7.3), no inverses of the form $[1 - \exp(-s : f :)]^{-1}$ are involved. Instead, we have benign relations like (8.11). Correspondingly, because they presuppose the existence of a single exponent form, Zassenhaus formulas can have better convergence properties than the BCH formula. Finally, we remark that the next few terms in (8.12), those terms proportional to powers of t^2, t^3, \dots , can also be found explicitly. See Exercises 10.3.* and 10.4.*.

Exercises

8.8.1. Derive (8.1) and (8.2) from (3.7.33) and (3.7.34).

8.8.2. Verify the expansion (8.5); derive (8.6) from (7.3) and (8.3).

8.8.3. Verify the integral identity (8.8), and the integral identity and expansion (8.9). Show that the Taylor series for $\text{iex}(w)$ has an infinite radius of convergence.

8.8.4. Derive the formula

$$\exp(s : f : + t : h :) = \exp[: O(t^2) :] \exp[\text{iex}(s \# f \#)(t : h :)] \exp(s : f :). \quad (8.8.13)$$

8.9 Ideals, Quotients, and Gradings

We know that the Lie algebra of all Lie operators, which we have called $ispm(2n, \mathbb{R})$, is infinite dimensional. Correspondingly $ISpM(2n, \mathbb{R})$, the group of symplectic maps, is infinite dimensional. Indeed, the factorization (7.7.23) gives a representation of the general analytic symplectic map. We see that the specification of a symplectic map generally requires an infinite number of parameters. This fact produces an awkward situation for human beings and computers, which can only work with a finite number of quantities (and often only with finite precision).

An optimistic perspective on the experimental and theoretical situation, for example in the field of accelerator physics, might be stated as follows: We know that a beam transport system, accelerator, storage ring, or any portion thereof may be described by a symplectic transfer map. However, because we cannot measure or control electromagnetic fields exactly, we are unsure of and unable to control exactly what this map is. Also, since it is impossible to perform computations with an infinite number of variables and to infinite precision, it is necessary to develop various approximation schemes. Thus, we are able to

study computationally (and probably theoretically) the detailed properties of only a subset of all symplectic maps. The hope is that if two symplectic maps are in some sense nearly the same, then their behavior [including, in some cases, long-term (repeated iteration) behavior] will be nearly the same. If this were not true from an experimental standpoint, then it would be impossible to build satisfactory storage rings, etc. If this were not true from a theoretical standpoint, then it would be impossible to design storage rings, etc., with any assurance that their actual performance would be satisfactory.

As just described, it is necessary to develop some sort of approximation scheme to treat symplectic maps in a practical way. In this section we will describe truncation schemes that maintain a Lie algebraic structure. We already know from Section 8.4 that the rules for multiplying symplectic maps can be expressed entirely in Lie algebraic terms. Thus, if the truncation scheme maintains a Lie algebraic structure, it follows that maps may either be truncated and then multiplied, or multiplied and then truncated. The results from both procedures are guaranteed to be the same.

For example, consider the Lie algebra spanned by the homogeneous polynomials f_2, f_3, f_4, \dots . Evidently, this Lie algebra is infinite dimensional. Let D be some integer. Suppose we decide to retain only the polynomials $f_2, f_3, f_4, \dots, f_{D-1}$, and discard all polynomials f_m with $m \geq D$. Correspondingly, in the map \mathcal{M} given by (7.6.3) we drop from the product all f_m with $m \geq D$. Is this a consistent procedure? The answer is yes. As we will see, the discarding of all f_m with $m \geq D$ amounts mathematically to working with a *quotient* Lie algebra and its corresponding quotient group. We note that since $:f_m : z_b$ consists of terms of degree $(m-1)$, the decision to drop the f_m with $m \geq D$ amounts physically to neglecting all aberrations of degree $(D-1)$ and higher.

As a second example, suppose ϵ is a (presumed small) parameter, and consider the quantities $f^{(0)}, \epsilon f^{(1)}, \epsilon^2 f^{(2)}, \epsilon^3 f^{(3)}, \dots$, where $f^{(0)}, f^{(1)}, f^{(2)}, f^{(3)} \dots$ are *arbitrary* functions. The quantities $f^{(0)}, \epsilon f^{(1)}, \epsilon^2 f^{(2)}, \epsilon^3 f^{(3)}, \dots$ also form (with the Poisson bracket as a Lie product) an infinite dimensional Lie algebra. Suppose, as a kind of perturbation theory, we decide to discard all $\epsilon^m f^{(m)}$ with $m > D$ where again D is some integer. Is this a consistent procedure? The answer again is yes. We will see that expanding in powers of ϵ is equivalent to introducing a *grading* into the Lie algebra, and that truncating the expansion is equivalent to using the grading to produce a quotient structure.

With this motivation as background, we are ready to develop some mathematical tools. The first concept we will need is that of an *ideal*. Let L be a Lie algebra, and let L' be a subalgebra of L . For L' to be a subalgebra means that the elements of L' must be in L , and must form a Lie algebra in their own right. That is, by themselves they must satisfy the properties 1 through 5 (as given in Section 3.7) required of a Lie algebra. Let x be any element in L and let x' be any element in L' . Suppose the elements of L' have the property

$$[x, x'] \in L' \text{ for all } x \in L, x' \in L'. \quad (8.9.1)$$

That is, no element of L' can be sent beyond L' by taking Lie products with arbitrary elements in L . In this case L' is said to be an *invariant* subalgebra. And if L' is a genuine invariant subalgebra, i.e. neither zero nor the full Lie algebra L , it is called an *ideal*.⁵

⁵Here is an opportunity for three more definitions: Recall that a Lie algebra is called *simple* if it has no ideals. Recall also that a Lie algebra or subalgebra is called *Abelian* if the Lie product of any two elements

Suppose a Lie algebra L has a subalgebra L' . Then L' can be used to set up an equivalence relation among the elements of L . Let x_1 and x_2 be any two elements in L . We say that x_2 is *equivalent* to x_1 (and write $x_2 \sim x_1$) if their difference $(x_2 - x_1)$ is in L' ,

$$x_2 \sim x_1 \Leftrightarrow (x_2 - x_1) \in L'. \quad (8.9.2)$$

(Here the symbol \Leftrightarrow is used to indicate logical implication in both directions.) This equivalence relation can be used to partition the elements of L into *disjoint* equivalence classes. Let the symbols $\{x\}$ denote all the elements of L that are equivalent to some element x . In a Lie algebraic (actually, vector space) context, the collection of these equivalence classes is called a quotient space, and is customarily denoted by the symbols L/L' . See Exercise 9.1.

To get a feeling for this construction, let 0 be the *zero* element in L and consider the set of elements $\{0\}$, the set of elements in L that are equivalent to 0. We see from (9.2) that the set $\{0\}$ is identical to the set of elements L' . Consequently, in the quotient space construction, all elements in L' are identified with (are equivalent to) the zero element in L . That is, we have the logical relation

$$x' \in L' \Leftrightarrow \{x'\} = \{0\}. \quad (8.9.3)$$

Moreover, suppose $x_2 \sim x_1$. Then by (9.2) we have a relation of the form

$$x_2 = x_1 + x' \text{ with } x' \in L'. \quad (8.9.4)$$

Thus, if x_1 and x_2 are equivalent, they differ only by an element that has been identified with zero.

As defined so far, the quotient space L/L' is simply a collection of equivalence classes. We now give it a vector space structure by a simple but ingenious (and, at first sight, confusing) construction. We begin by noting the logical implication

$$x_2 \sim x_1 \Rightarrow ax_2 \sim ax_1, \quad (8.9.5)$$

where a is any scalar. This result follows from (9.4) by noting that ax' belongs to L' if x' belongs to L' . (Remember that L' is an algebra.) Next, suppose that the elements x_1, x_2 and y_1, y_2 satisfy the equivalence relations

$$x_2 \sim x_1, \quad y_2 \sim y_1. \quad (8.9.6)$$

Then it follows from (9.4) and its y analog that we have the relation

$$x_2 + y_2 = x_1 + y_1 + x' + y'. \quad (8.9.7)$$

in it vanishes. Colloquially, we say that all elements in an Abelian Lie algebra or subalgebra *commute*. A Lie algebra is called *semisimple* if it has no Abelian ideals. By this definition, a simple Lie algebra is also semisimple. That is, simple Lie algebras form a subset of the set of semisimple Lie algebras. Suppose a Lie algebra L is semisimple but not simple. Then it can be shown that L is the *direct sum* of two or more simple Lie algebras. By direct sum it is meant the all the elements of any simple Lie subalgebra in the sum commute with all the elements of any other simple Lie subalgebra in the sum.

But if x' and y' belong to L' , then so must the sum $(x' + y')$. (Again, remember that L' is an algebra.) Thus, we also have the logical implication

$$x_2 \sim x_1 \text{ and } y_2 \sim y_1 \Rightarrow (x_2 + y_2) \sim (x_1 + y_1). \quad (8.9.8)$$

Now we are ready to give L/L' a *vector space* structure. First, we have to define scalar multiplication. Consider some equivalence class. Then, since equivalence classes are disjoint, each equivalence class may be labelled by any one of its members. Select some member of the equivalence class under consideration, and call it x_1 . Then the equivalence class may be given the label $\{x_1\}$. Now let a be any scalar. We define scalar multiplication acting on the element $\{x_1\}$ of L/L' by the rule

$$a\{x_1\} = \{ax_1\}. \quad (8.9.9)$$

Thus, by this definition, scalar multiplication sends equivalence classes into each other. But suppose x_2 also belongs to $\{x_1\}$, and that we had used x_2 to label $\{x_1\}$ instead of x_1 . Would this different choice affect the definition (9.9)? It would not. With the choice of x_2 as a label we would have the definition

$$a\{x_2\} = \{ax_2\}. \quad (8.9.10)$$

But, by (9.5), we have the relation

$$\{ax_2\} = \{ax_1\} \quad (8.9.11)$$

because an equivalence class is uniquely defined by any of its members. Thus, scalar multiplication is uniquely defined by the rule (9.9).

Next we define vector addition. Let $\{x_1\}$ and $\{y_1\}$ be two equivalence classes labelled by two members x_1 and y_1 . We define addition by the rule

$$\{x_1\} + \{y_1\} = \{(x_1 + y_1)\}. \quad (8.9.12)$$

By this definition, addition sends a pair of equivalence classes into some third (not necessarily different) equivalence class. Again, there is the question of uniqueness under the choice of labelling. However, thanks to (9.8), the definition (9.12) is in fact independent of labelling. Note that as a special case of (9.12) we have the relation

$$\{x\} + \{0\} = \{x\}. \quad (8.9.13)$$

That is, the equivalence class $\{0\}$ plays the role of the zero vector in L/L' .

We have given L/L' a vector space structure. What is the dimension of L/L' ? Suppose that L' has a basis b_1, b_2, \dots , and that a basis for L is constructed by taking the vectors b_1, b_2, \dots supplemented by the additional linearly independent vectors v_1, v_2, \dots, v_n . Then, we have the dimensional relations

$$\dim L = \dim L' + n, \text{ or } n = \dim L - \dim L'. \quad (8.9.14)$$

Suppose x is any vector in L . Then x has the unique decomposition

$$x = x' + \sum_{i=1}^n \xi_i v_i, \quad (8.9.15)$$

where x' is the portion of x spanned by the basis vectors b_1, b_2, \dots . Now form equivalence classes of both sides of (9.15). From the definition (9.12) and (9.3), (9.9), and (9.13) we find the result

$$\{x\} = \{x'\} + \left\{ \sum_{i=1}^n \xi_i v_i \right\} = \sum_{i=1}^n \xi_i \{v_i\}. \quad (8.9.16)$$

It is easily verified that the vectors $\{v_i\}$ are linearly independent. See Exercise 7.2. Consequently, the quotient space L/L' has dimension n ,

$$\dim(L/L') = n = \dim L - \dim L'. \quad (8.9.17)$$

So far we have assumed that L' is a subalgebra. Now make the further supposition that L' is an ideal. In this case we can give the quotient space a *Lie algebraic* structure. We have already seen that the quotient space can be given a vector space structure. What remains is to define a Lie product. Let $\{x_1\}$ and $\{y_1\}$ be two equivalence classes labelled by two members x_1 and y_1 . We define a *quotient space* Lie product, denoted by the symbols $[,]_{qs}$, by the rule

$$[\{x_1\}, \{y_1\}]_{qs} = \{[x_1, y_1]\}. \quad (8.9.18)$$

By this definition the quotient space Lie product sends a pair of equivalence classes into some third (not necessarily different) equivalence class. As before there is the question of uniqueness under the choice of labelling. Suppose we use instead labels x_2 and y_2 that satisfy (9.6). Then from (9.4) and its y counterpart we find the result

$$\begin{aligned} [\{x_2\}, \{y_2\}]_{qs} &= \{[x_1 + x', y_1 + y']\} \\ &= \{[x_1, y_1] + [x', y_1] + [x_1, y'] + [x', y']\} \\ &= \{[x_1, y_1]\} + \{[x', y_1]\} + \{[x_1, y']\} + \{[x', y']\} \\ &= [\{x_1\}, \{y_1\}]_{qs} + \{[x', y_1]\} + \{[x_1, y']\} + \{[x', y']\}. \end{aligned} \quad (8.9.19)$$

But, since L' is assumed to be an ideal, all the quantities $[x', y_1]$, $[x_1, y']$, and $[x', y']$ must be in L' . See (9.1). It follows from (9.3) and (9.13) that we have the relation

$$\{[x', y_1]\} + \{[x_1, y']\} + \{[x', y']\} = \{0\} + \{0\} + \{0\} = \{0\}. \quad (8.9.20)$$

Consequently, upon combining (9.19) with (9.20) and again using (9.3), we find the result

$$[\{x_2\}, \{y_2\}]_{qs} = [\{x_1\}, \{y_1\}]_{qs}. \quad (8.9.21)$$

Thus, the quotient space Lie product is uniquely defined by (9.18).

We claim that the addition rule (9.12) and Lie product rule (9.18) together satisfy requirements 1 through 5 for a Lie algebra as given in Section 3.7. For example, if $\{x\}$, $\{y\}$, and $\{z\}$ are any three equivalence classes, we have from (9.18) the relation

$$[\{x\}, [\{y\}, \{z\}]_{qs}]_{qs} = [\{x\}, \{[y, z]\}]_{qs} = \{[x, [y, z]]\}. \quad (8.9.22)$$

The Jacobi condition requirement follows immediately from (9.22). Verification of the remaining requirements is left as an exercise for the reader. We conclude that if L' is an ideal,

the elements of the quotient space L/L' can be viewed as elements of an n dimensional Lie algebra. This Lie algebra is called the *quotient* Lie algebra.

The discussion of the quotient Lie algebra L/L' that we have just worked through may seem overly abstract. It can be made more concrete by using structure constants. See (3.7.40). We know from our previous discussion that L is spanned by the basis vectors $b_1, b_2, b_3 \dots$ and v_1, v_2, \dots, v_n . There are therefore three kinds of Lie products: $[b_i, b_j]$, $[v_i, b_j]$, and $[v_i, v_j]$. Correspondingly, the structure constants are of six kinds: ${}^1c, {}^2c, \dots, {}^6c$. Consider first the Lie products $[b_i, b_j]$. Their results can be written in the form

$$[b_i, b_j] = \sum_k {}^1c_{ij}^k b_k + \sum_k {}^2c_{ij}^k v_k. \quad (8.9.23)$$

If L' is a Lie subalgebra spanned by the b_i , as we have assumed, then the structure constants 2c must vanish,

$${}^2c_{ij}^k = 0. \quad (8.9.24)$$

Consider next the Lie products $[v_i, b_j]$. Their results can be written in the form

$$[v_i, b_j] = \sum_k {}^3c_{ij}^k b_k + \sum_k {}^4c_{ij}^k v_k. \quad (8.9.25)$$

If L' is an ideal, as we have also assumed, then the structure constants 4c must also vanish,

$${}^4c_{ij}^k = 0. \quad (8.9.26)$$

See (9.1). Finally, the Lie products $[v_i, v_j]$ can be written in the form

$$[v_i, v_j] = \sum_k {}^5c_{ij}^k b_k + \sum_k {}^6c_{ij}^k v_k. \quad (8.9.27)$$

Now form equivalence classes of both sides of (9.27). Then, using (9.3), (9.13), and (9.18), we find the result

$$[\{v_i\}, \{v_j\}]_{qs} = \sum_k {}^6c_{ij}^k \{v_k\}. \quad (8.9.28)$$

We also know from (9.16) that the $\{v_i\}$ span L/L' . From (9.28) we conclude that the ${}^6c_{ij}^k$ are the structure constants of L/L' .

The next topic we need to discuss is that of *quotient groups*. We will see that for every quotient Lie algebra there is a corresponding quotient Lie group. To understand this connection we begin by describing the concept of a quotient group. Suppose G is a group, and suppose G' is a subgroup of G . We use the subgroup G' to set up an equivalence relation in G . Let g_1 and g_2 be any two elements in G . We say that g_2 is equivalent to g_1 (and again use the notation $g_2 \sim g_1$) if there exists a g' in G' such that $g_1^{-1}g_2 = g'$ or, put another way, $g_2 = g_1g'$:

$$g_2 \sim g_1 \Leftrightarrow g_1^{-1}g_2 = g' \in G' \Leftrightarrow g_2 = g_1g'. \quad (8.9.29)$$

This equivalence relation can be used to partition the elements of G into disjoint equivalence classes. These equivalence classes are called the (left) *cosets* of G with respect to G' . The collection of all of these cosets is called the *coset space*, and is customarily denoted by the

symbols G/G' . See Exercise 5.12.7. If g is an element in G , we use the notation $\{g\}$ to denote all the elements in G that are equivalent to g . Suppose e is the identity element in G . Then it is easily checked that

$$\{g'\} = \{e\}, \quad (8.9.30)$$

where g' is any element in G' .

We next assume that G' is a *normal* or *invariant* subgroup of G . Suppose g is any element of G , and g' is any element of G' . The subgroup G' is called invariant or normal if there is the relation

$$g^{-1}g'g \in G' \text{ for all } g \in G, g' \in G'. \quad (8.9.31)$$

If G' is normal, the collection of equivalence classes (coset space) G/G' can be made into a *group*. This group is called the *quotient* group.

To show that G/G' can be given a group structure, we must set up a rule for multiplying equivalence classes (cosets) in such a way that rules analogous to those given for matrices in Section 3.6 are satisfied. Suppose $\{g_1\}$ and $\{h_1\}$ are two equivalence classes labelled by representative elements g_1 and h_1 in G . We define their product, denoted by the symbols $\{g_1\}\{h_1\}$, to be the equivalence class given by the rule

$$\{g_1\}\{h_1\} = \{g_1h_1\}. \quad (8.9.32)$$

As a special case of (9.32) we find the results

$$\{g_1\}\{e\} = \{g_1e\} = \{g_1\}, \quad (8.9.33)$$

$$\{e\}\{h_1\} = \{eh_1\} = \{h_1\}. \quad (8.9.34)$$

Also, we define $\{g_1\}^{-1}$ by the rule

$$\{g_1\}^{-1} = \{g_1^{-1}\}. \quad (8.9.35)$$

Then, upon combining (9.32) and (9.35), we find the results

$$\{g_1\}\{g_1\}^{-1} = \{g_1g_1^{-1}\} = \{e\}, \quad (8.9.36)$$

$$\{g_1\}^{-1}\{g_1\} = \{g_1^{-1}g_1\} = \{e\}. \quad (8.9.37)$$

Both (9.32) and (9.35) are rules that send equivalence classes to equivalence classes. Of course, as usual, we must verify that the definitions (9.32) and (9.35) are in fact independent of the choice of representative elements selected to label the equivalence classes $\{g_1\}$ and $\{h_1\}$. For example, suppose we decide to designate the equivalence classes $\{g_1\}$ and $\{h_1\}$ by the representatives g_2 and h_2 so that we have the alternate labels $\{g_2\}$ and $\{h_2\}$. Of course g_2 and g_1 are related by (9.29), and h_2 and h_1 are related by an analogous equation. Then we find from (9.32) the result

$$\{g_2\}\{h_2\} = \{g_2h_2\} = \{g_1g'h_1h'\} = \{g_1h_1h_1^{-1}g'h_1h'\} = \{g_1h_1\} = \{g_1\}\{h_1\}. \quad (8.9.38)$$

Here we have used (9.31) and the fact that G' is a group to deduce that $h_1^{-1}g'h_1h'$ is in G' . We conclude that the equivalence class product is uniquely defined by (9.32). Similarly, it

can be shown that the equivalence class inverse is uniquely defined by (9.35). See Exercise 9.4. Thus, the quotient group is uniquely defined.

We are now ready to see the connection between quotient Lie algebras and quotient Lie groups. Suppose, for simplicity, that the Lie algebra L is realized as a set of linear operators, with the Lie product being a commutator. Then, as described in Section 3.7, there is (at least locally near the identity) an associated Lie group G obtained by exponentiating L . Also, suppose L has a subalgebra L' . Exponentiating L' gives a Lie subgroup G' . Suppose that L' is an ideal. Then we will discover that G' is normal. Also, since L' is an ideal, we can form the quotient Lie algebra L/L' . Correspondingly, since G' is normal, we can form the quotient group G/G' . We will discover that L/L' is the Lie algebra of G/G' .

To see how this comes about, suppose ℓ is an element of L , and ℓ' is an element of L' . Upon exponentiation we get elements g and g' of G and G' , respectively,

$$g = \exp(\ell), \quad g' = \exp(\ell'). \quad (8.9.39)$$

Now form the combination $g^{-1}g'g$. We find from (9.39) the result

$$g^{-1}g'g = \exp(-\ell)\exp(\ell')\exp(\ell). \quad (8.9.40)$$

Next use the adjoint operator $\#\ell\#$ and a relation of the form (2.16) to rewrite (9.40) in the form

$$g^{-1}g'g = \exp[\exp(-\#\ell\#)\ell']. \quad (8.9.41)$$

Since L' is assumed to be an ideal, we have from (8.1) the result

$$\#\ell\#\ell' = [\ell, \ell'] \in L', \quad (8.9.42)$$

from which it follows that $\exp(-\#\ell\#)\ell'$ is also in L' ,

$$\exp(-\#\ell\#)\ell' \in L'. \quad (8.9.43)$$

But, from (9.43) it follows that

$$g^{-1}g'g = \exp[\exp(-\#\ell\#)\ell'] \in G'. \quad (8.9.44)$$

Consequently G' is normal, as advertised. The converse can also be proved: If G' with Lie algebra L' is a normal subgroup of a Lie group G with Lie algebra L , then L' is an ideal in L .

To complete our demonstration, we must show that L/L' is the Lie algebra of G/G' . Suppose that x_1 is some element of L , and that it is used to label the equivalence class $\{x_1\}$, which is an element of L/L' . We define $\exp\{x_1\}$, which is supposed to be an element of G/G' , by the rule

$$\exp(\{x_1\}) = \{\exp(x_1)\}. \quad (8.9.45)$$

[Note that the $\{ \}$ on the left side of (9.45) refers to the Lie algebraic equivalence class, and that on the right side refers to the group equivalence class.] As a special case of (9.45) we have the relation

$$\exp(\{0\}) = \{\exp(0)\} = \{e\}. \quad (8.9.46)$$

Of course, as usual, we must check that our definition does not depend on the choice of equivalence class labels. Suppose we label $\{x_1\}$ by x_2 where $x_2 \sim x_1$. Then, we find the result

$$\begin{aligned}\exp(\{x_2\}) &= \{\exp(x_2)\} = \{\exp(x_1 + x')\} \\ &= \{\exp(x_1) \exp(-x_1) \exp(x_1 + x')\}. \end{aligned}\quad (8.9.47)$$

Here we have used (9.4). Now use the BCH series (3.7.33) and (3.7.34) to combine the exponents $(-x_1)$ and $(x_1 + x')$. According to (3.7.34), the first thing we must do is add them. We find the result

$$(-x_1) + (x_1 + x') = x' \in L'. \quad (8.9.48)$$

Next, according to (3.7.34), we must find their commutator. Doing so gives the result

$$[(-x_1), (x_1 + x')] = [(-x_1), x'] \in L'. \quad (8.9.49)$$

Here, because L' is an ideal, we have been able to use (9.1). Finally, we must compute an infinite number of higher-order commutators. See (3.7.34) and Appendix C. Examination of the contents of these commutators shows that each of them has a term of the form (9.49) buried inside, and we know this term is in L' . But since L' is an ideal, (9.1) shows that all further commutators will also be in L' . We have learned that all terms that arise when we combine the exponents in $\exp(-x_1) \exp(x_1 + x')$ are in L' . Consequently, the product $\exp(-x_1) \exp(x_1 + x')$ must be in G' ,

$$\exp(-x_1) \exp(x_1 + x') = g' \in G'. \quad (8.9.50)$$

It follows from (9.47), (9.50), (9.29), and (9.45) that we have the result

$$\begin{aligned}\exp(\{x_2\}) &= \{\exp(x_1) \exp(-x_1) \exp(x_1 + x')\} \\ &= \{\exp(x_1) g'\} = \{\exp(x_1)\} = \exp(\{x_1\}). \end{aligned}\quad (8.9.51)$$

Thus, the definition (9.45) is indeed independent of the choice of equivalence class labels.

The last thing we must show is that products of the form $\exp(\{x_1\}) \exp(\{y_1\})$ can be computed from a knowledge only of the quotient Lie algebra L/L' . Suppose $\{x_1\}$ and $\{y_1\}$ are two elements of L/L' . We then find from (9.45) and (9.32) the result

$$\begin{aligned}\exp(\{x_1\}) \exp(\{y_1\}) &= \{\exp(x_1)\} \{\exp(y_1)\} \\ &= \{\exp(x_1) \exp(y_1)\}. \end{aligned}\quad (8.9.52)$$

Let us use the BCH series to combine the exponents x_1 and y_1 into one grand exponent z_1 . Then we have the relation

$$\exp(x_1) \exp(y_1) = \exp(z_1). \quad (8.9.53)$$

Consequently, we find from (9.52) and (9.53) the result

$$\exp(\{x_1\}) \exp(\{y_1\}) = \{\exp(z_1)\} = \exp(\{z_1\}). \quad (8.9.54)$$

From the BCH formula (3.7.34) we know that z_1 is given by the series

$$\begin{aligned} z_1 = x_1 + y_1 &+ (1/2)[x_1, y_1] + (1/12)[x_1, [x_1, y_1]] \\ &+ (1/12)[y_1, [y_1, x_1]] + \dots \end{aligned} \quad (8.9.55)$$

Now, form equivalence classes of both sides of (9.55). By making repeated use of (9.12) and (9.18) we find from (9.55) the result

$$\begin{aligned} \{z_1\} = & \{x_1\} + \{y_1\} + (1/2)[\{x_1\}, \{y_1\}]_{qs} + (1/12)[\{x_1\}, [\{x_1\}, \{y_1\}]_{qs}]_{qs} \\ & + (1/12)[\{y_1\}, [\{y_1\}, \{x_1\}]_{qs}]_{qs} + \dots \end{aligned} \quad (8.9.56)$$

We see from (9.54) and (9.56) that the group multiplication rules for the quotient group G/G' are indeed determined by the quotient Lie algebra L/L' .

So far in this section our discussion has been devoted to the general concepts of quotient Lie algebras and their associated quotient Lie groups. We now turn to applying these concepts to $ispn(2n, \mathbb{R})$, the Lie algebra of the group of all symplectic maps acting on a $2n$ dimensional phase space. As mentioned at the beginning of this chapter, we will first restrict our attention to those symplectic maps that send the origin into itself. The general case will be treated at the end of this section. From Section 7.6 we know that the Lie algebra of maps that send the origin into itself is spanned by the Lie operators $:f_2: : f_3: : \dots$. Let us call this Lie algebra L_2 .

Let D be some integer satisfying $D > 2$, and let L_D be the set of Lie operators spanned by all $:f_m:$ with $m \geq D$. From (5.3.14) and (7.6.14) we find the result

$$[:f_m: : f_n:] =: [f_m, f_n] :=: \mathcal{P}_{m+n-2} :, \quad (8.9.57)$$

where we have used the notation \mathcal{P}_ℓ to denote the space spanned by all f_ℓ . Observe that if $m \geq D$ and $n \geq D$ (with $D > 2$), then $(m + n - 2) \geq D$. Thus, if $:f_m:$ and $:f_n:$ are in L_D , then so is their Lie product (9.57). It follows that L_D is a subalgebra of L_2 .

As a special case of (9.57) we have the result

$$[:f_2: : f_n:] =: [f_2, f_n] :=: \mathcal{P}_n : . \quad (8.9.58)$$

We see from (9.57) and (9.58) that if $:f_n:$ is in L_D , then so is $\{ :f_m: : f_n:\}$ for all $:f_m:$ in L_2 . It follows that L_D is an ideal in L_2 .

Suppose we form the quotient algebra L_2/L_D . From our discussion of quotient algebras, we know this construction is equivalent to discarding all $:f_m:$ with $m \geq D$, and retaining only the $:f_\ell:$ with $\ell = 2, 3, \dots, (D-1)$. We also discard all Lie products $\{ :f_m: : f_n:\}$ when $(m + n - 2) \geq D$. We have seen that dropping all $:f_m:$ with $m \geq D$ is equivalent to ignoring all aberrations of degree $(D-1)$ and higher. The result of this construction, the quotient algebra L_2/L_D , is a finite-dimensional Lie algebra whose dimension equals the number of monomials in the phase-space variables z of degrees $2, 3, \dots, (D-1)$. The number of monomials of degrees $1, 2, 3, \dots, (D-1)$ is given by $S(D-1, d)$. (Note that D as defined in this section differs by 2 from that defined in Section 7.9.) Also, we know that the number of monomials of degree 1 (in a d -dimensional phase space) is d . Thus, we conclude that the dimension of L_2/L_D is given by the relation

$$\dim(L_2/L_D) = S(D-1, d) - d. \quad (8.9.59)$$

This dimension is tabulated in Table 9.1 below for the cases of $d = 4$, $d = 6$, and $d = 8$ and various values of D . Finally, there is a finite-dimensional quotient group corresponding to L_2/L_D . The elements of this group are all symplectic maps of the form

$$\begin{aligned} \mathcal{M} = & \exp(: f_2^c :) \exp(: f_2^a :) \exp(: f_3 :) \exp(: f_4 :) \cdots \exp(: f_{D-1} :) \times \\ & \exp(: h_D :) \exp(: h_{D+1} :) \cdots, \end{aligned} \quad (8.9.60)$$

Table 8.9.1: Values of $\dim(L_2/L_D)$.

D	dim for $d = 4$	dim for $d = 6$	dim for $d = 8$
4	30	77	156
5	65	203	486
6	121	455	1278
7	205	917	2994
8	325	1709	6426
9	490	2996	12,861
10	710	4998	24,301
11	996	8001	43,749
12	1360	12,369	75,573
13	1815	18,557	125,961

where the f 's are specified and the homogeneous polynomial Lie operators $: h_D :, : h_{D+1} :, \dots$ can be anything since all such elements are in L'_D and their exponentials are in the normal subgroup G' . We note that (9.60) is a relation of the form $g_2 = g_1 g'$. See (9.29). Consequently, we may view all members of the quotient group as belonging to the equivalence classes $\{\mathcal{M}_f\}$ where

$$\mathcal{M}_f = \exp(: f_2^c :) \exp(: f_2^a :) \exp(: f_3 :) \exp(: f_4 :) \cdots \exp(: f_{D-1} :). \quad (8.9.61)$$

Since the Lie algebra L_2/L_D is finite dimensional, we might hope to be able to represent it by finite dimensional matrices. This is indeed possible. We will consider first the whole Lie algebra L_1 and show that it has a representation by infinite dimensional matrices.

We can take as a basis for L_1 the Lie operators $:G_t:$. Then, in accord with (3.38), these Lie operators have the associated matrices

$$O_{sr}(:G_t:) = \langle G_s, :G_t: G_r \rangle. \quad (8.9.62)$$

But we also have the result

$$:G_t: G_r = [G_t, G_r] = \sum_{s'} c_{tr}^{s'} G_{s'}, \quad (8.9.63)$$

where the coefficients $c_{tr}^{s'}$ are the structure constants of the underlying Poisson bracket Lie algebra. See (3.7.43). It follows from (9.62) and (9.63) that we have the relation

$$O_{sr}(:G_t:) = c_{tr}^s. \quad (8.9.64)$$

We see that the matrix representation of L_1 is determined by the structure constants. Evidently these matrices are infinite dimensional.

Let us pursue the relation (9.64) a bit further. From (5.3.14), (3.41), and (9.63) we have the result

$$\begin{aligned} \{O(:G_t:), O(:G_{t'}:)\} &= O(\{ :G_t:, :G_{t'}: \}) = O(:[G_t, G_{t'}]:) \\ &= O(: \sum_{t''} c_{tt'}^{t''} G_{t''}:) = \sum_{t''} c_{tt'}^{t''} O(:G_{t''}:). \end{aligned} \quad (8.9.65)$$

Next take r, s matrix elements of both sides of (9.65). Doing so and expanding the commutator on the left side of (9.65) gives the result

$$\sum_{t''} O_{rt''}(:G_t:) O_{t''s}(:G_{t'}:) - O_{rt''}(:G_{t'}:) O_{t''s}(:G_t:) = \sum_{t''} c_{tt'}^{t''} O_{rs}(:G_{t'}:). \quad (8.9.66)$$

Finally, use (9.64) in (9.66) to find the relation

$$\sum_{t''} c_{tt'}^r c_{t's}^{t''} - c_{t't''}^r c_{ts}^{t''} - c_{tt'}^{t''} c_{t''s}^r = 0. \quad (8.9.67)$$

It is easily verified with the aid of (3.7.44) that (9.67) is equivalent to (3.7.45), which is in turn a consequence of the Jacobi identity. [That the Jacobi identity should be involved should come as no surprise since (5.3.14), which was used in (9.65), is a consequence of the Jacobi identity.] Notice that the steps we have been following are quite general and hold, in fact, for any Lie algebra. We have learned from (9.65) that matrices defined in terms of the structure constants by the relation (9.64) provide a representation of the underlying Lie algebra. Indeed, a moment's reflection reveals that our present discussion is a recapitulation of that given at the end of Section 3.7. That is, we have found the *adjoint* representation of L_1 . Finally, we observe that if we exponentiate the matrices (9.64), we get a representation, again called the adjoint representation, of the group. It follows, in the case of the group of all symplectic maps, that the matrix representation given by (3.34) is just the adjoint representation. The matrices in this representation are also infinite dimensional.

We now turn to the case of the Lie algebra L_2/L_D . As in Section 8.6, define a *truncation* (by degree) operator $\mathcal{T}(> m)$ on the basis functions G_r by the rules

$$\deg(G_r) \leq m \Rightarrow \mathcal{T}(> m)G_r = G_r, \quad (8.9.68)$$

$$\deg(G_r) > m \Rightarrow \mathcal{T}(> m)G_r = 0, \quad (8.9.69)$$

and extend $\mathcal{T}(> m)$ to all functions by requiring that it be a *linear* operator. Given any $D > 2$ and any Lie operator $: f :$ in L_2 , we define an associated linear operator $\mathcal{L}_D(: f :)$ by the rule

$$\mathcal{L}_D(: f :) = \mathcal{T}(> D - 2) : f : \mathcal{T}(> D - 2). \quad (8.9.70)$$

We note that since (by hypothesis) $: f :$ is in L_2 , it can only preserve or raise the degree of any monomial on which it acts. Therefore the operator $\mathcal{T}(> D - 2)$ on the far right of (9.70) is actually redundant. We also define \mathcal{L}_D for other linear operators, for example products of Lie operators in L_2 , by rules of the form

$$\mathcal{L}_D(: f :: g :) = \mathcal{T}(> D - 2) : f :: g : \mathcal{T}(> D - 2). \quad (8.9.71)$$

Again, strictly speaking, the $\mathcal{T}(> D - 2)$ on the far right of (9.71) is redundant since $: f :$ and $: g :$ are assumed to be in L_2 .

This definition has several important properties: Suppose $: f :$ is also in L_D . Then, from (9.57) and (9.68) through (9.70), we find the result

$$: f : \in L_D \Rightarrow \mathcal{L}_D(: f :) = 0. \quad (8.9.72)$$

Next suppose $: f :$ and $: g :$ are in L_2 . Then, from (9.57) and (9.68) through (9.71), we have the product rule

$$\begin{aligned} : f :, : g : \in L_2 &\Rightarrow \mathcal{L}_D(: f :) \mathcal{L}_D(: g :) \\ &= \mathcal{T}(> D - 2) : f : \mathcal{T}(> D - 2) \mathcal{T}(> D - 2) : g : \mathcal{T}(> D - 2) \\ &= \mathcal{T}(> D - 2) : f :: g : \mathcal{T}(> D - 2) = \mathcal{L}_D(: f :: g :). \end{aligned} \quad (8.9.73)$$

Here we have used the fact that, since $: f :$ and $: g :$ are in L_2 , the truncation operator product $\mathcal{T}(> D - 2)\mathcal{T}(> D - 2)$ that occurs in the intermediate expression in (9.73) is redundant. From (9.73) it follows that the operators $\mathcal{L}_D(: f :)$ for $: f :$ in L_2 form a Lie algebra. Indeed, we have the result

$$\begin{aligned} : f :, : g : \in L_2 \Rightarrow \{\mathcal{L}_D(: f :), \mathcal{L}_D(: g :)\} &= \mathcal{L}_D(\{ : f :, : g :\}) \\ &= \mathcal{L}_D(: [f, g] :). \end{aligned} \quad (8.9.74)$$

Finally we find that all the $\mathcal{L}_D(: f :)$, for fixed D and $: f :$ in L_2 , have only a finite number of nonzero matrix elements. Suppose either G_r or G_s have degree greater than $(D - 2)$. Then from (9.57) and (9.68) through (9.70) we find the result

$$\begin{aligned} O_{sr}(\mathcal{L}_D(: f :)) &= \langle G_s, \mathcal{T}(> D - 2) : f : \mathcal{T}(> D - 2) G_r \rangle \\ &= 0 \text{ when } \deg(G_r) > D - 2 \text{ or } \deg(G_s) > D - 2. \end{aligned} \quad (8.9.75)$$

Recall (3.32) and the fact that each monomial has a definite degree.

We now claim that the matrices $O(\mathcal{L}_D(: f :))$ provide a faithful representation of the quotient algebra L_2/L_D . By construction we have the relation

$$O(\mathcal{L}_D(: f : + : g :)) = O(\mathcal{L}_D(: f :)) + O(\mathcal{L}_D(: g :)), \quad (8.9.76)$$

and from (3.41), (9.73), (9.74) we find the relations

$$O(\mathcal{L}_D(: f :))O(\mathcal{L}_D(: g :)) = O(\mathcal{L}_D(: f :: g :)), \quad (8.9.77)$$

$$\{O(\mathcal{L}_D(: f :)), O(\mathcal{L}_D(: g :))\} = O(\mathcal{L}_D(: [f, g] :)). \quad (8.9.78)$$

Consequently, the matrices $O(\mathcal{L}_D(: f :))$ provide a representation of L_2 . Also, according to (9.72), all elements in the ideal L_D are mapped to the zero matrix. Therefore, in view of (9.76), the matrix $O(\mathcal{L}_D(: f :))$ depends only on the equivalence class to which $: f :$ belongs. Thus, the matrices $O(\mathcal{L}_D(: f :))$ provide a representation of the quotient algebra L_2/L_D . Finally, it remains to be shown that this representation is *faithful*. That is, given a matrix $O(\mathcal{L}_D(: f :))$, we need to be able to determine the equivalence class to which $: f :$ belongs. Suppose that f is written in the form

$$f = f_2 + f_3 + \cdots + f_{D-1}. \quad (8.9.79)$$

Compute the action of $\mathcal{L}_D(: f :)$ on z_a . We find the result

$$\begin{aligned} \mathcal{L}_D(: f :)z_a &= \mathcal{L}_D(: f_2 :)z_a + \mathcal{L}_D(: f_3 :)z_a + \cdots + \mathcal{L}_D(: f_{D-1} :)z_a \\ &= \mathcal{T}(> D - 2)(: f_2 : z_a + : f_3 : z_a + \cdots + : f_{D-1} : z_a) \\ &= \mathcal{T}(> D - 2)[g_a(1, z) + g_a(2, z) + \cdots + g_a(D - 2, z)] \\ &= g_a(1, z) + g_a(2, z) + \cdots + g_a(D - 2, z). \end{aligned} \quad (8.9.80)$$

Here we have used the notation

$$g_a(m, z) = : f_{m+1} : z_a = [f_{m+1}, z_a], \quad (8.9.81)$$

and observe that the $g_a(m, z)$ are homogeneous polynomials of degree m . As is evident from (9.80), the polynomials $g_a(m, z)$ can be determined from a knowledge of the matrix elements

$$\begin{aligned} O_{ra} &= \langle G_r, \mathcal{L}_D(: f :)z_a \rangle \\ &= \langle G_r, g_a(1, z) \rangle + \langle G_r, g_a(2, z) \rangle + \cdots + \langle G_r, g_a(D - 2, z) \rangle \end{aligned} \quad (8.9.82)$$

with the G_r having degree

$$\deg(G_r) \leq D - 2. \quad (8.9.83)$$

Since the $g_a(m, z)$ are now known, we can determine the f_m from (9.81). See (7.6.24).

Is the matrix representation of L_2/L_D just described the adjoint representation? It is not. If it were, the matrix elements of $O(\mathcal{L}_D(: f :))$ would be related to the structure constants of the quotient algebra, which are the ${}^6c_{tr}^s$ where both the subscripts s and r refer to basis elements for the quotient algebra. See (9.28) and (9.64). But examination of (9.82) shows

that the functions z_a were used to compute matrix elements, and the Lie operators $:z_a:$ are not in L_2 , and therefore not even candidates for the quotient algebra L_2/L_D .

Consider the “truncated” analog of (9.61) written in the form

$$\begin{aligned} \mathcal{M}_f^T = & \exp(\mathcal{L}_D(:f_2^c:)) \exp(\mathcal{L}_D(:f_2^a:)) \exp(\mathcal{L}_D(:f_3:)) \times \\ & \exp(\mathcal{L}_D(:f_4:)) \cdots \exp(\mathcal{L}_D(:f_{D-1}:)). \end{aligned} \quad (8.9.84)$$

Also, arrange the labelling of the basis functions G_r so that G_0 corresponds to the constant function 1 (the monomial of degree zero), the G_a (with $a = 1 \cdots d$) correspond to the linear monomials z_a , and the subsequent monomials G_r [with $r = d + 1 \cdots S(D-2, d)$] correspond to the monomials of degrees $2, 3, \dots, (D-2)$. Take matrix elements of both sides of (9.84) using the basis G_r with $r = 1, 2 \cdots S(D-2, d)$. Doing so gives the result

$$\begin{aligned} M_f^T = & \exp(O(\mathcal{L}_D(:f_2^c:))) \exp(O(\mathcal{L}_D(:f_2^a:))) \exp(O(\mathcal{L}_D(:f_3:))) \times \\ & \exp(O(\mathcal{L}_D(:f_4:))) \cdots \exp(O(\mathcal{L}_D(:f_{D-1}:))). \end{aligned} \quad (8.9.85)$$

Here use has been made of relations of the form (9.76) and (9.77). We see that (9.85) provides a $S(D-2, d) \times S(D-2, d)$ matrix representation of the quotient Lie *group* associated with the quotient Lie algebra L_2/L_D . Moreover, this representation is faithful. To see the truth of this assertion, consider the matrix elements

$$(M_f^T)_{ra} = \langle G_r, \mathcal{M}_f^T G_a \rangle = \langle G_r, \mathcal{L}_D(\mathcal{M}_f) G_a \rangle, \quad (8.9.86)$$

with

$$r \in [1, S(D-2, d)] \text{ and } a \in [1, d]. \quad (8.9.87)$$

From these matrix elements we can determine the coefficients in the Taylor series (7.6.1) through terms of degree $(D-2)$, and from these coefficients we can determine the polynomials $f_2^c, f_2^a, f_3, \dots, f_{D-1}$.

The last concept to be discussed in this section is gradings. For our purposes, a *grading* of a vector space V is a decomposition of V into a direct sum of subspaces along with a function gr (called the *grading* function) that assigns an integer (called the *grade*) to all the elements of any subspace. For example, we may take as our vector space V the set of all analytic functions $f(z)$ on phase space. Any such function can be decomposed into homogeneous polynomials by writing

$$f = f_0 + f_1 + f_2 + \cdots, \quad (8.9.88)$$

and these polynomials are in the subspaces we have called \mathcal{P}_m . In this case we may define the function gr by the rule

$$gr(f_m) = \deg(f_m) = m. \quad (8.9.89)$$

Elements of V that satisfy (9.89) are called *homogeneous*. Suppose there is some multiplication rule, \circ , that makes V into an algebra. See Section 3.7. Suppose also that this multiplication rule, and the direct sum decomposition, are such that the product of any two homogeneous elements is also homogeneous. The multiplication rule and grading function are said to be *compatible* if, for all homogeneous elements, we have the relation

$$gr(f_m \circ g_n) = gr(f_m) + gr(g_n). \quad (8.9.90)$$

For example, in the case of functions, we may take for \circ the operation of ordinary function multiplication. Then, if we use the definition (9.89), we find the result

$$\text{gr}(f_m \circ g_n) = \deg(f_m g_n) = m + n = \text{gr}(f_m) + \text{gr}(g_n), \quad (8.9.91)$$

which shows that ordinary function multiplication and the grading function (9.89) are compatible.

Suppose we use Poisson bracket multiplication for \circ instead of ordinary multiplication. Then (4.28) shows that (9.89) is not compatible with Poisson bracket multiplication. However, if we define gr by the rule

$$\text{gr}(f_m) = m - 2, \quad (8.9.92)$$

we find the result

$$\begin{aligned} \text{gr}([f_m, g_n]) &= \text{gr}(\mathcal{P}_{m+n-2}) = m + n - 2 - 2 \\ &= m - 2 + n - 2 = \text{gr}(f_m) + \text{gr}(g_n). \end{aligned} \quad (8.9.93)$$

Thus, the grading function (9.92) is compatible with Poisson bracket multiplication. A Lie algebra equipped with a grading (compatible with the Lie product) is called a *graded* Lie algebra.

Given a graded Lie algebra, it is easy to find subalgebras and ideals. Consider the case of analytic functions defined on phase space. Introduce the notation

$$f^{n-2}(z) = f_n(z) \quad (8.9.94)$$

to indicate, in accord with (9.92), that homogeneous polynomials of degree n have grade $(n - 2)$. Equivalently, we have the relation

$$\text{gr}(f^m) = m. \quad (8.9.95)$$

We also introduce the notation

$$\mathcal{P}^{n-2} = \mathcal{P}_n \quad (8.9.96)$$

to indicate the subspace of polynomials of degree n and grade $(n - 2)$. Finally, we extend the concept of grade to Lie operators by the rule

$$\text{gr}(: f^m :) = \text{gr}(f^m) = m. \quad (8.9.97)$$

Now consider the space of all Lie operators spanned by basis elements of the form $: f^0 :, : f^1 :, : f^2 :, \dots$. Then, because of the relation

$$\begin{aligned} \text{gr}(\{ : f^m :, : f^n : \}) &= \text{gr}(: [f^m, f^n] :) = \text{gr}([f^m, f^n]) \\ &= \text{gr}(f^m) + \text{gr}(f^n) = m + n, \end{aligned} \quad (8.9.98)$$

we see that

$$\{ : f^m :, : f^n : \} \in : \mathcal{P}^{m+n} :, \quad (8.9.99)$$

and hence this space is a Lie algebra. This is just the Lie algebra that we found and called L_2 earlier, and that we now will also call L^0 . Similarly, we can use arguments based on

grading to show that $L^m = L_{m+2}$ (with $m > 0$) is a subalgebra of $L^0 = L_2$, and also an ideal in L^0 . (Indeed, the arguments we used earlier for this purpose were actually grading arguments without being identified as such.)

Note that the Lie algebra $L^{-1} = L_1$ also has L^m as a subalgebra. However L^m is not an ideal in L^{-1} since L^{-1} contains $:f^{-1} :=: f_1:$ which, according to (9.99), can lower the grade of elements in L^m until they are no longer in L^m . We have seen that L^0 can be approximated by using the finite dimensional quotient algebras $L^0/L^m = L_2/L_{m+2}$. Is there some way that we can approximate L^{-1} in a consistent Lie algebraic way even though L^m is not an ideal in L^{-1} ? We would certainly like to do so since maps of the form (7.8.1) are of interest when the origin is not mapped into itself. The answer to the question just posed is yes provided we are willing to treat the f_1 as being in some sense *small*. Fortunately this circumstance is the one usually encountered since, as we will see in Chapter 14, f_1 terms are usually associated with misalignment, misplacement, and mispowering errors, and these errors are generally small.

Before considering the inclusion of small f_1 terms, let us return to the case described at the beginning of this section. Let ϵ be a (presumably small) parameter, and define V to be the vector space spanned by all phase-space functions of the form $f^{(0)}, \epsilon f^{(1)}, \epsilon^2 f^{(2)} \dots$. Here the $f^{(n)}$ are *arbitrary* analytic functions not to be confused with the f^n defined earlier. Assign a grade to the subspaces $\epsilon^n f^{(n)}$ and the associated Lie operators $:\epsilon^n f^{(n)}:$ by the rules

$$\text{gr}(\epsilon^n f^{(n)}) = n, \quad (8.9.100)$$

$$\text{gr}(:\epsilon^n f^{(n)}:) = n. \quad (8.9.101)$$

This grading function is evidently compatible with Lie multiplication,

$$\begin{aligned} \text{gr}([\epsilon^m f^{(m)}, \epsilon^n f^{(n)}]) &= \text{gr}(\epsilon^{m+n}[f^{(m)}, f^{(n)}]) \\ &= m + n = \text{gr}(\epsilon^m f^{(m)}) + \text{gr}(\epsilon^n f^{(n)}), \end{aligned} \quad (8.9.102)$$

$$\begin{aligned} \text{gr}(:\epsilon^m f^{(m)} : , :\epsilon^n f^{(n)} :) &= \text{gr}(:[\epsilon^m f^{(m)}, \epsilon^n f^{(n)}] :) \\ &= m + n = \text{gr}(:\epsilon^m f^{(m)} :) + \text{gr}(:\epsilon^n f^{(n)} :). \end{aligned} \quad (8.9.103)$$

Here we have used the fact that $[f^{(m)}, f^{(n)}]$ is again an arbitrary analytic phase-space function. Thus we may use this grading to construct subalgebras, ideals, and quotient algebras. Suppose we define L^n to be the vector space spanned by $\epsilon^n f^{(n)}, \epsilon^{n+1} f^{(n+1)}, \epsilon^{n+2} f^{(n+2)} \dots$. Then it is easily verified that the L^n are Lie algebras. Moreover, any L^n for $n > 0$ is an ideal in L^0 , and each L^0/L^n is a quotient algebra. Finally, it is evident that working with the quotient algebra L^0/L^n is equivalent to doing perturbation theory in ϵ and retaining only those terms that carry powers of ϵ of the form $\epsilon^0, \epsilon^1, \dots, \epsilon^{n-1}$. What we have learned is that *finite* order perturbation theory is a consistent Lie algebraic procedure.

We now turn to the inclusion of small f_1 terms. Again let ϵ be a parameter. Consider now the vector space V spanned by all functions of the form $\epsilon^m f_n$. Assign a grade to these subspaces and their associated Lie operators $:\epsilon^m f_n:$ by the rules

$$\text{gr}(\epsilon^m f_n) = m + n - 2, \quad (8.9.104)$$

$$\text{gr}(: \epsilon^m f_n :) = m + n - 2. \quad (8.9.105)$$

With this definition we find that $\epsilon^2 f_0$, ϵf_1 , and f_2 have grade 0; $\epsilon^3 f_0$, $\epsilon^2 f_1$, ϵf_2 , and f_3 have grade 1; $\epsilon^4 f_0$, $\epsilon^3 f_1$, $\epsilon^2 f_2$, ϵf_3 , and f_4 have grade 2; etc. From the previous discussion it is easy to see that the grading functions (9.104) and (9.105) are compatible with Lie multiplication. Consequently, we may use it as before to construct subalgebras, ideals, and quotient algebras. Let ${}^\epsilon L^\ell$ denote the vector space of functions spanned by elements of the form $\epsilon^m f_n$ with $(m + n - 2) \geq \ell$ [that is, $\text{gr}(\epsilon^m f_n) \geq \ell$]. Then, following arguments given previously, it is easy to check that the ${}^\epsilon L^\ell$ (with $\ell \geq 0$) are Lie algebras, ${}^\epsilon L^\ell$ with $\ell > 0$ is an ideal in ${}^\epsilon L^0$, and each ${}^\epsilon L^0 / {}^\epsilon L^\ell$ is a quotient algebra. We also see that ${}^\epsilon L^0$ contains the element ϵf_1 , which can be interpreted as being a small f_1 . Finally, we observe that the quotient algebra ${}^\epsilon L^0 / {}^\epsilon L^\ell$, for fixed ℓ , is finite dimensional. (These dimensions are listed below in Table 9.2 for the cases of four and six-dimensional phase spaces. In computing these dimensions we set $\epsilon = 1$ and ignored all terms of the form $\epsilon^m f_0$.) Consequently, we have discovered a systematic and Lie algebraically consistent approximation scheme that includes small f_1 terms. This approximation scheme will be studied extensively in the next chapter.

Table 8.9.2: Values of $\dim({}^\epsilon L^0 / {}^\epsilon L^\ell)$.

ℓ	dim for $d = 4$	dim for $d = 6$
1	14	27
2	34	83
3	69	209
4	125	461
5	209	923
6	329	1715
7	494	3002
8	714	5004
9	1000	8007
10	1364	12,375
11	1819	18,563

Exercises

8.9.1. Review Exercise 5.12.7. Let L be a vector space having a vector subspace L' . Show that (9.2) defines (satisfies the properties of) an equivalence relation among the elements of L .

8.9.2. Verify that the vectors $\{v_i\}$ used in (9.16) are linearly independent. Hint: Assume there exist scalars α_i such that

$$\sum_{i=1}^n \alpha_i \{v_i\} = \{0\}.$$

Show that there exists a vector $x' \in L'$ such that

$$\sum_{i=1}^n \alpha_i v_i = x'.$$

Show that since $x' \in L'$, it must have an expansion of the form

$$x' = \sum_j \beta_j b_j.$$

Now show that all the coefficients α_i and β_j must vanish.

8.9.3. Verify the relation

$$[\{0\}, \{x\}]_{qs} = \{0\}.$$

Show that the addition rule (9.12) and Lie product rule (9.18) together satisfy requirements 1 through 5 for a Lie algebra as given in Section 3.7.

8.9.4. Verify (9.30). Verify (9.38) in detail. Show that the definition (9.35) is independent of equivalence class labeling if G' is normal. Verify that the definition (9.32) satisfies the associative property

$$\{f_1\}(\{g_1\}\{h_1\}) = (\{f_1\}\{g_1\})\{h_1\}.$$

8.9.5. If G is a group with a subgroup G' , the subgroup G' can be used to produce equivalence classes in G in two possibly different ways. First, there is the equivalence relation \sim defined by

$$g_2 \sim g_1 \Leftrightarrow g_1^{-1}g_2 \in G'.$$

Second, there is another equivalence relation, let us denote it by the symbol \leftrightarrow , defined by

$$g_2 \leftrightarrow g_1 \Leftrightarrow g_2g_1^{-1} \in G'.$$

We know that \sim decomposes G into *left* coset equivalence classes. Show that \leftrightarrow is indeed an equivalence relation, and that it decomposes G into *right* coset equivalence classes. Suppose g is any element of G . Let $\{g\}_\ell$ denote the set of all elements in G that are equivalent to g using the relation \sim , and let $\{g\}_r$ denote the set of all elements in G that are equivalent to g using the relation \leftrightarrow . Show that

$$\{g\}_\ell = \{g\}_r.$$

Next assume that G' is normal. In this case show that

$$\{g\}_\ell = \{g\}_r$$

for any g in G . Thus, left and right cosets are the same if G' is normal.

8.9.6. The *center* of a group G consists of those elements of G that commute with all elements of G . See Exercise 7.2.13. Show that the center of a group is a special case of an invariant (normal) subgroup. Review Exercise 5.11.1.

8.9.7. Verify (9.43).

8.9.8. Show that if G' with Lie algebra L' is a normal subgroup of a Lie group G with Lie algebra L , then L' is an ideal in L .

8.9.9. Starting with (9.64), verify (9.67) and show that it is equivalent to (3.7.42).

8.9.10. Find and describe the adjoint representation of L_2 . What is its dimension? Find and describe the adjoint representation of L_2/L_D . What is its dimension?

8.9.11. Verify (9.73) and (9.74).

8.9.12. Verify (9.77) and (9.78).

8.9.13. Verify (9.85).

8.9.14. Consider a 2-dimensional phase space with canonical variables q, p . Referring to (9.70), find the matrix elements of $\mathcal{L}_4(: q^3 :)$ and $\exp(\mathcal{L}_4(: q^3 :))$. See (9.75), (9.84), and (9.85).

8.9.15. Read Exercise 9.14. Find the matrix representations for the Lie algebras L_2/L_3 and L_2/L_4 in the case of a 2-dimensional phase space.

8.9.16. Verify (9.93).

8.9.17. Use the grading (9.104) and (9.105). Show that it is compatible with Lie multiplication.

8.9.18. Let ${}^\epsilon L^\ell$ denote the vector space of functions spanned by elements of the form ${}^\epsilon m f_n$ with $(m + n - 2) \geq \ell$. Show that the ${}^\epsilon L^\ell$ with $\ell \geq 0$ are Lie algebras, ${}^\epsilon L^\ell$ with $\ell > 0$ is an ideal in ${}^\epsilon L^0$, and each ${}^\epsilon L^0/{}^\epsilon L^\ell$ is a quotient Lie algebra. For a given ℓ and assuming a d -dimensional phase space, show that the dimension of ${}^\epsilon L^0/{}^\epsilon L^\ell$ is given in terms of (7.10.4) by the relation

$$\dim({}^\epsilon L^0/{}^\epsilon L^\ell) = S(\ell + 1, d). \quad (8.9.106)$$

In computing these dimensions, set $\epsilon = 1$ and ignore all terms of the form ${}^\epsilon m f_0$. Verify Table 9.2.

8.9.19. Review Exercise 8.2.12. This exercise is a continuation of that exercise. Our task is to show that the $L^\alpha = L(K^\alpha)$ form a basis for $so(6, \mathbb{R})$. This task could be carried out by computing all the L^α and verifying that they are indeed linearly independent. Instead, we will use an approach that is less tedious but also more abstract.

Suppose, to the contrary, that the $L^\alpha = L(K^\alpha)$ do not form a basis, and are therefore not linearly independent. Then there are constants λ_α , not all zero, such that

$$\sum_{\alpha} \lambda_\alpha L^\alpha = 0. \quad (8.9.107)$$

Use the constants λ_α to form an $su(4)$ element, call it $K(\lambda)$, by the rule

$$K(\lambda) = \sum_{\alpha} \lambda_\alpha K^\alpha. \quad (8.9.108)$$

We know that $K(\lambda) \neq 0$ because the K^α are linearly independent and not all the λ_α are zero. Show that

$$L[K(\lambda)] = \sum_{\alpha} \lambda_\alpha L(K^\alpha) = \sum_{\alpha} \lambda_\alpha L^\alpha = 0. \quad (8.9.109)$$

Next, let \mathcal{I} be the set of all elements in $su(4)$ of the form

$$K(\sigma) = \sum_{\alpha} \sigma_{\alpha} K^{\alpha} \quad (8.9.110)$$

such that

$$L[K(\sigma)] = 0. \quad (8.9.111)$$

Show that \mathcal{I} is a Lie subalgebra of $su(4)$. That is, show that \mathcal{I} is a linear vector space and verify that

$$L[\{K(\sigma), K(\sigma')\}] = \{L[K(\sigma)], L[K(\sigma')]\} = 0 \quad (8.9.112)$$

so that $\{K(\sigma), K(\sigma')\} \in \mathcal{I}$ if $K(\sigma) \in \mathcal{I}$ and $K(\sigma') \in \mathcal{I}$. Finally, show that

$$L[\{K^{\beta}, K(\sigma)\}] = \{L(K^{\beta}), L[K(\sigma)]\} = 0 \quad (8.9.113)$$

for any β and any $K(\sigma) \in \mathcal{I}$. It follows that \mathcal{I} is an *invariant* subalgebra of $su(4)$. Also, we know that \mathcal{I} is not empty because, by hypothesis, it contains $K(\lambda)$. Nor is it all of $su(4)$ because, for example, inspection of (2.164) shows that $L(K^1) \neq 0$. Therefore \mathcal{I} is an *ideal*. But, this is a contradiction because $su(4)$ is supposed to be *simple*, i.e. have no ideals. See Section 3.7.6.

Bibliography

Relations between Orthogonal and Unitary Groups

- [1] J. D. Louck and H. W. Galbraith, “Application of Orthogonal and Unitary Group Methods to the N -Body Problem”, *Reviews of Modern Physics* **44**, 540-601 (1972).

The Lorentz Group

- [2] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That*, W. A. Benjamin (1964).

Clifford Algebras, Spinors, and Lie Theory

- [3] P. Lounesto, *Clifford algebras and spinors* (2nd ed.), Cambridge University Press (2001).

- [4] R. Delanghe, F. Sommen, and V. Souček, *Clifford Algebra and Spinor-Valued Functions: A Function Theory For The Dirac Operator*, Springer Science (1992).

- [5] E. Meinrenken, *Clifford Algebras and Lie Theory*, Springer Verlag (2013).

Connection between Linear Operators and Matrices

- [6] J.M. Finn, “Integrals of Canonical transformations and Normal Forms for Mirror Machine Hamiltonians”, University of Maryland College Park Physics Department Ph.D. thesis (1974).

- [7] K. Kowalski and W-H. Steeb, *Nonlinear Dynamical Systems and Carleman Linearization*, World Scientific (1991).

- [8] K. Kowalski, *Methods of Hilbert Spaces in the Theory of Nonlinear Dynamical Systems*, World Scientific (1994).

- [9] P. Gralewicz and K. Kowalski, “Continuous time evolution from iterated maps and Carleman linearization”, *Chaos, Solitons, and Fractals* **14**, 563-572 (2002).

Transforming Expressions to Commutator Form

- [10] E.B. Dynkin, *Selected Papers of E.B. Dynkin with Commentary*, American Mathematical Society (2000).

- [11] A. Dragt and E. Forest, “Computation of nonlinear behavior of Hamiltonian systems using Lie algebraic methods”, *J. Math. Phys.* **24**, p. 2734 (1983). See Section 11 and Appendix.

Single Exponent Form

- [12] S. Habib and R. Ryne, “Symplectic Calculation of Lyapunov Exponents”, arXiv:acc-physics/9411001v1, (1994).
- [13] I. Gjaja, “Closed-Form Expressions for the Noncompact Part of $Sp(2n)$ ”, arXiv:chao-dyn/9602013v1, (1996).

Zassenhaus Formulas

- [14] R.M. Wilcox, “Exponential Operators and Parameter Differentiation in Quantum Physics”, *J. Math. Phys.* **8**, p. 962 (1967).
- [15] F. Casas, A. Murua, and M. Nadinic, “Efficient computation of the Zassenhaus formula”, available on arXiv (2012) at <http://arxiv.org/abs/1204.0389>.
- [16] F. Bayen, “On the convergence of the Zassenhaus formula”, *Lett. Math. Phys.* **3**, 161-167 (1979).

Operational Calculus for Noncommuting Operators

- [17] G. Johnson and M. Lapidus *The Feynman Integral and Feynman’s Operational Calculus*, Oxford University Press (2003).
- [18] G. Johnson, M. Lapidus, and L. Nielsen *Feynman’s Operational Calculus and Beyond: Noncommutativity and Time-Ordering*, Oxford University Press (2015).

Ideals and Quotient Groups

- [19] N. Jacobson, *Lectures in Abstract Algebra, Vol. I - Basic Concepts*, D. Van Nostrand (Princeton, 1951).

Gradings

- [20] I. Dorfman, *Dirac Structure and Integrability of Nonlinear Evolution Equations*, John Wiley and Sons (Chichester, 1993).

Chapter 9

Inclusion of Translations in the Calculus

9.1 Introduction

In Chapter 8 we dealt with, among other things, the restricted problem of concatenating maps, all of which had the property of sending the origin into itself. In this chapter we consider the general case. Let \mathcal{M}_f and \mathcal{M}_g denote the general symplectic maps given by the expressions

$$\mathcal{M}_f = \exp(: f_1 :) \exp(: f_2^c :) \exp(: f_2^a :) \exp(: f_3 :) \exp(: f_4 :) \cdots, \quad (9.1.1)$$

$$\mathcal{M}_g = \exp(: g_1 :) \exp(: g_2^c :) \exp(: g_2^a :) \exp(: g_3 :) \exp(: g_4 :) \cdots. \quad (9.1.2)$$

Also, let \mathcal{M}_h be the product of \mathcal{M}_f and \mathcal{M}_g ,

$$\mathcal{M}_h = \mathcal{M}_f \mathcal{M}_g. \quad (9.1.3)$$

Given \mathcal{M}_f and \mathcal{M}_g , our problem will be to find polynomials $h_1, h_2^c, h_2^a, h_3, h_4$, etc. such that

$$\mathcal{M}_h = \exp(: h_1 :) \exp(: h_2^c :) \exp(: h_2^a :) \exp(: h_3 :) \exp(: h_4 :) \cdots. \quad (9.1.4)$$

For future use it is convenient to use the notation

$$\mathcal{R}_f = \exp(: f_2^c :) \exp(: f_2^a :), \text{ etc.} \quad (9.1.5)$$

In this notation, our goal is to write \mathcal{M}_h in the form

$$\mathcal{M}_h = \exp(: h_1 :) \mathcal{R}_h \exp(: h_3 :) \exp(: h_4 :) \cdots. \quad (9.1.6)$$

Section 9.2 treats the special case where both the maps \mathcal{M}_f and \mathcal{M}_g produce only constant and linear terms when acting on z_a . This is the case of $ISp(2n, \mathbb{R})$ where all the nonlinear generators f_3, f_4, \dots and g_3, g_4, \dots are assumed to be zero. See Section 6.2 and Exercise 7.7.2. In this case we will be able to solve all possible problems to our hearts' content.

Subsequent sections will treat the general case where the nonlinear generators are also present. In this case, following the discussion in Section 8.9, we will find it necessary to introduce a grading in which f_1 and g_1 are treated as being small.

9.2 The Inhomogeneous Symplectic Group $ISp(2n, \mathbb{R})$

The *inhomogeneous* symplectic group $ISp(2n, \mathbb{R})$ consists of all transformations of phase space of the form

$$\bar{z}_a = \delta_a + \sum_b R_{ab} z_b \quad (9.2.1)$$

where the δ_a are arbitrary constants and R is a symplectic matrix. It is closely related to the *Jacobi* group. See Exercises 6.2.2, 7.7.2, and 7.7.3. We already know from Section 7.7 that there is a map \mathcal{M} such that

$$\bar{z} = \mathcal{M}z, \quad (9.2.2)$$

and \mathcal{M} has the factorization

$$\mathcal{M} = \exp(: f_2^c :) \exp(: f_2^a :) \exp(: g_1 :). \quad (9.2.3)$$

Indeed, f_2^a and f_2^c are determined by R using (7.6.17), (7.2.2), (7.2.3), and (7.2.8); and g_1 is given in terms of δ by (7.7.3). Lie operators of the form $:f_1:$ and $:f_2:$ provide a basis for $isp(2n, \mathbb{R})$, the Lie algebra of $ISp(2n, \mathbb{R})$. The purpose of this section is to state and prove various rearrangement, factorization, and concatenation formulas for the inhomogeneous symplectic group.

9.2.1 Rearrangement Formula

We begin with a *rearrangement* formula. Let us rewrite (2.3) in the form

$$\mathcal{M}_f = \mathcal{R}_f \exp(: g_1 :) \quad (9.2.4)$$

with \mathcal{R}_f being the linear map defined by (1.5). In terms of this notation we have the results

$$\mathcal{R}_f z = R_f z, \quad (9.2.5)$$

$$\exp(: g_1 :) z = z + \delta, \quad (9.2.6)$$

$$\bar{z} = \mathcal{M}_f z = \mathcal{R}_f \exp(: g_1 :) z = \mathcal{R}_f(z + \delta) = \delta + R_f z. \quad (9.2.7)$$

Next let δ' be another set of constants. Using (7.7.3), define a first-degree polynomial f_1 such that

$$\exp(: f_1 :) z = z + \delta'. \quad (9.2.8)$$

Consider the map $\exp(: f_1 :) \mathcal{R}_f$. It has the action

$$\exp(: f_1 :) \mathcal{R}_f z = \exp(: f_1 :) R_f z = R_f(z + \delta') = R_f \delta' + R_f z. \quad (9.2.9)$$

Upon comparing (2.7) and (2.9) we see that there will be the equality

$$\mathcal{R}_f \exp(: g_1 :) = \exp(: f_1 :) \mathcal{R}_f \quad (9.2.10)$$

provided there is the relation

$$R_f \delta' = \delta. \quad (9.2.11)$$

We will call the relation specified by (2.10) and (2.11) a rearrangement formula.

There is another way to obtain the rearrangement formula. Insert the identity factor $\mathcal{R}_f^{-1}\mathcal{R}_f$ on the right side of (2.4) to get the result

$$\mathcal{M}_f = \mathcal{R}_f \exp(: g_1 :) \mathcal{R}_f^{-1} \mathcal{R}_f. \quad (9.2.12)$$

Next make use of (8.2.25) to write

$$\mathcal{R}_f \exp(: g_1 :) \mathcal{R}_f^{-1} = \exp(: \mathcal{R}_f g_1 :). \quad (9.2.13)$$

Now define a first-degree polynomial f_1 by the rule

$$f_1 = \mathcal{R}_f g_1. \quad (9.2.14)$$

We have found the result

$$\mathcal{M}_f = \mathcal{R}_f \exp(: g_1 :) = \exp(: f_1 :) \mathcal{R}_f \quad (9.2.15)$$

with f_1 and g_1 related by (2.14).

What remains is to make (2.15) more explicit. According to the work of Section 7.7 there are the relations

$$f_1(z) = (J\delta', z), \quad (9.2.16)$$

$$g_1(z) = (J\delta, z). \quad (9.2.17)$$

Consequently, (2.15) can be rewritten in the form

$$(J\delta', z) = f_1(z) = \mathcal{R}_f g_1 = .\mathcal{R}_f(J\delta, z) = (J\delta, R_f z) = (R_f^T J\delta, z). \quad (9.2.18)$$

Upon comparing the left and right sides of (2.18) we conclude that

$$J\delta' = R_f^T J\delta. \quad (9.2.19)$$

Now multiply both sides of (2.19) by $-R_f J$. Doing so to the left side of (2.19) yields the result

$$-R_f J J\delta' = R_f \delta'. \quad (9.2.20)$$

And doing so to the right side of (2.19) yields the result

$$-R_f J R_f^T J\delta = -J J\delta = \delta. \quad (9.2.21)$$

Upon comparing the right sides of (2.20) and (2.21) we see that (2.11) has been recovered.

9.2.2 Factorization Formula

The next result to obtain is a *factorization* formula. Given any two homogeneous polynomials h_1 and h_2 (of degrees 1 and 2, respectively), there exist related polynomials f_1, f_2 that satisfy the *factorization formula*

$$\exp(: h_1 + h_2 :) = \exp(: f_2 :) \exp(: f_1 :). \quad (9.2.22)$$

There are at least two ways to prove (2.22). The first amounts to combining the two exponents on the right side of (2.22), and then matching terms with the left side. Consider all Lie products made from f_1 and f_2 . We observe that all Lie products containing two or more f_1 factors, for example $[f_1, [f_1, f_2]]$, must give only constant terms, and hence their associated Lie operators vanish. Let s and t be small parameters. According to formula (8.7.3), we have the result

$$\begin{aligned} \exp(s : f_2 :) \exp(t : f_1 :) = \\ \exp[s : f_2 : + : \{s : f_2 : [1 - \exp(-s : f_2 :)]^{-1}(tf_1)\} : + : O(t^2) :], \end{aligned} \quad (9.2.23)$$

where the notation $: O(t^2) :$ indicates Lie operators that contain at least *two* factors of t . But the presence of two factors of t requires the presence of two factors of f_1 and hence, by the previous observation these Lie operators must all vanish. It follows that (2.23) is *exact* for f_1 and f_2 . Now set $s = t = 1$ in (2.23) and combine the result so obtained with (2.22) to get the relation

$$\exp(: h_1 + h_2 :) = \exp[: f_2 : + : \{ : f_2 : [1 - \exp(- : f_2 :)]^{-1}f_1\} :]. \quad (9.2.24)$$

Upon comparing like terms in (2.6), we find the relations

$$f_2 = h_2, \quad (9.2.25)$$

$$\begin{aligned} h_1 &= : f_2 : [1 - \exp(- : f_2 :)]^{-1}f_1 \\ &= : h_2 : [1 - \exp(- : h_2 :)]^{-1}f_1. \end{aligned} \quad (9.2.26)$$

Finally, (2.26) may be solved for f_1 to give the relation

$$\begin{aligned} f_1 &= \{[1 - \exp(- : h_2 :)]/[: h_2 :] \} h_1 \\ &= \text{iex}(- : h_2 :) h_1. \end{aligned} \quad (9.2.27)$$

See (8.8.9). We note that f_1 and f_2 are well defined by (2.27) and (2.25) for all h_1 and h_2 .

The converse question is more difficult: given f_1 and f_2 , does (2.26) always define an h_1 ? Or equivalently, in view of (2.27), does $[\text{iex}(- : f_2 :)]^{-1}f_1$ always exist? If not, then it is not possible to write every inhomogeneous symplectic group element in terms of a single exponential. See Exercise 2.3 for a discussion of this question.

There is a second derivation of (2.25) and (2.27) that is worth knowing. Consider the functions \bar{z}_a defined by the relation

$$\bar{z}_a = \exp(: sh_2 + th_1 :) z_a. \quad (9.2.28)$$

As before, s and t are parameters. Expanding (2.28) in a power series gives the result

$$\bar{z}_a = \sum_{n=0}^{\infty} [(: sh_2 + th_1 :)^n / n!] z_a. \quad (9.2.29)$$

Next expand the powers in (2.29). For $n \geq 1$ we have the result

$$\begin{aligned} (: sh_2 + th_1 :)^n &= (s : h_2 : + t : h_1 :)^n \\ &= s^n : h_2 :^n + ts^{n-1} \sum_{m=0}^{n-1} : h_2 :^m : h_1 :: h_2 :^{n-m-1} + O(t^2). \end{aligned} \quad (9.2.30)$$

Here we have kept track of the fact that $: h_1 :$ and $: h_2 :$ may not commute. Observe that the terms in (2.30) proportional to t^2 must have two factors of $: h_1 :$ and are therefore of the form

$$t^2 \text{ terms} \sim (: h_2 :)^{\alpha} : h_1 : (: h_2 :)^{\beta} : h_1 : (: h_2 :)^{\gamma} \quad (9.2.31)$$

where α, β, γ satisfy the relations

$$\begin{aligned} \alpha + \beta + \gamma &= n - 2, \\ \alpha \geq 0, \beta \geq 0, \gamma \geq 0. \end{aligned} \quad (9.2.32)$$

From (7.6.16) it is easy to verify the relation

$$(: h_2 :)^{\alpha} : h_1 : (: h_2 :)^{\beta} : h_1 : (: h_2 :)^{\gamma} z_a = 0. \quad (9.2.33)$$

Similarly, analogous results hold for terms having three or more factors of $: h_1 :$. It follows that all the $O(t^2)$ terms in (2.30) annihilate the z_a . Thus, we find the exact result

$$\begin{aligned} \exp(: sh_2 + th_1 :) z_a &= \\ \left\{ \sum_{n=0}^{\infty} s^n : h_2 :^n / n! + t \sum_{n=1}^{\infty} (s^{n-1}/n!) \sum_{m=0}^{n-1} : h_2 :^m : h_1 :: h_2 :^{n-m-1} \right\} z_a. \end{aligned} \quad (9.2.34)$$

The first term on the right side of (2.34) sums to the exponential function,

$$\sum_{n=0}^{\infty} s^n : h_2 :^n / n! = \exp(s : h_2 :). \quad (9.2.35)$$

The second term sums to a relation involving the exponential and integrated exponential functions,

$$\sum_{n=1}^{\infty} \sum_{m=0}^{n-1} (1/n!) : sh_2 :^m : th_1 :: sh_2 :^{n-m-1} = \exp(s : h_2 :) \text{iex}(-\#sh_2\#) : th_1 :. \quad (9.2.36)$$

See Appendix C. As a consequence of relations of the form (8.2.22), we have the result

$$\text{iex}(-\#sh_2\#) : th_1 :=: \text{iex}(- : sh_2 :) th_1 :. \quad (9.2.37)$$

Putting (2.34) through (2.37) together gives the relation

$$\exp(: sh_2 + th_1 :) z_a = \exp(s : h_2 :)[1 + : \text{iex}(- : sh_2 :) th_1 :] z_a. \quad (9.2.38)$$

Suppose we define f_1 by writing

$$f_1 = \text{iex}(- : sh_2 :) h_1. \quad (9.2.39)$$

Again from (7.6.16) we have the relation

$$[1 + t : f_1 :] z_a = \exp(t : f_1 :) z_a. \quad (9.2.40)$$

We conclude that (2.38) can be rewritten in the form

$$\exp(: sh_2 + th_1 :) z_a = \exp(s : h_2 :) \exp(t : f_1 :) z_a. \quad (9.2.41)$$

Since the operator factors on both sides of (2.23) are manifestly group elements (symplectic maps), we get the group (map) factorization relation

$$\exp(: sh_2 + th_1 :) = \exp(s : h_2 :) \exp(t : f_1 :) \quad (9.2.42)$$

with f_1 defined by (2.39). Now put $s = t = 1$ in (2.39) and (2.42) and make the definition (2.25) to recover (2.22), (2.25), and (2.27).

9.2.3 Concatenation Formulas

We have found the rearrangement formula given by (2.10) and (2.11) and the factorization formula given by (2.22), (2.25, and (2.27). We now turn to the easier subjects of concatenation formulas. Let \mathcal{M}_f and \mathcal{M}_g be the inhomogeneous symplectic group maps

$$\mathcal{M}_f = \exp(: f_1 :) \exp(: f_2^c :) \exp(: f_2^a :) = \exp(: f_1 :) \mathcal{R}_f, \quad (9.2.43)$$

$$\mathcal{M}_g = \exp(: g_1 :) \exp(: g_2^c :) \exp(: g_2^a :) = \exp(: g_1 :) \mathcal{R}_g. \quad (9.2.44)$$

Also, let \mathcal{M}_h be the product of \mathcal{M}_f and \mathcal{M}_g as in (1.3). We wish to find polynomials h_1 , h_2^c , and h_2^a such that

$$\mathcal{M}_h = \mathcal{M}_f \mathcal{M}_g = \exp(: h_1 :) \exp(: h_2^c :) \exp(: h_2^a :) = \exp(: h_1 :) \mathcal{R}_h. \quad (9.2.45)$$

From (2.43) and (2.44) we have the result

$$\mathcal{M}_f \mathcal{M}_g = \exp(: f_1 :) \mathcal{R}_f \exp(: g_1 :) \mathcal{R}_g. \quad (9.2.46)$$

Insert an identity factor of the form $\mathcal{R}_f^{-1} \mathcal{R}_f$ into (2.46) to find the relation

$$\mathcal{M}_f \mathcal{M}_g = \exp(: f_1 :) \mathcal{R}_f \exp(: g_1 :) \mathcal{R}_f^{-1} \mathcal{R}_f \mathcal{R}_g. \quad (9.2.47)$$

From (2.13) we conclude that (2.47) can be rewritten in the form

$$\begin{aligned} \mathcal{M}_f \mathcal{M}_g &= \exp(: f_1 :) \exp(: \mathcal{R}_f g_1 :) \mathcal{R}_f \mathcal{R}_g \\ &= \exp(: f_1 + \mathcal{R}_f g_1 :) \mathcal{R}_f \mathcal{R}_g. \end{aligned} \quad (9.2.48)$$

Here we have used the fact that the Lie operators associated with first-degree polynomials commute. (See Exercise 2.4.) Now compare (2.45) and (2.48) to get the concatenation formulas

$$h_1 = f_1 + \mathcal{R}_f g_1, \quad (9.2.49)$$

$$\mathcal{R}_h = \mathcal{R}_f \mathcal{R}_g. \quad (9.2.50)$$

We note that (2.50) is identical to (8.4.19), as expected, and consequently we also have the corresponding matrix relation (8.4.20) as before.

Exercises

9.2.1. Read (2.22) from right to left. That is, assume that f_1 and f_2 are two given homogeneous polynomials (of degrees 1 and 2, respectively), and we wish to find h_1 and h_2 . From Exercise 7.7.2 we know that $:f_1:$ and $:f_2:$ generate a Lie algebra. Also, we know from the BCH formula (3.7.33) and (3.7.34) that all terms that occur when we combine the exponents on the right side of (2.22) must be in this Lie algebra. Show that the most general such term is of the form $:h_1 + h_2:$ where h_1 and h_2 have degrees 1 and 2, respectively. Show from (8.7.3) that h_1 as defined by the first line in (2.26) does indeed have degree 1, so that terms of like degree have indeed been equated.

9.2.2. The purpose of this exercise is to convert (2.27) into an explicit matrix equation. Since h_1 is a given degree 1 polynomial, it can be written in the form

$$h_1 = \sum_a h_1^a z_a \quad (9.2.51)$$

where the h_1^a are known coefficients. From (7.6.16) the action of $:h_2:$ on the z_a is a relation of the form

$$:h_2: z_a = \sum_{a'} H_{a'a} z_{a'}. \quad (9.2.52)$$

The matrix H is given in terms of the scalar product (7.3.8) by the relation

$$H_{a'a} = (z_{a'}, :h_2: z_a). \quad (9.2.53)$$

Define a matrix O in terms of H by the rule

$$O = \text{iex}(-H) = \sum_{m=0}^{\infty} (-H)^m / (m+1)! \quad (9.2.54)$$

Show that the series (2.54) converges for any matrix H , and therefore O is well defined. Next verify the relation

$$\text{iex}(-:h_2:) z_a = \sum_{a'} O_{a'a} z_{a'}. \quad (9.2.55)$$

Suppose we write f_1 in the form

$$f_1 = \sum_{a'} f_1^{a'} z_{a'}. \quad (9.2.56)$$

Show that (2.27) implies the relation

$$f_1^{a'} = \sum_a O_{a'a} h_1^a. \quad (9.2.57)$$

What remains is to determine more explicitly the matrix H . Following (7.2.3), let us write h_2 in the form

$$h_2 = -(1/2) \sum_{de} S_{de} z_d z_e \quad (9.2.58)$$

where S is a symmetric matrix. Following (7.2.4), show that

$$H = (JS)^T = -SJ = J(JSJ) = JS' \quad (9.2.59)$$

where

$$S' = JSJ. \quad (9.2.60)$$

Show that S' is symmetric.

9.2.3. This exercise examines the following question: given f_2^c , f_2^a , and f_1 , when do there exist h_1 and h_2 such that there is the relation

$$\exp(: f_2^c :)\exp(: f_2^a :)\exp(: f_1 :)=\exp(: h_1+h_2 :)? \quad (9.2.61)$$

Thanks to Section 8.7 we know that such a relation is not always possible even when f_1 is absent (and its presence never helps). So we should phrase the question more narrowly: given f_2 and f_1 , when do there exist an h_1 and h_2 such that (2.22) holds? Even this question is too broad. Given an \mathcal{R} such as in (1.5), we are in effect given a symplectic matrix R and the requirement

$$\mathcal{R}z_a = \sum_b R_{ab}z_b. \quad (9.2.62)$$

There may be *many* f_2 such that

$$\mathcal{R} = \exp(: f_2 :), \quad (9.2.63)$$

and (2.62) holds. Give an explicit example of this fact. [Hint: look at (7.2.23).] Thus, the question should be made still narrower: Given f_1 and a symplectic matrix R with the property that there exists some f_2 such that (2.62) and (2.63) hold, when do there exist h_1 and h_2 such that

$$\mathcal{R}\exp(: f_1 :)=\exp(: h_1+h_2 :)? \quad (9.2.64)$$

We will begin with the case of 2-dimensional phase space. Following (8.7.25), let us write h_2 in the form

$$h_2 = -(bp^2 + 2aqp + cq^2)/2. \quad (9.2.65)$$

With this definition, show from (8.7.27) and (8.7.28) that the H matrix of (2.53) is given by the relation

$$H = F^T = \begin{pmatrix} a & -c \\ b & -a \end{pmatrix}. \quad (9.2.66)$$

Like F , the matrix H has the property

$$H^2 = \Delta I \quad (9.2.67)$$

where Δ is the discriminant (8.7.30). We are now prepared to compute O as given by (2.54). From (2.54) and (8.7.9) derive the formal result

$$\begin{aligned} O &= \text{iex}(-H) = \sum_{m=0}^{\infty} (-H)^m / (m+1)! = -[\exp(-H) - I]/H \\ &= -[\cosh(H) - \sinh(H) - I]/H = [\sinh(H)]/H - [\cosh(H) - I]/H. \end{aligned} \quad (9.2.68)$$

Show using (2.67) that the series in (2.68) have the sums

$$[\sinh(H)]/H = I[\sinh(\Delta^{1/2})]/\Delta^{1/2}, \quad (9.2.69)$$

$$- [\cosh(H) - I]/H = -H[\cosh(\Delta^{1/2}) - 1]/\Delta. \quad (9.2.70)$$

Thus, show that O has the explicit matrix form

$$O = \begin{pmatrix} [\sinh(\Delta^{1/2})]/\Delta^{1/2} - a[\cosh(\Delta^{1/2}) - 1]/\Delta & c[\cosh(\Delta^{1/2}) - 1]/\Delta \\ -b[\cosh(\Delta^{1/2}) - 1]/\Delta & [\sinh(\Delta^{1/2})]/\Delta^{1/2} + a[\cosh(\Delta^{1/2}) - 1]/\Delta \end{pmatrix}. \quad (9.2.71)$$

Note that O is well defined for all values of a, b, c as expected. Verify that O has the determinant

$$\det(O) = 2[\cosh(\Delta^{1/2}) - 1]/\Delta. \quad (9.2.72)$$

It follows that O is not invertible when $\Delta = -4\pi^2, -16\pi^2, \dots$. Indeed, inspection of (2.71) shows that O vanishes identically for these Δ values! Give a geometrical argument showing that this must be the case. From looking at (2.57) we conclude that, given f_1, h_1 cannot be determined when O is not invertible. Therefore, we are tempted to conclude that our single-exponent goal (2.64) cannot be accomplished in this case. However, we see from (8.7.34) that R is the *identity* matrix when $\Delta = -4\pi^2, -16\pi^2, \dots$. In this case we may take $f_2 = 0$ and $h_2 = f_2 = 0$. Show that O is then the identity matrix, and the single-exponent goal is trivially achieved. We conclude, despite our fears to the contrary, that the single-exponent goal (2.64) can always be achieved in the 2-dimensional phase-space case. What about the cases of higher dimensional phase spaces, say 4 and 6-dimensional phase spaces? These cases appear to be more difficult. They will be treated after we have developed a theory of normal forms for quadratic polynomials.

9.2.4. Verify that Lie operators associated with first-degree polynomials commute. That is, let f_1 and g_1 be any two first-degree polynomials. Show that

$$\{ : f_1 : , : g_1 : \} =: [f_1, g_1] := 0 \quad (9.2.73)$$

using (5.3.14), (5.3.21), and (7.6.14). From Exercise 7.7.2 we know that Lie operators of the form $: f_1 :$ and $: f_2 :$ comprise the Lie algebra $isp(2n, \mathbb{R})$. Show that the subalgebra composed of Lie operators of the form $: f_1 :$ comprises a subalgebra that is an ideal in $isp(2n, \mathbb{R})$. Show that $isp(2n, \mathbb{R})$ is not semisimple. See Section 8.9. It is in fact the *semi-direct sum* of the Lie subalgebra generated by Lie operators of the form $: f_2 :$, namely the subalgebra $sp(2n, \mathbb{R})$, and the Lie subalgebra generated by Lie operators of the form $: f_1 :$, namely the Lie subalgebra of the translation group. The sum is called *semi-direct* because Lie operators of the form $: f_2 :$ and do not commute with Lie operators of the form $: f_1 :$. Instead, Lie operators of the form $: f_2 :$ have a nontrivial action on first-order polynomials and, correspondingly, on Lie operators of the form $: f_1 :$. See Section 25.2.1.

9.2.5. Equation (2.22) gives a reverse factorization. Consider the problem of making the forward factorization

$$\exp(: h_1 + h_2 :) = \exp(: g_1 :) \exp(: g_2 :). \quad (9.2.74)$$

Show that in this case one has the results

$$g_2 = h_2, \quad (9.2.75)$$

$$g_1 = \text{iex}(: h_2 :)h_1. \quad (9.2.76)$$

Hint: Either invert both sides of (2.22), or use (2.22), (2.10), and (2.11).

9.2.6. Review (2.1) through (2.3) and (7.7.3). Consider the general $ISp(2n, \mathbb{R})$ element \mathcal{M} given by (2.3). When $g_1 = 0$ we know that \mathcal{M} has the origin as a fixed point. The purpose of this exercise is to show that when $g_1 \neq 0$ the map \mathcal{M} generally still has a fixed point. Suppose that \mathcal{M} has a fixed point z^f . Show from (2.1) that there must then be the relation

$$z^f = \delta + Rz^f, \quad (9.2.77)$$

from which it follows that z^f is uniquely defined by the relation

$$z^f = -(R - I)^{-1}\delta \quad (9.2.78)$$

provided the matrix $(R - I)$ has an inverse. For $(R - I)$ to have an inverse it must be the case that

$$\det(R - I) \neq 0. \quad (9.2.79)$$

Consequently, provided R does not have $+1$ as an eigenvalue, \mathcal{M} has a fixed point z^f given by (2.78), and this fixed point is unique. In particular, if the eigenvalues of R lie on the unit circle so that tunes are defined, no tune may be integer.

Provide an example of a symplectic matrix R , with some eigenvalue equal $+1$, for which \mathcal{M} does not have a fixed point for some $g_1 \neq 0$. For example, study in detail the 2×2 case for which

$$R = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix} \quad (9.2.80)$$

where λ is an arbitrary parameter. Show that, when $\delta_2 \neq 0$, \mathcal{M} has no fixed point. Show that, when $\delta_2 = 0$ and $\lambda \neq 0$, \mathcal{M} has a whole line of fixed points and therefore does not have a unique fixed point. What happens when $\lambda = 0$?

Given \mathcal{M} and any degree one polynomial h_1 , define a map \mathcal{N} by the relation

$$\mathcal{N} = \exp(: h_1 :) \mathcal{M} \exp(- : h_1 :). \quad (9.2.81)$$

Show that, when

$$h_1 = (z, Jz^f), \quad (9.2.82)$$

there is the relation

$$\mathcal{N} = \mathcal{R}. \quad (9.2.83)$$

Thus, show that generally \mathcal{M} has the factorization

$$\mathcal{M} = \exp(- : h_1 :) \mathcal{R} \exp(: h_1 :) \quad (9.2.84)$$

with h_1 given by (2.82), z^f being the fixed point of \mathcal{M} , and \mathcal{R} being the linear part of \mathcal{M} . However, not all \mathcal{M} can be written in the form (2.84) because we know from the work of the previous paragraph that there are some \mathcal{M} that do not have a fixed point.

9.2.7. Consider a generating function of the form

$$g(u) = (k, u) + (1/2)(u, Wu) \quad (9.2.85)$$

where k is a vector with $2n$ entries and W is a general $2n \times 2n$ symmetric matrix. That is, g consists of arbitrary linear and quadratic parts. Using (6.7.21), find the associated symplectic map \mathcal{M} for any Darboux matrix α , and show that \mathcal{M} is an element of $ISp(2n, \mathbb{R})$.

9.3 Lie Concatenation in the General Nonlinear Case

We now turn to the general case $ISpM(2n, \mathbb{R})$ where we have to take into account the presence of the nonlinear generators $f_3, f_4 \dots$ and $g_3, g_4 \dots$. In this case, the Lie algebras are infinite dimensional and, as described in Section 8.9, we will introduce a quotient-space structure in order to produce an approximation scheme. This quotient-space structure will be based on the grading given by (8.9.78) and (8.9.79). As we will see, it amounts to treating the quantities f_1 and g_1 in (1.1) and (1.2) as being small.

For purposes of illustration, we will begin our discussion with the case in which we retain only f_3 and f_4 in (1.1) and g_3 and g_4 in (1.2). Correspondingly, we will only retain h_3 and h_4 in (1.3). In addition, we will let ϵ be a parameter which we will initially regard as small, but will eventually set equal to one. Now consider terms of the form $\epsilon^m f_n$, and corresponding terms for the g 's and h 's. We assign these terms a grade following (8.9.78) and (8.9.79),

$$\text{grade } 0 : \epsilon^2 f_0, \epsilon f_1, f_2; \quad (9.3.1)$$

$$\text{grade } 1 : \epsilon^3 f_0, \epsilon^2 f_1, \epsilon f_2, f_3; \quad (9.3.2)$$

$$\text{grade } 2 : \epsilon^4 f_0, \epsilon^3 f_1, \epsilon^2 f_2, \epsilon f_3, f_4. \quad (9.3.3)$$

We have already seen that these elements span the quotient Lie algebra ${}^\epsilon L^0 / {}^\epsilon L^3$, and we will work within this quotient Lie algebra and its corresponding quotient group. Note that since we will be working with Lie operators, and the Lie operator for a constant function vanishes, terms of the form $\epsilon^m f_0$ actually play no role.

Following the notation (2.26) and (8.4.14), let us rewrite (1.1), (1.2), and (1.4) in the form

$$\mathcal{M}_f = \exp(: \epsilon f_1 :)\mathcal{R}_f \exp(: f_3 :)\exp(: f_4 :), \quad (9.3.4)$$

$$\mathcal{M}_g = \exp(: \epsilon g_1 :)\mathcal{R}_g \exp(: g_3 :)\exp(: g_4 :), \quad (9.3.5)$$

$$\mathcal{M}_h = \exp(: \epsilon h_1 :)\mathcal{R}_h \exp(: h_3 :)\exp(: h_4 :). \quad (9.3.6)$$

Here we have replaced f_1 by ϵf_1 , etc. Upon comparing (3.4) through (3.6), we see that performing the multiplication (1.3) requires that all first-order exponents be moved to the extreme left. We will do this in steps. Let us write the product (1.3) in the form

$$\mathcal{M}_h = \exp(: \epsilon f_1 :)\mathcal{R}_f \exp(: f_3 :)\exp(: f_4 :)\exp(: \epsilon g_1 :)\mathcal{R}_g \exp(: g_3 :)\exp(: g_4 :). \quad (9.3.7)$$

The first step will be to move the first-order exponent, g_1 in this case, to the left of f_4 . We begin with the relation

$$\begin{aligned} \exp(: f_4 :)\exp(: \epsilon g_1 :) &= \exp(: \epsilon g_1 :)\exp(-: \epsilon g_1 :)\exp(: f_4 :)\exp(: \epsilon g_1 :) \\ &= \exp(: \epsilon g_1 :)\exp(: \exp(-: \epsilon g_1 :)f_4 :). \end{aligned} \quad (9.3.8)$$

Here use has been made of (8.2.20). Now we use the relation

$$\begin{aligned} : \exp(-\epsilon g_1) f_4 &:= \left[\sum_{m=0}^{\infty} (-\epsilon g_1)^m / m! \right] f_4 : \\ &= : [f_4 - \epsilon : g_1 : f_4 + (\epsilon^2/2!) : g_1 :^2 f_4 - (\epsilon^3/3!) : g_1 :^3 f_4] : . \end{aligned} \quad (9.3.9)$$

Note that the apparently infinite series in (3.9) actually terminates because of (7.6.16). Also, use has been made of (5.3.21).

Let us rewrite (3.8) and (3.9) in the form

$$\exp(: f_4 :) \exp(: \epsilon g_1 :) = \exp(: \epsilon g_1 :) \exp(: j_1^{(2)} + j_2^{(2)} + j_3^{(2)} + j_4^{(2)} :) \quad (9.3.10)$$

where the $j_i^{(2)}$ are the homogeneous polynomials

$$j_1^{(2)} = -(\epsilon^3/3!) : g_1 :^3 f_4, \quad (9.3.11)$$

$$j_2^{(2)} = (\epsilon^2/2!) : g_1 :^2 f_4, \quad (9.3.12)$$

$$j_3^{(2)} = -\epsilon : g_1 : f_4, \quad (9.3.13)$$

$$j_4^{(2)} = f_4. \quad (9.3.14)$$

Here a subscript on a j indicates the *degree* of the polynomial. Observe from (3.3) that all the j polynomials have *grade* two. Hence they also all carry a superscript 2 in parentheses to indicate this fact. Next we use the factorization theorem to write the product representation

$$\exp(: j_1^{(2)} + j_2^{(2)} + j_3^{(2)} + j_4^{(2)} :) = \exp(: k_1 :) \exp(: k_2 :) \exp(: k_3 :) \exp(: k_4 :). \quad (9.3.15)$$

Without loss of generality we may require that the k_i are in the Lie algebra generated by the $j_i^{(2)}$ and are also in ${}^\epsilon L^0 / {}^\epsilon L^3$. It follows that we have the relations

$$k_i = j_i^{(2)}. \quad (9.3.16)$$

Now combine (3.10) and (3.15) to obtain the result

$$\begin{aligned} \exp(: f_4 :) \exp(: \epsilon g_1 :) &= \exp(: \epsilon g_1 :) \exp(: k_1 :) \exp(: k_2 :) \exp(: k_3 :) \exp(: k_4 :) \\ &= \exp(: \epsilon g_1 + k_1 :) \exp(: k_2 :) \exp(: k_3 :) \exp(: k_4 :). \end{aligned} \quad (9.3.17)$$

Observe that in relations such as (3.11) through (3.14), powers of ϵ are correlated with powers of $: g_1 :$. Thus, we may simply view the introduction of ϵ as a way of counting powers of g_1 . Correspondingly, after obtaining final results, we may set $\epsilon = 1$ to obtain, under the assumption that g_1 itself is small, a set of formulas that make systematic expansions in the size of g_1 . The final result of this process for the work done thus far is a formula that can be written as

$$\exp(: f_4 :) \exp(: g_1 :) = \exp(: h_1^4 :) \exp(: h_2^4 :) \exp(: h_3^4 :) \exp(: h_4^4 :) \quad (9.3.18)$$

where the h_i^4 are given by the relations

$$h_1^4 = g_1 - (1/3!) : g_1 :^3 f_4, \quad (9.3.19)$$

$$h_2^4 = (1/2!) : g_1 :^2 f_4, \quad (9.3.20)$$

$$h_3^4 = - : g_1 : f_4, \quad (9.3.21)$$

$$h_4^4 = f_4. \quad (9.3.22)$$

Here the subscript on h denotes its degree, and the superscript 4 indicates that it is associated with f_4 . (Unlike the notation used earlier, the superscript on h is *not* a grade.) Note that in moving g_1 to the left of f_4 we have *generated* the third and second-degree polynomials h_3^4 and h_2^4 as well as the additional first-degree term in h_1^4 . This generation of lower-order terms is a nonlinear *feed-down* effect. It shows, for example, that a misplaced octupole can produce sextupole, quadrupole, and steering-like effects.

We have seen how to move a first-order exponent to the left of f_4 . The second step is to move such an exponent to the left of f_3 . Let us now call the first-order exponent \tilde{g}_1 . Then, in analogy to (3.8) and (3.9) find the relations

$$\exp(: f_3 :) \exp(: \epsilon \tilde{g}_1 :) = \exp(: \epsilon \tilde{g}_1 :) \exp(: \exp(- : \epsilon \tilde{g}_1 :) f_3 :), \quad (9.3.23)$$

$$: \exp(- : \epsilon \tilde{g}_1 :) f_3 := [f_3 - \epsilon : \tilde{g}_1 : f_3 + (\epsilon^2/2!) : \tilde{g}_1 :^2 f_3] : . \quad (9.3.24)$$

As before, we rewrite these relations in the form

$$\exp(: f_3 :) \exp(: \epsilon \tilde{g}_1 :) = \exp(: \epsilon \tilde{g}_1 :) \exp(: \tilde{j}_1^{(1)} + \tilde{j}_2^{(1)} + \tilde{j}_3^{(1)} :), \quad (9.3.25)$$

where the $\tilde{j}_i^{(1)}$ are now the homogeneous polynomials

$$\tilde{j}_1^{(1)} = (\epsilon^2/2!) : \tilde{g}_1 :^2 f_3, \quad (9.3.26)$$

$$\tilde{j}_2^{(1)} = -\epsilon : \tilde{g}_1 : f_3, \quad (9.3.27)$$

$$\tilde{j}_3^{(1)} = f_3. \quad (9.3.28)$$

We note that all the terms on the left sides of the relations (3.26) through (3.28) are of grade one. Hence all the \tilde{j} 's carry, within parentheses, a superscript of 1. Again, as before, we use the factorization theorem to write the representation

$$\exp(: \tilde{j}_1^{(1)} + \tilde{j}_2^{(1)} + \tilde{j}_3^{(1)} :) = \exp(: \tilde{k}_1 :) \exp(: \tilde{k}_2 :) \exp(: \tilde{k}_3 :) \exp(: \tilde{k}_4 :). \quad (9.3.29)$$

At this point there are two new features to the calculation: First, we have included a fourth-degree polynomial \tilde{k}_4 on the right side of (3.29) even though the highest degree polynomial on the left of (3.29), namely $\tilde{j}_3^{(1)}$, is of degree three. This is done for the sake of consistency since our calculation is being carried out in the quotient algebra ${}^\epsilon L^0 / {}^\epsilon L^3$, which contains fourth-degree generators. Second, since the \tilde{k}_i are in the Lie algebra generated by the $\tilde{j}_i^{(1)}$ and also in ${}^\epsilon L^0 / {}^\epsilon L^3$, they may contain terms of grade 1 and grade 2. Therefore we make the decompositions

$$\tilde{k}_1 = \tilde{k}_1^{(1)} + \tilde{k}_1^{(2)}, \quad (9.3.30)$$

$$\tilde{k}_2 = \tilde{k}_2^{(1)} + \tilde{k}_2^{(2)}, \quad (9.3.31)$$

$$\tilde{k}_3 = \tilde{k}_3^{(1)} + \tilde{k}_3^{(2)}, \quad (9.3.32)$$

$$\tilde{k}_4 = \tilde{k}_4^{(2)}. \quad (9.3.33)$$

[Note that according to (3.1) through (3.3) there is no fourth-degree polynomial of grade 1.] Now use the BCH formula and the decompositions (3.30) through (3.33) to combine all exponents on the right side of (3.29) into one grand exponent. We find, through terms of grade 2, the result

$$\exp(:\tilde{k}_1:) \exp(:\tilde{k}_2:) \exp(:\tilde{k}_3:) \exp(:\tilde{k}_4:) = \exp(:\ell_1 + \ell_2 + \ell_3 + \ell_4:) \quad (9.3.34)$$

where the ℓ_i are given by the relations

$$\ell_1 = \tilde{k}_1^{(1)} + \tilde{k}_1^{(2)} + [\tilde{k}_1^{(1)}, \tilde{k}_2^{(1)}]/2, \quad (9.3.35)$$

$$\ell_2 = \tilde{k}_2^{(1)} + \tilde{k}_2^{(2)} + [\tilde{k}_1^{(1)}, \tilde{k}_3^{(1)}]/2, \quad (9.3.36)$$

$$\ell_3 = \tilde{k}_3^{(1)} + \tilde{k}_3^{(2)} + [\tilde{k}_2^{(1)}, \tilde{k}_3^{(1)}]/2, \quad (9.3.37)$$

$$\ell_4 = \tilde{k}_4^{(2)}. \quad (9.3.38)$$

Upon comparing (3.29) and (3.34) we find the results

$$\ell_i = \tilde{j}_i \text{ for } i = 1 \text{ to } 3, \quad (9.3.39)$$

$$\ell_4 = 0. \quad (9.3.40)$$

We thus have the relations

$$\tilde{j}_1^{(1)} = \tilde{k}_1^{(1)} + \tilde{k}_1^{(2)} + [\tilde{k}_1^{(1)}, \tilde{k}_2^{(1)}]/2, \quad (9.3.41)$$

$$\tilde{j}_2^{(1)} = \tilde{k}_2^{(1)} + \tilde{k}_2^{(2)} + [\tilde{k}_1^{(1)}, \tilde{k}_3^{(1)}]/2, \quad (9.3.42)$$

$$\tilde{j}_3^{(1)} = \tilde{k}_3^{(1)} + \tilde{k}_3^{(2)} + [\tilde{k}_2^{(1)}, \tilde{k}_3^{(1)}]/2, \quad (9.3.43)$$

$$0 = \tilde{k}_4^{(2)}, \quad (9.3.44)$$

and these relations must be solved for the $\tilde{k}_i^{(1)}$ and $\tilde{k}_i^{(2)}$. To carry out this task, we equate terms of like grade in (3.41) through (3.44) to find the results

$$\tilde{k}_1^{(1)} = \tilde{j}_1^{(1)}, \quad (9.3.45)$$

$$\tilde{k}_2^{(1)} = \tilde{j}_2^{(1)}, \quad (9.3.46)$$

$$\tilde{k}_3^{(1)} = \tilde{j}_3^{(1)}, \quad (9.3.47)$$

$$\tilde{k}_1^{(2)} = -[\tilde{k}_1^{(1)}, \tilde{k}_2^{(1)}]/2 = -[\tilde{j}_1^{(1)}, \tilde{j}_2^{(1)}]/2, \quad (9.3.48)$$

$$\tilde{k}_2^{(2)} = -[\tilde{k}_1^{(1)}, \tilde{k}_3^{(1)}]/2 = -[\tilde{j}_1^{(1)}, \tilde{j}_3^{(1)}]/2, \quad (9.3.49)$$

$$\tilde{k}_3^{(2)} = -[\tilde{k}_2^{(1)}, \tilde{k}_3^{(1)}]/2 = -[\tilde{j}_2^{(1)}, \tilde{j}_3^{(1)}]/2, \quad (9.3.50)$$

$$\tilde{k}_4^{(2)} = 0. \quad (9.3.51)$$

Now put the results (3.26) through (3.28), (3.30) through (3.33), and (3.45) through (3.51) together to get the relations

$$\tilde{k}_1 = (\epsilon^2/2) : \tilde{g}_1 :^2 f_3 - (\epsilon^3/4) [: \tilde{g}_1 :^2 f_3, : \tilde{g}_1 : f_3], \quad (9.3.52)$$

$$\tilde{k}_2 = -\epsilon : \tilde{g}_1 : f_3 - (\epsilon^2/4) [: \tilde{g}_1 :^2 f_3, f_3], \quad (9.3.53)$$

$$\tilde{k}_3 = f_3 + (\epsilon/2) [: \tilde{g}_1 : f_3, f_3], \quad (9.3.54)$$

$$\tilde{k}_4 = 0. \quad (9.3.55)$$

Finally, as before, these relations should be evaluated with $\epsilon = 1$. The net result is the formula

$$\exp(: f_3 :) \exp(: \tilde{g}_1 :) = \exp(: h_1^3 :) \exp(: h_2^3 :) \exp(: h_3^3 :) \exp(: h_4^3 :) \quad (9.3.56)$$

where the h_i^3 are given by the relations

$$h_1^3 = \tilde{g}_1 + (1/2) : \tilde{g}_1 :^2 f_3 - (1/4) [: \tilde{g}_1 :^2 f_3, : \tilde{g}_1 : f_3], \quad (9.3.57)$$

$$h_2^3 = - : \tilde{g}_1 : f_3 - (1/4) [: \tilde{g}_1 :^2 f_3, f_3], \quad (9.3.58)$$

$$h_3^3 = f_3 + (1/2) [: \tilde{g}_1 : f_3, f_3], \quad (9.3.59)$$

$$h_4^3 = 0. \quad (9.3.60)$$

We note that moving \tilde{g}_1 past f_3 is more difficult than moving g_1 past f_4 ! This greater difficulty occurs because, as is evident from (3.57) through (3.60), the feed-down terms are more complicated.

We have seen how to move a first-degree exponent past f_4 and f_3 , and how to calculate (within the quotient algebra ${}^\epsilon L^0 / {}^\epsilon L^3$) the feed-down terms left in its wake. The penultimate step is to move the first-degree exponent past \mathcal{R}_f and then combine it with f_1 . But this we know how to do exactly using the concatenation formulas (2.34) through (2.38) for the inhomogeneous symplectic group. Finally, we have to combine the results obtained so far with the remaining factors $\mathcal{R}_g \exp(: g_3 :) \exp(: g_4 :)$ in (3.7). This we also know how to do exactly using the concatenation formulas of Section 8.4. Thus, we have explored in some detail how to perform concatenation within the quotient group generated by the Lie algebra ${}^\epsilon L^0 / {}^\epsilon L^3$.

Two tasks remain. The first is to find a convenient way of evaluating the symplectic matrices associated with the feed-down linear transformations of the form $\exp(: h_2^n :)$. This subject has already been treated in Chapter 4.

The second task is to find results for the larger quotient algebras ${}^\epsilon L^0 / {}^\epsilon L^\ell$ with $\ell > 3$. A suitable Mathematica program for this purpose is presented in Appendix E. Essentially two problems must be solved to carry out the second task. First, we need formulas of the kind (3.15) and (3.29). Given a set of graded polynomials $j_i^{(n)}$, we need formulas for the k_m that appear in the product representation

$$\exp(: j_1^{(n)} + j_2^{(n)} + j_3^{(n)} + \dots :) = \exp(: k_1 :) \exp(: k_2 :) \exp(: k_3 :) \dots . \quad (9.3.61)$$

Second, let us write a relation of the form

$$\exp(: f_n :) \exp(: g_1 :) = \exp(: h_1^n :) \exp(: h_2^n :) \exp(: h_3^n :) \cdots, \quad (9.3.62)$$

where, as before, we have used the notation h_i^n to denote the polynomial of degree i that results from moving $: g_1 :$ past $: f_n :$. For this relation we need formulas for the h_i^n in terms of g_1 and f_n . We summarize below the results we have already found for the quotient algebra ${}^\epsilon L^0 / {}^\epsilon L^3$ and, as a more complicated example found using the program in Appendix E, the results for the quotient algebra ${}^\epsilon L^0 / {}^\epsilon L^5$.

Formulas for the k_i in (3.61) in the quotient algebra ${}^\epsilon L^0 / {}^\epsilon L^3$.

$n=2$

$$k_i = j_1^{(2)}, \quad i = 1 \text{ to } 4. \quad (9.3.63)$$

$n=1$

$$k_1 = j_1^{(1)} - [j_1^{(1)}, j_2^{(1)}]/2, \quad (9.3.64)$$

$$k_2 = j_2^{(1)} - [j_1^{(1)}, j_3^{(1)}]/2, \quad (9.3.65)$$

$$k_3 = j_3^{(1)} - [j_2^{(1)}, j_3^{(1)}]/2, \quad (9.3.66)$$

$$k_4 = 0. \quad (9.3.67)$$

Formulas for the h_i^n in (3.62) in the quotient algebra ${}^\epsilon L^0 / {}^\epsilon L^3$.

$$h_1^4 = g_1 - (1/3!) : g_1 :^3 f_4, \quad (9.3.68)$$

$$h_2^4 = (1/2!) : g_1 :^2 f_4, \quad (9.3.69)$$

$$h_3^4 = - : g_1 : f_4, \quad (9.3.70)$$

$$h_4^4 = f_4. \quad (9.3.71)$$

$$h_1^3 = g_1 + (1/2!) : g_1 :^2 f_3 - (1/4)[: g_1 :^2 f_3, : g_1 : f_3], \quad (9.3.72)$$

$$h_2^3 = - : g_1 : f_3 - (1/4)[: g_1 :^2 f_3, f_3], \quad (9.3.73)$$

$$h_3^3 = f_3 + (1/2)[: g_1 : f_3, f_3], \quad (9.3.74)$$

$$h_4^3 = 0. \quad (9.3.75)$$

Formulas for the k_i in (3.61) in the quotient algebra ${}^\epsilon L^0 / {}^\epsilon L^5$.

$n=4$

$$k_i = j_i^{(4)}, \quad i = 1, 6. \quad (9.3.76)$$

$n=3$

$$k_1 = j_1^{(3)}, \quad (9.3.77)$$

$$k_2 = j_2^{(3)}, \quad (9.3.78)$$

$$k_3 = j_3^{(3)}, \quad (9.3.79)$$

$$k_4 = j_4^{(3)}, \quad (9.3.80)$$

$$k_5 = j_5^{(3)}, \quad (9.3.81)$$

$$k_6 = 0. \quad (9.3.82)$$

$n=2$

$$k_1 = j_1^{(2)} + [j_2^{(2)}, j_1^{(2)}]/2, \quad (9.3.83)$$

$$k_2 = j_2^{(2)} + [j_3^{(2)}, j_1^{(2)}], \quad (9.3.84)$$

$$k_3 = j_3^{(2)} + [j_3^{(2)}, j_2^{(2)}]/2 + [j_4^{(2)}, j_1^{(2)}], \quad (9.3.85)$$

$$k_4 = j_4^{(2)} + [j_4^{(2)}, j_2^{(2)}], \quad (9.3.86)$$

$$k_5 = [j_4^{(2)}, j_3^{(2)}], \quad (9.3.87)$$

$$k_6 = 0. \quad (9.3.88)$$

$n=1$

$$\begin{aligned} k_1 &= j_1^{(1)} + [j_2^{(1)}, j_1^{(1)}]/2 - [j_1^{(1)}, j_3^{(1)}, j_1^{(1)}]/6 + [j_2^{(1)}, j_2^{(1)}, j_1^{(1)}]/6 \\ &- [j_1^{(1)}, j_3^{(1)}, j_2^{(1)}, j_1^{(1)}]/8 - [j_2^{(1)}, j_1^{(1)}, j_3^{(1)}, j_1^{(1)}]/24 \\ &+ [j_2^{(1)}, j_2^{(1)}, j_2^{(1)}, j_1^{(1)}]/24, \end{aligned} \quad (9.3.89)$$

$$\begin{aligned} k_2 &= j_2^{(1)} + [j_3^{(1)}, j_1^{(1)}]/2 - [j_2^{(1)}, j_3^{(1)}, j_1^{(1)}]/12 + [j_3^{(1)}, j_2^{(1)}, j_1^{(1)}]/6 \\ &- [j_2^{(1)}, j_3^{(1)}, j_2^{(1)}, j_1^{(1)}]/24 - [j_3^{(1)}, j_1^{(1)}, j_3^{(1)}, j_1^{(1)}]/24 \\ &+ [j_3^{(1)}, j_2^{(1)}, j_2^{(1)}, j_1^{(1)}]/24, \end{aligned} \quad (9.3.90)$$

$$\begin{aligned} k_3 &= j_3^{(1)} + [j_3^{(1)}, j_2^{(1)}]/2 - [j_2^{(1)}, j_3^{(1)}, j_2^{(1)}]/6 + [j_3^{(1)}, j_3^{(1)}, j_1^{(1)}]/6 \\ &+ [j_2^{(1)}, j_2^{(1)}, j_3^{(1)}, j_2^{(1)}]/24 - [j_2^{(1)}, j_3^{(1)}, j_3^{(1)}, j_1^{(1)}]/8 + [j_3^{(1)}, j_2^{(1)}, j_3^{(1)}, j_1^{(1)}]/24 \\ &+ [j_3^{(1)}, j_3^{(1)}, j_2^{(1)}, j_1^{(1)}]/24, \end{aligned} \quad (9.3.91)$$

$$k_4 = -[j_3^{(1)}, j_3^{(1)}, j_2^{(1)}]/12 + [j_3^{(1)}, j_2^{(1)}, j_3^{(1)}, j_2^{(1)}]/24 - [j_3^{(1)}, j_3^{(1)}, j_3^{(1)}, j_1^{(1)}]/24, \quad (9.3.92)$$

$$k_5 = [j_3^{(1)}, j_3^{(1)}, j_3^{(1)}, j_2^{(1)}]/24, \quad (9.3.93)$$

$$k_6 = 0. \quad (9.3.94)$$

Formulas for the h_i^n in (3.62) in the quotient algebra ${}^\epsilon L^0 / {}^\epsilon L^5$.

$$h_1^6 = g_1 - (1/120) : g_1 :^5 f_6, \quad (9.3.95)$$

$$h_2^6 = (1/24) : g_1 :^4 f_6, \quad (9.3.96)$$

$$h_3^6 = -(1/6) : g_1 :^3 f_6, \quad (9.3.97)$$

$$h_4^6 = (1/2) : g_1 :^2 f_6, \quad (9.3.98)$$

$$h_5^6 = - : g_1 : f_6, \quad (9.3.99)$$

$$h_6^6 = f_6; \quad (9.3.100)$$

$$h_1^5 = g_1 + (1/24) : g_1 :^4 f_5, \quad (9.3.101)$$

$$h_2^5 = -(1/6) : g_1 :^3 f_5, \quad (9.3.102)$$

$$h_3^5 = (1/2) : g_1 :^2 f_5, \quad (9.3.103)$$

$$h_4^5 = - : g_1 : f_5, \quad (9.3.104)$$

$$h_5^5 = f_5, \quad (9.3.105)$$

$$h_6^5 = 0; \quad (9.3.106)$$

$$h_1^4 = g_1 + j_1^4 + [j_2^4, j_1^4]/2, \quad (9.3.107)$$

$$h_2^4 = j_2^4 + [j_3^4, j_1^4]/2, \quad (9.3.108)$$

$$h_3^4 = j_3^4 + [j_3^4, j_2^4]/2 + [j_4^4, j_1^4]/2, \quad (9.3.109)$$

$$h_4^4 = j_4^4 + [j_4^4, j_2^4]/2, \quad (9.3.110)$$

$$h_5^4 = [j_4^4, j_3^4]/2, \quad (9.3.111)$$

$$h_6^4 = 0, \quad (9.3.112)$$

where

$$j_1^4 = - : g_1 :^3 f_4 / 6, \quad (9.3.113)$$

$$j_2^4 = : g_1 :^2 f_4 / 2, \quad (9.3.114)$$

$$j_3^4 = - : g_1 : f_4, \quad (9.3.115)$$

$$j_4^4 = f_4; \quad (9.3.116)$$

$$\begin{aligned} h_1^3 &= g_1 + j_1^3 + [j_2^3, j_1^3]/2 - [j_1^3, j_3^3, j_1^3]/6 + [j_2^3, j_2^3, j_1^3]/6 - [j_1^3, j_3^3, j_2^3, j_1^3]/8 \\ &\quad - [j_2^3, j_1^3, j_3^3, j_1^3]/24 + [j_2^3, j_2^3, j_2^3, j_1^3]/24, \end{aligned} \quad (9.3.117)$$

$$\begin{aligned} h_2^3 &= j_2^3 + [j_3^3, j_1^3]/2 - [j_2^3, j_3^3, j_1^3]/12 + [j_3^3, j_2^3, j_1^3]/6 - [j_2^3, j_3^3, j_2^3, j_1^3]/24 \\ &\quad - [j_3^3, j_1^3, j_3^3, j_1^3]/24 + [j_3^3, j_2^3, j_2^3, j_1^3]/24, \end{aligned} \quad (9.3.118)$$

$$\begin{aligned} h_3^3 &= j_3^3 + [j_3^3, j_2^3]/2 - [j_2^3, j_3^3, j_2^3]/6 + [j_3^3, j_3^3, j_1^3]/6 + [j_2^3, j_2^3, j_3^3, j_2^3]/24 \\ &\quad - [j_2^3, j_3^3, j_3^3, j_1^3]/8 + [j_3^3, j_2^3, j_3^3, j_1^3]/24 + [j_3^3, j_3^3, j_2^3, j_1^3]/24, \end{aligned} \quad (9.3.119)$$

$$h_4^3 = -[j_3^3, j_3^3, j_2^3]/12 + [j_3^3, j_2^3, j_3^3, j_2^3]/24 - [j_3^3, j_3^3, j_3^3, j_1^3]/24, \quad (9.3.120)$$

$$h_5^3 = [j_3^3, j_3^3, j_3^3, j_2^3]/24, \quad (9.3.121)$$

$$h_6^3 = 0, \quad (9.3.122)$$

where

$$j_1^3 = : g_1 :^2 f_3 / 2, \quad (9.3.123)$$

$$j_2^3 = - : g_1 : f_3, \quad (9.3.124)$$

$$j_3^3 = f_3. \quad (9.3.125)$$

Here for multiple Poisson brackets we have used the notation

$$[a_1, a_2, a_3] = [a_1, [a_2, a_3]], \quad (9.3.126)$$

$$[a_1, a_2, a_3, a_4] = [a_1, [a_2, [a_3, a_4]]]. \quad (9.3.127)$$

Note that for $\epsilon L^0/\epsilon L^5$ the feed-down terms are quite complicated. Also, note that the terms h_5^4 , h_4^3 , and h_5^3 are nonzero. Consequently there can also be, in effect, nonlinear feed-up terms due to translations in phase space. Finally, we see that the $\epsilon L^0/\epsilon L^3$ formulas (3.68) through (3.75) are special cases of the $\epsilon L^0/\epsilon L^5$ formulas for h_i^4 and h_i^3 in which higher-power terms in g_1 are neglected.

We close this section with an observation that will be of relevance for the work of the next section. Observe that the relations (3.25) and (3.29) can be written in the form

$$\exp(: f_3 :) \exp(: \epsilon g_1 :) = \exp(: k_1^{(0)} :) \exp(: k_1^{(1)} + k_1^{(2)} :) \exp(: k_2^{(1)} + k_2^{(2)} :) \exp(: k_3^{(1)} + k_3^{(2)} :). \quad (9.3.128)$$

Here, for convenience, we have dropped the tildes and we have defined the $k_i^{(n)}$ by the relations

$$k_1^{(0)} = \epsilon g_1, \quad (9.3.129)$$

$$k_1^{(1)} = (\epsilon^2/2) : g_1 :^2 f_3, \quad (9.3.130)$$

$$k_1^{(2)} = -(\epsilon^3/4) [: g_1 :^2 f_3, : g_1 : f_3], \quad (9.3.131)$$

$$k_2^{(1)} = -\epsilon : g_1 : f_3, \quad (9.3.132)$$

$$k_2^{(2)} = -(\epsilon^2/4) [: g_1 :^2 f_3, f_3], \quad (9.3.133)$$

$$k_3^{(1)} = f_3, \quad (9.3.134)$$

$$k_3^{(2)} = (\epsilon/2) [: g_1 : f_3, f_3]. \quad (9.3.135)$$

See (3.1) through (3.3) and (3.52) through (3.55). We see that several of the exponents contain terms of different grades. We will therefore refer to expressions of the form (3.128) as *mixed* grade factorizations.

From the Lie algebraic perspective of working within $\epsilon L^0/\epsilon L^3$ we could as well sought relations of the form

$$\begin{aligned} \exp(: f_3 :) \exp(: \epsilon g_1 :) &= \exp(: \hat{k}_1^{(0)} :) \exp(: \hat{k}_1^{(1)} :) \exp(: \hat{k}_1^{(2)} :) \times \\ \exp(: \hat{k}_2^{(0)} :) \exp(: \hat{k}_2^{(1)} :) \exp(: \hat{k}_2^{(2)} :) \times \\ \exp(: \hat{k}_3^{(1)} :) \exp(: \hat{k}_3^{(2)} :) \exp(: \hat{k}_4^{(2)} :). \end{aligned} \quad (9.3.136)$$

Here we have used hats to indicate that the $\hat{k}_i^{(n)}$ may differ from the $k_i^{(n)}$. (They are actually the same for the algebra $\epsilon L^0/\epsilon L^3$, but they may differ for relations analogous to (3.136) in the case of algebras having a larger maximum grade.) Note that in the factorization (3.136) each exponent has a single grade. [And, according to (3.1) through (3.3), each exponent carries a single power of ϵ .] We will call factorizations of this kind *single* grade factorizations.

Once again ϵ serves as a counting parameter and, having obtained a single grade factorization of the form (3.136), we may set $\epsilon = 1$. Evidently, since we are working within a quotient algebra based on grade, setting $\epsilon = 1$ in either a mixed grade or a single grade factorization gives equivalent results.

Exercises

9.3.1. Verify the relations (3.8) through (3.22).

9.3.2. Verify the relations (3.30) through (3.38). Note that Poisson brackets between grade one and grade two polynomials, and Poisson brackets between two grade two polynomials, vanish in the quotient algebra ${}^e L^0 / {}^e L^3$.

9.3.3. Verify the relations (3.39) through (3.55).

9.3.4.

9.3.5.

9.4 Canonical Treatment of Translations

The Lie concatenation formulas for maps that send the origin into itself, see Sections 8.4 and 10.12, were relatively easy to derive. By contrast, the concatenation formulas just derived in the previous section, where translations were included, seem much more complicated. In this section we will show how the translation case can be handled using a concatenator for the simpler origin-preserving case. This will be done by *enlarging* the $2n$ -dimensional phase space to include the extra variables q_{n+1} and p_{n+1} . For this reason, the method will be referred to as a *canonical* treatment of translations.

9.4.1 Preliminaries

Since the origin-preserving concatenation formulas do not depend on the phase-space dimension, the only costs associated with increased phase-space dimension are those of increased storage and slower execution when a concatenator is realized as computer code. The advantages of this approach are simplicity and reliability. If one has a reliable origin-preserving concatenator, one can construct from it a self-checking concatenator for general maps.

Those who have been meticulous to do the Exercises in this book will recognize that Exercise 7.7.2 showed that the Jacobi Lie algebra $j(2n, \mathbb{R})$ is homomorphic to the inhomogeneous symplectic Lie algebra $isp(2n, \mathbb{R})$ of Section 9.2, and Exercise 7.7.3 treated, among other things, the relation between $j(2n, \mathbb{R})$ and the symplectic group Lie algebra $sp[2(n + 1), \mathbb{R}]$ for a phase space having two additional dimensions. We begin our discussion here by further elaborating on this theme. Subsequently we will treat $ISpM(2n, \mathbb{R})$ and $ispm(2n, \mathbb{R})$, the full nonlinear case of all symplectic maps.

Let us use the symbol \hat{z} to denote the coordinates in the $(2n + 2)$ -dimensional enlarged phase space,

$$\hat{z} = (q_1 \cdots q_n, q_{n+1}, p_1 \cdots p_n, p_{n+1}). \quad (9.4.1)$$

We will also use the notation $\hat{\mathcal{M}}_f$ to denote a symplectic (and origin-preserving) map acting on the enlarged phase space. For what follows we will want to consider special maps $\hat{\mathcal{M}}_{\hat{h}}$ on the enlarged phase space that have the property

$$\hat{\mathcal{M}}_{\hat{h}} q_{n+1} = q_{n+1}. \quad (9.4.2)$$

Such maps obviously form a group. Moreover, they have a factorization of the form

$$\hat{\mathcal{M}}_{\hat{h}} = \hat{\mathcal{R}}_{\hat{h}} \exp(:\hat{h}_3:) \exp(:\hat{h}_4:) \cdots \quad (9.4.3)$$

where the linear part $\hat{\mathcal{R}}_{\hat{h}}$ has the property

$$\hat{\mathcal{R}}_{\hat{h}} q_{n+1} = q_{n+1}, \quad (9.4.4)$$

and the $\hat{h}_m(\hat{z})$ that describe the nonlinear part are independent of p_{n+1} ,

$$\partial \hat{h}_m(\hat{z}) / \partial p_{n+1} = 0. \quad (9.4.5)$$

To see that this is so, equate terms of like degree on both sides of (4.2). From (7.6.14) there is the result

$$[\exp(:\hat{h}_3:) \exp(:\hat{h}_4:) \cdots] q_{n+1} = q_{n+1} + O(\hat{z}^2). \quad (9.4.6)$$

Therefore (4.2) requires the relation

$$q_{n+1} = \hat{\mathcal{M}}_{\hat{h}} q_{n+1} = \hat{\mathcal{R}}_{\hat{h}} [q_{n+1} + O(\hat{z}^2)] = \hat{\mathcal{R}}_{\hat{h}} q_{n+1} + O(\hat{z}^2), \quad (9.4.7)$$

and equating terms of first degree yields (4.4).

Now that (4.4) is established, multiply both sides of (4.2) by $\hat{\mathcal{R}}_{\hat{h}}^{-1}$ to find the result

$$\hat{\mathcal{R}}_{\hat{h}}^{-1} \hat{\mathcal{M}}_{\hat{h}} q_{n+1} = \hat{\mathcal{R}}_{\hat{h}}^{-1} q_{n+1} = q_{n+1} \quad (9.4.8)$$

from which it follows that

$$[\exp(:\hat{h}_3:) \exp(:\hat{h}_4:) \cdots] q_{n+1} = q_{n+1}. \quad (9.4.9)$$

Evaluate both sides of (4.9) through terms of degree 2. Doing so gives the relation

$$q_{n+1} + : \hat{h}_3 : q_{n+1} + O(\hat{z}^3) = q_{n+1}, \quad (9.4.10)$$

and equating terms of like degree gives the relation

$$0 = : \hat{h}_3 : q_{n+1} = [\hat{h}_3, q_{n+1}] = -\partial \hat{h}_3 / \partial p_{n+1}, \quad (9.4.11)$$

which establishes (4.5) for the case $m = 3$. From (4.11) we also conclude that

$$\exp(:\hat{h}_3:) q_{n+1} = q_{n+1}. \quad (9.4.12)$$

Finally, multiply both sides of (4.9) by $\exp(-:\hat{h}_3:)$ to find the result

$$[\exp(:\hat{h}_4:) \exp(:\hat{h}_5:) \cdots] q_{n+1} = \exp(-:\hat{h}_3:) q_{n+1} = q_{n+1}. \quad (9.4.13)$$

Expanding both sides of (4.13) and equating terms of like degree shows that (4.5) also holds for the case $m = 4$, etc.

Let us explore the consequences of (4.4) in detail. For simplicity, consider the case of a phase space that is initially two dimensional ($n = 1$) and is enlarged to become four dimensional. The results for general n can easily be inferred from what we will find for the

$n = 1$ case. Suppose $\hat{R}^{\hat{h}}$ is the matrix associated with $\hat{\mathcal{R}}_{\hat{h}}$. In the 4×4 case we have decided to consider, and in view of (4.4) with $n = 1$, it has the general form

$$\hat{R}^{\hat{h}} = \begin{pmatrix} \hat{R}_{11}^{\hat{h}} & \hat{R}_{12}^{\hat{h}} & \hat{R}_{13}^{\hat{h}} & \hat{R}_{14}^{\hat{h}} \\ \hat{R}_{21}^{\hat{h}} & \hat{R}_{22}^{\hat{h}} & \hat{R}_{23}^{\hat{h}} & \hat{R}_{24}^{\hat{h}} \\ 0 & 0 & 1 & 0 \\ \hat{R}_{41}^{\hat{h}} & \hat{R}_{42}^{\hat{h}} & \hat{R}_{43}^{\hat{h}} & \hat{R}_{44}^{\hat{h}} \end{pmatrix}. \quad (9.4.14)$$

Here we have used the ordering

$$\hat{z} = (q_1, p_1; q_2, p_2). \quad (9.4.15)$$

However, since $\hat{R}_{\hat{h}}$ is a symplectic map, $\hat{R}^{\hat{h}}$ must be a symplectic matrix. Enforcing the symplectic condition (3.1.2) or (3.1.10) gives, among others, the relations

$$\hat{R}_{a4}^{\hat{h}} = \delta_{a4}. \quad (9.4.16)$$

It follows that $\hat{R}^{\hat{h}}$ has the more specific form

$$\hat{R}^{\hat{h}} = \begin{pmatrix} \hat{R}_{11}^{\hat{h}} & \hat{R}_{12}^{\hat{h}} & \hat{R}_{13}^{\hat{h}} & 0 \\ \hat{R}_{21}^{\hat{h}} & \hat{R}_{22}^{\hat{h}} & \hat{R}_{23}^{\hat{h}} & 0 \\ 0 & 0 & 1 & 0 \\ \hat{R}_{41}^{\hat{h}} & \hat{R}_{42}^{\hat{h}} & \hat{R}_{43}^{\hat{h}} & 1 \end{pmatrix}. \quad (9.4.17)$$

Introduce the 2×2 matrix $\bar{R}^{\hat{h}}$ defined by the upper-left 2×2 block in $\hat{R}^{\hat{h}}$,

$$\bar{R}^{\hat{h}} = \begin{pmatrix} \hat{R}_{11}^{\hat{h}} & \hat{R}_{12}^{\hat{h}} \\ \hat{R}_{21}^{\hat{h}} & \hat{R}_{22}^{\hat{h}} \end{pmatrix}, \quad (9.4.18)$$

and let $\check{R}^{\hat{h}}$ be the 4×4 matrix

$$\check{R}^{\hat{h}} = \begin{pmatrix} \bar{R}^{\hat{h}} & 0 \\ 0 & I \end{pmatrix}. \quad (9.4.19)$$

Also, define quantities $\alpha_2^{\hat{h}}$ and $\delta_a^{\hat{h}}$ for $a = 1$ to 2 by the rules

$$\alpha_2^{\hat{h}} = \hat{R}_{43}^{\hat{h}}/2, \quad (9.4.20)$$

$$\delta_1^{\hat{h}} = -\hat{R}_{42}^{\hat{h}}, \quad (9.4.21)$$

$$\delta_2^{\hat{h}} = \hat{R}_{41}^{\hat{h}}, \quad (9.4.22)$$

and define associated polynomials h_1 and \hat{h}_1^2 by the rules

$$h_1(z) = \delta_2^{\hat{h}} q_1 - \delta_1^{\hat{h}} p_1 = \delta_2^{\hat{h}} z_1 - \delta_1^{\hat{h}} z_2 = (z, (\delta^{\hat{h}})^*), \quad (9.4.23)$$

$$\hat{h}_1^2(\hat{z}) = q_2 h_1(z). \quad (9.4.24)$$

Here we have used the notation of Section 7.7, and z denotes the original phase-space variables,

$$z = (q_1, p_1).$$

The superscript indicates that \hat{h}_1^2 is homogeneous of degree *two* in the variables \hat{z} , and the subscript indicates that it is homogeneous of degree *one* with respect to the variables z .

Note that h_1 has the property

$$: h_1 : z_a = \delta_a^{\hat{h}}. \quad (9.4.25)$$

Correspondingly, \hat{h}_1^2 has the properties

$$: \hat{h}_1^2 : z_a =: q_2 h_1(z) : z_a = [q_2 h_1(z), z_a] = q_2 [h_1(z), z_a] = q_2 : h_1 : z_a = q_2 \delta_a^{\hat{h}}, \quad (9.4.26)$$

$$: \hat{h}_1^2 :^m z_a = 0 \text{ for } m > 1, \quad (9.4.27)$$

$$: \hat{h}_1^2 : q_2 = [q_2 h_1(z), q_2] = 0, \quad (9.4.28)$$

$$: \hat{h}_1^2 : p_2 = [q_2 h_1(z), p_2] = h_1(z), \quad (9.4.29)$$

$$: \hat{h}_1^2 :^m p_2 = 0 \text{ for } m > 1. \quad (9.4.30)$$

Then, with these definitions, we assert that $\hat{\mathcal{R}}_{\hat{h}}$ has the unique factorization

$$\hat{\mathcal{R}}_{\hat{h}} = \hat{\mathcal{F}}_{\hat{h}} \hat{\mathcal{R}}_{\hat{h}_1^2} \check{\mathcal{R}}_{\hat{h}} \quad (9.4.31)$$

where

$$\hat{\mathcal{F}}_{\hat{h}} = \exp(: \alpha_2^{\hat{h}} q_2^2 :), \quad (9.4.32)$$

$$\hat{\mathcal{R}}_{\hat{h}_1^2} = \exp(: \hat{h}_1^2 :), \quad (9.4.33)$$

and $\check{\mathcal{R}}_{\hat{h}}$ is a linear symplectic map whose associated matrix $\check{R}^{\hat{h}}$ is given by (4.19).

If correct, the operator assertion (4.31) is equivalent to the matrix assertion

$$\hat{R}^{\hat{h}} = \check{R}^{\hat{h}} \hat{R}^{\hat{h}_1^2} \hat{F}^{\hat{h}} \quad (9.4.34)$$

where $\hat{R}^{\hat{h}_1^2}$ and $\hat{F}^{\hat{h}}$ are the matrices associated with $\hat{\mathcal{R}}_{\hat{h}_1^2}$ and $\hat{\mathcal{F}}_{\hat{h}}$. Let us find these matrices. From (4.26) through (4.30) we see that $\hat{\mathcal{R}}_{\hat{h}_1^2}$ has the property

$$\hat{\mathcal{R}}_{\hat{h}_1^2} z_a = z_a + q_2 \delta_a^{\hat{h}}, \quad (9.4.35)$$

$$\hat{\mathcal{R}}_{\hat{h}_1^2} q_2 = q_2, \quad (9.4.36)$$

$$\hat{\mathcal{R}}_{\hat{h}_1^2} p_2 = p_2 + h_1(z) = p_2 + (z, (\delta^{\hat{h}})^*). \quad (9.4.37)$$

It follows that the matrix $\hat{R}^{\hat{h}_1^2}$ is given by the relation

$$\hat{R}^{\hat{h}_1^2} = \begin{pmatrix} 1 & 0 & \delta_1^{\hat{h}} & 0 \\ 0 & 1 & \delta_2^{\hat{h}} & 0 \\ 0 & 0 & 1 & 0 \\ \delta_2^{\hat{h}} & -\delta_1^{\hat{h}} & 0 & 1 \end{pmatrix}. \quad (9.4.38)$$

Finding $\hat{F}^{\hat{h}}$ is easier. A simple calculation gives the result

$$\hat{F}^{\hat{h}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \hat{R}_{43}^{\hat{h}} & 1 \end{pmatrix}. \quad (9.4.39)$$

Let us solve (4.34) for $\check{R}^{\hat{h}}$ to find the relation

$$\check{R}^{\hat{h}} = \hat{R}^{\hat{h}}(\hat{F}^{\hat{h}})^{-1}(\hat{R}^{\hat{h}^2})^{-1}. \quad (9.4.40)$$

Carrying out the indicated multiplications gives the result

$$\check{R}^{\hat{h}} = \begin{pmatrix} \hat{R}_{11}^{\hat{h}} & \hat{R}_{12}^{\hat{h}} & \epsilon_1 & 0 \\ \hat{R}_{21}^{\hat{h}} & \hat{R}_{22}^{\hat{h}} & \epsilon_2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (9.4.41)$$

where

$$\epsilon_1 = \hat{R}_{13}^{\hat{h}} + \hat{R}_{11}^{\hat{h}}\hat{R}_{42}^{\hat{h}} - \hat{R}_{12}^{\hat{h}}\hat{R}_{41}^{\hat{h}} = \hat{R}_{13}^{\hat{h}} - \hat{R}_{11}^{\hat{h}}\delta_1^{\hat{h}} - \hat{R}_{12}^{\hat{h}}\delta_2^{\hat{h}}, \quad (9.4.42)$$

$$\epsilon_2 = \hat{R}_{23}^{\hat{h}} + \hat{R}_{21}^{\hat{h}}\hat{R}_{42}^{\hat{h}} - \hat{R}_{22}^{\hat{h}}\hat{R}_{41}^{\hat{h}} = \hat{R}_{23}^{\hat{h}} - \hat{R}_{21}^{\hat{h}}\delta_1^{\hat{h}} - \hat{R}_{22}^{\hat{h}}\delta_2^{\hat{h}}. \quad (9.4.43)$$

Next, because $\hat{R}^{\hat{h}}$, $(\hat{F}^{\hat{h}})^{-1}$, and $(\hat{R}^{\hat{h}^2})^{-1}$ are symplectic matrices ($\mathcal{R}_{\hat{h}}$, $\mathcal{F}_{\hat{h}}$, and $\mathcal{R}_{\hat{h}^2}$ are symplectic maps), $\check{R}^{\hat{h}}$ must be a symplectic matrix. The symplectic condition (3.1.10) yields for $\check{R}^{\hat{h}}$ as given by (4.41) the relations

$$\epsilon_a = 0 \text{ for } a = 1 \text{ to } 2, \quad (9.4.44)$$

$$\hat{R}_{11}^{\hat{h}}\hat{R}_{22}^{\hat{h}} - \hat{R}_{12}^{\hat{h}}\hat{R}_{21}^{\hat{h}} = 1. \quad (9.4.45)$$

From (4.44) we conclude that the ϵ_a entries in (4.41) must vanish, and therefore (4.34) is correct with $\check{R}^{\hat{h}}$ given by (4.19). The relation (4.45) is the condition that $\check{R}^{\hat{h}}$ and hence $\check{R}^{\hat{h}}$ be symplectic matrices. Finally (4.42) and (4.43), when combined with (4.44), show that the matrix $\hat{R}^{\hat{h}}$ in (4.17) must have the form

$$\hat{R}^{\hat{h}} = \begin{pmatrix} \hat{R}_{11}^{\hat{h}} & \hat{R}_{12}^{\hat{h}} & (\bar{R}^{\hat{h}}\delta^{\hat{h}})_1 & 0 \\ \hat{R}_{21}^{\hat{h}} & \hat{R}_{22}^{\hat{h}} & (\bar{R}^{\hat{h}}\delta^{\hat{h}})_2 & 0 \\ 0 & 0 & 1 & 0 \\ \delta_2^{\hat{h}} & -\delta_1^{\hat{h}} & \hat{R}_{43}^{\hat{h}} & 1 \end{pmatrix}. \quad (9.4.46)$$

With what we have learned we are now prepared to show how general maps (including translations) for the case of $2n$ -dimensional phase space can be imbedded in the set of $(2n+2)$ -dimensional origin preserving maps. We will first consider maps with no nonlinear part, and then move on to the general case.

9.4.2 Case of Maps with No Nonlinear Part

Enlarging

Let \mathcal{M}_f be an inhomogeneous symplectic group map acting on the original $2n$ -dimensional phase space z . As in (2.31), it may be written in the form

$$\mathcal{M}_f = \exp(: f_1 :) \mathcal{R}_f. \quad (9.4.47)$$

Define a function $\hat{f}_1^2(\hat{z})$ by the rule

$$\hat{f}_1^2(\hat{z}) = (q_{n+1})f_1(z). \quad (9.4.48)$$

As before, the superscript indicates that \hat{f}_1^2 is homogeneous of degree *two* in the variables \hat{z} , and the subscript indicates that it is homogeneous of degree *one* with respect to the variables z . Now define a map $\hat{\mathcal{M}}_{\hat{f}}$ on the enlarged phase space by the rule

$$\hat{\mathcal{M}}_{\hat{f}} = \exp(: \hat{f}_1^2 :) \check{\mathcal{R}}_{\hat{f}}. \quad (9.4.49)$$

Here $\check{\mathcal{R}}_{\hat{f}}$ is a linear map with the associated matrix $\check{R}_{\hat{f}}$ given by

$$\check{R}_{\hat{f}} = \begin{pmatrix} R^f & 0 \\ 0 & I \end{pmatrix} \quad (9.4.50)$$

where R^f is the matrix associated with \mathcal{R}_f , I denotes the 2×2 identity matrix acting on the q_{n+1}, p_{n+1} space, and the 0's denote rectangular matrices of zeroes. Evidently, we have mapped an element of the inhomogeneous symplectic group in $2n$ dimensions into an element of the homogeneous (origin-preserving) symplectic group in $(2n + 2)$ dimensions. This process will be called *enlarging*.

What is the effect of $\hat{\mathcal{M}}_{\hat{f}}$ on the enlarged phase space? Evidently we immediately have the relation

$$\hat{\mathcal{M}}_{\hat{f}} q_{n+1} = q_{n+1}. \quad (9.4.51)$$

To explore matters further suppose, in analogy with (4.25), that f_1 has the property

$$: f_1 : z_a = \delta_a^f. \quad (9.4.52)$$

See (7.7.1) through (7.7.6). Then \hat{f}_1^2 has the properties

$$\begin{aligned} : \hat{f}_1^2 : z_a &= : (q_{n+1})f_1(z) : z_a = [(q_{n+1})f_1(z), z_a] \\ &= (q_{n+1})[f_1(z), z_a] = (q_{n+1}) : f_1 : z_a = (q_{n+1})\delta_a^f, \end{aligned} \quad (9.4.53)$$

$$: \hat{f}_1^2 :^m z_a = 0 \text{ for } m > 1, \quad (9.4.54)$$

$$: \hat{f}_1^2 : (q_{n+1}) = [(q_{n+1})f_1(z), (q_{n+1})] = 0, \quad (9.4.55)$$

$$: \hat{f}_1^2 : (p_{n+1}) = [(q_{n+1})f_1(z), (p_{n+1})] = f_1(z), \quad (9.4.56)$$

$$: \hat{f}_1^2 :^m (p_{n+1}) = 0 \text{ for } m > 1. \quad (9.4.57)$$

Let $\hat{\mathcal{R}}_{\hat{f}_1^2}$ denote the map

$$\hat{\mathcal{R}}_{\hat{f}_1^2} = \exp(:\hat{f}_1^2:). \quad (9.4.58)$$

From (4.53) through (4.57) we see that $\hat{\mathcal{R}}_{\hat{f}_1^2}$ has the property

$$\hat{\mathcal{R}}_{\hat{f}_1^2} z_a = z_a + (q_{n+1})\delta_a^f, \quad (9.4.59)$$

$$\hat{\mathcal{R}}_{\hat{f}_1^2} q_{n+1} = q_{n+1}, \quad (9.4.60)$$

$$\hat{\mathcal{R}}_{\hat{f}_1^2} p_{n+1} = p_{n+1} + f_1(z) = p_{n+1} + (z, (\delta^f)^*). \quad (9.4.61)$$

See also (7.7.3).

It follows from (4.59) through (4.61) that the action of $\hat{\mathcal{R}}_{\hat{f}_1^2}$ can be represented by a matrix $\hat{R}_{\hat{f}_1^2}$. In the simplest case that the original phase space is two dimensional, this matrix is 4×4 and is given by the relation

$$\hat{R}_{\hat{f}_1^2} = \begin{pmatrix} 1 & 0 & \delta_1^f & 0 \\ 0 & 1 & \delta_2^f & 0 \\ 0 & 0 & 1 & 0 \\ \delta_2^f & -\delta_1^f & 0 & 1 \end{pmatrix}. \quad (9.4.62)$$

Here we have used the ordering (4.15) and the J of (3.2.11). Evidently (4.62) is analogous to (4.38).

Finally, let us find the effect of $\hat{\mathcal{M}}_{\hat{f}}$. According to (4.49) and (4.58), it can be written in the form

$$\hat{\mathcal{M}}_{\hat{f}} = \hat{\mathcal{R}}_{\hat{f}_1^2} \check{\mathcal{R}}_{\hat{f}}. \quad (9.4.63)$$

It follows that the action of $\hat{\mathcal{M}}_{\hat{f}}$ can be represented by the matrix $\hat{R}^{\hat{f}}$ given by the relation

$$\hat{R}^{\hat{f}} = \check{R}^{\hat{f}} \hat{R}_{\hat{f}_1^2}. \quad (9.4.64)$$

See (8.4.19) and (8.4.20). Carrying out the indicated multiplication gives (in the 4×4 case) the result

$$\hat{R}^{\hat{f}} = \begin{pmatrix} R_{11}^f & R_{12}^f & (R^f \delta^f)_1 & 0 \\ R_{21}^f & R_{22}^f & (R^f \delta^f)_2 & 0 \\ 0 & 0 & 1 & 0 \\ \delta_2^f & -\delta_1^f & 0 & 1 \end{pmatrix}. \quad (9.4.65)$$

which is analogous to (4.46).

Shrinking

We have defined enlarged maps and have studied their effect on the enlarged phase space. Let us now explore how they behave under multiplication. Suppose \mathcal{M}_f and \mathcal{M}_g are any two inhomogeneous symplectic group maps, and $\hat{\mathcal{M}}_{\hat{f}}$ and $\hat{\mathcal{M}}_{\hat{g}}$ are their enlargements. We can form corresponding maps \mathcal{M}_h and $\hat{\mathcal{M}}_{\hat{h}}$ by the products

$$\mathcal{M}_h = \mathcal{M}_f \mathcal{M}_g, \quad (9.4.66)$$

$$\hat{\mathcal{M}}_{\hat{h}} = \hat{\mathcal{M}}_f \hat{\mathcal{M}}_g. \quad (9.4.67)$$

The product (4.66) involves the concatenation of maps that include translations, and its calculation entails the derivation and use of complicated (and only partially known) feed-down formulae as described in the previous section. By contrast, the product (4.67) is for origin-preserving maps in the enlarged 8-dimensional phase space. Its computation involves only the use of far simpler universal dimension-independent origin-preserving concatenation rules. What we wish to learn is whether \mathcal{M}_h can be deduced from a knowledge of $\hat{\mathcal{M}}_{\hat{h}}$. The process of constructing \mathcal{M}_h from $\hat{\mathcal{M}}_{\hat{h}}$ will be called *shrinking*. See Figure 9.4.1 for a pictorial presentation of this question.

To answer this question, let us compute $\hat{R}^{\hat{h}}$, the matrix corresponding to $\hat{\mathcal{M}}_{\hat{h}}$. It is given by the relation

$$\begin{aligned} \hat{R}^{\hat{h}} = \hat{R}^g \hat{R}^f &= \begin{pmatrix} R_{11}^g & R_{12}^g & (R^g \delta^g)_1 & 0 \\ R_{21}^g & R_{22}^g & (R^g \delta^g)_2 & 0 \\ 0 & 0 & 1 & 0 \\ \delta_2^g & -\delta_1^g & 0 & 1 \end{pmatrix} \begin{pmatrix} R_{11}^f & R_{12}^f & (R^f \delta^f)_1 & 0 \\ R_{21}^f & R_{22}^f & (R^f \delta^f)_2 & 0 \\ 0 & 0 & 1 & 0 \\ \delta_2^f & -\delta_1^f & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} R_{11}^h & R_{12}^h & (R^h \delta^f + R^g \delta^g)_1 & 0 \\ R_{21}^h & R_{22}^h & (R^h \delta^f + R^g \delta^g)_2 & 0 \\ 0 & 0 & 1 & 0 \\ * & * & * & 1 \end{pmatrix}. \end{aligned} \quad (9.4.68)$$

Here,

$$R^h = R^g R^f. \quad (9.4.69)$$

As before, for simplicity, we have treated the case where the original phase space is two dimensional, and the enlarged phase space is four dimensional. Again, the result in this case is sufficient to deduce the result in any dimension. Finally, we have not computed the starred entries in the bottom row of $\hat{R}^{\hat{h}}$. See Exercise 4.10.

We observe, as a consequence of (4.69), that the matrix R^h can be read off from the upper-left corner of $\hat{R}^{\hat{h}}$. Also, upon comparison of (4.68) with (4.46), we expect the upper two entries of the penultimate column of $\hat{R}^{\hat{h}}$ to be the entries of the vector $(R^h \delta^h)$. Therefore, from (4.46) and (4.68), we get the relation

$$R^h \delta^h = R^h \delta^f + R^g \delta^g. \quad (9.4.70)$$

In view of (4.69), the relation (4.70) can also be written in the form

$$\delta^h = \delta^f + (R^f)^{-1} \delta^g, \quad (9.4.71)$$

from which it follows that

$$(z, J \delta^h) = (z, J \delta^f) + (z, J(R^f)^{-1} \delta^g). \quad (9.4.72)$$

But from the symplectic condition (3.1.2) there is the relation

$$J(R^f)^{-1} = (R^f)^T J. \quad (9.4.73)$$

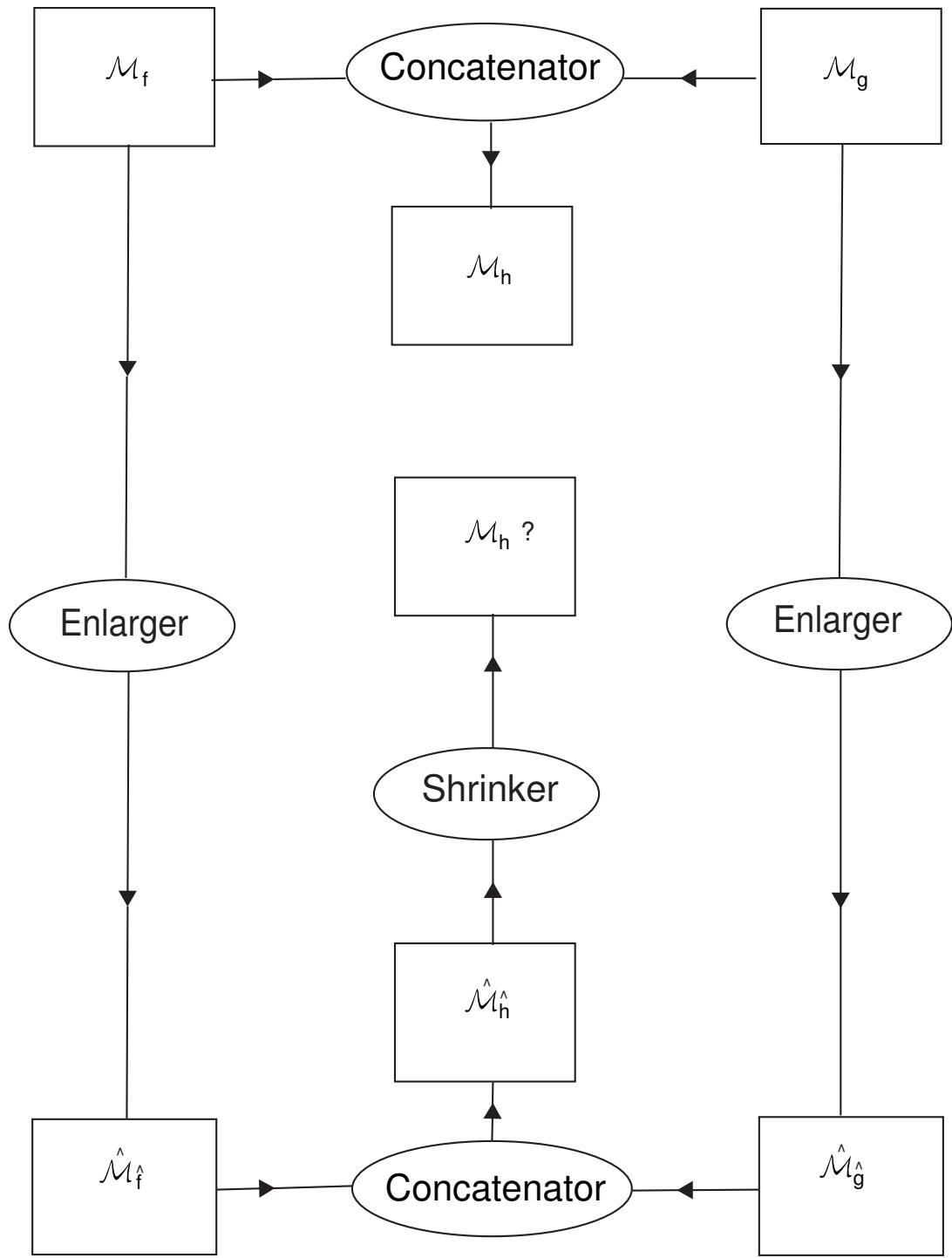


Figure 9.4.1: Concatenation of origin-preserving maps in an enlarged phase space to find equivalent results for maps, including translations, in the original phase space. The concatenator depicted at the top of the figure works with the usual phase space. When translations are taken into account, it involves the use of complicated feed-down formulae as illustrated in Section 9.3. The concatenator at the bottom of the figure works in an enlarged phase space, and employs the far-simpler concatenation rules for origin preserving maps.

Consequently, (4.72) can also be written in the form

$$(z, J\delta^h) = (z, J\delta^f) + (R^f z, J\delta^g). \quad (9.4.74)$$

Finally, use of (7.7.3) gives the result

$$h_1(z) = f_1(z) + g_1(R^f z) \quad (9.4.75)$$

or, equivalently,

$$h_1(z) = f_1(z) + \mathcal{R}_f g_1(z). \quad (9.4.76)$$

But (4.75), along with (4.69), are the rules (2.37) and (2.38) for concatenating inhomogeneous symplectic group maps. We conclude that, in the case of inhomogeneous symplectic group maps, the map \mathcal{M}_h can indeed be deduced from $\hat{\mathcal{M}}_{\hat{h}}$.

9.4.3 Case of General Maps

Enlarging

We now turn to the general case $ISpM(2n, \mathbb{R})$ for maps \mathcal{M}_f and \mathcal{M}_g of the form (1.1) and (1.2). The enlargement process will be carried out as before to yield the maps $\hat{\mathcal{M}}_{\hat{f}}$ and $\hat{\mathcal{M}}_{\hat{g}}$. For example, the map $\hat{\mathcal{M}}_{\hat{f}}$ is given by

$$\hat{\mathcal{M}}_{\hat{f}} = \exp(: \hat{f}_1^2 :)\check{\mathcal{R}}_{\hat{f}} \exp(: \hat{f}_3^3 :)\exp(: \hat{f}_4^4 :)\cdots \quad (9.4.77)$$

where \hat{f}_1^2 and $\check{\mathcal{R}}_{\hat{f}}$ are given by (4.48) and (4.50) as before, and

$$\hat{f}_m^m(\hat{z}) = f_m(z), \quad m = 3, 4, \dots. \quad (9.4.78)$$

Next form the product map

$$\hat{\mathcal{M}}_{\hat{h}} = \hat{\mathcal{M}}_{\hat{f}} \hat{\mathcal{M}}_{\hat{g}}. \quad (9.4.79)$$

Since $\hat{\mathcal{M}}_{\hat{h}}$ sends the origin into itself, it has a factorization of the form

$$\hat{\mathcal{M}}_{\hat{h}} = \hat{\mathcal{R}}_{\hat{h}} \hat{\mathcal{N}}_{\hat{h}}. \quad (9.4.80)$$

The linear map $\hat{\mathcal{R}}_{\hat{h}}$ will be described by a matrix $\hat{R}^{\hat{h}}$, and from the relation

$$\hat{\mathcal{R}}_{\hat{h}} = \hat{\mathcal{R}}_{\hat{f}} \hat{\mathcal{R}}_{\hat{g}} \quad (9.4.81)$$

we have the rule

$$\hat{R}^{\hat{h}} = \hat{R}^{\hat{g}} \hat{R}^{\hat{f}}. \quad (9.4.82)$$

The nonlinear map $\hat{\mathcal{N}}_{\hat{h}}$ will have a representation of the form

$$\hat{\mathcal{N}}_{\hat{h}} = \exp(: \hat{h}_3 :)\exp(: \hat{h}_4 :)\cdots \quad (9.4.83)$$

with the \hat{h}_m given by the relations of the form (8.4.31) through (8.4.36) already found in Section 8.4. Our task now is to extract \mathcal{M}_h from $\hat{\mathcal{M}}_{\hat{h}}$.

Let us first examine $\hat{\mathcal{R}}_h$ and its associated matrix \hat{R}^h . We know that $\hat{\mathcal{M}}_h$ has the property (4.2) since by construction the maps $\hat{\mathcal{M}}_{\hat{f}}$ and $\hat{\mathcal{M}}_{\hat{g}}$ have this property, and such maps form a group. Consequently $\hat{\mathcal{R}}_h$ satisfies (4.4). It follows, in analogy with (4.46) when written out for the full 8×8 case, that the matrix \hat{R}^h has the form

$$\hat{R}^h = \begin{pmatrix} R_{11}^h & R_{12}^h & R_{13}^h & R_{14}^h & R_{15}^h & R_{16}^h & (R^h \delta^h)_1 & 0 \\ R_{21}^h & R_{22}^h & R_{23}^h & R_{24}^h & R_{25}^h & R_{26}^h & (R^h \delta^h)_2 & 0 \\ R_{31}^h & R_{32}^h & R_{33}^h & R_{34}^h & R_{35}^h & R_{36}^h & (R^h \delta^h)_3 & 0 \\ R_{41}^h & R_{42}^h & R_{43}^h & R_{44}^h & R_{45}^h & R_{46}^h & (R^h \delta^h)_4 & 0 \\ R_{51}^h & R_{52}^h & R_{53}^h & R_{54}^h & R_{55}^h & R_{56}^h & (R^h \delta^h)_5 & 0 \\ R_{61}^h & R_{62}^h & R_{63}^h & R_{64}^h & R_{65}^h & R_{66}^h & (R^h \delta^h)_6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \delta_2^h & -\delta_1^h & \delta_4^h & -\delta_3^h & \delta_6^h & -\delta_5^h & \hat{R}_{87}^h & 1 \end{pmatrix}. \quad (9.4.84)$$

We can read off the entries in δ^h from the bottom row of (4.84), and from these entries construct $h_1(z)$. Specifically, if we write

$$h_1(z) = (z, (\delta^h)^*), \quad (9.4.85)$$

then we have the relation

$$[(\delta^h)^*]_b = (\hat{R}^h)_{8b} \text{ for } b \in [1, 6]. \quad (9.4.86)$$

Again see Section 7.7.

Next let $\check{\mathcal{R}}^h$ be the symplectic map described by the matrix \check{R}^h with

$$\check{R}^h = \begin{pmatrix} R^h & 0 \\ 0 & I \end{pmatrix}. \quad (9.4.87)$$

Here R^h is the 6×6 matrix obtained earlier, and I denotes the 2×2 identity matrix. It follows that $\hat{\mathcal{R}}_h$ can be rewritten in the form

$$\hat{\mathcal{R}}_h = \exp(\alpha_2^h : q_{n+1}^2 :) \exp(: q_{n+1} h_1 :) \check{\mathcal{R}}_h \quad (9.4.88)$$

with $n = 3$. Here α_2^h is given by the relation

$$\alpha_2^h = R_{87}^h / 2. \quad (9.4.89)$$

The quantity α_2^h is not presently of direct interest to us, but if desired it can be computed from the entries in $\hat{R}^{\hat{f}}$ and $\hat{R}^{\hat{g}}$. See Exercise 4.10.

We next turn to the nonlinear part $\hat{\mathcal{N}}_h$. We know from (4.5) that the \hat{h}_m are independent of p_{n+1} . Therefore the \hat{h}_m must have expansions of the form

$$\hat{h}_m(\hat{z}) = h_m^m(z) + (q_{n+1}) h_{m-1}^m(z) + \cdots + (q_{n+1})^m h_0^m(z). \quad (9.4.90)$$

Here the superscript m on the quantity h_ℓ^m indicates that the quantity is associated with \hat{h}_m , and the subscript ℓ indicates that the quantity is homogeneous of degree ℓ in the variables z . Let us explore the consequences of this expansion. In Section 9.3 we employed an expansion

in powers of ϵ where ϵ was a measure of the smallness of the first-degree generators. See, for example, relations of the kind (3.10) through (3.14). From this perspective, the expansion (4.90) is an expansion in powers of q_{n+1} with q_{n+1} playing the role of ϵ . Compare also (3.1) and (4.48). (See also Exercise 4.11.) Moreover, the use of the standard concatenator for origin preserving maps in the enlarged phase space produces power series expansions in the quantity q_{n+1} automatically!

Shrinking by Concatenation

Equally pleasant is the fact that this concatenator can be used to construct a *shrinker*. Since the quantities $[(q_{n+1}^{m-\ell})h_\ell^m]$ form a basis for the (p_{n+1} independent) polynomials of degree m in the enlarged phase space, $\hat{\mathcal{N}}_{\hat{h}}$ must have a factorization of the form

$$\begin{aligned}\hat{\mathcal{N}}_{\hat{h}} = & \exp(:\alpha_3 q_{n+1}^3:) \exp(:\alpha_4 q_{n+1}^4:) \exp(:\alpha_5 q_{n+1}^5:) \cdots \times \\ & \exp(:q_{n+1}^2 \tilde{h}_1^3:) \exp(:q_{n+1}^3 \tilde{h}_1^4:) \exp(:q_{n+1}^4 \tilde{h}_1^5:) \cdots \times \\ & \exp(:q_{n+1} \tilde{h}_2^3:) \exp(:q_{n+1}^2 \tilde{h}_2^4:) \exp(:q_{n+1}^3 \tilde{h}_2^5:) \cdots \times \\ & \exp(:\tilde{h}_3^3:) \exp(:q_{n+1} \tilde{h}_3^4:) \exp(:q_{n+1}^2 \tilde{h}_3^5:) \cdots \times \\ & \exp(:\tilde{h}_4^4:) \exp(:q_{n+1} \tilde{h}_4^5:) \exp(:q_{n+1}^2 \tilde{h}_4^6:) \cdots \times \\ & \exp(:\tilde{h}_5^5:) \exp(:q_{n+1} \tilde{h}_5^6:) \exp(:q_{n+1}^2 \tilde{h}_5^7:) \cdots ,\end{aligned}\tag{9.4.91}$$

where the quantities \tilde{h}_ℓ^m are yet to be determined. In a moment we will see that the \tilde{h}_ℓ^m can be computed using the concatenator. But first we observe that (4.91) is a *single grade* factorization with q_{n+1} playing the role of ϵ . See the end of Section 9.3. We may therefore set $q_{n+1} = 1$ in (4.91) and (4.88) to obtain \mathcal{M}_h from $\hat{\mathcal{M}}_{\hat{h}}$. So doing gives the result

$$\begin{aligned}\mathcal{M}_h = & \exp(:h_1^2:) \mathcal{R}_h \exp(:\tilde{h}_1^3 + \tilde{h}_1^4 + \tilde{h}_1^5 + \cdots:) \times \\ & \exp(:\tilde{h}_2^3:) \exp(:\tilde{h}_2^4:) \exp(:\tilde{h}_2^5:) \cdots \times \\ & \exp(:\tilde{h}_3^3:) \exp(:\tilde{h}_3^4:) \exp(:\tilde{h}_3^5:) \cdots \times \\ & \exp(:\tilde{h}_4^4:) \exp(:\tilde{h}_4^5:) \exp(:\tilde{h}_4^6:) \cdots \times \\ & \exp(:\tilde{h}_5^5:) \exp(:\tilde{h}_5^6:) \exp(:\tilde{h}_5^7:) \cdots .\end{aligned}\tag{9.4.92}$$

Here \mathcal{R}_h is a linear map whose associated matrix is R^h , and

$$h_1^2(z) = h_1(z).\tag{9.4.93}$$

The second two factors in (4.92) can be rearranged using the results of Section 9.2,

$$\mathcal{R}_h \exp(:\tilde{h}_1^3 + \tilde{h}_1^4 + \tilde{h}_1^5 + \cdots:) = \exp(:\check{h}_1:) \mathcal{R}_h\tag{9.4.94}$$

where

$$\check{h}_1 = \mathcal{R}_h(\tilde{h}_1^3 + \tilde{h}_1^4 + \tilde{h}_1^5 + \cdots).\tag{9.4.95}$$

Consequently, \mathcal{M}_h can also be written in the form

$$\begin{aligned} \mathcal{M}_h = & \exp(: h_1^2 + \check{h}_1 :) \times \\ & \mathcal{R}_h \exp(: \tilde{h}_2^3 :) \exp(: \tilde{h}_2^4 :) \exp(: \tilde{h}_2^5 :) \cdots \times \\ & \exp(: \tilde{h}_3^3 :) \exp(: \tilde{h}_3^4 :) \exp(: \tilde{h}_3^5 :) \cdots \times \\ & \exp(: \tilde{h}_4^4 :) \exp(: \tilde{h}_4^5 :) \exp(: \tilde{h}_4^6 :) \cdots \times \\ & \exp(: \tilde{h}_5^5 :) \exp(: \tilde{h}_5^6 :) \exp(: \tilde{h}_5^7 :) \cdots . \end{aligned} \quad (9.4.96)$$

Finally, the factors appearing in each of the lines in (4.96) beyond the first line may be combined using the concatenator for origin preserving maps in the original $2n$ -dimensional phase space to obtain \mathcal{M}_h in the final form (1.6). We have constructed a shrinker based on the assumption that the terms \tilde{h}_ℓ^m appearing in (4.91) can be found.

Illustration for the quotient algebra L^0/L^3

What remains to be shown is how the \tilde{h}_ℓ^m can be computed from the \hat{h}_m in (4.83) and (4.90) using the concatenator for origin preserving maps in the enlarged $(2n+2)$ -dimensional phase space. Since the procedure requires several steps, it is best illustrated first for a relatively simple example. Suppose we are working in the quotient algebra L^0/L^3 . Then $\hat{\mathcal{N}}_{\hat{h}}$ has the form

$$\begin{aligned} \hat{\mathcal{N}}_{\hat{h}} = & \exp(: h_3^3 + q_{n+1} h_2^3 + q_{n+1}^2 h_1^3 + q_{n+1}^3 h_0^3 :) \times \\ & \exp(: h_4^4 + q_{n+1} h_3^4 + q_{n+1}^2 h_2^4 + q_{n+1}^3 h_1^4 + q_{n+1}^4 h_0^4 :). \end{aligned} \quad (9.4.97)$$

Since the generators have *no* p_{n+1} dependence, they are in involution with powers of q_{n+1} , and these powers may be removed to the far right so that we may also write

$$\begin{aligned} \hat{\mathcal{N}}_{\hat{h}} = & \exp(: h_3^3 + q_{n+1} h_2^3 + q_{n+1}^2 h_1^3 :) \times \\ & \exp(: h_4^4 + q_{n+1} h_3^4 + q_{n+1}^2 h_2^4 + q_{n+1}^3 h_1^4 :) \times \\ & \exp(: q_{n+1}^3 h_0^3 :) \exp(: q_{n+1}^4 h_0^4 :). \end{aligned} \quad (9.4.98)$$

Now we are ready to begin.

Isolation of linear in z generators

The *linear* in z generator $q_{n+1}^2 h_1^3$, which produces a translation in the $2n$ -dimensional phase space, may be isolated by writing the identity

$$\hat{\mathcal{N}}_{\hat{h}} = \exp(: q_{n+1}^2 h_1^3 :) [\exp(- : q_{n+1}^2 h_1^3 :) \hat{\mathcal{N}}_{\hat{h}}] \quad (9.4.99)$$

and making the definition

$${}^1\hat{\mathcal{N}}_{\hat{h}} = [\exp(- : q_{n+1}^2 h_1^3 :) \hat{\mathcal{N}}_{\hat{h}}]. \quad (9.4.100)$$

Upon manipulating exponents using the BCH theorem we find that ${}^1\hat{\mathcal{N}}_{\hat{h}}$ has the form

$$\begin{aligned} {}^1\hat{\mathcal{N}}_{\hat{h}} = & \exp(: {}^1h_3^3 + q_{n+1} {}^1h_2^3 :) \times \\ & \exp(: {}^1h_4^4 + q_{n+1} {}^1h_3^4 + q_{n+1}^2 {}^1h_2^4 + q_{n+1}^3 {}^1h_1^4 :) \times \\ & \exp(: q_{n+1}^3 {}^1h_0^3 :) \exp(: q_{n+1}^4 {}^1h_0^4 :), \end{aligned} \quad (9.4.101)$$

and we conclude that

$$\tilde{h}_1^3 = h_1^3. \quad (9.4.102)$$

Here the superscript 1 in ${}^1\hat{\mathcal{N}}_{\hat{h}}$ indicates that one isolation step has been taken; and the superscript 1 in ${}^1h_{\ell}^m$ indicates that one isolation step has been taken and that the ${}^1h_{\ell}^m$ may differ from the previous h_{ℓ}^m .

Next the linear in z generator $q_{n+1}^3 {}^1h_1^4$, which which also produces a translation in the $2n$ -dimensional phase space, may be isolated by writing the identity

$${}^1\hat{\mathcal{N}}_{\hat{h}} = \exp(: q_{n+1}^3 {}^1h_1^4 :)[\exp(- : q_{n+1}^3 {}^1h_1^4 :) {}^1\hat{\mathcal{N}}_{\hat{h}}] \quad (9.4.103)$$

and making the definition

$${}^2\hat{\mathcal{N}}_{\hat{h}} = [\exp(- : q_{n+1}^3 {}^1h_1^4 :) {}^1\hat{\mathcal{N}}_{\hat{h}}]. \quad (9.4.104)$$

Again, upon manipulating exponents using the BCH theorem, we find that ${}^2\hat{\mathcal{N}}_{\hat{h}}$ has the form

$$\begin{aligned} {}^2\hat{\mathcal{N}}_{\hat{h}} = & \exp(: {}^2h_3^3 + q_{n+1} {}^2h_2^3 :) \times \\ & \exp(: {}^2h_4^4 + q_{n+1} {}^2h_3^4 + q_{n+1}^2 {}^2h_2^4 :) \times \\ & \exp(: q_{n+1}^3 {}^2h_0^3 :) \exp(: q_{n+1}^4 {}^2h_0^4 :), \end{aligned} \quad (9.4.105)$$

and we conclude that

$$\tilde{h}_1^4 = {}^1h_1^4. \quad (9.4.106)$$

Here the superscript 2 in ${}^2\hat{\mathcal{N}}_{\hat{h}}$ indicates that a second isolation step has been taken; and the superscript 2 in ${}^2h_{\ell}^m$ indicates that a second isolation step has been taken and that the ${}^2h_{\ell}^m$ may differ from the previous ${}^1h_{\ell}^m$.

Inspection of (4.105) indicates that all linear in z generators have now been isolated away. We are ready to begin isolating the *quadratic* in z generators.

Isolation of quadratic in z generators

The quadratic in z generator $q_{n+1} {}^2h_2^3$, which produces a linear transformation in the $2n$ -dimensional phase space, may be isolated by writing the identity

$${}^2\hat{\mathcal{N}}_{\hat{h}} = \exp(: q_{n+1} {}^2h_2^3 :)[\exp(- : q_{n+1} {}^2h_2^3 :) {}^2\hat{\mathcal{N}}_{\hat{h}}] \quad (9.4.107)$$

and making the definition

$${}^3\hat{\mathcal{N}}_{\hat{h}} = [\exp(- : q_{n+1} {}^2h_2^3 :) {}^2\hat{\mathcal{N}}_{\hat{h}}]. \quad (9.4.108)$$

Now use of the BCH theorem shows that ${}^3\hat{\mathcal{N}}_{\hat{h}}$ has the form

$$\begin{aligned} {}^3\hat{\mathcal{N}}_{\hat{h}} = & \exp(: {}^3h_3^3 :) \times \\ & \exp(: {}^3h_4^4 + q_{n+1} {}^3h_3^4 + q_{n+1}^2 {}^3h_2^4 :) \times \\ & \exp(: q_{n+1}^3 {}^3h_0^3 :) \exp(: q_{n+1}^4 {}^3h_0^4 :), \end{aligned} \quad (9.4.109)$$

and we conclude that

$$\tilde{h}_2^3 = {}^2h_2^3. \quad (9.4.110)$$

Next the quadratic in z generator $q_{n+1}^2 {}^3h_2^4$, which also produces a linear transformation in the $2n$ -dimensional phase space, may be isolated by writing the identity

$${}^3\hat{\mathcal{N}}_{\hat{h}} = \exp(: q_{n+1}^2 {}^3h_2^4 :)[\exp(- : q_{n+1}^2 {}^3h_2^4 :) {}^3\hat{\mathcal{N}}_{\hat{h}}] \quad (9.4.111)$$

and making the definition

$${}^4\hat{\mathcal{N}}_{\hat{h}} = [\exp(- : q_{n+1}^2 {}^3h_2^4 :) {}^3\hat{\mathcal{N}}_{\hat{h}}]. \quad (9.4.112)$$

Now use of the BCH theorem shows that ${}^4\hat{\mathcal{N}}_{\hat{h}}$ has the form

$$\begin{aligned} {}^4\hat{\mathcal{N}}_{\hat{h}} = & \exp(: {}^4h_3^3 :) \times \\ & \exp(: {}^4h_4^4 + q_{n+1} {}^4h_3^4 :) \times \\ & \exp(: q_{n+1}^3 {}^4h_0^3 :) \exp(: q_{n+1}^4 {}^4h_0^4 :), \end{aligned} \quad (9.4.113)$$

and we conclude that

$$\tilde{h}_2^4 = {}^3h_2^4. \quad (9.4.114)$$

Inspection of (4.113) indicates that all quadratic in z generators have now been isolated away. We are ready to begin isolating the *cubic* in z generators.

Isolation of cubic in z generators

The cubic in z generator ${}^4h_3^3$, which produces a quadratic plus higher-order transformation in the $2n$ -dimensional phase space, may be isolated by writing the identity

$${}^4\hat{\mathcal{N}}_{\hat{h}} = \exp(: {}^4h_3^3 :)[\exp(- : {}^4h_3^3 :) {}^4\hat{\mathcal{N}}_{\hat{h}}] \quad (9.4.115)$$

and making the definition

$${}^5\hat{\mathcal{N}}_{\hat{h}} = [\exp(- : {}^4h_3^3 :) {}^4\hat{\mathcal{N}}_{\hat{h}}]. \quad (9.4.116)$$

Now use of the BCH theorem shows that ${}^5\hat{\mathcal{N}}_{\hat{h}}$ has the form

$$\begin{aligned} {}^5\hat{\mathcal{N}}_{\hat{h}} = & \exp(: {}^5h_4^4 + q_{n+1} {}^5h_3^4 :) \times \\ & \exp(: q_{n+1}^3 {}^5h_0^3 :) \exp(: q_{n+1}^4 {}^5h_0^4 :), \end{aligned} \quad (9.4.117)$$

and we conclude that

$$\tilde{h}_3^3 = {}^4h_3^3. \quad (9.4.118)$$

Next the cubic in z generator $q_{n+1} {}^5h_3^4$, which also produces a quadratic plus higher-order transformation in the $2n$ -dimensional phase space, may be isolated by writing the identity

$${}^5\hat{\mathcal{N}}_{\hat{h}} = \exp(: q_{n+1} {}^5h_3^4 :)[\exp(- : q_{n+1} {}^5h_3^4 :) {}^5\hat{\mathcal{N}}_{\hat{h}}] \quad (9.4.119)$$

and making the definition

$${}^6\hat{\mathcal{N}}_{\hat{h}} = [\exp(- : q_{n+1} {}^5h_3^4 :) {}^5\hat{\mathcal{N}}_{\hat{h}}]. \quad (9.4.120)$$

Now use of the BCH theorem shows that ${}^6\hat{\mathcal{N}}_{\hat{h}}$ has the form

$$\begin{aligned} {}^6\hat{\mathcal{N}}_{\hat{h}} = & \exp(: {}^6h_4^4 :) \times \\ & \exp(: q_{n+1}^3 {}^6h_0^3 :) \exp(: q_{n+1}^4 {}^6h_0^4 :), \end{aligned} \quad (9.4.121)$$

and we conclude that

$$\tilde{h}_3^4 = {}^5h_3^4. \quad (9.4.122)$$

Inspection of (4.121) indicates that all cubic in z generators have now been isolated away. We are ready to begin isolating the *quartic* in z generators.

Isolation of quartic in z generators

The remaining quartic in z generator ${}^6h_4^4$, which produces a cubic plus higher-order transformation in the $2n$ -dimensional phase space, may be isolated by writing the identity

$${}^6\hat{\mathcal{N}}_{\hat{h}} = \exp(: {}^6h_4^4 :) [\exp(- : {}^6h_4^4 :) {}^6\hat{\mathcal{N}}_{\hat{h}}] \quad (9.4.123)$$

and making the definition

$${}^7\hat{\mathcal{N}}_{\hat{h}} = [\exp(- : {}^6h_4^4 :) {}^6\hat{\mathcal{N}}_{\hat{h}}]. \quad (9.4.124)$$

Now use of the BCH theorem shows that ${}^7\hat{\mathcal{N}}_{\hat{h}}$ has the form

$${}^7\hat{\mathcal{N}}_{\hat{h}} = \exp(: q_{n+1}^3 {}^7h_0^3 :) \exp(: q_{n+1}^4 {}^7h_0^4 :), \quad (9.4.125)$$

and we conclude that

$$\tilde{h}_4^4 = {}^6h_4^4. \quad (9.4.126)$$

Overview

Inspection of (4.125) indicates that, in the case of L^0/L^3 , we have achieved our goal. Namely, by repeated isolating and concatenating, we eventually achieved the factorization (4.91). All z -dependent generators have been isolated away, and all that remains are generators that depend solely on q_{n+1} . These remaining generators have no effect on the the $2n$ -dimensional phase space variables z , and therefore ${}^7\hat{\mathcal{N}}_{\hat{h}}$ acts as the identity map \mathcal{I} on the z phase space.

We note that while these steps are somewhat difficult to characterize analytically (which is to be expected because the results of Section 9.3 were complicated), they are easy to implement numerically.

Is there a pattern in what we have done? Review of the steps we have taken shows that we have extracted the \tilde{h}_ℓ^m in a particular order. Let r be a *running* index. Table 4.1 below shows the order in which we have extracted the \tilde{h}_ℓ^m . Here we have defined the difference d by the relation

$$d = m - \ell \quad (9.4.127)$$

so that q_{n+1} and \tilde{h}_ℓ^m occur in the combination $q_{n+1}^d \tilde{h}_\ell^m$.

Let us make the definition

$${}^0\hat{\mathcal{N}}_{\hat{h}} = \hat{\mathcal{N}}_{\hat{h}}. \quad (9.4.128)$$

Table 9.4.1: Order in which the \tilde{h}_ℓ^m are to be extracted for the case L^0/L^3 .

r	ℓ	m	d
1	1	3	2
2	1	4	3
3	2	3	1
4	2	4	2
5	3	3	0
6	3	4	1
7	4	4	0

Then we may view the whole process as being recursive. At any given stage a map ${}^{r-1}\hat{\mathcal{N}}_h$ and a pair of indices $\ell(r)$ and $m(r)$ are provided as input, and a map ${}^r\hat{\mathcal{N}}_h$ and polynomial \tilde{h}_ℓ^m are produced as output. See Figure 4.2. The polynomial \tilde{h}_ℓ^m is determined by examination of ${}^{r-1}\hat{\mathcal{N}}_h$ and given by the rule

$$\tilde{h}_\ell^m = {}^{r-1}h_\ell^m. \quad (9.4.129)$$

The map ${}^r\hat{\mathcal{N}}_h$ is given by carrying out the concatenation

$${}^r\hat{\mathcal{N}}_h = \exp(- : q_{n+1}^d \tilde{h}_\ell^m :) \times {}^{r-1}\hat{\mathcal{N}}_h. \quad (9.4.130)$$



Figure 9.4.2: A recursive step that takes a map ${}^{r-1}\hat{\mathcal{N}}_h$ and a pair of indices $\ell(r)$ and $m(r)$ as input, and produces a map ${}^r\hat{\mathcal{N}}_h$ and polynomial \tilde{h}_ℓ^m as output.

Shrinking in the General Case

It is now a simple matter to generalize to higher-order cases. For example, Table 4.2 shows the extraction order to be used when working with maps through 7th order, the order that is available using the concatenation rules provided by (8.4.31) through (8.4.36).

Finally, we note that the procedure is self checking. For the final value of r the corresponding map ${}^r\hat{\mathcal{N}}_h$ will have the property that all its generators have no z dependence; they can depend only on q_{n+1} . We have already seen this for the map ${}^7\hat{\mathcal{N}}_h$ when working with third-order maps. See (4.125). The same will be true of the map ${}^{33}\hat{\mathcal{N}}_h$ when working with

maps through 7th order. Conversely, if all the factors in (4.91) are concatenated together, the result must be the original map $\hat{\mathcal{N}}_{\hat{h}}$.

Table 9.4.2: Order in which the \tilde{h}_ℓ^m are to be extracted for the case L^0/L^7 .

r	ℓ	m	d	r	ℓ	m	d
1	1	3	2	19	4	4	0
2	1	4	3	20	4	5	1
3	1	5	4	21	4	6	2
4	1	6	5	22	4	7	3
5	1	7	6	23	4	8	4
6	1	8	7	24	5	5	0
7	2	3	1	25	5	6	1
8	2	4	2	26	5	7	2
9	2	5	3	27	5	8	3
10	2	6	4	28	6	6	0
11	2	7	5	29	6	7	1
12	2	8	6	30	6	8	2
13	3	3	0	31	7	7	0
14	3	4	1	32	7	8	1
15	3	5	2	33	8	8	0
16	3	6	3				
17	3	7	4				
18	3	8	5				

Exercises

9.4.1. Verify that maps $\hat{\mathcal{M}}$ satisfying (4.2) form a group.

9.4.2. Prove (4.5) in detail.

9.4.3. Verify that the symplectic condition requires the relations (4.16).

9.4.4. Verify (4.25) through (4.30).

9.4.5. Verify (4.38) and (4.39).

9.4.6. Verify (4.41) through (4.43).

9.4.7. Verify (4.44) through (4.46).

9.4.8. Verify (4.53) through (4.62).

9.4.9. Verify (4.65).

9.4.10. Consider two *linear* symplectic maps $\hat{\mathcal{M}}_{\hat{f}}$ and $\hat{\mathcal{M}}_{\hat{g}}$ of the factored form

$$\hat{\mathcal{M}}_{\hat{f}} = \exp(: \alpha_2^{\hat{f}} q_{n+1}^2 :) \exp(: q_{n+1} f_1(z) :) \check{\mathcal{R}}_{\hat{f}}, \quad (9.4.131)$$

$$\hat{\mathcal{M}}_{\hat{g}} = \exp(: \alpha_2^{\hat{g}} q_{n+1}^2 :) \exp(: q_{n+1} g_1(z) :) \check{\mathcal{R}}_{\hat{g}}, \quad (9.4.132)$$

where $\check{\mathcal{R}}_{\hat{f}}$ and $\check{\mathcal{R}}_{\hat{g}}$ leave the q_{n+1}, p_{n+1} subspace invariant,

$$\check{\mathcal{R}}_{\hat{f}} q_{n+1} = q_{n+1}, \text{ etc.,} \quad (9.4.133)$$

$$\check{\mathcal{R}}_{\hat{f}} p_{n+1} = p_{n+1}, \text{ etc.} \quad (9.4.134)$$

Show that (4.106) and (4.107) plus the symplectic condition require that the associated matrices $\check{R}^{\hat{f}}$ etc. be of the general form (4.50). Let $\hat{\mathcal{M}}_{\hat{h}}$ be the product of $\hat{\mathcal{M}}_{\hat{f}}$ and $\hat{\mathcal{M}}_{\hat{g}}$,

$$\begin{aligned} \hat{\mathcal{M}}_{\hat{h}} &= \hat{\mathcal{M}}_{\hat{f}} \hat{\mathcal{M}}_{\hat{g}} = \exp(: \alpha_2^{\hat{f}} q_{n+1}^2 :) \exp(: q_{n+1} f_1(z) :) \check{\mathcal{R}}_{\hat{f}} \times \\ &\quad \exp(: \alpha_2^{\hat{g}} q_{n+1}^2 :) \exp(: q_{n+1} g_1(z) :) \check{\mathcal{R}}_{\hat{g}}. \end{aligned} \quad (9.4.135)$$

Show, by manipulating the various factors involved, that $\hat{\mathcal{M}}_{\hat{h}}$ can be re-expressed in the factored form

$$\begin{aligned} \hat{\mathcal{M}}_{\hat{h}} &= \exp\{ : q_{n+1}^2 : (\alpha_2^{\hat{f}} + \alpha_2^{\hat{g}} + [f_1(z), \check{\mathcal{R}}_{\hat{f}} g_1(z)]/2) : \} \times \\ &\quad \exp\{ : q_{n+1} (f_1(z) + \check{\mathcal{R}}_{\hat{f}} g_1(z)) : \} \check{\mathcal{R}}_{\hat{f}} \check{\mathcal{R}}_{\hat{g}}. \end{aligned} \quad (9.4.136)$$

Thus, if we write $\hat{\mathcal{M}}_{\hat{h}}$ as

$$\hat{\mathcal{M}}_{\hat{h}} = \exp(: \alpha_2^{\hat{h}} q_{n+1}^2 :) \exp(: q_{n+1} h_1(z) :) \check{\mathcal{R}}_{\hat{h}}, \quad (9.4.137)$$

there are the relations

$$\alpha_2^{\hat{h}} = \alpha_2^{\hat{f}} + \alpha_2^{\hat{g}} + [f_1(z), \check{\mathcal{R}}_{\hat{f}} g_1(z)]/2, \quad (9.4.138)$$

$$h_1(z) = f_1(z) + \check{\mathcal{R}}_{\hat{f}} g_1(z), \quad (9.4.139)$$

$$\check{\mathcal{R}}_{\hat{h}} = \check{\mathcal{R}}_{\hat{f}} \check{\mathcal{R}}_{\hat{g}}. \quad (9.4.140)$$

Carry out the indicated multiplication in (4.68) and verify that your results are equivalent to those found above.

9.4.11. Problem about the relation between the inhomogeneous symplectic map Lie algebra $\epsilon L^0/\epsilon L^\ell$ in $2n$ dimensions and the subgroup of the homogeneous symplectic map Lie algebra L^0/L^ℓ in $2n + 2$ dimensions produced by all p_{n+1} independent generators.

9.4.12. Verify that a factorization of the form (4.91) is possible.

9.4.13. Verify the relations (4.93) through (4.96).

9.4.14. Verify that $\hat{\mathcal{N}}_h$ has the form (4.97) when one works in the quotient algebra L^0/L^3 .

9.4.15. Verify (4.98) through (4.103).

9.4.16. Verify that extracting the \tilde{h}_ℓ^m in the order given by Table 4.1 or 4.2 never results, during the extraction process, in the "reappearance" in the subsequent maps ${}^r\hat{\mathcal{N}}_h$ of any of the previously removed terms.

9.5 Map Inversion and Reverse and Mixed Factorizations

Much of the discussion of this section is analogous to that of Section 8.5. Suppose, as in (1.1), that the map \mathcal{M}_f is written in the standard factored product form

$$\mathcal{M}_f = \exp(: f_1 :) \mathcal{R}_f \exp(: f_3 :) \exp(: f_4 :) \cdots . \quad (9.5.1)$$

Here, as in Section 8.5, \mathcal{R}_f denotes the map

$$\mathcal{R}_f = \exp(: f_2^c :) \exp(: f_2^a :) \quad (9.5.2)$$

with the associated matrix R^f given by the relation

$$R^f = \exp(JS^a) \exp(JS^c). \quad (9.5.3)$$

It follows immediately from (5.1) that the *inverse* of \mathcal{M}_f has the representation

$$(\mathcal{M}_f)^{-1} = \cdots \exp(- : f_4 :) \exp(- : f_3 :)(\mathcal{R}_f)^{-1} \exp(- : f_1 :). \quad (9.5.4)$$

As before we observe that (5.4) gives a representation for the inverse of \mathcal{M}_f in the form of a reverse factorization, and that we would also like to have a representation in the standard forward factorization

$$(\mathcal{M}_f)^{-1} = \exp(: h_1 :) \mathcal{R}_h \exp(: h_3 :) \exp(: h_4 :) \cdots . \quad (9.5.5)$$

See Section 7.8. This is easily accomplished with the aid of the concatenation formulas of the previous section. We simply write (5.4) and (5.5) in the form

$$\begin{aligned} & \cdots [\exp(- : f_4 :)][\exp(- : f_3 :)][(\mathcal{R}_f)^{-1}][\exp(- : f_1 :)] = \\ & \exp(: h_1 :) \mathcal{R}_h \exp(: h_3 :) \exp(: h_4 :) \cdots \end{aligned} \quad (9.5.6)$$

where we have used square brackets to indicate that the various maps are to be concatenated together. Note that in this case (8.5.7) no longer holds because of the feed-down terms produced by moving the f_1 factor in (5.6) to the left.

The relation (5.5) also provides a procedure for reverse factorizing a map. Suppose we wish to represent \mathcal{M}_f in reverse factorized form. That is, we wish to find generators g_m such that

$$\begin{aligned} \mathcal{M}_f = & \exp(: f_1 :) \mathcal{R}_f \exp(: f_3 :) \exp(: f_4 :) \cdots = \\ & \cdots \exp(: g_4 :) \exp(: g_3 :) \mathcal{R}_g \exp(: g_1 :). \end{aligned} \quad (9.5.7)$$

Simply take the inverse of both sides of (5.7) and use (5.5) to get the relation

$$\begin{aligned} & \exp(- : g_1 :)(\mathcal{R}_g)^{-1} \exp(- : g_3 :)\exp(- : g_4 :)\cdots = \\ & \exp(: h_1 :) \mathcal{R}_h \exp(: h_3 :) \exp(: h_4 :) \cdots . \end{aligned} \quad (9.5.8)$$

From (5.8) we find the desired results

$$\mathcal{R}_g = (\mathcal{R}_h)^{-1}, \quad (9.5.9)$$

$$g_m = -h_m. \quad (9.5.10)$$

We close this section with a remark about mixed factorizations. See Section 7.8. Suppose, for example, we desire a mixed factorization for \mathcal{M}_f of the form

$$\mathcal{M}_f = \mathcal{R}_{f'} \exp(: f'_3 :) \exp(: f'_4 :) \cdots \exp(: f'_1 :). \quad (9.5.11)$$

That is, we wish to move the f_1 term in (5.1) to the far right, but keep the remaining factors in ascending order. Comparison of (5.7) and (5.11) gives the result

$$f'_1 = g_1, \quad (9.5.12)$$

and the relation

$$\mathcal{R}_{f'} \exp(: f'_3 :) \exp(: f'_4 :) \cdots = \cdots \exp(: g_4 :) \exp(: g_3 :) \mathcal{R}_g. \quad (9.5.13)$$

The remaining factors $\mathcal{R}_{f'}, : f'_3 :, : f'_4 : \cdots$ can be gotten by applying the concatenation formulas of Section 8.4 to the right side of (5.13). That is, we write (5.13) in the form

$$\mathcal{R}_{f'} \exp(: f'_3 :) \exp(: f'_4 :) = \cdots [\exp(: g_4 :)][\exp(: g_3 :)][\mathcal{R}_g]. \quad (9.5.14)$$

where the square brackets indicate that the various maps are to be concatenated together.

There is a relation between f_1 and f'_1 that could have been examined in the previous section, or even in Section 7.7, but can just as conveniently be discussed here. See Exercises 5.1 and 5.2. Let us write (5.1) and (5.11) in the forms

$$\mathcal{M}_f = \exp(: f_1 :) \mathcal{S}_f, \quad (9.5.15)$$

$$\mathcal{M}_f = \mathcal{S}_{f'} \exp(: f'_1 :), \quad (9.5.16)$$

with

$$\mathcal{S}_f = \mathcal{R}_f \exp(: f_3 :) \exp(: f_4 :) \cdots, \quad (9.5.17)$$

$$\mathcal{S}_{f'} = \mathcal{R}_{f'} \exp(: f'_3 :) \exp(: f'_4 :) \cdots. \quad (9.5.18)$$

Following Section 7.7, let us also write f_1 and f'_1 in the forms

$$f_1(z) = -(\delta, Jz), \quad (9.5.19)$$

$$f'_1(z) = -(\delta', Jz). \quad (9.5.20)$$

Then we have the relations

$$\delta'_a = \mathcal{S}_f z_a|_{z=\delta}, \quad (9.5.21)$$

$$\delta_a = (\mathcal{S}_f)^{-1} z_a|_{z=\delta'}. \quad (9.5.22)$$

To see the truth of (5.21) and (5.22), apply \mathcal{M}_f in both its representations (5.15) and (5.16) to the origin. Consider first the representation (5.16). We know that by construction,

see (5.18), the map $\mathcal{S}_{f'}$ sends the origin into itself. Therefore, since maps act in the order in which they occur when read from left to right (see Section 8.3), the first factor in (5.16) acts on the origin and leaves it in peace. Also, from Section 7.7, we know that $\exp(: f'_1 :)$ sends the origin into δ' . Consequently, we find the result

$$\mathcal{M}_f z_a|_{z=0} = \delta'_a. \quad (9.5.23)$$

Consider next the representation (5.15). The first factor, $\exp(: f_1 :)$, sends the origin into δ . Subsequently \mathcal{S}_f acts on δ to give the net result

$$\mathcal{M}_f z_a|_{z=0} = \mathcal{S}_f z_a|_{z=\delta}. \quad (9.5.24)$$

Upon comparing (5.23) and (5.24) we see that (5.21) and (5.22) are indeed correct. For another set of similar relations, see Exercise 5.3.

Exercises

9.5.1. Derive (5.21) starting with (7.7.13) and (7.7.23).

9.5.2. Consider a 2-dimensional phase space and suppose that $\mathcal{S}_{f'}$ has the simple form

$$\mathcal{S}_{f'} = \exp(: q^4 :). \quad (9.5.25)$$

Working within the quotient group generated by the Lie algebra L^0/L^3 , use (3.18) and (3.22) to verify (5.21).

9.5.3. Verify the relations

$$\mathcal{M}_f z_a|_{z=-\delta} = 0, \quad (9.5.26)$$

$$(\mathcal{M}_f)^{-1} z_a|_{z=0} = -\delta_a. \quad (9.5.27)$$

Derive the relations

$$-\delta'_a = \mathcal{S}_{f'} z_a|_{z=-\delta}, \quad (9.5.28)$$

$$-\delta_a = (\mathcal{S}_{f'})^{-1} z_a|_{z=-\delta'}. \quad (9.5.29)$$

Hint: Apply \mathcal{M}_f as given by (5.15) and (5.16) to the phase-space point $(-\delta)$.

9.6 Taylor and Hybrid Taylor-Lie Concatenation and Inversion

This section extends the results of Section 8.6 to the case where the map to be treated may have constant terms. Again all the possibilities illustrated in Figure 8.6.1 may arise, and we will discuss those of greatest interest.

Suppose, as described in the beginning of Section 8.6, that both the maps \mathcal{M}_1 and \mathcal{M}_2 are in Taylor form, and we also desire to represent their product in Taylor form. We consider first the best of all possible circumstances. In that circumstance \mathcal{M}_1 sends the phase-space point z^0 to the intermediate point \bar{z}^0 ,

$$\mathcal{M}_1 : z^0 \rightarrow \bar{z}^0, \quad (9.6.1)$$

and we assume that \mathcal{M}_1 has a known Taylor expansion about z^0 . Also, \mathcal{M}_2 sends the intermediate point \bar{z}^0 to the final point $\bar{\bar{z}}^0$,

$$\mathcal{M}_2 : \bar{z}^0 \rightarrow \bar{\bar{z}}^0, \quad (9.6.2)$$

and we assume that \mathcal{M}_2 has a known Taylor expansion about \bar{z}^0 . What we desire is a Taylor expansion about the point z^0 for the product map \mathcal{M}_3 that sends z^0 immediately to $\bar{\bar{z}}^0$,

$$\mathcal{M}_3 : z^0 \rightarrow \bar{\bar{z}}^0. \quad (9.6.3)$$

This desire is easily met. Introduce the deviation variables ζ_a by writing

$$z_a = z_a^0 + \zeta_a. \quad (9.6.4)$$

Then, by assumption, \mathcal{M}_1 has a known truncated Taylor expansion of the form

$$\bar{z}_a = \bar{z}_a(z) = \bar{z}_a^0 + \sum_{m=1}^D g_a^1(m; \zeta). \quad (9.6.5)$$

Introduce as well the deviation variables $\bar{\zeta}_a$ by writing

$$\bar{z}_a = \bar{z}_a^0 + \bar{\zeta}_a. \quad (9.6.6)$$

Then \mathcal{M}_2 is assumed to have the known truncated Taylor expansion

$$\bar{\bar{z}}_a = \bar{\bar{z}}_a(\bar{z}) = \bar{\bar{z}}_a^0 + \sum_{m'=1}^D g_a^2(m'; \bar{\zeta}). \quad (9.6.7)$$

Now use (6.5) and (6.6) to write the relation

$$\bar{\zeta}_a = \bar{\zeta}_a(\zeta) = \sum_{m=1}^D g_a^1(m; \zeta). \quad (9.6.8)$$

Also introduce the deviation variables $\bar{\zeta}_a$, defined by

$$\bar{z}_a = \bar{z}_a^0 + \bar{\zeta}_a, \quad (9.6.9)$$

to rewrite (6.7) in the form

$$\bar{\zeta}_a = \bar{\zeta}_a(\bar{\zeta}) = \sum_{m'=1}^D g_a^2(m'; \bar{\zeta}). \quad (9.6.10)$$

Then \mathcal{M}_3 has the expansion

$$\bar{z}_a = \bar{z}_a(z) = \bar{z}_a^0 + \bar{\zeta}_a(\zeta) = \bar{z}_a^0 + \sum_{m''=1}^D g_a^3(m''; \zeta) \quad (9.6.11)$$

where the polynomials g_a^3 are given by the relations

$$g_a^3(m''; \zeta) = P_{m''} \sum_{m'=1}^D g_a^2(m'; \bar{\zeta}(\zeta)). \quad (9.6.12)$$

As before, $P_{m''}$ denotes a projection operator that retains only terms of degree m'' in the variables ζ . Also, all the required operations can again be carried out using TPSA. We see that the relations (6.8), (6.10), and (6.12) are completely analogous to the relations (8.6.2), (8.6.4), and (8.6.7) for the case of no translations. Indeed, with the use of deviation variables, all the methods of Section 8.6 can be employed. For example, a deviation variable map of the form (6.8) can be inverted by the recursion method.

Often the optimal circumstance we have just treated does not hold. It may be that \mathcal{M}_1 sends z^0 to \bar{z}^0 and \mathcal{M}_2 sends \bar{z}^0 to $\bar{\bar{z}}^0$ as before and as described by (6.1) and (6.2). However, it may happen that \mathcal{M}_2 does not have a known Taylor expansion about \bar{z}^0 . Instead, we assume that \mathcal{M}_2 has a known Taylor expansion about a point \bar{z}' that is near \bar{z}^0 . With the introduction of suitable deviation variables if necessary, and without loss of generality, we may consider truncated Taylor series of the form

$$\bar{z}_a = \bar{z}_a(z) = \sum_{m=0}^D g_a^1(m; z) \quad (9.6.13)$$

and

$$\bar{z}_a = \bar{z}_a(\bar{z}) = \sum_{m'=0}^D g_a^2(m'; \bar{z}). \quad (9.6.14)$$

The relations (8.6.2) and (8.6.4) have simply been modified so that all summations over m and m' begin with 0 instead of 1, and we assume that the constant term $g_a^1(0; z)$ is small. See Exercise 6.*. Finally, we make the expansion

$$\bar{z}_a = \bar{z}_a(z) = \sum_{m''=0}^D g_a^3(m''; z) \quad (9.6.15)$$

and find that the polynomials g_a^3 are given by the relations

$$g_a^3(m''; z) = P_{m''} \sum_{m'=0}^D g_a^2(m'; \bar{z}(z)). \quad (9.6.16)$$

Suppose the maps \mathcal{M}_1 and \mathcal{M}_2 are symplectic. Then, in the terminology of Section 7.5, the truncated Taylor series (6.13) and (6.14) are symplectic D -jets (about the origin). It is important to remark at this juncture that the concatenation of two symplectic D -jets does not generally yield a *symplectic* D -jet if \mathcal{M}_1 has nonvanishing constant terms $g_a^1(0; z)$ so that \mathcal{M}_1 does not send the origin into itself. Correspondingly, the factorization theorems of Sections 7.6 and 7.7 generally do not apply to the D -jet (6.15) for \mathcal{M}_3 . The problem is that truncation of the Taylor expansion of a symplectic map, in this case the map \mathcal{M}_2 , generally violates the symplectic condition, and this violation can feed down to low orders in the presence of translations. See, for example, Exercise 6.*.

There is a second point that we should also recognize. Suppose that the Taylor expansion for \mathcal{M}_2 is not truncated. That is, consider letting $D \rightarrow \infty$ in (6.14) and (6.16). Then it may happen that the series (5.16) diverges. This will happen if the point $\bar{z}_a(0)$ lies outside the convergence domain of the homogeneous polynomial expansion for \mathcal{M}_2 . See Exercise 1.4.4 and Chapter 26. We conclude that translations must be handled with care.

Let \mathcal{J}_3 denote the D -jet (6.15). We expect that if the translation part of \mathcal{M}_1 is small, then \mathcal{J}_3 will be nearly symplectic. It should therefore be possible to construct a D -jet that is symplectic and near \mathcal{J}_3 in the sense that the two jets differ only by appropriate powers of the small translation terms. Indeed, this is what the method of Section 9.3 accomplishes when maps are represented in Lie form. That is, suppose the two maps \mathcal{M}_1 and \mathcal{M}_2 are written in Lie form and are concatenated using the method of Section 9.3, and suppose that the resulting map is then expanded as a Taylor series about the origin and truncated beyond terms of degree D . This resulting D -jet will be symplectic, and will be near \mathcal{J}_3 .

Given a D -jet that is nearly a symplectic jet, there are many procedures for constructing nearby jets that are symplectic. The method just described is only one such procedure. Another convenient procedure is to employ methods analogous to those used in Section 7.6 to prove the factorization theorem.

Let \mathcal{J}'_3 be the jet obtained from \mathcal{J}_3 by removing its translation part. That is, \mathcal{J}'_3 sends z to \bar{z}' according to the rule

$$\bar{z}'_a = \sum_{m=1}^D g_a^3(m; z), \quad (9.6.17)$$

and thus sends the origin into itself. Examine the matrix (linear) part of \mathcal{J}'_3 described by the terms $g_a^3(1; z)$. These terms correspond to a matrix that is nearly symplectic. Replace this matrix by a matrix that is exactly symplectic using one of the matrix symplectification methods of Chapter 4. Call this matrix R , and let \mathcal{R} be its corresponding linear symplectic map. Finally, let \mathcal{J}''_3 be the jet that results from replacing the $g_a^3(1; z)$ terms in (6.17) by $(Rz)_a$.

Next apply \mathcal{R}^{-1} to \mathcal{J}''_3 to get a result of the form

$$(\mathcal{R}^{-1} \mathcal{J}''_3 z)_a = z_a + r_a (> 1), \quad (9.6.18)$$

which is analogous to (7.6.19). As before, the remainder term $r_a(> 1)$ will have a quadratic piece and still higher degree terms,

$$r_a(> 1) = \hat{g}_a(2; z) + r_a(> 2). \quad (9.6.19)$$

Because \mathcal{J}_3 is not a symplectic jet, the quadratic piece will generally not satisfy the analog of (7.6.23). However, we may still define an f_3 by the rule

$$f_3 = -(1/3) \sum_{ab} \hat{g}_a(2; z) J_{ab} z_b, \quad (9.6.20)$$

which is analogous to (7.6.26). (Indeed, Section 17.11 shows that this prescription is unique.) From this f_3 we produce the quadratic polynomials $\tilde{g}_a(2; z)$ by the rules

$$\tilde{g}_a(2; z) =: f_3 : z_a. \quad (9.6.21)$$

Because \mathcal{J}_3 is nearly symplectic, the polynomials $\hat{g}_a(2; z)$ and $\tilde{g}_a(2; z)$ will be nearly the same. We therefore may replace $\hat{g}_a(2; z)$ by $\tilde{g}_a(2; z)$ and, in so doing, obtain a nearby map that is more nearly symplectic.

It should now be as clear to the reader as it is to the writer that the steps just described can be applied repeatedly to yield a sequence of homogeneous polynomials $f_3, f_4 \cdots f_{D+1}$. Also, by Section 7.7, there is an f_1 polynomial that will reproduce the translation part $g_a^3(0; z)$ in (6.15). Consequently, we have found the approximate result

$$\mathcal{M}_3 \simeq \mathcal{R} \exp(: f_3 :) \cdots \exp(: f_{D+1} :) \exp(: f_1 :). \quad (9.6.22)$$

Finally, the map on the right side of (6.22) may be expanded in a Taylor series and truncated beyond terms of degree D . Doing so yields a symplectic D -jet that is close to \mathcal{J}_3 .

After this pleasant digression, let us return to the subject of map concatenation. In analogy to the discussion of Section 8.6, the next topic to be treated is the case where \mathcal{M}_1 is in Lie form and \mathcal{M}_2 is in Taylor form. See Figure 8.6.2. In this case the definition (8.6.11) must be extended to become

$$T_a^D(\bar{z}) = \sum_{m'=0}^D g_a^2(m'; \bar{z}) \quad (9.6.23)$$

to include the possibility that \mathcal{M}_2 may have a translation part. The remaining relations (8.6.10) and (8.6.12) through (8.6.14) continue to hold. In particular, we still have the result

$$\bar{\bar{z}}_a(z) = \mathcal{M}_1 T_a^D(z). \quad (9.6.24)$$

Again, there are three common ways that \mathcal{M}_1 may be specified in Lie form. First suppose, as before, that \mathcal{M}_1 is given in terms of a single exponent,

$$\mathcal{M}_1 = \exp(: h :), \quad (9.6.25)$$

where h now has a homogeneous polynomial expansion of the form

$$h = h_1 + h_2 + \cdots h_{D+1}. \quad (9.6.26)$$

We still have the relation (8.6.17) and the result

$$g_a^3(m; z) = P_m \sum_{m'=0}^D \sum_{\ell=0}^{\infty} (1/\ell!) : h :^\ell g_a^2(m'; z). \quad (9.6.27)$$

As before, there are caveats about the rate at which the sum over ℓ converges. Moreover, as illustrated for a special example in Section 10.5, the sum over ℓ may also fail to converge. As in Section 10.5, this possible divergence is not due to any defect in the method of direct Taylor summation, but rather indicates that \mathcal{M}_1 may fail to exist, and shows that Hamiltonians for which $h_1 \neq 0$ must be treated with care.

Next we suppose, as a second possibility, that \mathcal{M}_1 is given in the factored product form

$$\mathcal{M}_1 = \exp(: f_1 :) \mathcal{R}_f \exp(: f_3 :) \cdots \exp(: f_{D+1} :). \quad (9.6.28)$$

Handling this possibility is straight forward. Suppose that $\exp(: f_1 :)$ has the effect

$$\exp(: f_1 :) z_a = z_a + k_a. \quad (9.6.29)$$

Then the results (8.6.20) through (8.6.25) continue to hold except that the sums over m and m' begin at 0 instead of 1, and (8.6.27) is modified to become

$$g_a^3(m; z) = P_m \sum_{m'=0}^D \tilde{g}_a[m'; R^f(z + k)]. \quad (9.6.30)$$

The third possibility is that \mathcal{M}_1 arises as a result of some symplectic integration approximation and is therefore given as a product of Lie transformations of the form

$$\mathcal{M}_1 = \exp[(w_1 h : A :) \exp(w_2 h : B :) \cdots \exp(w_m h : A :)]. \quad (9.6.31)$$

As before, B typically has a homogeneous polynomial expansion consisting of terms of degree three and higher. However, if \mathcal{M}_1 has a translation part, A will contain terms of degree one as well as a second-degree terms. In this case we make use of (2.4) or (2.62) to factorize the terms of the form $\exp(w_j h : A :)$. With this accomplished, we may proceed as before using the tools already developed.

At this point we should remark that if \mathcal{M}_1 has a translation part, then the D -jet produced for \mathcal{M}_3 in each of the three possibilities just described will again not be a symplectic D -jet, and for the same reason as before. Also, nearby symplectic D -jets can again be constructed. For example, the procedure based on the methods of the factorization theorem will work as before.

The last topic to be discussed in this section is the inversion of maps in Taylor form including the possibility of translations. When translations are included, (8.6.40) takes the form

$$\bar{z}_{b'} = k_{b'} + \sum_b R_{b'b} z_b + N_{b'}(z). \quad (9.6.32)$$

This equation can be partially solved to give the result

$$z_a = [R^{-1}(\bar{z} - k)]_a + \tilde{N}_a(z) \quad (9.6.33)$$

where \tilde{N}_a is again given by (8.6.41), and therefore contains terms only of degree 2 and higher. Now form the recursion relation

$$z_a^{(m+1)}(\bar{z}) = [R^{-1}(\bar{z} - k)]_a + \mathcal{T}^{m+1}\tilde{N}_a[z^{(m)}(\bar{z})] \quad (9.6.34)$$

with the starting relation

$$z_a^{(1)}(\bar{z}) = [R^{-1}(\bar{z} - k)]_a. \quad (9.6.35)$$

Here the translation quantities k_b are to be treated as small, and a monomial in *all* the variables z_a and k_b is regarded as having degree d if the sum of *all* the exponents in the monomial adds up to d . Correspondingly, the operator \mathcal{T}^d in (6.34) is now defined in terms of this total degree. Application of the recursion relation (6.34) D times produces a D -jet representation for the map \mathcal{M}_1^{-1} .

As the reader should expect by now, if \mathcal{M}_1 has a translation part (as we have assumed), then the D -jet for \mathcal{M}_1^{-1} obtained in this way will generally not be symplectic. But again, nearby symplectic D -jets can be constructed from this D -jet.

Finally, we remark that the concatenation and inversion methods of this section can, if desired, be employed in the formulas of Section 9.4 to compute reverse and mixed factorizations.

Exercises

9.6.1.

9.7 The Lie Algebra of the Group of all Symplectic Maps Is Simple

Section 8.9 described what it means for a Lie algebra to be simple. In this section we will show that $ispm(2n, \mathbb{R})$, the Lie algebra of the group of all symplectic maps, is simple.

Bibliography

- [1] L.M. Healy and A.J. Dragt, Concatentation of Lie Algebraic Maps, in *Lie Methods in Optics II*, K.B. Wolf, Ed., Lecture Notes in Physics **352**, Springer-Verlag (1989).

Chapter 10

Computation of Transfer Maps

Much of the material in the previous chapters dealt with the general problem of representing and manipulating symplectic maps. This chapter, along with some that follow, deals with the *computation* of transfer maps. For the most part we will deal with the symplectic case, but there are ready extensions to the general case that can be found by replacing Hamiltonian vector fields by general vector fields.

10.1 Equation of Motion

10.1.1 Background and Derivation

Let $H(z, t)$ be a general, possibly time-dependent, Hamiltonian. We know from Theorem 6.4.1 that following the flow specified by H produces a symplectic transfer map $\mathcal{M}(t)$. Let z^i denote a general initial condition. Then we have the relations

$$z(t) = \mathcal{M}(t)z^i, \quad (10.1.1)$$

$$\mathcal{M}(t^i) = \mathcal{I}. \quad (10.1.2)$$

Our goal is to find an equation of motion for \mathcal{M} .

Suppose $g(z)$ is any function of the phase-space variables z (but not explicitly of the time t). By (8.3.52) we have the relation

$$g(z) = g(\mathcal{M}z^i) = \mathcal{M}g(z^i). \quad (10.1.3)$$

Now differentiate both sides of (1.3) along the flow specified by H . We find the result

$$\dot{g}(z) = \dot{\mathcal{M}}g(z^i). \quad (10.1.4)$$

But from (1.7.4) we also have the relation

$$\dot{g}(z) = [g(z), H(z, t)]. \quad (10.1.5)$$

With the aid of (1.1), (8.3.52), and (8.3.53) this relation can be rewritten in the form

$$\begin{aligned} \dot{g}(z) &= [g(z), H(z, t)] = [g(\mathcal{M}z^i), H(\mathcal{M}z^i, t)] \\ &= [\mathcal{M}g(z^i), \mathcal{M}H(z^i, t)] = \mathcal{M}[g(z^i), H(z^i, t)] \\ &= \mathcal{M}[-H(z^i, t), g(z^i)] = \mathcal{M} : -H(z^i, t) : g(z^i). \end{aligned} \quad (10.1.6)$$

Now compare (1.4) and (1.6). Doing so gives the result

$$\dot{\mathcal{M}}g(z^i) = \mathcal{M} : -H(z^i, t) : g(z^i). \quad (10.1.7)$$

However, g is an arbitrary function. We conclude that (1.7) is equivalent to the *operator* equation of motion

$$\dot{\mathcal{M}} = \mathcal{M} : -H : . \quad (10.1.8)$$

We note that this result agrees with the result (7.4.9) that was obtained earlier for the special case of autonomous Hamiltonians.

10.1.2 Perturbation/Splitting Theory and Reverse Factorization

In some cases the Hamiltonian can be split into the sum of two terms so that it can be written in the form

$$H(z, t) = H_0(z, t) + H_1(z, t). \quad (10.1.9)$$

Often the motion governed by H_0 can be determined and the effect of H_1 may be viewed as a perturbation. Let \mathcal{M}_0 be the map produced by H_0 . That is, \mathcal{M}_0 satisfies the equation of motion

$$\dot{\mathcal{M}}_0 = \mathcal{M}_0 : -H_0 : \quad (10.1.10)$$

with the initial condition

$$\mathcal{M}_0(t^i) = \mathcal{I}. \quad (10.1.11)$$

For the \mathcal{M} produced by H let us write the representation

$$\mathcal{M} = \mathcal{M}_1 \mathcal{M}_0 \quad (10.1.12)$$

where the map \mathcal{M}_1 remains to be determined. We will call the Ansatz (1.12) a *reverse* factorization.

What is the equation of motion for \mathcal{M}_1 ? From (1.8), (1.9), and (1.12) we find the result

$$\begin{aligned} \dot{\mathcal{M}} &= \dot{\mathcal{M}}_1 \mathcal{M}_0 + \mathcal{M}_1 \dot{\mathcal{M}}_0 = \mathcal{M} : -H := \mathcal{M}_1 \mathcal{M}_0 : -H : \\ &= \mathcal{M}_1 \mathcal{M}_0 : -H_0 : + \mathcal{M}_1 \mathcal{M}_0 : -H_1 : . \end{aligned} \quad (10.1.13)$$

Use of (1.10) gives the relation

$$\mathcal{M}_1 \dot{\mathcal{M}}_0 = \mathcal{M}_1 \mathcal{M}_0 : -H_0 :, \quad (10.1.14)$$

and consequently (1.13) can be reduced to the relation

$$\dot{\mathcal{M}}_1 \mathcal{M}_0 = \mathcal{M}_1 \mathcal{M}_0 : -H_1 : \quad (10.1.15)$$

from which it follows that

$$\dot{\mathcal{M}}_1 = \mathcal{M}_1 \mathcal{M}_0 : -H_1 : \mathcal{M}_0^{-1}. \quad (10.1.16)$$

Given H_1 and \mathcal{M}_0 , let us define an *interaction* Hamiltonian H_1^{int} by the rule

$$H_1^{\text{int}}(z^i, t) = H_1(\mathcal{M}_0 z^i, t). \quad (10.1.17)$$

[We note that, because $\mathcal{M}_0(t)$ is time dependent, in general H_1^{int} will depend on time even if H_1 happens to be time independent.] Then, as a consequence of (8.2.25), we have the result

$$\mathcal{M}_0 : -H_1 : \mathcal{M}_0^{-1} =: -H_1^{\text{int}} : . \quad (10.1.18)$$

Upon combining (1.16) and (1.18) we find the equation of motion

$$\dot{\mathcal{M}}_1 = \mathcal{M}_1 : -H_1^{\text{int}} : . \quad (10.1.19)$$

Finally, we observe from (1.2) and (1.11) that \mathcal{M}_1 also has the initial condition

$$\mathcal{M}_1(t^i) = \mathcal{I}. \quad (10.1.20)$$

10.1.3 Perturbation/Splitting Theory and Forward Factorization

For the \mathcal{M} produced by H let us write, instead of (1.12), the representation

$$\mathcal{M} = \mathcal{M}_0 \mathcal{N}_1 \quad (10.1.21)$$

where the map \mathcal{N}_1 remains to be determined. We will call the Ansatz (1.21) a *forward* factorization.

Note that the \mathcal{N}_1 in (1.21) and the \mathcal{M}_1 in (1.12) are generally different. Indeed, we may rewrite (1.12) in the form

$$\mathcal{M} = \mathcal{M}_1 \mathcal{M}_0 = \mathcal{M}_0 \mathcal{M}_0^{-1} \mathcal{M}_1 \mathcal{M}_0 \quad (10.1.22)$$

from which it follows that

$$\mathcal{N}_1 = \mathcal{M}_0^{-1} \mathcal{M}_1 \mathcal{M}_0. \quad (10.1.23)$$

We see that if a reverse factorization has been found so that both \mathcal{M}_0 and \mathcal{M}_1 are known, then (1.21) and (1.23) provide the associated forward factorization.

10.2 Series (Dyson) Solution

Readers familiar with Quantum Mechanics will recognize a similarity between equations (1.8), (1.17), and (1.19) and analogous quantum mechanical results. This is to be expected because Quantum Mechanics and Classical Mechanics have closely related Lie algebraic structures. Because of this similarity, many mathematical tools originally developed for Quantum Mechanics can also be applied in Classical Mechanics. Indeed, these tools could have (and, in retrospect, should have) been developed first in the context of Classical Mechanics.

One such tool is *Neumann* iteration. Suppose both sides of (1.8) are integrated with respect to time from the initial time t^i to the variable time t . So doing, and making use of (1.2), gives the integral equation

$$\mathcal{M}(t) = \mathcal{I} + \int_{t^i}^t dt' \mathcal{M}(t') : -H(t') : . \quad (10.2.1)$$

(Here, for notational simplicity, we have suppressed the fact that H also depends on z^i .) Now iterate (2.1) by substituting the right side back into the integral. If this is done once, we obtain the result

$$\mathcal{M}(t) = \mathcal{I} + \int_{t^i}^t dt' : -H(t') : + \int_{t^i}^t dt' \int_{t^i}^{t'} dt'' \mathcal{M}(t'') : -H(t'') :: -H(t') : . \quad (10.2.2)$$

Evidently, repeated iteration gives the result

$$\mathcal{M}(t) = \mathcal{I} + \int_{t^i}^t dt' : -H(t') : + \int_{t^i}^t dt' \int_{t^i}^{t'} dt'' : -H(t'') :: -H(t') : + \dots \quad (10.2.3)$$

Note that (2.3) is in effect an expansion in powers of H , and that the factors in the integrands occur in *chronological* order – with earlier times preceding later times. We conclude that \mathcal{M} can be expressed as an infinite sum of multiple *time-ordered* integrals over the Lie operators $: -H(t) : ..$. In the context of Quantum Mechanics, the analog of the series (2.3) is the *Dyson* series. Also note that Neumann iteration is the map counterpart of Picard iteration. Compare (1.3.8) and (1.3.9) with (2.1) through (2.3).

The series (2.3) bears a certain resemblance to the exponential series. Consider m -dimensional Euclidean space. It is easily verified that the volume of the region $t^i \leq t_m \leq t_{m-1} \leq \dots \leq t_2 \leq t_1 \leq t$ is related to the volume of the region $[t^i, t] \times [t^i, t] \times [t^i, t] \dots$ (m factors) by the proportionality constant ($m!$). Indeed, we have the relation

$$\begin{aligned} \int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 \int_{t^i}^{t_2} dt_3 \dots \int_{t^i}^{t_{m-1}} dt_m &= (1/m!) \int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 \dots \int_{t^i}^{t_{m-1}} dt_m \\ &= (1/m!) \left[\int_{t^i}^t dt' \right]^m. \end{aligned} \quad (10.2.4)$$

Consequently, we may write the identity

$$\begin{aligned} &\int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 \dots \int_{t^i}^{t_{m-1}} dt_m : -H(t_m) :: -H(t_{m-1}) : \dots : -H(t_1) : \\ &= (1/m!) \int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 \dots \int_{t^i}^{t_{m-1}} dt_m T : -H(t_m) :: -H(t_{m-1}) : \dots : -H(t_1) : \\ &= (1/m!) T \left[\int_{t^i}^t dt' : -H(t') : \right]^m. \end{aligned} \quad (10.2.5)$$

Here the *time-ordering* symbol T indicates that the factors in the operator product $: -H(t_m) : \dots : -H(t_1) :$ are to be rearranged so that operators with earlier times precede those with later times. Finally, with the aid of (2.5), we may write the series (2.3) in the form

$$\mathcal{M}(t) = \mathcal{I} + T \sum_{m=1}^{\infty} (1/m!) \left[\int_{t^i}^t dt' : -H(t') : \right]^m = T \exp \left[\int_{t^i}^t dt' : -H(t') : \right]. \quad (10.2.6)$$

The right side of (2.6) is often called a *time-ordered exponential*. However, as neat as this expression may appear to be, in reality it is simply the series (2.3).

We close this section by noting that our discussion of the series solution to (1.8) applies equally well to (1.19). Consequently, \mathcal{M}_1 has the series solution

$$\mathcal{M}_1(t) = T \exp \left[\int_{t^i}^t dt' : -H_1^{\text{int}}(t') : \right]. \quad (10.2.7)$$

This result for \mathcal{M}_1 may be viewed as an expansion in powers of H_1^{int} .

Exercises

10.2.1. Verify (2.4) by performing the indicated integrations on each side.

10.2.2. Suppose $F(z, t)$ and $G(z, t)$ are two Hamiltonians that are in *involution*. That is, we assume

$$[F(z, t), G(z, t')] = 0 \text{ for all } t, t'. \quad (10.2.8)$$

Correspondingly, there will be the Lie operator commutation relation

$$\{ : F(z, t) :, : G(z, t') : \} = 0 \text{ for all } t, t'. \quad (10.2.9)$$

Let $\mathcal{M}_F(t^{\text{in}}, t^{\text{fin}})$ and $\mathcal{M}_G(t^{\text{in}}, t^{\text{fin}})$ be the maps generated by F and G , respectively. Define a sum Hamiltonian $H(z, t)$ by writing

$$H(z, t) = F(z, t) + G(z, t), \quad (10.2.10)$$

and let $\mathcal{M}_H(t^{\text{in}}, t^{\text{fin}})$ be the map generated by H . Your task is to show that

$$\mathcal{M}_H(t^{\text{in}}, t^{\text{fin}}) = \mathcal{M}_F(t^{\text{in}}, t^{\text{fin}})\mathcal{M}_G(t^{\text{in}}, t^{\text{fin}}) = \mathcal{M}_G(t^{\text{in}}, t^{\text{fin}})\mathcal{M}_F(t^{\text{in}}, t^{\text{fin}}). \quad (10.2.11)$$

Begin by making the Ansatz

$$\mathcal{M}_H(t^{\text{in}}, t) = \mathcal{M}_?(t^{\text{in}}, t)\mathcal{M}_G(t^{\text{in}}, t) \quad (10.2.12)$$

where the map $\mathcal{M}_?(t^{\text{in}}, t)$ remains to be determined. Verify that taking the time derivative of both sides of (2.12) produces the result

$$\begin{aligned} \dot{\mathcal{M}}_H &= \dot{\mathcal{M}}_?\mathcal{M}_G + \mathcal{M}_?\dot{\mathcal{M}}_G = \mathcal{M}_H : -H := \mathcal{M}_?\mathcal{M}_G : -H : \\ &= \mathcal{M}_?\mathcal{M}_G : -F : + \mathcal{M}_?\mathcal{M}_G : -G : . \end{aligned} \quad (10.2.13)$$

Next find the equation of motion for $\mathcal{M}_?$. To do so, verify that

$$\mathcal{M}_?\dot{\mathcal{M}}_G = \mathcal{M}_?\mathcal{M}_G : -G :, \quad (10.2.14)$$

and consequently (2.13) can be reduced to the relation

$$\dot{\mathcal{M}}_?\mathcal{M}_G = \mathcal{M}_?\mathcal{M}_G : -F : . \quad (10.2.15)$$

At this point imagine that the series solution (2.3) is used to represent \mathcal{M}_G . Verify that doing so gives the result

$$\mathcal{M}_G(t^{\text{in}}, t) = \mathcal{I} + \int_{t^{\text{in}}}^t dt' : -G(t') : + \int_{t^{\text{in}}}^t dt' \int_{t^{\text{in}}}^{t'} dt'' : -G(t'') :: -G(t') : + \cdots . \quad (10.2.16)$$

Employ the assumption (2.9) and the representation (2.16) to conclude that

$$\mathcal{M}_G : -F :=: -F : \mathcal{M}_G, \quad (10.2.17)$$

from which it follows that (2.15) can be rewritten in the form

$$\dot{\mathcal{M}}_? \mathcal{M}_G = \mathcal{M}_? : -F : \mathcal{M}_G, \quad (10.2.18)$$

and consequently

$$\dot{\mathcal{M}}_? = \mathcal{M}_? : -F : . \quad (10.2.19)$$

Finally, observe from (2.12) that $\mathcal{M}_?$ has the initial condition

$$\mathcal{M}_?(t^{\text{in}}, t^{\text{in}}) = \mathcal{I}. \quad (10.2.20)$$

It follows from (2.19) and (2.20) that

$$\mathcal{M}_?(t^{\text{in}}, t^{\text{fin}}) = \mathcal{M}_F(t^{\text{in}}, t^{\text{fin}}), \quad (10.2.21)$$

and insertion of (2.21) into the Ansatz (2.12) proves the first part of (2.11).

In an analogous way, prove the second part of (2.11) by making the Ansatz

$$\mathcal{M}_H(t^{\text{in}}, t) = \mathcal{M}_?(t^{\text{in}}, t) \mathcal{M}_F(t^{\text{in}}, t). \quad (10.2.22)$$

10.3 Exponential (Magnus) Solution

The series solutions (2.6) and (2.7) for the transfer map, or equivalently (2.3), have the defect that they are not manifestly symplectic. Indeed, if the series are truncated, the resulting maps are generally not symplectic. Moreover, maps in series form are somewhat difficult to concatenate. For these reasons, it is also useful to have solutions in exponential form. Possibilities include the single exponential form and various factored product forms. Here we consider the single exponential form.

Let us seek a solution to the equation of motion (1.8), or (1.19), of the form

$$\mathcal{M}(t) = \exp(: F(z^i, t) :). \quad (10.3.1)$$

We know that in general there is no such solution. Indeed, even in the simplest linear case of $Sp(2)$, maps cannot generally be written in single exponent form. See Section 8.7. Nevertheless we will pursue the assumption (3.1) to see where it leads. We will find an expansion for F in terms of powers of H , and the convergence of this expansion will determine the validity of the assumption.

Let us differentiate both sides of (3.1). Doing so gives the relation

$$\dot{\mathcal{M}} = \exp(:F:) \text{iex}(-\#F\#) : \dot{F} := \mathcal{M} \text{iex}(-\#F\#) : \dot{F} : . \quad (10.3.2)$$

Here we have used results from Appendix C. Now insert the equation of motion (1.8) into (3.2) to get the relation

$$: -H := \text{iex}(-\#F\#) : \dot{F} : . \quad (10.3.3)$$

This relation can be solved to produce an equation of motion for the Lie operator $:F:$,

$$: \dot{F} := [\text{iex}(-\#F\#)]^{-1} : -H : . \quad (10.3.4)$$

According to Appendix C, the operator $[\text{iex}(-\#F\#)]^{-1}$ has an expansion in powers of $\#F\#$ of the form

$$[\text{iex}(-\#F\#)]^{-1} = \sum_{m=0}^{\infty} b_m (\#F\#)^m. \quad (10.3.5)$$

Consequently, (3.4) can be rewritten in the form

$$: \dot{F} := \sum_{m=0}^{\infty} b_m (\#F\#)^m : -H : = : \left[\sum_{m=0}^{\infty} b_m : F :^m (-H) \right] : . \quad (10.3.6)$$

Here we have also used (8.2.2). Now remove the outside colons from both sides of (3.6). So doing gives an equation of motion for F ,

$$\dot{F} = \sum_{m=0}^{\infty} b_m : F :^m (-H) = [\text{iex}(-:F:)]^{-1} (-H), \quad (10.3.7)$$

which could have been deduced directly by “decolonizing” (3.4). This equation is to be solved with the initial condition

$$F(t^i) = 0, \quad (10.3.8)$$

which follows from (1.2).

Let us try to solve (3.7) by perturbation theory. Replace H by ϵH , and assume F has an expansion of the form

$$F = \sum_{n=1}^{\infty} \epsilon^n F_n. \quad (10.3.9)$$

(This procedure is equivalent to the introduction of a grading, and the expansion obtained is often called the *Magnus* expansion. See Section 8.9.) Put the expansion (3.9) into (3.7) to get the result

$$\sum_{n=1}^{\infty} \epsilon^n \dot{F}_n = \sum_{m=0}^{\infty} b_m \left(\sum_{n=1}^{\infty} \epsilon^n F_n : \right)^m (-\epsilon H). \quad (10.3.10)$$

Now equate powers of ϵ . So doing gives the results

$$\dot{F}_1 = -H, \quad (10.3.11)$$

$$\dot{F}_2 = (1/2) : F_1 : (-H), \quad (10.3.12)$$

$$\dot{F}_3 = (1/12) : F_1 :^2 (-H) + (1/2) : F_2 : (-H), \quad (10.3.13)$$

$$\dot{F}_4 = (1/2) : F_3 : (-H) + (1/12) (: F_1 :: F_2 : + : F_2 :: F_1 :)(-H), \quad (10.3.14)$$

$$\dot{F}_n = \text{something involving } H \text{ and the } : F_m : \text{ with } m < n. \quad (10.3.15)$$

Here we have used the values for the coefficients b_m given in Appendix C. The equations (3.11) through (3.15) are to be solved with the initial conditions

$$F_n(t^i) = 0. \quad (10.3.16)$$

The equations for the \dot{F}_n can be integrated numerically or solved by quadrature. Evidently (3.11) with the initial condition (3.16) has the solution

$$F_1(t) = \int_{t^i}^t dt_1 [-H(t_1)]. \quad (10.3.17)$$

Now substitute (3.17) into (3.12) to get the result

$$\begin{aligned} \dot{F}_2(t_1) &= (1/2) \int_{t^i}^{t_1} dt_2 : -H(t_2) : [-H(t_1)] \\ &= (1/2) \int_{t^i}^{t_1} dt_2 [-H(t_2), -H(t_1)]. \end{aligned} \quad (10.3.18)$$

This equation can be integrated, again with the initial condition (3.16), to give the result

$$F_2(t) = (1/2) \int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 [-H(t_2), -H(t_1)]. \quad (10.3.19)$$

Let us introduce the short-hand notation $-j$ for the quantity $-H(t_j)$. With this notation, (3.19) can be written in the more compact form

$$F_2(t) = (1/2) \int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 [-2, -1]. \quad (10.3.20)$$

Similarly, again using this notation, we find the result

$$\begin{aligned} F_3(t) &= (1/6) \int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 \int_{t^i}^{t_2} dt_3 \times \\ &\quad \{2[-3, [-2, -1]] - [-2, [-3, -1]]\} \\ &= (1/6) \int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 \int_{t^i}^{t_2} dt_3 \times \\ &\quad \{[-1, [-2, -3]] + [-3, [-2, -1]]\}. \end{aligned} \quad (10.3.21)$$

Here $2[-3, [-2, -1]]$ is short hand for $2[-H(t_3), [-H(t_2), -H(t_1)]]$, etc. And for F_4 we find the result

$$\begin{aligned} F_4(t) &= (1/12) \int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 \int_{t^i}^{t_2} dt_3 \int_{t^i}^{t_3} dt_4 \times \\ &\quad \{+3[-4, [-3, [-2, -1]]] - [-4, [-2, [-3, -1]]] \\ &\quad - [-3, [-4, [-2, -1]]] - [-3, [-2, [-4, -1]]] \\ &\quad - [-2, [-4, [-3, -1]]] + [-2, [-3, [-4, -1]]]\}. \end{aligned} \quad (10.3.22)$$

See Exercise 3.3.

At this point at least three comments are in order. First, we recognize from the structure of the equations (3.11) through (3.15) that the quantity ϵ serves only as a “counting” parameter that counts powers of H , and therefore we may now set $\epsilon = 1$. The net result is an expansion of F in powers of H . Next we see from (3.15) that the F_n can be determined successively, and consequently a solution by numerical methods or by quadrature is always possible. Third, we see that all the quantities on the right side of the equations (3.11) through (3.15) lie in the Lie algebra generated by the $H(z^i, t)$ for different values of t . (That is, all the quantities consist of H and Poisson brackets involving only factors of H .) In particular, if all the $H(z^i, t)$ are in *involution*,

$$[H(z^i, t), H(z^j, t')] = 0, \quad (10.3.23)$$

then we have the results

$$F_n = 0 \text{ for } n > 1, \quad (10.3.24)$$

and

$$\mathcal{M}(t) = \exp \left[\int_{t^i}^t dt' : -H(t') : \right]. \quad (10.3.25)$$

Note that (3.25) is consistent with (2.6) because if (3.23) holds, then time ordering makes no difference. Finally, if the Hamiltonian is autonomous, then the integral in (3.25) can be done to give the result (7.4.7).

We close this section with the reminder that the discussion we have just given for the solution of (1.8) for \mathcal{M} applies equally well to the solution of (1.19) for \mathcal{M}_1 . In the case of \mathcal{M}_1 , the associated exponential quantity F can be developed as an expansion in powers of H_1^{int} . And, since H_1^{int} may be small, it could be the case that \mathcal{M}_1 can be written in single exponent form.

Exercises

10.3.1. Verify (3.11) through (3.15).

10.3.2. Verify (3.17) through (3.20).

10.3.3. Verify (3.21) through (3.22). Hint: To do so, use the identity (5.3.14) and the Jacobi identity; and verify and employ integral identities of the form

$$\int_{t^i}^t dt_1 \int_{t^i}^t dt_2 ** = \int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 ** + \int_{t^i}^t dt_2 \int_{t^i}^{t_2} dt_1 **, \quad (10.3.26)$$

where $**$ denotes some common integrand.

10.4 Factored Product Solution: Powers of H Expansion

As before, let us replace H by ϵH so that the equation of motion (1.8) becomes

$$\dot{\mathcal{M}} = \mathcal{M} : -\epsilon H : . \quad (10.4.1)$$

Suppose also that we factor \mathcal{M} in the form

$$\mathcal{M} = \mathcal{M}_1 \mathcal{M}_2 \mathcal{M}_3 \mathcal{M}_4 \dots \quad (10.4.2)$$

where

$$\mathcal{M}_m = \exp(:\epsilon^m G_m:) \quad (10.4.3)$$

and the functions G_m are to be determined. [Note that (4.2) is a forward factorization.] Next differentiate the Ansatz (4.2) to find the result

$$\begin{aligned} \dot{\mathcal{M}} = & \dot{\mathcal{M}}_1 \mathcal{M}_2 \mathcal{M}_3 \mathcal{M}_4 \dots + \mathcal{M}_1 \dot{\mathcal{M}}_2 \mathcal{M}_3 \mathcal{M}_4 \dots \\ & + \mathcal{M}_1 \mathcal{M}_2 \dot{\mathcal{M}}_3 \mathcal{M}_4 \dots + \mathcal{M}_1 \mathcal{M}_2 \mathcal{M}_3 \dot{\mathcal{M}}_4 \dots \\ & + \dots, \end{aligned} \quad (10.4.4)$$

from which it follows using (4.1) and (4.2) that

$$\begin{aligned} \mathcal{M}^{-1} \dot{\mathcal{M}} = & \dots \mathcal{M}_4^{-1} \mathcal{M}_3^{-1} \mathcal{M}_2^{-1} \mathcal{M}_1^{-1} \dot{\mathcal{M}}_1 \mathcal{M}_2 \mathcal{M}_3 \mathcal{M}_4 \dots \\ & + \dots \mathcal{M}_4^{-1} \mathcal{M}_3^{-1} \mathcal{M}_2^{-1} \dot{\mathcal{M}}_2 \mathcal{M}_3 \mathcal{M}_4 \dots \\ & + \dots \mathcal{M}_4^{-1} \mathcal{M}_3^{-1} \dot{\mathcal{M}}_3 \mathcal{M}_4 \dots \\ & + \dots \mathcal{M}_4^{-1} \dot{\mathcal{M}}_4 \dots \\ & + \dots = : -\epsilon H : . \end{aligned} \quad (10.4.5)$$

The various terms in (4.5) can be simplified by the use of adjoint operators. For example, we have the result

$$\begin{aligned} \dots \mathcal{M}_4^{-1} \mathcal{M}_3^{-1} \mathcal{M}_2^{-1} \mathcal{M}_1^{-1} \dot{\mathcal{M}}_1 \mathcal{M}_2 \mathcal{M}_3 \mathcal{M}_4 \dots = \\ \dots \exp(-\#\epsilon^4 G_4 \#) \exp(-\#\epsilon^3 G_3 \#) \exp(-\#\epsilon^2 G_2 \#) \mathcal{M}_1^{-1} \dot{\mathcal{M}}_1. \end{aligned} \quad (10.4.6)$$

Also, there are relations of the form

$$\mathcal{M}_m^{-1} \dot{\mathcal{M}}_m = \text{iex}(-\#\epsilon^m G_m \#) : \epsilon^m \dot{G}_m : . \quad (10.4.7)$$

Upon using (4.7) and relations of the form (4.6) we find that (4.5) can be rewritten in the form

$$\begin{aligned} \dots \exp(-\#\epsilon^4 G_4 \#) \exp(-\#\epsilon^3 G_3 \#) \exp(-\#\epsilon^2 G_2 \#) \text{iex}(-\#\epsilon G_1 \#) : \epsilon \dot{G}_1 : \\ + \dots \exp(-\#\epsilon^4 G_4 \#) \exp(-\#\epsilon^3 G_3 \#) \text{iex}(-\#\epsilon^2 G_2 \#) : \epsilon^2 \dot{G}_2 : \\ + \dots \exp(-\#\epsilon^4 G_4 \#) \text{iex}(-\#\epsilon^3 G_3 \#) : \epsilon^3 \dot{G}_3 : \\ + \dots \text{iex}(-\#\epsilon^4 G_4 \#) : \epsilon^4 \dot{G}_4 : + \dots = : -\epsilon H : . \end{aligned} \quad (10.4.8)$$

The colons can now be removed from both sides of (4.8) to give the equivalent result

$$\begin{aligned} \dots \exp(-:\epsilon^4 G_4:) \exp(-:\epsilon^3 G_3:) \exp(-:\epsilon^2 G_2:) \text{iex}(-:\epsilon G_1:) \epsilon \dot{G}_1 \\ + \dots \exp(-:\epsilon^4 G_4:) \exp(-:\epsilon^3 G_3:) \text{iex}(-:\epsilon^2 G_2:) \epsilon^2 \dot{G}_2 \\ + \dots \exp(-:\epsilon^4 G_4:) \text{iex}(-:\epsilon^3 G_3:) \epsilon^3 \dot{G}_3 \\ + \dots \text{iex}(-:\epsilon^4 G_4:) \epsilon^4 \dot{G}_4 + \dots = -\epsilon H. \end{aligned} \quad (10.4.9)$$

Recall that the integrated exponential function has the expansion.

$$\begin{aligned} \text{iex}(w) &= \int_0^1 d\tau \exp(\tau w) = (e^w - 1)/w = \sum_{m=0}^{\infty} w^m / (m+1)! \\ &= 1 + w/2 + w^2/3 + w^3/4 + w^4/5 + \dots . \end{aligned} \quad (10.4.10)$$

Using this expansion we may expand each of the lines in (4.9) as a Taylor series in ϵ . We find, for example, the results

$$\begin{aligned} \dots \exp(- : \epsilon^4 G_4 :) \exp(- : \epsilon^3 G_3 :) \exp(- : \epsilon^2 G_2 :) \text{iex}(- : \epsilon G_1 :) \epsilon \dot{G}_1 = \\ \epsilon \dot{G}_1 + \epsilon^2 [-(1/2) : G_1 : \dot{G}_1] + \epsilon^3 [*] + \epsilon^4 [*] + \dots , \end{aligned} \quad (10.4.11)$$

$$\begin{aligned} \dots \exp(- : \epsilon^4 G_4 :) \exp(- : \epsilon^3 G_3 :) \text{iex}(- : \epsilon^2 G_2 :) \epsilon^2 \dot{G}_2 = \\ \epsilon^2 \dot{G}_2 + \epsilon^3 [*] + \epsilon^4 [*] + \dots , \end{aligned} \quad (10.4.12)$$

$$\begin{aligned} \dots \exp(- : \epsilon^4 G_4 :) \text{iex}(- : \epsilon^3 G_3 :) \epsilon^3 \dot{G}_3 = \\ \epsilon^3 \dot{G}_3 + \epsilon^4 [*] + \dots , \end{aligned} \quad (10.4.13)$$

$$\begin{aligned} \dots \text{iex}(- : \epsilon^4 G_4 :) \epsilon^4 \dot{G}_4 = \\ \epsilon^4 \dot{G}_4 + \dots . \end{aligned} \quad (10.4.14)$$

Next equate powers of ϵ on both sides of (4.9). So doing gives, for example through powers of ϵ^4 , the results

$$\epsilon \dot{G}_1 = -\epsilon H, \quad (10.4.15)$$

$$\epsilon^2 [\dot{G}_2 - (1/2) : G_1 : \dot{G}_1] = 0, \quad (10.4.16)$$

$$\epsilon^3 [\dot{G}_3+] = 0, \quad (10.4.17)$$

$$\epsilon^4 [\dot{G}_4+] = 0. \quad (10.4.18)$$

We conclude that the G_n obey the equations of motion

$$\dot{G}_1 = -H, \quad (10.4.19)$$

$$\dot{G}_2 = (1/2) : G_1 : \dot{G}_1 = (1/2) : G_1 : (-H), \quad (10.4.20)$$

$$\dot{G}_3 = ?(1/12) : G_1 :^2 (-H) + (1/2) : G_2 : (-H), \quad (10.4.21)$$

$$\dot{G}_4 = ?(1/2) : G_3 : (-H) + (1/12) (: G_1 :: G_2 : + : G_2 :: G_1 :)(-H), \dots \quad (10.4.22)$$

$$\dot{G}_n = \text{something involving } H \text{ and the } : G_m : \text{ with } m < n. \quad (10.4.23)$$

These equations of motion are to be solved with the initial conditions

$$G_n(t^i) = 0. \quad (10.4.24)$$

The equations for the \dot{G}_n can be integrated numerically or solved by quadrature. Evidently (4.19) with the initial condition (4.24) has the solution

$$G_1(t) = \int_{t^i}^t dt_1 [-H(t_1)]. \quad (10.4.25)$$

Now substitute (4.25) into (4.20) to get the result

$$\begin{aligned} \dot{G}_2(t_1) &= (1/2) \int_{t^i}^{t_1} dt_2 : -H(t_2) : [-H(t_1)] \\ &= (1/2) \int_{t^i}^{t_1} dt_2 [-H(t_2), -H(t_1)]. \end{aligned} \quad (10.4.26)$$

This equation can be integrated, again with the initial condition (4.24), to give the result

$$G_2(t) = (1/2) \int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 [-H(t_2), -H(t_1)]. \quad (10.4.27)$$

Let us introduce the short-hand notation $-j$ for the quantity $-H(t_j)$. With this notation, (4.27) can be written in the more compact form

$$G_2(t) = (1/2) \int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 [-2, -1]. \quad (10.4.28)$$

Similarly, again using this notation, we find the result

$$\begin{aligned} G_3(t) &=? (1/6) \int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 \int_{t^i}^{t_2} dt_3 \times \\ &\quad \{2[-3, [-2, -1]] - [-2, [-3, -1]]\} \\ &= (1/6) \int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 \int_{t^i}^{t_2} dt_3 \times \\ &\quad \{[-1, [-2, -3]] + [-3, [-2, -1]]\}. \end{aligned} \quad (10.4.29)$$

Here $2[-3, [-2, -1]]$ is short hand for $2[-H(t_3), [-H(t_2), -H(t_1)]]$, etc. And for G_4 we find the result

$$\begin{aligned} G_4(t) &=? (1/12) \int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 \int_{t^i}^{t_2} dt_3 \int_{t^i}^{t_3} dt_4 \times \\ &\quad \{+3[-4, [-3, [-2, -1]]] - [-4, [-2, [-3, -1]]] \\ &\quad - [-3, [-4, [-2, -1]]] - [-3, [-2, [-4, -1]]] \\ &\quad - [-2, [-4, [-3, -1]]] + [-2, [-3, [-4, -1]]]\}. \end{aligned} \quad (10.4.30)$$

See Exercise 4.1.

At this point at least three comments are in order. First, we recognize from the structure of the equations (4.19) through (4.23) that the quantity ϵ serves only as a “counting” parameter that counts powers of H , and therefore we may now set $\epsilon = 1$. Next we see from (4.23) that the G_n can be determined successively, and consequently a solution by

numerical methods or by quadrature is always possible. Third, we see that all the quantities on the right side of the equations (4.19) through (4.23) lie in the Lie algebra generated by the $H(z^i, t)$ for different values of t . (That is, all the quantities consist of H and Poisson brackets involving only factors of H .) In particular, if all the $H(z^i, t)$ are in *involution*,

$$[H(z^i, t), H(z^j, t')] = 0, \quad (10.4.31)$$

then we have the results

$$G_n = 0 \text{ for } n > 1, \quad (10.4.32)$$

and we again find the result

$$\mathcal{M}(t) = \exp \left[\int_{t^i}^t dt' : -H(t') : \right]. \quad (10.4.33)$$

Finally, if the Hamiltonian is autonomous, then the integral in (4.33) can be done to give the result (7.4.7).

We close this section with the reminder that the discussion we have just given for the solution of (1.8) for \mathcal{M} applies equally well to the solution of (1.19) for \mathcal{M}_1 . In the case of \mathcal{M}_1 , the associated quantities G_n can be developed as expansions in powers of H_1^{int} . Note that in this context we have put ourselves in a notationally awkward position: The \mathcal{M}_1 appearing in (1.12) is different from and should not be confused with the \mathcal{M}_1 appearing in (4.2).

Exercises

10.4.1. Suppose that \mathcal{M} obeys the equation of motion (4.1) and that \mathcal{M} is factored in the *reversed* product form

$$\mathcal{M} = \cdots \mathcal{M}_4 \mathcal{M}_3 \mathcal{M}_2 \mathcal{M}_1 \quad (10.4.34)$$

where

$$\mathcal{M}_m = \exp(: \epsilon^m G_m^{\text{rev}} :), \quad (10.4.35)$$

and the functions G_m^{rev} are to be determined. Find equations of motion for these functions.

10.5 Factored Product Solution: Taylor Expansion about Design Orbit

10.5.1 Background

The discussion so far has been quite general in that no particular use has been made of Taylor expansions in the phase-space variables z . In this section we will explore the use of factored product representations such as those described in Sections 7.6 and 7.8.

Let $H(z, t)$ be a general, possibly time-dependent, Hamiltonian. Suppose that $z^d(t)$ is some given trajectory (which is assumed to be known and will be called the *design* trajectory), and that our task is to characterize all trajectories near z^d . Introduce $2n$ new *deviation* variables ζ by the rule

$$z = z^d + \zeta. \quad (10.5.1)$$

The transformation (5.1) is canonical. Consequently, the time evolution of the deviation variables ζ will also be described by some Hamiltonian. Call this Hamiltonian $H^{\text{new}}(\zeta, t)$. Evidently, the problem of studying trajectories near z^d is equivalent to studying the trajectories governed by $H^{\text{new}}(\zeta, t)$ in the case where ζ is small.

What is the relation between $H(z, t)$ and $H^{\text{new}}(\zeta, t)$? Define a function $\bar{H}(\zeta, t)$ by the rule

$$\bar{H}(\zeta, t) = H[z^d(t) + \zeta, t]. \quad (10.5.2)$$

Here the time dependence of $\bar{H}(\zeta, t)$ arises both from the possible time dependence of H and the time dependence of the design orbit $z^d(t)$. Next suppose that the quantity $\bar{H}(\zeta, t)$ is expressed as a power series in ζ by making the expansion

$$\bar{H}(\zeta, t) = \sum_{m=0}^{\infty} \bar{H}_m(\zeta, t). \quad (10.5.3)$$

Here each quantity $\bar{H}_m(\zeta, t)$ is a homogeneous polynomial of degree m in the components of ζ . Then we claim that $H^{\text{new}}(\zeta, t)$ is given by the relation

$$H^{\text{new}}(\zeta, t) = \sum_{m=2}^{\infty} \bar{H}_m(\zeta, t). \quad (10.5.4)$$

There are at least two ways to verify the truth of (5.4). In analogy with (4.8.2), let us write

$$z^d = (\beta_1 \cdots \beta_n, \alpha_1 \cdots \alpha_n), \quad (10.5.5)$$

$$\zeta = (Q_1 \cdots Q_n, P_1 \cdots P_n). \quad (10.5.6)$$

Introduce the mixed-variable generating function $F_2(q, P, t)$ by the rule

$$F_2(q, P, t) = \sum_{\ell=1}^n [\alpha_\ell(t)q_\ell - \beta_\ell(t)P_\ell + q_\ell P_\ell]. \quad (10.5.7)$$

Then, following the rules (4.8.4) and (4.8.5), we find the results

$$Q_\ell = \partial F_2 / \partial P_\ell = q_\ell - \beta_\ell(t), \quad (10.5.8)$$

$$p_\ell = \partial F_2 / \partial q_\ell = P_\ell + \alpha_\ell(t), \quad (10.5.9)$$

which are equivalent to the relation (5.1). Also, following the standard rules, the transformed Hamiltonian produced by the symplectic (canonical) transformation associated with (5.7) is given by the relation

$$H^{\text{new}}(Q, P, t) = H^{\text{new}}(\zeta, t) = H^{\text{old}}(z^d + \zeta, t) + \partial F_2 / \partial t = \bar{H}(\zeta, t) + \partial F_2 / \partial t. \quad (10.5.10)$$

But from (5.7) we find the result

$$\partial F_2 / \partial t = \sum_{\ell=1}^n [\dot{\alpha}_\ell(t)q_\ell - \dot{\beta}_\ell(t)P_\ell] = \sum_{\ell=1}^n [\dot{\alpha}_\ell(t)\beta_\ell(t) + \dot{\alpha}_\ell(t)Q_\ell - \dot{\beta}_\ell P_\ell] \quad (10.5.11)$$

so that

$$H^{\text{new}}(Q, P, t) = H^{\text{new}}(\zeta, t) = \bar{H}(\zeta, t) + \sum_{\ell=1}^n [\dot{\alpha}_\ell(t)\beta_\ell(t) + \dot{\alpha}_\ell(t)Q_\ell - \dot{\beta}_\ell P_\ell]. \quad (10.5.12)$$

We see that the second term on the far right of (5.12), the $\partial F_2 / \partial t$ component of H^{new} , consists only of terms independent of ζ and terms linear in ζ . Consequently, consistent with the claim made in (5.2) through (5.4), the quadratic and higher-order terms in ζ that appear in the expansions of $\bar{H}(\zeta, t)$ and $H^{\text{new}}(\zeta, t)$ agree. Also, again consistent with (5.4), we know that $H^{\text{new}}(\zeta, t)$ cannot contain terms linear in ζ , for otherwise $\zeta = 0$ would not be a possible trajectory for the equations of motion generated by H^{new} . Finally, terms in H^{new} independent of ζ make no contribution to the equations of motion and, consistent with (5.4), can simply be dropped.

A second way to verify the truth of (5.4) is simply to examine equations of motion. According to (5.2.3) a general trajectory satisfies the equation of motion

$$\dot{z} = J\partial_z H(z, t), \quad (10.5.13)$$

and the given trajectory z^d satisfies the equation of motion

$$\dot{z}^d = J\partial_z H(z, t)|_{z=z^d}. \quad (10.5.14)$$

Also, as is easily verified from (5.2) and (5.3), we have the result

$$\partial_z H(z, t)|_{z=z^d} = \partial_\zeta \bar{H}_1(\zeta, t). \quad (10.5.15)$$

Now let us insert (5.1) through (5.3) into (5.13) to get the relation

$$\dot{z} = \dot{z}^d + \dot{\zeta} = J\partial_z H(z, t) = J\partial_\zeta H(z^d + \zeta, t) = J\partial_\zeta \bar{H}(\zeta, t) = J\partial_\zeta \sum_{m=1}^{\infty} \bar{H}_m(\zeta, t). \quad (10.5.16)$$

Finally, subtract (5.14) from (5.16) and use (5.15) and (5.4) to find the result

$$\dot{\zeta} = J\partial_\zeta \sum_{m=1}^{\infty} \bar{H}_m(\zeta, t) - J\partial_\zeta \bar{H}_1(\zeta, t) = J\partial_\zeta \sum_{m=2}^{\infty} \bar{H}_m(\zeta, t) = J\partial_\zeta H^{\text{new}}(\zeta, t). \quad (10.5.17)$$

We see that the time evolution of the deviation variables ζ is indeed governed by H^{new} , as claimed. We also note that, according to (5.14) and (5.15), the equation of motion for the design trajectory itself is provided by the relation

$$\dot{z}^d = J\partial_\zeta \bar{H}_1(\zeta, t). \quad (10.5.18)$$

We close this subsection by observing that there is a variant definition of H^{new} that is often convenient. The relation (5.4) can be rewritten in the form

$$H^{\text{new}}(\zeta, t) = \sum_{m=2}^{\infty} \bar{H}_m(\zeta, t) = \bar{H}(\zeta, t) - \bar{H}_0(\zeta, t) - \bar{H}_1(\zeta, t). \quad (10.5.19)$$

Since term $\bar{H}_0(\zeta, t)$ is independent of ζ , it makes no contribution to the equations of motion and therefore may be dropped. Consequently we may make the alternate definition

$$H^{\text{new}}(\zeta, t) = \bar{H}(\zeta, t) - \bar{H}_1(\zeta, t). \quad (10.5.20)$$

Note that all the definitions (5.12), (5.19), and (5.20) are in closed form and do not actually involve the summation of infinite series.

10.5.2 Term by Term Procedure

To continue the general discussion, let us write H^{new} as given by (5.4) in the form

$$H^{\text{new}} = H_2 + H_3 + H_4 + \dots = H_2 + H_r. \quad (10.5.21)$$

Alternatively, we may expand H^{new} as given by (5.10) or (5.20) in homogeneous polynomials (in the components of ζ) and omit any irrelevant H_0 term to obtain the same result. Let \mathcal{M} be the transfer map associated with H^{new} . In accord with the spirit of (1.9) and (1.12), let us factor \mathcal{M} in the form

$$\mathcal{M} = \mathcal{M}_r \mathcal{M}_2. \quad (10.5.22)$$

Here, as in (5.21), we use the subscript “ r ” to denote “remaining” terms. [Note that (5.22) is a reverse factorization.] Following the discussion of Section 10.1, we will require that \mathcal{M}_2 obey the equation of motion

$$\dot{\mathcal{M}}_2 = \mathcal{M}_2 : -H_2 : . \quad (10.5.23)$$

Correspondingly, \mathcal{M}_r will obey the equation of motion

$$\dot{\mathcal{M}}_r = \mathcal{M}_r : -H_r^{\text{int}} : , \quad (10.5.24)$$

where the interaction Hamiltonian H_r^{int} is given by rule

$$H_r^{\text{int}}(\zeta^i, t) = H_r(\mathcal{M}_2 \zeta^i, t). \quad (10.5.25)$$

We will now describe how to compute \mathcal{M}_2 and \mathcal{M}_r . Let us begin with \mathcal{M}_2 . Since H_2 is a quadratic Hamiltonian, its associated transfer map \mathcal{M}_2 must be linear. Let $\bar{\zeta}(t)$ be the result of $\mathcal{M}_2(t)$ acting on ζ^i . Then there is a symplectic matrix R such that

$$\bar{\zeta}_a(t) = \mathcal{M}_2(t) \zeta_a^i = \sum_b R_{ab}(t) \zeta_b^i, \quad (10.5.26)$$

or, in more compact vector and matrix notation,

$$\bar{\zeta}(t) = R(t) \zeta^i. \quad (10.5.27)$$

Thus, the computation of \mathcal{M}_2 is equivalent to finding the matrix R . Since H_2 is quadratic, there is an associated symmetric matrix $S(t)$ such that H_2 is given by the relation

$$H_2(\zeta^i, t) = (1/2) \sum_{a,b} S_{ab}(t) \zeta_a^i \zeta_b^i. \quad (10.5.28)$$

In analogy with (8.3.64) and (8.3.66), we have the result

$$: -H_2 : \zeta^i = JS\zeta^i. \quad (10.5.29)$$

Suppose both sides of (5.23) are applied to the quantity ζ^i . For the left side we find the result

$$\dot{\mathcal{M}}_2 \zeta^i = \dot{\bar{\zeta}} = \dot{R} \zeta^i. \quad (10.5.30)$$

For the right side we find the result

$$\mathcal{M}_2 : -H_2 : \zeta^i = \mathcal{M}_2 JS\zeta^i = JS\mathcal{M}_2\zeta^i = JSR\zeta^i. \quad (10.5.31)$$

Now compare the right sides of (5.30) and (5.31). Since ζ^i is an arbitrary vector, we conclude that R must obey the matrix differential equation

$$\dot{R} = JSR. \quad (10.5.32)$$

Also, the requirement that \mathcal{M}_2 be the identity operator \mathcal{I} when $t = t^i$ makes R subject to the initial condition

$$R(t^i) = I. \quad (10.5.33)$$

The differential equation (5.32) with the initial condition (5.33) has a unique solution whose computation, in most cases, requires numerical integration. [The system (5.32) is equivalent to $(2n)^2$ first-order coupled and time-dependent linear equations.] In the special case when the matrices $JS(t)$ and $JS(t')$ commute for all times t and t' , one has, in analogy to (3.25), the explicit solution

$$R(t) = \exp \left[\int_{t^i}^t JS(t') dt' \right]. \quad (10.5.34)$$

[Indeed, it can be shown that the solution to (5.32) depends entirely upon the Lie algebra generated by the matrices $JS(t)$.] In the even more special case that S (and therefore H_2) is time independent, the integration required in (5.34) is immediate, and one obtains the result

$$R = \exp[(t - t^i)JS]. \quad (10.5.35)$$

In either of the cases corresponding to (5.34) and (5.35) one can write

$$\mathcal{M}_2 = \exp(: f_2 :), \quad (10.5.36)$$

with f_2 given by the relation

$$f_2 = - \int_{t^i}^t H_2(t') dt'. \quad (10.5.37)$$

In the general case, if desired, one may polar decompose R in analogy to (6.2.2) and, in analogy to (7.2.10), write \mathcal{M}_2 in the form

$$\mathcal{M}_2 = \exp(: f_2^c :) \exp(: f_2^a :). \quad (10.5.38)$$

We turn now to the calculation of \mathcal{M}_r . We begin by making the computation of H_r^{int} more explicit. By definition, H_r consists of terms of degree 3 and higher,

$$H_r = H_3 + H_4 + \dots. \quad (10.5.39)$$

Also, in view of (5.25) and the fact that \mathcal{M}_2 produces a linear transformation when acting on ζ^i [see (5.26) and (5.27)], it follows that H_r^{int} has the decomposition

$$H_r^{\text{int}} = H_3^{\text{int}} + H_4^{\text{int}} + \dots, \quad (10.5.40)$$

where each term H_m^{int} is a homogeneous polynomial of degree m given by the relation

$$H_m^{\text{int}}(\zeta^i, t) = H_m(\mathcal{M}_2 \zeta^i, t). \quad (10.5.41)$$

[We note in passing that the operations involved in computing (5.41) using (5.26) are analogous to those employed in (8.4.23) except that R^{-1} is replaced by R .]

To see how this works out in a specific case, consider the computation of H_3^{int} . The terms of still higher degree are handled analogously. Suppose that H_3 is written in the explicit form

$$H_3(\zeta^i, t) = \sum_{abc} T_{abc}(t) \zeta_a^i \zeta_b^i \zeta_c^i, \quad (10.5.42)$$

where T_{abc} is a set of (possibly time-dependent) coefficients. Then use of (5.41) gives the relation

$$H_3^{\text{int}}(\zeta^i, t) = \sum_{abc} T_{abc} (\mathcal{M}_2 \zeta_a^i) (\mathcal{M}_2 \zeta_b^i) (\mathcal{M}_2 \zeta_c^i). \quad (10.5.43)$$

However, thanks to (5.26), the terms on the right side of (5.43) may be evaluated explicitly so that H_3^{int} can be expressed in the form

$$H_3^{\text{int}}(\zeta^i, t) = \sum_{abc} \sum_{a'b'c'} T_{abc} R_{aa'} R_{bb'} R_{cc'} \zeta_{a'}^i \zeta_{b'}^i \zeta_{c'}^i. \quad (10.5.44)$$

Finally, the sums in (5.44) can be grouped so that H_3^{int} can be written in the final form

$$H_3^{\text{int}}(\zeta^i, t) = \sum_{a'b'c'} T_{a'b'c'}^{\text{int}}(t) \zeta_{a'}^i \zeta_{b'}^i \zeta_{c'}^i, \quad (10.5.45)$$

where T^{int} is defined by the equation

$$T_{a'b'c'}^{\text{int}}(t) = \sum_{abc} T_{abc}(t) R_{aa'}(t) R_{bb'}(t) R_{cc'}(t). \quad (10.5.46)$$

As mentioned earlier, because of the time dependence of R , note that H_3^{int} is in general *time dependent* even if H_3 is not.

Let us write \mathcal{M}_r in reversed factorized form. See Section 7.8. Since H_r^{int} consists of terms of degree 3 and higher, \mathcal{M}_r can be written as the product

$$\mathcal{M}_r = \cdots \mathcal{M}_5 \mathcal{M}_4 \mathcal{M}_3, \quad (10.5.47)$$

where each factor \mathcal{M}_m is generated by the homogeneous polynomial function $f_m(\zeta^i)$,

$$\mathcal{M}_m = \exp(: f_m :). \quad (10.5.48)$$

[Note that (5.47) is also a reversed factorization.] Our goal is to find equations of motion for the f_m .

From the factorization (5.47) and the product rule for differentiation, it follows that $\dot{\mathcal{M}}_r$ can be written in the form

$$\dot{\mathcal{M}}_r = \cdots + \cdots \dot{\mathcal{M}}_5 \mathcal{M}_4 \mathcal{M}_3 + \cdots \mathcal{M}_5 \dot{\mathcal{M}}_4 \mathcal{M}_3 + \cdots \mathcal{M}_5 \mathcal{M}_4 \dot{\mathcal{M}}_3. \quad (10.5.49)$$

Suppose (5.49) is substituted into the equation of motion (5.24) and both sides of the resulting relation are multiplied by \mathcal{M}_r^{-1} . So doing gives the result

$$\begin{aligned}\mathcal{M}_r^{-1} \dot{\mathcal{M}}_r &= \cdots + \mathcal{M}_3^{-1} \mathcal{M}_4^{-1} \mathcal{M}_5^{-1} \dot{\mathcal{M}}_5 \mathcal{M}_4 \mathcal{M}_3 + \mathcal{M}_3^{-1} \mathcal{M}_4^{-1} \dot{\mathcal{M}}_4 \mathcal{M}_3 + \mathcal{M}_3^{-1} \dot{\mathcal{M}}_3 \\ &=: -H_r^{\text{int}} :.\end{aligned}\quad (10.5.50)$$

The various terms appearing in (5.50) can be simplified by the use of adjoint operators. For example, we have the result

$$\mathcal{M}_3^{-1} \mathcal{M}_4^{-1} \mathcal{M}_5^{-1} \dot{\mathcal{M}}_5 \mathcal{M}_4 \mathcal{M}_3 = \exp(-\#f_3\#) \exp(-\#f_4\#) \mathcal{M}_5^{-1} \dot{\mathcal{M}}_5. \quad (10.5.51)$$

See (8.2.23). Also, we have the relation

$$\mathcal{M}_m^{-1} \dot{\mathcal{M}}_m = \text{iex}(-\#f_m\#) : \dot{f}_m :. \quad (10.5.52)$$

See Appendix C. Upon using (5.51) and (5.52) in (5.50), we find that (5.50) can be rewritten in the form

$$\begin{aligned}\cdots &+ \exp(-\#f_3\#) \exp(-\#f_4\#) \text{iex}(-\#f_5\#) : \dot{f}_5 : \\ &+ \exp(-\#f_3\#) \text{iex}(-\#f_4\#) : \dot{f}_4 : \\ &+ \text{iex}(-\#f_3\#) : \dot{f}_3 :=: -H_r^{\text{int}} :.\end{aligned}\quad (10.5.53)$$

At this stage the colons can be removed from both sides of (5.53) to give the result

$$\begin{aligned}\cdots &+ \exp(-:f_3:) \exp(-:f_4:) \text{iex}(-:f_5:) \dot{f}_5 \\ &+ \exp(-:f_3:) \text{iex}(-:f_4:) \dot{f}_4 \\ &+ \text{iex}(-:f_3:) \dot{f}_3 = -H_r^{\text{int}}.\end{aligned}\quad (10.5.54)$$

Let us examine both sides of (5.54) with the aim of equating terms of like degree. From the expansion (8.8.9) we find the result

$$\text{iex}(-:f_m:) \dot{f}_m = (1 - :f_m:/2! + :f_m:/^2/3! - \dots) \dot{f}_m. \quad (10.5.55)$$

According to (7.6.16), the terms of the right side of (5.55) have degrees m , $2m-2$, $3m-4$, etc. Consequently, upon using (5.40), and equating terms of like degree in (5.54), we find the result

$$\begin{aligned}P_m[\cdots &+ \exp(-:f_3:) \exp(-:f_4:) \text{iex}(-:f_5:) \dot{f}_5 \\ &+ \exp(-:f_3:) \text{iex}(-:f_4:) \dot{f}_4 \\ &+ \text{iex}(-:f_3:) \dot{f}_3] = -H_m^{\text{int}}.\end{aligned}\quad (10.5.56)$$

Here P_m denotes a *projection* operator that projects out terms of degree m . For example, we have the results

$$P_3[\cdots + \text{iex}(-:f_3:) \dot{f}_3] = \dot{f}_3, \quad (10.5.57)$$

$$P_4[\cdots + \text{iex}(-:f_3:) \dot{f}_3] = \dot{f}_4 - (:f_3:/2!) \dot{f}_3, \quad (10.5.58)$$

$$P_5[\cdots + \text{iex}(-:f_3:) \dot{f}_3] = \dot{f}_5 - :f_3:\dot{f}_4 + (:f_3:/^2/3!) \dot{f}_3. \quad (10.5.59)$$

The relations (5.56) can now be solved for the various \dot{f}_m . We find, for example, through $m = 8$, the results

$$\dot{f}_3 = -H_3^{\text{int}}, \quad (10.5.60)$$

$$\dot{f}_4 = -H_4^{\text{int}} + (: f_3 : /2)(-H_3^{\text{int}}), \quad (10.5.61)$$

$$\dot{f}_5 = -H_5^{\text{int}} + : f_3 : (-H_4^{\text{int}}) + (1/3) : f_3 :^2 (-H_3^{\text{int}}), \quad (10.5.62)$$

$$\begin{aligned} \dot{f}_6 = & - H_6^{\text{int}} + : f_3 : (-H_5^{\text{int}}) + (1/2) : f_4 : (-H_4^{\text{int}}) \\ & + (1/4) : f_4 :: f_3 : (-H_3^{\text{int}}) + (1/2) : f_3 :^2 (-H_4^{\text{int}}) \\ & + (1/8) : f_3 :^3 (-H_3^{\text{int}}), \end{aligned} \quad (10.5.63)$$

$$\begin{aligned} \dot{f}_7 = & - H_7^{\text{int}} + : f_3 : (-H_6^{\text{int}}) + : f_4 : (-H_5^{\text{int}}) + : f_4 :: f_3 : (-H_4^{\text{int}}) \\ & + (1/3) : f_4 :: f_3 :^2 (-H_3^{\text{int}}) + (1/2) : f_3 :^2 (-H_5^{\text{int}}) \\ & + (1/6) : f_3 :^3 (-H_4^{\text{int}}) + (1/30) : f_3 :^4 (-H_3^{\text{int}}), \end{aligned} \quad (10.5.64)$$

$$\begin{aligned} \dot{f}_8 = & - H_8^{\text{int}} + : f_3 : (-H_7^{\text{int}}) + : f_4 : (-H_6^{\text{int}}) + : f_4 :: f_3 : (-H_5^{\text{int}}) \\ & + (1/2) : f_4 :: f_3 :^2 (-H_4^{\text{int}}) + (1/8) : f_4 :: f_3 :^3 (-H_3^{\text{int}}) \\ & + (1/2) : f_5 : (-H_5^{\text{int}}) + (1/2) : f_5 :: f_3 : (-H_4^{\text{int}}) \\ & + (1/6) : f_5 :: f_3 :^2 (-H_3^{\text{int}}) + (1/2) : f_3 :^2 (-H_6^{\text{int}}) \\ & + (1/3) : f_4 :^2 (-H_4^{\text{int}}) + (1/6) : f_4 :^2 : f_3 : (-H_3^{\text{int}}) \\ & + (1/6) : f_3 :^3 (-H_5^{\text{int}}) + (1/24) : f_3 :^4 (-H_4^{\text{int}}) \\ & + (1/144) : f_3 :^5 (-H_3^{\text{int}}), \end{aligned} \quad (10.5.65)$$

$$\dot{f}_m = \text{something involving } H_m \text{ and the } f_\ell \text{ and } H_\ell \text{ with } \ell < m. \quad (10.5.66)$$

What have we accomplished? From (5.22), (5.38), (5.47), and (5.48) we see that \mathcal{M} has been computed in a reverse factorized product form; and we have found how to calculate the \mathcal{M}_m . To find \mathcal{M}_2 we need to integrate the equations (5.32) with the initial condition (5.33). Here it is assumed that $z^d(t)$ is known so that H_2 and hence S is known. See (5.28). If this is not the case, then we must also integrate the equations (5.14) or (5.18). That is, we must integrate the equations (5.14) [or (5.18)] and (5.32) as a coupled set. To find the f_m that define \mathcal{M}_r according to (5.47) and (5.48), we must also integrate the equations of motion for the f_m as given by (5.60) through (5.66). The requirement that \mathcal{M}_r be the identity operator \mathcal{I} when $t = t^i$ makes the f_m subject to the initial condition

$$f_m(t^i) = 0. \quad (10.5.67)$$

We note from the form of equations (5.60) through (5.66) that the computation of the f_m requires a knowledge of the H_ℓ^{int} with $\ell \leq m$. The computation of $H_\ell^{\text{int}}(t)$ in turn requires a knowledge of $R(t)$. [See, for example, (5.43) through (5.46).] Consequently, the equations of motion (5.60) through (5.66) for the f_m must be integrated simultaneously with the equations (5.32) for R . [Moreover, in general $z^d(t)$ must also be known to compute $H_\ell^{\text{int}}(t)$.]

Thus, if $z^d(t)$ is not known explicitly, then the equations (5.14) or (5.18) must in general also be in the set of equations to be integrated.]

We close this section with the observation that the equations of motion (5.60) through (5.66) for the f_m , like equations (3.11) through (3.15) for the F_n , have the property that the f_m can be determined successively. This property has two consequences. First, a solution of the equations of motion for R , f_3 , f_4 , \dots , f_m by numerical methods is always possible for any m . Moreover, solution by *quadrature* is also possible. For example, assuming that R has already been determined, (5.60) can be integrated immediately to give the result

$$f_3(\zeta^i, t) = \int_{t^i}^t dt_1 [-H_3^{\text{int}}(\zeta^i, t_1)]. \quad (10.5.68)$$

Here we have used (5.67). Next, this result for f_3 can be substituted into (5.61) and the resulting expression for f_4 can be integrated to give the relation

$$f_4(\zeta^i, t) = \int_{t^i}^t dt_1 [-H_4^{\text{int}}(\zeta^i, t_1)] + (1/2) \int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 [-H_3^{\text{int}}(\zeta^i, t_2), -H_3^{\text{int}}(\zeta^i, t_1)]. \quad (10.5.69)$$

Similarly, one finds for f_5 the result

$$\begin{aligned} f_5 &= \int_{t^i}^t dt_1 [-H_5^{\text{int}}(t_1)] \\ &+ \int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 [-H_3^{\text{int}}(t_2), -H_4^{\text{int}}(t_1)] \\ &+ (1/3) \int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 \int_{t^i}^{t_2} dt_3 [-H_3^{\text{int}}(t_3), [-H_3^{\text{int}}(t_2), -H_3^{\text{int}}(t_1)]] \\ &+ (1/3) \int_{t^i}^t dt_1 \int_{t^i}^{t_1} dt_2 \int_{t^i}^{t_2} dt_3 [-H_3^{\text{int}}(t_2), [-H_3^{\text{int}}(t_3), -H_3^{\text{int}}(t_1)]]. \end{aligned} \quad (10.5.70)$$

Evidently, one can find explicit integral representations for all the f_m . We note that to determine any f_m it is necessary to know only the H_ℓ^{int} with $\ell \leq m$. Moreover, if the H_ℓ^{int} do not commute (are not in involution) at different times, which is usually the case, there are feed-up effects: lower-order terms in the Hamiltonian can contribute to higher-order Lie generators. Specifically, we see that the f_m all lie in the Lie algebra generated by the H_ℓ^{int} . Finally, suppose the time dependencies of the various H_ℓ^{int} are sufficiently simple that all the integrations occurring in (5.68), (5.69), (5.70), and analogous expressions for the other f_m can be carried out analytically. Then the equations of motion (5.60) through (5.66) can in principle be solved directly by symbolic manipulation to obtain complete analytic expressions for all desired f_m .

We close this subsection by noting that for simplicity we have derived formulas for R and the f_m when the map \mathcal{M} is written in *reversed* factorized product form. That is, we have written \mathcal{M} in the form

$$\mathcal{M} = \cdots \mathcal{M}_5 \mathcal{M}_4 \mathcal{M}_3 \mathcal{M}_2. \quad (10.5.71)$$

See (5.22) and (5.47). Of course, once we have found \mathcal{M} is reversed factorized product form, we can convert it to the *forward* factorized product form

$$\mathcal{M} = \mathcal{M}'_2 \mathcal{M}'_3 \mathcal{M}'_4 \mathcal{M}'_5 \dots \quad (10.5.72)$$

by means of concatenation formulas. Recall Section 8.4. Alternatively, as will be seen in the next section, there are formulas for the associated R' and f'_m , analogous to those for R and the f_m , when \mathcal{M} is written in forward factorized product form. Finally we note, from the work of Section 8.5, that there is a standard procedure for passing between forward and reverse factorized forms.

10.5.3 Summary and GENMAP Nomenclature

Given a deviation variable Hamiltonian expanded in the form (5.3), we have formulated the equation (5.18) for the design trajectory and the equation (5.72) for *generating* the map \mathcal{M} about the design trajectory, with \mathcal{M} written in forward factorized form. Intermediate steps involved formulation and (usually numerical) integration of the equations (5.32) and (5.60) through (5.66). We will refer to the combined execution of all these steps as the GENMAP algorithm, and all the differential equations and Lie-algebraic procedures involved as the GENMAP equations.

Exercises

10.5.1. The purpose of this exercise is to verify the relation (5.18) from another perspective. Begin by showing that by Taylor's theorem there is the relation

$$\bar{H}_1(\zeta, t) = \sum_a [\partial H(z, t)/\partial z_a]|_{z=z^d(t)} \zeta_a, \quad (10.5.73)$$

and consequently

$$\partial \bar{H}_1(\zeta, t)/\partial \zeta_a = [\partial H(z, t)/\partial z_a]|_{z=z^d(t)}. \quad (10.5.74)$$

Now verify it follows that

$$J \partial_\zeta \bar{H}_1(\zeta, t) = J[\partial_z H(z, t)]|_{z=z^d(t)} = \dot{z}^d. \quad (10.5.75)$$

10.5.2. Consider a general Hamiltonian system with Hamiltonian $H(z, t)$, and suppose $z^d(t)$ is any particular trajectory for (solution to) the equations of motion associated with this Hamiltonian. Form the variational equations about the trajectory $z^d(t)$. Show that the variational equations also arise from a Hamiltonian, and that this Hamiltonian is the quadratic Hamiltonian \bar{H}_2 that appears in the sum (5.3). Show that integrating (5.32) with the initial condition (5.33) provides all possible solutions to the variational equations.

10.6 Forward Factorization and Lie Concatenation Revisited

10.6.1 Preliminary Discussion

The previous section derived a reversed factorized product solution to the equation of motion (1.8) for \mathcal{M} . For this present section, and other purposes as well, it is useful to also have formulas for a forward factorized product solution. We begin by finding them.

The main purpose of this section, which is really a diversion from the logical presentation of methods for the computation of maps, is to provide another derivation of the Lie concatenation formulas of Section 8.4. Subsequent sections will return to the main subject of this chapter.

10.6.2 Forward Factorization

As before we write

$$\mathcal{M} = \mathcal{M}_r \mathcal{M}_2 \quad (10.6.1)$$

and require that \mathcal{M}_2 again satisfy (5.23). Then we find, as before, that \mathcal{M}_r obeys the equation of motion

$$\dot{\mathcal{M}}_r = \mathcal{M}_r : -H_r^{\text{int}} : \quad (10.6.2)$$

with H_r^{int} again given by (5.25). Now, however, we write \mathcal{M}_r in the forward factorized form

$$\mathcal{M}_r = \mathcal{M}_3 \mathcal{M}_4 \mathcal{M}_5 \cdots \quad (10.6.3)$$

with

$$\mathcal{M}_m = \exp(: f_m :). \quad (10.6.4)$$

We will obtain equations of motion for the f_m shortly, but imagine for the moment that they have already been found and solved. Then we may write \mathcal{M} in the form

$$\mathcal{M} = \exp(: f_3 :) \exp(: f_4 :) \exp(: f_5 :) \cdots \mathcal{M}_2. \quad (10.6.5)$$

Algebraic manipulation now gives the equivalent result

$$\begin{aligned} \mathcal{M} &= \mathcal{M}_2 \mathcal{M}_2^{-1} \exp(: f_3 :) \mathcal{M}_2 \mathcal{M}_2^{-1} \exp(: f_4 :) \mathcal{M}_2 \mathcal{M}_2^{-1} \exp(: f_5 :) \cdots \mathcal{M}_2 \\ &= \mathcal{M}_2 \exp(: f_3^{\text{tr}} :) \exp(: f_4^{\text{tr}} :) \exp(: f_5^{\text{tr}} :) \cdots \end{aligned} \quad (10.6.6)$$

where

$$f_m^{\text{tr}} = \mathcal{M}_2^{-1} f_m. \quad (10.6.7)$$

We see that \mathcal{M} as given by (6.6) and (6.7) is in the desired forward factorized product form.

It remains to find the f_m . Differentiating (6.3) gives the result

$$\dot{\mathcal{M}}_r = \dot{\mathcal{M}}_3 \mathcal{M}_4 \mathcal{M}_5 \cdots + \mathcal{M}_3 \dot{\mathcal{M}}_4 \mathcal{M}_5 \cdots + \mathcal{M}_3 \mathcal{M}_4 \dot{\mathcal{M}}_5 \cdots + \cdots. \quad (10.6.8)$$

Next, substitution of (6.8) into the equation of motion (6.2) produces the relation

$$\begin{aligned} \mathcal{M}_r^{-1} \dot{\mathcal{M}}_r &= \cdots \mathcal{M}_5^{-1} \mathcal{M}_4^{-1} \mathcal{M}_3^{-1} \dot{\mathcal{M}}_3 \mathcal{M}_4 \mathcal{M}_5 \cdots + \cdots \mathcal{M}_5^{-1} \mathcal{M}_4^{-1} \dot{\mathcal{M}}_4 \mathcal{M}_5 \cdots \\ &+ \cdots \mathcal{M}_5^{-1} \dot{\mathcal{M}}_5 \cdots + \cdots =: -H_r^{\text{int}} : . \end{aligned} \quad (10.6.9)$$

As in the previous section, the various terms in (6.9) can be simplified by the use of adjoint operators. For example, we have the result

$$\cdots \mathcal{M}_5^{-1} \mathcal{M}_4^{-1} \mathcal{M}_3^{-1} \dot{\mathcal{M}}_3 \mathcal{M}_4 \mathcal{M}_5 \cdots = \cdots \exp(-\#f_5\#) \exp(-\#f_4\#) \mathcal{M}_3^{-1} \dot{\mathcal{M}}_3. \quad (10.6.10)$$

Again we recall the relation

$$\mathcal{M}_m^{-1} \dot{\mathcal{M}}_m = \text{iex}(-\#f_m\#) : \dot{f}_m : . \quad (10.6.11)$$

Upon using (6.10) and (6.11) in (6.9) we find that it can be rewritten in the form

$$\begin{aligned} & \cdots \exp(-\#f_5\#) \exp(-\#f_4\#) \text{iex}(-\#f_3\#) : \dot{f}_3 : \\ & + \cdots \exp(-\#f_5\#) \text{iex}(-\#f_4\#) : \dot{f}_4 : \\ & + \cdots \text{iex}(-\#f_5\#) : \dot{f}_5 := -H_r^{\text{int}} : . \end{aligned} \quad (10.6.12)$$

The colons can now be removed from both sides of (6.12) to give the equivalent result

$$\begin{aligned} & \cdots \exp(-: f_5:) \exp(-: f_4:) \text{iex}(-: f_3:) \dot{f}_3 \\ & + \cdots \exp(-: f_5:) \text{iex}(-: f_4:) \dot{f}_4 \\ & + \cdots \text{iex}(-: f_5:) \dot{f}_5 \\ & + \cdots = -H_r^{\text{int}}. \end{aligned} \quad (10.6.13)$$

Finally, similar to the procedure in the previous section, we equate terms of like degree in (6.13). So doing gives the equations of motion

$$\dot{f}_3 = -H_3^{\text{int}}, \quad (10.6.14)$$

$$\dot{f}_4 = -H_4^{\text{int}} + (: f_3 : / 2)(-H_3^{\text{int}}), \quad (10.6.15)$$

$$\dot{f}_5 = -H_5^{\text{int}} - (1/6) : f_3 :^2 (-H_3^{\text{int}}) + : f_4 : (-H_3^{\text{int}}), \quad (10.6.16)$$

$$\begin{aligned} \dot{f}_6 = & -H_6^{\text{int}} + (1/24) : f_3 :^3 (-H_3^{\text{int}}) + (1/2) : f_4 : (-H_4^{\text{int}}) \\ & - (1/4) : f_4 :: f_3 : (-H_3^{\text{int}}) + : f_5 : (-H_3^{\text{int}}), \end{aligned} \quad (10.6.17)$$

$$\begin{aligned} \dot{f}_7 = & -H_7^{\text{int}} - (1/120) : f_3 :^4 (-H_3^{\text{int}}) + (1/6) : f_4 :: f_3 :^2 (-H_3^{\text{int}}) \\ & - (1/2) : f_4 :^2 (-H_3^{\text{int}}) + : f_5 : (-H_4^{\text{int}}) + : f_6 : (-H_3^{\text{int}}), \end{aligned} \quad (10.6.18)$$

$$\begin{aligned} \dot{f}_8 = & -H_8^{\text{int}} + (1/720) : f_3 :^5 (-H_3^{\text{int}}) - (1/24) : f_4 :: f_3 :^3 (-H_3^{\text{int}}) \\ & - (1/6) : f_4 :^2 (-H_4^{\text{int}}) + (1/6) : f_4 :^2 : f_3 : (-H_3^{\text{int}}) + (1/2) : f_5 : (-H_5^{\text{int}}) \\ & + (1/12) : f_5 :: f_3 :^2 (-H_3^{\text{int}}) - (1/2) : f_5 :: f_4 : (-H_3^{\text{int}}) + : f_6 : (-H_4^{\text{int}}) \\ & + : f_7 : (-H_3^{\text{int}}), \end{aligned} \quad (10.6.19)$$

$$\dot{f}_m = \text{expression involving } H_m^{\text{int}} \text{ and the } f_\ell \text{ and } H_\ell^{\text{int}} \text{ with } \ell < m. \quad (10.6.20)$$

As before, these equations can be integrated with the initial condition (5.67).

10.6.3 Alternate Derivation of Lie Concatenation Formulas

The tools are now in hand to carry out the main task of this section: an alternate derivation of the Lie concatenation formulas. According to (8.4.26) in Section 8.4, the problem is to find the h_m in the relation

$$\exp(: h_3 :) \exp(: h_4 :) \cdots = \exp(: f_3^{tr} :) \exp(: f_4^{tr} :) \cdots \exp(: g_3 :) \exp(: g_4 :) \cdots. \quad (10.6.21)$$

[Note that here the f_m^{tr} are given by (8.4.23) and not by (6.7).] To do this, we employ a trick. Define a one-parameter family of maps $\mathcal{N}(\lambda)$ by the rule

$$\mathcal{N}(\lambda) = \exp(\lambda : f_3^{tr} :) \exp(\lambda : f_4^{tr} :) \cdots \exp(\lambda : g_3 :) \exp(\lambda : g_4 :) \cdots. \quad (10.6.22)$$

Then, by construction, $\mathcal{N}(\lambda)$ has the properties

$$\mathcal{N}(0) = \mathcal{I}, \quad (10.6.23)$$

$$\mathcal{N}(1) = \exp(: h_3 :) \exp(: h_4 :) \cdots, \quad (10.6.24)$$

$$\mathcal{N}^{-1}(\lambda) = \cdots \exp(-\lambda : g_4 :) \exp(-\lambda : g_3 :) \cdots \exp(-\lambda : f_4^{tr} :) \exp(-\lambda : f_3^{tr} :). \quad (10.6.25)$$

From Section 6.4 we know that there is an associated Hamiltonian $H(\lambda)$. In this case, in analogy to (1.8), it can be found from the relation

$$: -H := \mathcal{N}^{-1} \dot{\mathcal{N}} \quad (10.6.26)$$

where a dot denotes differentiation with respect to λ . From (6.22) we find the result

$$\begin{aligned} \dot{\mathcal{N}} &= \exp(\lambda : f_3^{tr} :) : f_3^{tr} : \exp(\lambda : f_4^{tr} :) \cdots \exp(\lambda : g_3 :) \exp(\lambda : g_4 :) \cdots \\ &+ \exp(\lambda : f_3^{tr} :) \exp(\lambda : f_4^{tr} :) : f_4^{tr} : \cdots \exp(\lambda : g_3 :) \exp(\lambda : g_4 :) \cdots \\ &+ \cdots \\ &+ \exp(\lambda : f_3^{tr} :) \exp(\lambda : f_4^{tr} :) \cdots \exp(\lambda : g_3 :) : g_3 : \exp(\lambda : g_4 :) \cdots \\ &+ \exp(\lambda : f_3^{tr} :) \exp(\lambda : f_4^{tr} :) \cdots \exp(\lambda : g_3 :) \exp(\lambda : g_4 :) : g_4 : \cdots \\ &+ \cdots. \end{aligned} \quad (10.6.27)$$

Consequently the Hamiltonian Lie operator $: -H :$ is given by

$$\begin{aligned} : -H : &= \mathcal{N}^{-1} \dot{\mathcal{N}} \\ &= \cdots \exp(-\lambda : g_4 :) \exp(-\lambda : g_3 :) \cdots \times \\ &\quad \exp(-\lambda : f_4^{tr} :) : f_3^{tr} : \exp(\lambda : f_4^{tr} :) \cdots \exp(\lambda : g_3 :) \exp(\lambda : g_4 :) \cdots \\ &+ \cdots \exp(-\lambda : g_4 :) \exp(-\lambda : g_3 :) \cdots : f_4^{tr} : \cdots \exp(\lambda : g_3 :) \exp(\lambda : g_4 :) \cdots \\ &+ \cdots \\ &+ \cdots \exp(-\lambda : g_4 :) : g_3 : \exp(\lambda : g_4 :) \cdots \\ &+ \cdots : g_4 : \cdots \\ &+ \cdots. \end{aligned} \quad (10.6.28)$$

This result can also be written more compactly with the aid of adjoint operators to take the form

$$\begin{aligned}
 : -H : &= \cdots \exp(-\lambda \# g_4 \#) \exp(-\lambda \# g_3 \#) \cdots \exp(-\lambda \# f_4^{tr} \#) : f_3^{tr} : \\
 &+ \cdots \exp(-\lambda \# g_4 \#) \exp(-\lambda \# g_3 \#) \cdots : f_4^{tr} : \\
 &+ \cdots \\
 &+ \cdots \exp(-\lambda \# g_4 \#) : g_3 : \\
 &+ \cdots : g_4 : \\
 &+ \cdots .
 \end{aligned} \tag{10.6.29}$$

The colons can now be removed from both sides of (6.29) to give the equivalent result

$$\begin{aligned}
 -H &= \cdots \exp(-\lambda : g_4 :) \exp(-\lambda : g_3 :) \cdots \exp(-\lambda : f_4^{tr} :) f_3^{tr} \\
 &+ \cdots \exp(-\lambda : g_4 :) \exp(-\lambda : g_3 :) \cdots f_4^{tr} \\
 &+ \cdots \\
 &+ \cdots \exp(-\lambda : g_4 :) g_3 \\
 &+ \cdots g_4 \\
 &+ \cdots .
 \end{aligned} \tag{10.6.30}$$

Finally, equating terms of like degree on both sides of (6.30) yields the results

$$-H_2 = 0, \tag{10.6.31}$$

$$-H_3 = f_3^{tr} + g_3, \tag{10.6.32}$$

$$-H_4 = -\lambda : g_3 : f_3^{tr} + f_4^{tr} + g_4, \tag{10.6.33}$$

$$\begin{aligned}
 -H_5 &= f_5^{tr} + g_5 - \lambda : f_4^{tr} : f_3^{tr} - \lambda : g_3 : f_4^{tr} \\
 &+ (1/2)\lambda^2 : g_3 :^2 f_3^{tr} - \lambda : g_4 : f_3^{tr} - \lambda : g_4 : g_3, \text{ etc.}
 \end{aligned} \tag{10.6.34}$$

Evidently the h_m in (6.24) can be regarded as the solutions to differential equations of the form (6.14) through (6.20) with H given by (6.31) through (6.34). Also, as a consequence of (6.31), we have the result

$$H_m^{\text{int}} = H_m. \tag{10.6.35}$$

It follows that the h_m satisfy the differential equations

$$\dot{h}_3 = -H_3 = f_3^{tr} + g_3, \tag{10.6.36}$$

$$\begin{aligned}
 \dot{h}_4 &= -H_4 + (: h_3 : /2)(-H_3) \\
 &= -\lambda : g_3 : f_3^{tr} + f_4^{tr} + g_4 + (: h_3 : /2)(f_3^{tr} + g_3),
 \end{aligned} \tag{10.6.37}$$

$$\begin{aligned}
 \dot{h}_5 &= -H_5 - (1/6) : h_3 :^2 (-H_3) + : h_4 : (-H_3) \\
 &= f_5^{tr} + g_5 - \lambda : f_4^{tr} : f_3^{tr} - \lambda : g_3 : f_4^{tr} + (1/2)\lambda^2 : g_3 :^2 f_3^{tr} - \lambda : g_4 : f_3^{tr} \\
 &- \lambda : g_4 : g_3 - (1/6) : h_3 :^2 (f_3^{tr} + g_3) + : h_4 : (f_3^{tr} + g_3), \text{ etc.}
 \end{aligned} \tag{10.6.38}$$

These equations can now be solved. Equation (6.36) has the immediate solution

$$h_3(\lambda) = \lambda(f_3^{tr} + g_3). \quad (10.6.39)$$

When this solution is substituted into (6.37), the result is that h_4 satisfies the differential equation

$$\dot{h}_4 = -\lambda : g_3 : f_3^{tr} + f_4^{tr} + g_4 \quad (10.6.40)$$

with the solution

$$h_4(\lambda) = \lambda(f_4^{tr} + g_4) - (\lambda^2/2) : g_3 : f_3^{tr}. \quad (10.6.41)$$

Next, with this knowledge of $h_3(\lambda)$ and $h_4(\lambda)$, the equation of motion for h_5 becomes

$$\dot{h}_5 = f_5^{tr} + g_5 - 2\lambda : g_3 : f_4^{tr} + \lambda^2 : g_3 :^2 f_3^{tr} - (\lambda^2/2) : f_3^{tr} :^2 g_3, \quad (10.6.42)$$

with the solution

$$h_5(\lambda) = \lambda(f_5^{tr} + g_5) - \lambda^2 : g_3 : f_4^{tr} + (\lambda^3/3) : g_3 :^2 f_3^{tr} - (\lambda^3/6) : f_3^{tr} :^2 g_3. \quad (10.6.43)$$

The $h_m(\lambda)$ for $m = 6, 7, \dots$ can be found in an analogous way. We conclude that all the $h_m(\lambda)$ can be computed recursively for any value of m .

Finally, the quantities $h_m(1)$ yield the desired h_m in (6.21). Compare, for example, (8.4.31) through (8.4.33) with the $h_m(1)$ computed from (6.39), (6.41), and (6.43). The virtue of this method for deriving the Lie concatenation formulas is that no knowledge of the BCH series coefficients is required and the results are immediately obtained in Lie form. Only the Taylor coefficients in the *exp* and *iex* functions are needed, and these are known to all orders. And since the equations to be integrated are polynomial in λ , all calculations can be carried out to any desired order by symbolic manipulation.

Exercises

10.6.1.

10.7 Direct Taylor Summation

We now return to the task of computing maps. Suppose the Hamiltonian H is autonomous (time independent) and is analytic in z about the origin $z = 0$. Then it has an expansion in homogeneous polynomials of the form

$$H = \sum_{\ell=1}^{\infty} H_\ell(z). \quad (10.7.1)$$

(Here we have omitted a possible constant term since it has no dynamic effect.) We know from Section 7.4 that in this case the associated transfer map \mathcal{M} can be written formally as

$$\mathcal{M} = \exp(-\tau : H :) \quad (10.7.2)$$

where we have used the short-hand notation

$$\tau = t - t^i. \quad (10.7.3)$$

The purpose of this section, among other things, is to explore what happens when there are singularities in τ .

Let us write

$$z_a^f = \mathcal{M}z_a^i \quad (10.7.4)$$

and make the Taylor expansion

$$z_a^f = k_a + \sum_b R_{ab} z_b^i + \sum_{bc} T_{abc} z_b^i z_c^i + \sum_{bcd} U_{abcd} z_b^i z_c^i z_d^i + \dots \quad (10.7.5)$$

We also have the result

$$z_a^f = \exp(-\tau : H :) z_a^i = \sum_{j=0}^{\infty} [(-\tau)^j / j!] : H :^j z_a^i. \quad (10.7.6)$$

Here H is to be regarded as a function of the z^i ,

$$H = H(z^i) = H_1(z^i) + H_2(z^i) + H_3(z^i) + \dots \quad (10.7.7)$$

If we substitute (7.7) into (7.6), carry out all the indicated Poisson brackets, and then group terms by degree, we will find the Taylor coefficients k , R , T , U , etc. that appear in (7.5). Once the Taylor coefficients are known, there is (according to the Factorization Theorem of Sections 7.6 through 7.8) a standard procedure for finding the homogeneous polynomials f_1 , f_2^c , f_2^a , f_3 , f_4 , \dots such that \mathcal{M} can be written in the form

$$\mathcal{M} = \exp(: f_1 :) \exp(: f_2^c :) \exp(: f_2^a :) \exp(: f_3 :) \exp(: f_4 :) \dots \quad (10.7.8)$$

What we wish to investigate at this point is the convergence of the series (7.6) *in the sense* that it produces *series* for the Taylor coefficients k , R , T , U , etc. Note that this issue is separate from the convergence of the Taylor series (7.5) with regard to the variables z^i .

To facilitate this investigation, it is convenient to employ the basis monomials G_r introduced in Sections 7.3 and 8.3. With the aid of these monomials and the inner product (8.3.32), introduce the infinite dimensional matrices M and H by the rules

$$M_{sr} = \langle G_s, \mathcal{M}G_r \rangle, \quad (10.7.9)$$

$$H_{sr} = \langle G_s, : H : G_r \rangle. \quad (10.7.10)$$

Then, as a result of (8.3.41), the operator relation (7.2) has the equivalent matrix formulation

$$M = \exp(-\tau H) = \sum_{j=0}^{\infty} (\tau^j / j!) H^j. \quad (10.7.11)$$

[Here, to test the reader's mental agility, the symbol H stands not for the Hamiltonian (7.7) but rather for the matrix (7.10).] Moreover, the various Taylor coefficients k , R , T , U , etc.

in (7.5) are just matrix elements of the form $\langle G_s, \mathcal{M}G_a^1 \rangle$ where we have used the notation G_a^1 to denote the *linear* functions z_a .

In any practical calculation with power series it is necessary to truncate them at some stage. Operations performed on and with truncated power series will be referred to as *Truncated Power Series Algebra* (TPSA). In this context it is useful to introduce a *projection* operator \mathcal{P}^D . Let $m(r)$ denote the degree of the monomial G_r . It is defined by the relation

$$m(r) = \sum (u_i + v_i). \quad (10.7.12)$$

See (7.3.33) and (7.3.34). We now define \mathcal{P}^D to be the *linear* operator with the property

$$\begin{aligned} \mathcal{P}^D G_r &= G_r \text{ if } m(r) \leq D, \\ \mathcal{P}^D G_r &= 0 \text{ if } m(r) > D. \end{aligned} \quad (10.7.13)$$

That is, \mathcal{P}^D retains monomials having degree D or less, and discards those with degree larger than D . In terms of the earlier notation associated with (8.9.68) and (8.9.69), \mathcal{P}^D is just the truncation operator $\mathcal{T}(> D)$. Let θ be a step function given by the rule

$$\begin{aligned} \theta(x) &= 0 \text{ for } x < 0, \\ \theta(x) &= 1 \text{ for } x \geq 0. \end{aligned} \quad (10.7.14)$$

Then we also have the equivalent definition

$$\mathcal{P}^D G_r = \theta(D - m) G_r. \quad (10.7.15)$$

Evidently \mathcal{P}^D has the property

$$(\mathcal{P}^D)^2 = \mathcal{P}^D. \quad (10.7.16)$$

Now let \mathcal{A} be any linear operator. In TPSA this operator is realized by its *truncated* counterpart ${}^D\mathcal{A}$ defined by the rule

$${}^D\mathcal{A} = \mathcal{P}^D \mathcal{A} \mathcal{P}^D. \quad (10.7.17)$$

Following (7.3.33) through (7.3.35), let us also use the modified notation G_r^m for the basis monomials. Consider the vector space V^D spanned by the monomials $G_r^0, G_r^1, \dots, G_r^D$. From the definition (7.17) we evidently have the result

$$\langle G_r^m, {}^D\mathcal{A} G_{r'}^{m'} \rangle = 0 \text{ if either } m > D \text{ or } m' > D. \quad (10.7.18)$$

It follows that ${}^D\mathcal{A}$ maps V^D into itself, and consequently has a representation as a *finite* dimensional matrix ${}^D\mathcal{A}$ acting on V^D . Indeed, the matrix ${}^D\mathcal{A}$ has the entries

$${}^D A_{sr} = \langle G_s, {}^D\mathcal{A} G_r \rangle \text{ with } m(s), m(r) \leq D. \quad (10.7.19)$$

Suppose \mathcal{A} , \mathcal{B} , and \mathcal{C} are three linear operators related by the equation

$$\mathcal{C} = \mathcal{A}\mathcal{B}. \quad (10.7.20)$$

Then in general one has the inequalities

$${}^D\mathcal{C} \neq {}^D\mathcal{A} {}^D\mathcal{B}, \quad (10.7.21)$$

$${}^D\mathcal{C} \neq {}^D\mathcal{A} {}^D\mathcal{B}. \quad (10.7.22)$$

However, if either \mathcal{A} or \mathcal{B} commute with \mathcal{P}^D , then (7.21) and (7.22) become equalities. Indeed, from (7.16) and the definitions (7.13) we have the result

$${}^D\mathcal{A} {}^D\mathcal{B} = \mathcal{P}^D \mathcal{A} \mathcal{P}^D \mathcal{P}^D \mathcal{B} \mathcal{P}^D = \mathcal{P}^D \mathcal{A} \mathcal{P}^D \mathcal{B} \mathcal{P}^D. \quad (10.7.23)$$

Evidently if one can move the middle factor \mathcal{P}^D either to the left or to the right with impunity, then (7.21) and (7.22) become equalities.

The truncated counterparts of Lie operators are also not derivations. Consider, for example, the case of a 2-dimensional phase space. We have in accord with (5.3.7) the relation

$$:q:p^4 = (:q:p^2)p^2 + p^2 :q:p^2. \quad (10.7.24)$$

Suppose $D = 3$. Then the counterpart of the left side of (7.24) has the value

$$({}^3:q:)p^4 = (\mathcal{P}^3 : q : \mathcal{P}^3)p^4 = 0. \quad (10.7.25)$$

On the other hand, the counterpart of the right side of (7.24) is the quantity

$$[({}^3:q:)p^2]p^2 + p^2 ({}^3:q:)p^2 = 4p^3. \quad (10.7.26)$$

Thus $({}^3:q:)$ is not a derivation. It follows from analogous calculations that no truncated Lie operator of the form $({}^D:f_1:)$ is a derivation.

However, any $({}^D:f_\ell:)$ with $\ell \geq 2$ does at least enjoy the property

$$({}^D:f_\ell:) = \mathcal{P}^D : f_\ell : \mathcal{P}^D = \mathcal{P}^D : f_\ell : \text{ when } \ell \geq 2. \quad (10.7.27)$$

Let G_r^m be any monomial. From (7.6.16) and (7.15) we find the results

$$\begin{aligned} ({}^D:f_\ell:)G_r^m &= \mathcal{P}^D : f_\ell : \mathcal{P}^D G_r^m = \mathcal{P}^D : f_\ell : \theta(d-m)G_r^m \\ &= \theta(D-m)\mathcal{P}^D \sum_{r'} c_{r'} G_{r'}^{m+\ell-2} \\ &= \theta(D-m)\theta(D+2-\ell-m) \sum_{r'} c_{r'} G_{r'}^{m+\ell-2}, \end{aligned} \quad (10.7.28)$$

$$\mathcal{P}^D : f_\ell : G_r^m = \mathcal{P}^D \sum_{r'} c_{r'} G_{r'}^{m+\ell-2} = \theta(D+2-\ell-m) \sum_{r'} c_{r'} G_{r'}^{m+\ell-2}. \quad (10.7.29)$$

Here the $c_{r'}$ are certain coefficients whose exact values need not concern us. But from (7.14) we conclude that

$$\theta(D-m)\theta(D+2-\ell-m) = \theta(D+2-\ell-m) \text{ for } \ell \geq 2. \quad (10.7.30)$$

Consequently, (7.27) holds when $\ell \geq 2$.

How close does $(^D : f_\ell :)$ with $\ell \geq 2$ come to being a derivation? Let g and h be any two polynomials and suppose $\ell \geq 2$. We find from (4.3.7) and (7.7) the result

$$\begin{aligned} (^D : f_\ell :)(gh) &= \mathcal{P}^D : f_\ell : (gh) = \mathcal{P}^D[(: f_\ell : g)h + g : f_\ell : h] \\ &= \mathcal{P}^D\{[(^D : f_\ell :)g]h + g (^D : f_\ell :)h\} \text{ when } \ell \geq 2. \end{aligned} \quad (10.7.31)$$

We see in general that the factor of \mathcal{P}^D cannot be removed from the right side of (7.31), and thus no $(^D : f_\ell :)$ is a derivation. Finally, as the counter example (7.24) through (7.26) shows, neither (7.22) nor (7.31) hold when $\ell = 1$.

The stage is now set to study the question of convergence. Consider the operator $\hat{\mathcal{M}}$ defined by the equation

$$\hat{\mathcal{M}} = \exp[-\tau (^D : H :)], \quad (10.7.32)$$

which is the truncated analog of (7.2). Let us compute the matrix element $\langle G_s, \hat{\mathcal{M}}G_r \rangle$ with the assumption that

$$m(s), m(r) \leq D. \quad (10.7.33)$$

We find the result

$$\langle G_s, \hat{\mathcal{M}}G_r \rangle = \sum_{j=0}^{\infty} [(-\tau^j / j!)] \langle G_s, (^D : H :)^j G_r \rangle. \quad (10.7.34)$$

Let ${}^D H$ be the matrix associated with $(^D : H :)$,

$$({}^D H)_{sr} = \langle G_s, (^D : H :)G_r \rangle. \quad (10.7.35)$$

With this notation (7.17) takes the form

$$\langle G_s, \hat{\mathcal{M}}G_r \rangle = [\exp(-\tau {}^D H)]_{sr}. \quad (10.7.36)$$

Since ${}^D H$ is a finite dimensional matrix, the exponential series for $\exp(-\tau {}^D H)$ converges for all τ . See Section 3.7. It follows that all the matrix elements $\langle G_s, \hat{\mathcal{M}}G_r \rangle$, and in particular the matrix elements $\langle G_s, \hat{\mathcal{M}}G_a^1 \rangle$, are well defined for all H of the form (7.7) and any τ and any D .

What happens to the matrix elements in the limit $D \rightarrow \infty$ when truncation no longer occurs? We need to distinguish the two cases $H_1 = 0$ and $H_1 \neq 0$. Suppose $H_1 = 0$. Then, according to (7.27), $(^D : H :)$ has the property

$$(^D : H :) = \mathcal{P}^D : H :. \quad (10.7.37)$$

Consequently, from (7.16) and (7.37), $(^D : H :)$ also has the property

$$({}^D : H :)^j = \mathcal{P}^D : H :^j \mathcal{P}^D = [{}^D (: H :^j)]. \quad (10.7.38)$$

Now make use of this property in either (7.32) or its series and matrix element equivalent. Doing so gives the result

$$\hat{\mathcal{M}} = {}^D \mathcal{M}. \quad (10.7.39)$$

It follows that all matrix elements

$$({}^D M)_{rs} = \langle G_r, {}^D \mathcal{M} G_s \rangle = \langle G_r, \hat{\mathcal{M}} G_s \rangle \quad (10.7.40)$$

are well defined and are independent of D once D is large enough to satisfy (7.33).

In particular, suppose we wish to compute the coefficients in the Taylor series (7.5) through terms of degree D . Then we need the matrix elements $\langle G_s^m, \mathcal{M}G_a^1 \rangle$ for $m \leq D$. We know that all the constant terms k will vanish since we have assumed $H_1 = 0$. Also, all terms $H_\ell(z)$ in (7.1) with $\ell > (D+1)$ may be discarded since, in view of (7.6) and (7.6.16), they make no contribution to the desired terms having degree D and lower. Similarly, if we compute the terms $(: H :^j)z_a^i$ in (7.6) recursively by the relation

$$: H :^{j+1} z_a^i =: H : (: H :^j z_a^i) = [H, : H :^j z_a^i], \quad (10.7.41)$$

then we may discard at each step all those terms that would produce results having degree larger than D . Thus, all operations may be carried out within TPSA. Finally, since we know that the exponential series is convergent, it is only necessary to carry out the Poisson bracket operation (7.41) and the summation in (7.6) for successive values of j until some convergence criterion is met. Of course, all the caveats described in Section 4.1 concerning the use of Taylor series to evaluate the exponential function also apply here. Thus, as described in the next section, we may wish to consider approaches other than direct Taylor summation.

The case $H_1 \neq 0$ is more complicated. In this case there are simple examples for which not even the constant terms k in (7.5) are well defined. Consider the example of two-dimensional phase space and the Hamiltonian

$$\begin{aligned} H &= p^2/2 - (q + q_0)^4/2 + q_0^4/2 = -2qq_0^3 + (p^2/2 - 3q^2q_0^2) - 2q^3q_0 - q^4/2 \\ &= H_1 + H_2 + H_3 + H_4 \end{aligned} \quad (10.7.42)$$

where q_0 is some constant. Assume that

$$q_0 > 0, \quad (10.7.43)$$

and examine the trajectory with the initial conditions

$$t^i = q^i = p^i = 0. \quad (10.7.44)$$

From energy conservation we have the result

$$\dot{q} = [(q + q_0)^4 - q_0^4]^{1/2}, \quad (10.7.45)$$

and hence the time t along the trajectory is given by the integral

$$t(q) = \int_0^q dq' [(q' + q_0)^4 - q_0^4]^{-1/2}. \quad (10.7.46)$$

This integral is well defined and finite for all $q \geq 0$ including $q = +\infty$. That is, the trajectory reaches $q = \infty$ in *finite* time. Moreover, the trajectory also has infinite momentum at this time. Thus both k_1 and k_2 in (7.5) are divergent as τ approaches $t(\infty)$. Put other way, if in this example we set $\tau = t(\infty)$ in (7.36) and then let $D \rightarrow \infty$, we will get divergent results for at least some of the matrix elements, including the matrix elements that yield k_1 and k_2 . Note that this nonexistence of at least some of the matrix elements of \mathcal{M} is not due to any defect in the method of direct Taylor summation. It is inherent in \mathcal{M} , and must occur no matter how \mathcal{M} is computed. We conclude that Hamiltonians for which $H_1 \neq 0$ must be handled with care and on a case by case basis.

Exercises

10.7.1.

10.8 Scaling, Splitting, and Squaring

Let us assume that H is autonomous and that $H_1 = 0$. Then we know that all matrix elements are in principle well defined. However, we also know from Section 4.1 that direct computation of $\exp(-\tau {}^D H)$ by simply summing the exponential series can be problematic. We therefore wish to explore alternatives.

One approach is to use scaling and squaring. In this section we will try to generalize, for the case of operators, the method used to exponentiate matrices in Section 4.1. The matrix elements corresponding to the Taylor coefficients in (7.5), and in fact all the matrix elements of ${}^D M$, can be computed reliably from the exponential series if τ is sufficiently small. Indeed, as described earlier, to compute the Taylor coefficients one simply carries out within TPSA the Poisson brackets indicated in (7.6) for successive values of j until some convergence criterion is met; and if τ is sufficiently small we know that this convergence criterion is met for modest values of j . However, successively squaring the resulting map is not so easy: We might consider computing the full matrix ${}^D M$ for a small (scaled) value of τ and then successively squaring the result. The full matrix ${}^D M$ has dimension $[S(D, d) + 1] \times [S(D, d) + 1]$, and this number can be very large. (See Section 7.9 and Table 7.2.) Thus, computing and successively squaring it requires considerable effort. Alternatively, one might consider squaring the map using the Taylor form (7.5). In this case one must successively substitute Taylor series into themselves. This process too might be quite time consuming. Yet another approach is to compute the Taylor series for the scaled map, factorize the scaled map, and then successively square the map in the factorized Lie form (7.6.3). This process might be faster.

At this point, and after some thought, one wonders if it might be possible to compute the Lie generators for the scaled map directly without going through the intermediate steps of computing the scaled Taylor map and then factorizing the result. If so, such an approach might both be considerably faster and require less storage. The first part of this section is devoted to describing such a procedure. It is then applied to scaling and squaring.

We know that, as a special case of the previous discussion in Section 10.5, the map (7.2) has the representation

$$\mathcal{M} = \exp(t : -H :) = \cdots \exp(: f_5 :) \exp(: f_4 :) \exp(: f_3 :) \mathcal{R}. \quad (10.8.1)$$

Here we have replaced the symbol τ by t since we will soon need τ for other purposes. We have also assumed $H_1 = 0$. Equation (8.1) may be viewed as a kind of *splitting* formula that writes $\exp(t : -H :)$ as a product of factors having desirable properties. (We will learn more about other splitting formulas in Section 10.10.) As such, it has three advantages: First, (as a consequence of the Factorization Theorem) its form is fixed and potentially exact. See Section 7.6. Second, it can be concatenated easily with other maps of the same form. See Section 8.4. Consequently, it can be squared repeatedly with relative ease. Third, the exact

f_ℓ are *entire* (analytic everywhere except at ∞) functions of t , and have rapidly convergent Taylor expansions in t for small t .

We will now describe how to find the Taylor expansions (in t) for the f_ℓ . Let us begin with f_2 , which is equivalent to determining \mathcal{R} . From (5.18), (5.31), and (5.32) we find the result

$$\mathcal{R} = \mathcal{M}_2 = \exp(:f_2:) \quad (10.8.2)$$

with

$$f_2(z^i, t) = -tH_2(z^i). \quad (10.8.3)$$

Evidently f_2 is entire in t .

According to the equations of motion (5.60) through (5.66) for the remaining f_ℓ , we need to know the interaction Hamiltonian terms H_m^{int} . Following (5.41), they are given by the relations

$$H_m^{\text{int}}(z^i, t) = H_m(\mathcal{M}_2 z^i) = \mathcal{M}_2 H_m(z^i) = \exp(t : -H_2 :) H_m(z^i). \quad (10.8.4)$$

We note that because H is assumed to be autonomous, the time dependence of the H_m^{int} comes entirely from the factor $\exp(t : -H_2 :)$. From (8.4) we see that each H_m^{int} has the Taylor expansion

$$H_m^{\text{int}} = \sum_{\ell=0}^{\infty} (1/\ell!)(-t)^\ell : H_2 :^\ell H_m. \quad (10.8.5)$$

From (8.4) and (5.25) through (5.35) we know that H_m^{int} can be written as well in the form

$$H_m^{\text{int}} = H_m(\mathcal{M}_2 z^i) = H_m(Rz^i) \quad (10.8.6)$$

with

$$R = \exp(tJS). \quad (10.8.7)$$

We know that the matrix exponential function converges and therefore is analytic for all t . Finally, H_m is a polynomial in z^i . It follows that H_m^{int} is entire in t and consequently (8.5) converges for all t .

Let us next compute f_3 . From (5.60) and the initial condition $f_3(0) = 0$ we obtain the result

$$f_3 = - \int_0^t dt' H_3^{\text{int}}. \quad (10.8.8)$$

This integral can be done using the representation (8.5) to give the series

$$f_3 = \sum_{\ell=1}^{\infty} (1/\ell!)(-t)^\ell : H_2 :^{\ell-1} H_3. \quad (10.8.9)$$

Since the right side of (5.60), namely H_3^{int} , is entire in t , it follows from Poincaré's Theorem 1.3.3 that f_3 is entire in t . More simply, we just observe that the integral of an entire function is also an entire function. Either way, we conclude that (8.9) converges for all t .

The computation of f_4 is a bit more involved. From (5.61) and the initial condition $f_4(0) = 0$ we find that f_4 contains two terms:

$$f_4 = - \int_0^t dt' H_4^{\text{int}} - (1/2) \int_0^t dt' : f_3 : H_3^{\text{int}}. \quad (10.8.10)$$

We will refer to the first term as the *direct* term since it is produced by H_4^{int} , which is of the same degree as f_4 . The second term will be called a *feed-up* term since it arises from the combined effect of the lower-degree term H_3^{int} and the lower-degree term f_3 (which comes from H_3^{int}). Thus we write

$$f_4 = f_4^{\text{d}} + f_4^{\text{fu}}. \quad (10.8.11)$$

For f_4^{d} we use (8.5) to get a result analogous to that for f_3 ,

$$f_4^{\text{d}} = \sum_{\ell=1}^{\infty} (1/\ell!)(-t)^{\ell} : H_2 :^{\ell} H_4. \quad (10.8.12)$$

Again we know that f_4^{d} is entire in t and the series (8.12) converges for all t . For the feed-up term we use the series representations (8.5) and (8.9) to get the result

$$\begin{aligned} f_4^{\text{fu}} &= -(1/2) \int_0^t dt' : f_3 : H_3^{\text{int}} = -(1/2) \int_0^t dt' [f_3, H_3^{\text{int}}] \\ &= -(1/2) \int_0^t dt' \sum_{\ell=1}^{\infty} \sum_{m=1}^{\infty} (1/\ell!)(1/m!)(-t')^{\ell+m} [: H_2^{\ell-1} : H_3, : H_2 :^m H_3] \\ &= (1/2) \sum_{\ell=1}^{\infty} \sum_{m=0}^{\infty} (1/\ell!)(1/m!)[1/(\ell+m+1)](-t)^{\ell+m+1} [: H_2 :^{\ell-1} H_3, : H_2 :^m H_3] \\ &= -(1/12)t^3[H_3, : H_2 : H_3] + (1/24)t^4[H_3, : H_2 :^2 H_3] - t^5\{(1/80)[H_3, : H_2 :^3 H_3] \\ &\quad + (1/120)[: H_2 : H_3, : H_2 :^2 H_3]\} + \dots \end{aligned} \quad (10.8.13)$$

The quantity f_4^{fu} is also entire in t and the series (8.13) converges for all t .

The computation of the remaining f_j is similar. In each case there is a direct term analogous to (8.9) and (8.12). There are also feed-up terms arising from multiple Poisson brackets involving lower degree terms in the Hamiltonian. By contrast, there are no *feed-down* terms. To find a given f_{ℓ} , it is only necessary to know the $H_{\ell'}$ with $\ell' \leq \ell$. We also observe that all the formulas for the f_{ℓ} are expressible entirely in terms of Poisson brackets. All formulas involve only operations within the Poisson bracket Lie algebra generated by the H_m . Such results are to be expected in general as a consequence of the BCH theorem. Analogous formulas, but involving instead commutators of vector fields, are to be expected in the non-Hamiltonian case. The coefficients in the various series should be universal. Also, all the series represent entire functions and therefore converge for all values of t . Finally, we note that the rate of convergence of the various series depends only on the properties of $t : H_2 :$ (and hence tH_2), because that is the only term that appears infinitely often in the series.

To fix these ideas more clearly in the mind, we will also consider in some detail the computation of f_5 . From (5.62) we find the result

$$f_5 = f_5^{\text{d}} + f_5^{\text{fu}} \quad (10.8.14)$$

where

$$f_5^{\text{d}} = - \int_0^t dt' H_5^{\text{int}} \quad (10.8.15)$$

and

$$f_5^{\text{fu}} = \int_0^t dt' [: f_3 : (-H_4^{\text{int}}) + (1/3) : f_3 :^2 (-H_3^{\text{int}})]. \quad (10.8.16)$$

The direct term has the expansion

$$f_5^{\text{d}} = \sum_{\ell=1}^{\infty} (1/\ell!) (-t)^{\ell} : H_2 :^{\ell-1} H_5. \quad (10.8.17)$$

To find the expansion for the feed-up term we insert the previously obtained expressions for H_3^{int} , H_4^{int} , f_3 , and f_4 in (8.16) to obtain the result

$$\begin{aligned} f_5^{\text{fu}} &= \sum_{\ell=1}^{\infty} \sum_{m=0}^{\infty} (1/\ell!) (1/m!) [1/(\ell+m+1)] (-t)^{\ell+m+1} [: H_2 :^{\ell-1} H_3, : H_2 :^m H_4] \\ &- (1/3) \sum_{\ell=1}^{\infty} \sum_{m=0}^{\infty} \sum_{n=1}^{\infty} (1/\ell!) (1/m!) (1/n!) [1/(\ell+m+n+1)] \\ &\quad (-t)^{\ell+m+n+1} [: H_2 :^{\ell-1} H_3, [: H_2 :^m H_3, : H_2 :^{n-1} H_3]] \\ &= (1/2)t^2[H_3, H_4] - t^3\{(1/3)[H_3, : H_2 : H_4] + (1/6)[: H_2 : H_3, H_4]\} \\ &+ t^4\{(-1/24)[H_3, [: H_2 : H_3, H_3]] + (1/8)[H_3, : H_2 :^2 H_4] \\ &+ (1/8)[: H_2 : H_3, : H_2 : H_4] + (1/24)[: H_2 :^2 H_3, H_4]\} \\ &+ t^5\{(1/45)[H_3, [: H_2 :^2 H_3, H_3]] - (1/30)[H_3, : H_2 :^3 H_4] \\ &+ (1/60)[: H_2 : H_3, [: H_2 : H_3, H_3]] - (1/20)[: H_2 : H_3, : H_2 :^2 H_4] \\ &- (1/30)[: H_2 :^2 H_3, : H_2, H_4] - (1/120)[: H_2 :^3 H_3, H_4]\} + \dots. \end{aligned} \quad (10.8.18)$$

At this point we might quote formulas for the f_j for the next few higher values of j and up to some power in t . Instead we observe that, rather than the reverse factorization (8.1) which we write in the form

$$\mathcal{M} = \exp(t : -H :) = \dots \exp[: f_5(t) :] \exp[: f_4(t) :] \exp[: f_3(t) :] \mathcal{R}(t), \quad (10.8.19)$$

it is often more convenient to have results for the forward factorization

$$\mathcal{M} = \exp(t : -H :) = \mathcal{R}'(t) \exp[: g_3(t) :] \exp[: g_4(t) :] \exp[: g_5(t) :] \dots. \quad (10.8.20)$$

The relation between the two factorizations is immediate. Suppose we invert both sides of (8.19). Doing so gives the result

$$\exp(t : +H :) = \mathcal{R}^{-1}(t) \exp[- : f_3(t) :] \exp[- : f_4(t) :] \exp[- : f_5(t) :] \dots. \quad (10.8.21)$$

Next replace t by $-t$ in (8.21) to find the relation

$$\exp(t : -H :) = \mathcal{R}^{-1}(-t) \exp[- : f_3(-t) :] \exp[- : f_4(-t) :] \exp[- : f_5(-t) :] \dots. \quad (10.8.22)$$

Since the factorization (8.20) is unique, comparison of the first factors on the right sides of (8.20) and (8.22) shows that

$$\mathcal{R}'(t) = \mathcal{R}^{-1}(-t). \quad (10.8.23)$$

But from (8.2) and (8.3) we find that

$$\mathcal{R}^{-1}(-t) = \mathcal{R}(t) \quad (10.8.24)$$

so that

$$\mathcal{R}'(t) = \mathcal{R}(t). \quad (10.8.25)$$

This result is also evident on general grounds. Finally, upon comparing the remaining factors in (8.20) and (8.22), we conclude that

$$g_m(t) = -f_m(-t). \quad (10.8.26)$$

We now quote formulas, obtained by symbolic manipulation, for the first few g_m through terms of order t^5 . Again each g_m is the sum of a direct and a feed-up term,

$$g_m = g_m^d + g_m^{fu}. \quad (10.8.27)$$

For the direct terms we have in general the formula

$$g_m^d = - \sum_{\ell=1}^{\infty} (1/\ell!) t^{\ell} : H_2 :^{\ell-1} H_m. \quad (10.8.28)$$

For the feed-up terms g_3^{fu} through g_6^{fu} we find, through terms of order t^5 , the result

$$g_3^{fu} = 0, \quad (10.8.29)$$

$$\begin{aligned} g_4^{fu} &= -(1/2) \sum_{\ell=1}^{\infty} \sum_{m=0}^{\infty} (1/\ell!) (1/m!) [1/(\ell+m+1)] t^{\ell+m+1} [: H_2 :^{\ell-1} H_3, : H_2 :^m H_3] \\ &= (1/12)t^3 [: H_2 : H_3, H_3] + (1/24)t^4 [: H_2 :^2 H_3, H_3] \\ &\quad + t^5 \{(1/80) [: H_2 :^3 H_3, H_3] + (1/120) [: H_2 :^2 H_3, : H_2 : H_3]\} + \dots, \end{aligned} \quad (10.8.30)$$

$$\begin{aligned} g_5^{fu} &= -(1/2)t^2 [H_3, H_4] - t^3 \{(1/3)[H_3, : H_2 : H_4] + (1/6)[: H_2 : H_3, H_4]\} \\ &\quad - t^4 \{(-1/24)[H_3, [: H_2 : H_3, H_3]] + (1/8)[H_3, : H_2 :^2 H_4] \\ &\quad + (1/8)[: H_2 : H_3, : H_2 : H_4] + (1/24)[: H_2 :^2 H_3, H_4]\} \\ &\quad + t^5 \{(1/45)[H_3, [: H_2 :^2 H_3, H_3]] - (1/30)[H_3, : H_2 :^3 H_4] \\ &\quad + (1/60)[: H_2 : H_3, [: H_2 : H_3, H_3]] - (1/20)[: H_2 : H_3, : H_2 :^2 H_4] \\ &\quad - (1/30)[: H_2 :^2 H_3, : H_2, H_4 :] - (1/120)[: H_2 :^3 H_3, H_4]\} + \dots, \end{aligned} \quad (10.8.31)$$

$$\begin{aligned}
g_6^{\text{fu}} = & -(1/2)t^2[H_3, H_5] \\
& + t^3\{(-1/6)[H_3, [H_3, H_4]] - (1/3)[H_3, :H_2:H_5] \\
& - (1/3)[H_3, :H_2:H_5] - (1/6)[:H_2:H_3, H_5] \\
& + (1/12)[:H_2:H_4, H_4]\} \\
& - t^4\{(1/8)[H_3, [H_3, :H_2:H_4]] + (1/6)[H_3, [:H_2:H_3, H_4]] \\
& + (1/8)[H_3 : H_2 :^2 H_5] - (1/48)[H_4, [:H_2:H_3, H_3]] \\
& + (1/16)[:H_2:H_3, [H_3, H_4]] + (1/8)[:H_2:H_3, :H_2:H_5] \\
& + (1/24)[:H_2 :^2 H_3, H_5] - (1/24)[:H_2 :^2 H_4, H_4]\} \\
& + t^5\{(1/80)[H_3, [H_3, [:H_2:H_3, H_3]]] - (1/20)[H_3, [H_3, :H_2 :^2 H_4]] \\
& - (1/20)[H_3, [:H_2:H_3, :H_2:H_4]] - (1/60)[H_3, [:H_2 :^2 H_3, H_4]] \\
& - (1/30)[H_3, :H_2 :^3 H_5] + (1/80)[H_4, [:H_2 :^2 H_3, H_3]] \\
& - (1/20)[:H_2:H_3, [H_3, :H_2:H_4]] - (1/40)[:H_2:H_3, [:H_2:H_3, H_4]] \\
& - (1/20)[:H_2:H_3, :H_2 :^2 H_5] + (1/240)[:H_2:H_4, [:H_2:H_3, H_3]] \\
& - (1/60)[:H_2 :^2 H_3, [H_3, H_4]] - (1/30)[:H_2 :^2 H_3, :H_2:H_5] \\
& + (1/20)[:H_2 :^2 H_4, :H_2:H_4] - (1/120)[:H_2 :^3 H_3, H_5] \\
& + (1/180)[:H_2 :^3 H_4, H_4]\} + \dots
\end{aligned} \tag{10.8.32}$$

These results were obtained using a Mathematica program. See Appendix E.

We have derived the splitting formula (8.1) with the f_m given by (8.9), (8.11) through (8.18), and analogous expressions. Equivalently, we have also derived the splitting formula (8.20) with the g_m given by (8.27) through (8.32) and analogous expressions. How are these formulas to be used? With the aid of scaling and squaring we have the result

$$\begin{aligned}
\mathcal{M} = & \exp(t : -H :) = \{\exp[(t/2^n) : -H :]\}^{2^n} \\
= & \{\dots \{\{\exp[(t/2^n) : -H :]\}^2\}^2 \dots\}^2 \text{ (n squarings).}
\end{aligned} \tag{10.8.33}$$

See Section 4.1 and (4.1.6). Now define a quantity τ by writing

$$\tau = t/2^n. \tag{10.8.34}$$

Next insert, for example, the splitting formula (8.20) into (8.33). Doing so gives the result

$$\begin{aligned}
\mathcal{M} = & \exp(t : -H :) = \mathcal{R}(t) \exp[:g_3(t):] \exp[:g_4(t):] \exp[:g_5(t):] \dots \\
= & \{\mathcal{R}(\tau) \exp[:g_3(\tau):] \exp[:g_4(\tau):] \exp[:g_5(\tau):] \dots\}^{2^n} \\
= & \{\dots \{\{\mathcal{R}(\tau) \exp[:g_3(\tau):] \exp[:g_4(\tau):] \exp[:g_5(\tau):] \dots\}^2\}^2 \dots\}^2 \text{ (n squarings).}
\end{aligned} \tag{10.8.35}$$

Finally, suppose we choose n to be large enough so that τ is sufficiently small that the truncated series (8.28) through (8.32) give accurate results for the $g_m(\tau)$. Then (8.35) gives an accurate result for \mathcal{M} .

How large must n be (or, equivalently, how small must τ be) for the truncated series to give accurate results for the $g_m(\tau)$? We have already observed that the convergence of these series depends only on the properties of τH_2 . Let us explore what can be said about terms of the form $(\tau : H_2 :)^{\ell} H_m$, which are the common ingredient of all the series. As before we

write H_2 in a form analogous to (5.28) except that S is now a constant time-independent matrix. We then find, in analogy to (5.29), the relation

$$: H_2 : z_a = - \sum_b (JS)_{ab} z_b. \quad (10.8.36)$$

Consider the matrix $(-JS)^T$. In general, barring degeneracy, it has $2n$ eigenvectors v^j with eigenvalues λ_j :

$$(-JS)^T v^j = \lambda_j v^j. \quad (10.8.37)$$

(Here, for the moment, n is the number of degrees of freedom, and *not* the number of squarings.) Define $2n$ first-degree polynomials h_1^j by the rule

$$h_1^j = \sum_a v_a^j z_a. \quad (10.8.38)$$

In general, again barring degeneracy, the h_1^j will be functionally independent and will span the space of all first-degree polynomials. Let us compute the effect of $: H_2 :$ on the h_1^j . From (8.36) through (8.38) we find the result

$$\begin{aligned} : H_2 : h_1^j &= \sum_a v_a^j : H_2 : z_a = \sum_{a,b} v_a^j (-JS)_{ab} z_b \\ &= \sum_b \left\{ \sum_a [(-JS)^T]_{ba} v_a^j \right\} z_b = \lambda_j \sum_b v_b^j z_b \\ &= \lambda_j h_1^j. \end{aligned} \quad (10.8.39)$$

We know from (7.6.14) that the (linear) operator $: H_2 :$ maps the space \mathcal{P}_m into itself. (Note that here \mathcal{P}_m denotes a vector space and not a projection operator.) What is its largest eigenvalue when acting on this space? Consider for example, the degree 3 homogeneous polynomials h_3^{ijk} defined by the relation

$$h_3^{ijk} = h_1^i h_1^j h_1^k. \quad (10.8.40)$$

In general, these polynomials will span the space of third-degree polynomials. Since $: H_2 :$ is a derivation, and in view of (8.39), we find the result

$$\begin{aligned} : H_2 : h_3^{ijk} &= (: H_2 : h_1^i) h_1^j h_1^k + h_1^i (: H_2 : h_1^j) h_1^k + h_1^i h_1^j (: H_2 : h_1^k) \\ &= (\lambda_i + \lambda_j + \lambda_k) h_3^{ijk}. \end{aligned} \quad (10.8.41)$$

Let λ_{\max} be the modulus of the eigenvalue with the largest absolute value,

$$\lambda_{\max} = \max_j |\lambda_j|. \quad (10.8.42)$$

We conclude from relations of the form (8.41) that the eigenvalues of $: H_2 :$ when acting on \mathcal{P}_m are bounded by the quantity $(m\lambda_{\max})$.

In general, given the matrix $(-JS)^T$, one can compute its eigenvalues. However, this computation requires some work, and often an estimate that requires less computation is

sufficient. Let λ_k be the eigenvalue of $(-JS)^T$ having the largest absolute value. Then from (8.39) we have the result

$$\lambda_k v^k = (-JS)^T v^k = (SJ)v^k. \quad (10.8.43)$$

Here we have used the fact that S is symmetric and J is antisymmetric. Now take norms of both sides of (8.43) to get the result

$$\|\lambda_k v^k\| = \|SJv^k\| \leq \|SJ\| \|v^k\|. \quad (10.8.44)$$

The left side of (8.44) can be manipulated using (3.7.8) and (8.42) to give the relation

$$\|\lambda_k v^k\| = |\lambda_k| \|v^k\| = \lambda_{\max} \|v^k\|, \quad (10.8.45)$$

and we conclude upon comparison with (8.44) that λ_{\max} has the bound

$$\lambda_{\max} \leq \|SJ\|. \quad (10.8.46)$$

Moreover, we may also write the equation

$$SJ = -JJSJ. \quad (10.8.47)$$

Now take the norm of both sides of (8.47) to get the bound

$$\|SJ\| = \| -JJSJ \| = \|JJSJ\| \leq \|J\| \|JS\| \|J\|. \quad (10.8.48)$$

Suppose the matrix norm $\| \cdot \|$ to be employed has the property

$$\|J\| = 1, \quad (10.8.49)$$

which is true of the norm (3.7.15). Then (8.46) through (8.49) may be combined to give the bound

$$\lambda_{\max} \leq \|JS\|. \quad (10.8.50)$$

We now have all the tools in hand to examine the convergence of series that involve terms of the form $(\tau : H_2 :)^{\ell} H_m$. According to the previous discussion the convergence of such series is governed by the size of the terms $(m\tau\lambda_{\max})^{\ell}$ with λ_{\max} bounded by (8.50). Suppose, for example, we truncate the series (8.28) for $g_m^d(\tau)$ by selecting some N and discarding all terms with $\ell > N$. Let us estimate the error committed in doing so by examining the size of the first neglected term,

$$\text{first neglected term} = -[1/(N+1)!]\tau^{N+1} : H_2 :^N H_m. \quad (10.8.51)$$

We know from our previous discussion that $(: H_2 :^N H_m)$ behaves at worst according to the estimate

$$: H_2 :^N H_m \sim (m\lambda_{\max})^N H_m. \quad (10.8.52)$$

Also, we expect from the first term in (8.28) that $g_m^d(\tau)$ itself will be of order (τH_m) . Consequently, using (8.51) and (8.52), we should make the comparison

$$\tau H_m \stackrel{?}{\leftrightarrow} [1/(N+1)!]\tau^{N+1}(m\lambda_{\max})^N H_m. \quad (10.8.53)$$

We conclude that the *relative* error in computing $g_m^d(\tau)$, and hence also $\mathcal{M}(t)$, has the estimate

$$\text{relative error} \sim [1/(N+1)!](\tau m \lambda_{\max})^N. \quad (10.8.54)$$

Finally, let us define a quantity λ by the rule

$$\lambda = t \lambda_{\max}. \quad (10.8.55)$$

By (8.50) it has the estimate

$$\lambda \leq \|tJS\|, \quad (10.8.56)$$

and is dimensionless. With the aid of (8.34), (8.54), and (8.55) we see that the relative error can be written in the form

$$\text{relative error} \sim [1/(N+1)!](m\lambda/2^n)^N. \quad (10.8.57)$$

Suppose, for example, we set $N = 5$, limit our attention to the cases $m \leq 8$, and select n such that

$$(8\lambda/2^n) < (1/20). \quad (10.8.58)$$

[Note that λ and hence the required n can be computed in advance using (8.58).] Then we find from (8.57) that the relative error has the estimate

$$\text{relative error} \sim (1/6!)(1/20)^5 \simeq 4 \times 10^{-10}. \quad (10.8.59)$$

Although we have only estimated the error due to truncating the series for g_m^d , we expect that the result of truncating the other series at the same N will be comparable as long as (8.58) is satisfied. We conclude from (8.59) that (just as scaling and squaring works well for matrix exponentiation) scaling, splitting, and squaring works well for computing \mathcal{M} in the factorized product forms (8.1) and (8.20). It has high, controllable, and predictable accuracy. Of course, the n required to satisfy a relation of the form (8.58) varies from Hamiltonian to Hamiltonian. However, just as in the matrix case, the n required to achieve some specific accuracy grows only logarithmically with the norm of (tJS) , and for any given Hamiltonian the accuracy increases very rapidly for increasing n . Correspondingly, because the number of required operations is relatively small and no cancellations are required to occur between large terms, problems with round-off error are minimized. Finally, with regard to computational speed, the method of scaling, splitting, and squaring is far faster than numerical integration. (Of course, numerical integration is required in the nonautonomous case.) It is also faster and, as expected, far more reliable than direct use of the Taylor series (7.6). The price that has been paid for this good performance is that one must know the expansions of the form (8.28) through (8.32) as well as concatenation formulas of the form (8.4.31) through (8.4.36). By contrast, the implementation of the Taylor series (7.6) to any order is straightforward.

Exercises

10.8.1. Show that if the matrix norm has the property (8.49), then one has the equation

$$\|JS\| = \|SJ\| = \|S\|. \quad (10.8.60)$$

10.9 Canonical Treatment of Errors

Let $H(z, t)$ be a general, possibly time-dependent, Hamiltonian that is analytic about the origin and consequently has an expansion in homogeneous polynomials in z of the form

$$H(z, t) = \sum_{m=1}^{\infty} H_m(z, t). \quad (10.9.1)$$

(Here we drop a possible z independent term H_0 since it has no effect on the equations of motion.) Moreover, we shall assume that H_1 is *small* so that $z = 0$ is close to being a solution to the equation of motion generated by H . Such Hamiltonians often arise in connection with the description of errors. For example, we will see in Chapter 26 that both mispowered dipole bending magnets and dipole steering magnets are described by Hamiltonians of this form.

Hamiltonians of the form (9.1) with H_1 small can be treated by the method of Section 10.5 and, in the autonomous case, also by the method of Section 10.7. The method of Section 10.5 requires determination of the design trajectory $z^d(t)$, and then provides an expansion about this trajectory. However, since $z = 0$ is nearly a trajectory, we may prefer an expansion of the transfer map \mathcal{M} about $z = 0$. Such an expansion is provided, in the autonomous case, by the method of Section 10.7, but requires the summation of Taylor series for the exponential function. We have seen that the use of such series may be problematic.

The purpose of this section is to develop a method for expanding the transfer map about $z = 0$ under the assumption that H_1 is small. In essence, we will produce a simultaneous expansion both in z and in powers of some parameter that characterizes the smallness of H_1 . This method is applicable to both the time dependent and time independent cases.

The method is based on an *enlargement* of $2n$ dimensional phase space to include the extra variables q_{n+1} and p_{n+1} . This is the same enlargement that was used in Section 9.4, and the method based on it will be referred to as a *canonical* treatment of errors. As before, let us use the symbol \hat{z} to denote the coordinates in this enlarged phase space,

$$\hat{z} = (q_1 \cdots q_n, q_{n+1}, p_1 \cdots p_n, p_{n+1}). \quad (10.9.2)$$

Next, modify the Hamiltonian (9.1) to obtain the Hamiltonian \hat{H} defined by the rule

$$\hat{H}(\hat{z}, t) = (q_{n+1})H_1(z, t) + \sum_{m=2}^{\infty} H_m(z, t). \quad (10.9.3)$$

We will see that the transfer map for the Hamiltonian \hat{H} , which we will call $\hat{\mathcal{M}}$, can be computed using the methods we have developed previously, and that $\hat{\mathcal{M}}$ contains the information we seek.

We begin by noting that $\hat{z} = 0$ is a trajectory for \hat{H} , and hence the transfer map $\hat{\mathcal{M}}$ produced by \hat{H} maps the origin of the enlarged phase space into itself. Indeed, with respect to the enlarged phase-space variables, the Hamiltonian \hat{H} has the expansion

$$\hat{H}(\hat{z}, t) = \sum_{m=2}^{\infty} \hat{H}_m(\hat{z}, t) \quad (10.9.4)$$

with

$$\hat{H}_2(\hat{z}, t) = (q_{n+1})H_1(z, t) + H_2(z, t), \quad (10.9.5)$$

$$\hat{H}_m(\hat{z}, t) = H_m(z, t) \text{ for } m > 2. \quad (10.9.6)$$

Observe that this expansion begins with homogeneous polynomials of degree *two*. Correspondingly, the transfer map $\hat{\mathcal{M}}$ can be written in the factored product form

$$\hat{\mathcal{M}} = \hat{\mathcal{R}} \exp(:\hat{f}_3:) \exp(:\hat{f}_4:) \exp(:\hat{f}_5:) \cdots. \quad (10.9.7)$$

Here the \hat{f}_m denote homogeneous polynomials of degree m in the enlarged phase-space variables \hat{z} . The map $\hat{\mathcal{M}}$ can be computed in both the time dependent and time independent cases using the method of Section 10.5, and in the time independent case using the methods of Sections 10.7 and 10.8.

By construction \hat{H} is independent of p_{n+1} , and therefore there is the obvious equation of motion

$$\dot{q}_{n+1} = \partial \hat{H} / \partial p_{n+1} = 0. \quad (10.9.8)$$

It follows that $\hat{\mathcal{M}}$ leaves q_{n+1} unchanged,

$$q_{n+1}^f = \hat{\mathcal{M}} q_{n+1}^i = q_{n+1}^i. \quad (10.9.9)$$

But now the reasoning presented in the beginning of Section 9.4 applies. We conclude that $\hat{\mathcal{R}}$ must satisfy the relation

$$\hat{\mathcal{R}} q_{n+1} = q_{n+1}, \quad (10.9.10)$$

and its associated matrix \hat{R} must (for the case $n = 3$) be of the general form (9.4.84). Also, the \hat{f}_m in (9.7) must be independent of p_{n+1} ,

$$\partial \hat{f}_m / \partial p_{n+1} = 0. \quad (10.9.11)$$

They will depend only on the z^i and q_{n+1}^i . Finally, we see that the relation

$$\hat{z}^f = \hat{\mathcal{M}} \hat{z}^i \quad (10.9.12)$$

provides an expansion of z^f in terms of z^i and q_{n+1}^i , and p_{n+1}^i does not occur in this expansion.

Evidently the size of q_{n+1}^i governs the effect of the term $(q_{n+1})H_1$ in (9.3), and an expansion in powers of q_{n+1} is, in effect, an expansion in terms of powers of H_1 . Thus, after some final result for $\hat{\mathcal{M}}$ (or some consequence of $\hat{\mathcal{M}}$) has been obtained as a series in q_{n+1} up to some order, we may set $q_{n+1} = 1$ in this series to obtain a result appropriate to the original Hamiltonian (9.1) when H_1 in this Hamiltonian is treated as a perturbation through the same order.

A simple example helps clarify this approach. Consider the *displaced* one-dimensional harmonic oscillator described by the Hamiltonian

$$H = (p_1^2 + q_1^2)/2 + \delta q_1 \quad (10.9.13)$$

where δ is a small quantity, not to be confused with the δ_a in Section 9.4. It is easily verified that the equations of motion associated with (9.13) have the solution

$$q_1(t) = -\delta + (q_1^i + \delta) \cos t + p_1^i \sin t, \quad (10.9.14)$$

$$p_1(t) = -(q_1^i + \delta) \sin t + p_1^i \cos t, \quad (10.9.15)$$

where the initial time is taken to be $t^i = 0$. According to (9.3) the modified Hamiltonian associated with (9.13) is given by the relation

$$\hat{H} = (p_1^2 + q_1^2)/2 + \delta q_1 q_2. \quad (10.9.16)$$

Since \hat{H} is time independent, and all the \hat{H}_m with $m > 2$ happen to vanish in this simple case, we have the immediate result

$$\hat{\mathcal{M}} = \exp(-t : \hat{H} :) = \exp(-t : \hat{H}_2 :) = \hat{\mathcal{R}}(t). \quad (10.9.17)$$

The map $\hat{\mathcal{R}}(t)$ is in turn described by the matrix $\hat{R}(t)$ given by

$$\hat{R} = \exp(t \hat{J} \hat{S}). \quad (10.9.18)$$

Here we have placed a hat over J to indicate that the 4×4 J is to be used. Also, for convenience, we have used the ordering (9.4.15). Correspondingly, \hat{J} is of the form (3.2.10). With this convention, and according to (5.28), \hat{S} is the matrix

$$\hat{S} = \begin{pmatrix} 1 & 0 & \delta & 0 \\ 0 & 1 & 0 & 0 \\ \delta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (10.9.19)$$

The exponentiation (9.18) can be carried out to give the result

$$\hat{R} = \begin{pmatrix} \cos t & \sin t & \delta(\cos t - 1) & 0 \\ -\sin t & \cos t & -\delta \sin t & 0 \\ 0 & 0 & 1 & 0 \\ -\delta \sin t & \delta(\cos t - 1) & \delta^2(t - \sin t) & 1 \end{pmatrix}. \quad (10.9.20)$$

Correspondingly, in analogy with (5.27), we find the relations

$$q_1(t) = q_1^i \cos t + p_1^i \sin t + q_2^i \delta(\cos t - 1), \quad (10.9.21)$$

$$p_1(t) = -q_1^i \sin t + p_1^i \cos t - q_2^i \delta \sin t, \quad (10.9.22)$$

$$q_2(t) = q_1^i, \quad (10.9.23)$$

$$p_2(t) = -q_1^i \delta \sin t + p_1^i \delta(\cos t - 1) + q_2^i \delta^2(t - \sin t) + p_2^i. \quad (10.9.24)$$

We see, as advertised, that (9.21) and (9.22) reduce to (9.14) and (9.15), respectively, when we set $q_2 = 1$. Also, (9.23) is consistent with (9.8) and (9.9). Finally, although of no particular interest for our purposes, (9.24) is consistent with the equation of motion

$$\dot{p}_2 = -\partial \hat{H} / \partial q_2 = -\delta q_1. \quad (10.9.25)$$

In this example we have been able to solve for the map $\hat{\mathcal{M}}$ exactly, and have found that the result for $z(t)$ when q_{n+1} is set to one agrees with the corresponding trajectory, which we were also able to find exactly, for the original problem. For most cases we will only compute

$\hat{\mathcal{M}}$ to some order in the phase-space variables including the variable q_{n+1} . In such cases we expect that the correspondingly results for $z(t)$ will be correct through the same order in δ where δ measures the size of H_1 .

There is one last observation to be made. In the example just described the computation of \hat{R} was somewhat complicated because it was 4×4 , and the information about $p_2(t)$, which was contained in \hat{R} , was of no interest. Is there a way of bypassing such complications? In some cases the answer is yes. For the example problem let us rewrite (9.17) in the form

$$\hat{\mathcal{R}}(t) = \exp(: \hat{f}_2 :) \quad (10.9.26)$$

with \hat{f}_2 defined by

$$\hat{f}_2 = -t\hat{H}_2 = -t[(p_1^i)^2 + (q_1^i)^2]/2 - t\delta q_1^i q_2^i. \quad (10.9.27)$$

When \hat{f}_2 is regarded as a function of q_1^i and p_1^i , which are the dynamical variables of interest, we see that it can be written in the form

$$\hat{f}_2 = f_1^{(2)} + f_2^{(2)} \quad (10.9.28)$$

where $f_1^{(2)}$ is a homogeneous polynomial of degree 1 in the variables of interest,

$$f_1^{(2)} = -t\delta q_2^i q_1^i, \quad (10.9.29)$$

and $f_2^{(2)}$ is a polynomial of degree 2 in these variables,

$$f_2^{(2)} = -t[(p_1^i)^2 + (q_1^i)^2]/2. \quad (10.9.30)$$

Here we have used a subscript on f to indicate degree in the variables of interest, and a superscript to indicate overall degree. This notation is similar to that of Section 9.3 where the subscript served the same purpose and the superscript indicated overall grade. Indeed, at this stage, we may simply view q_2^i as a parameter that plays the same role as ϵ did in Section 9.3.

With the decomposition (9.28) we may rewrite $\hat{\mathcal{M}}$, which we now simply call \mathcal{M} because we are no longer treating q_2 and p_2 as dynamical variables, in the form

$$\mathcal{M} = \exp(: f_1^{(2)} + f_2^{(2)} :). \quad (10.9.31)$$

At this stage we use (9.2.4) to rewrite \mathcal{M} in the form

$$\mathcal{M} = \exp(: f_2 :) \exp(: f_1 :) \quad (10.9.32)$$

where, according to (9.2.7) and (9.2.9),

$$f_2 = f_2^{(2)} = -(t/2)[(p_1^i)^2 + (q_1^i)^2], \quad (10.9.33)$$

$$f_1 = \text{iex} (- : f_2^{(2)} :) f_1^{(2)}. \quad (10.9.34)$$

Simple calculation gives the result

$$\begin{aligned} \exp(-\tau : f_2^{(2)} :) f_1^{(2)} &= \exp[(\tau t/2) : (p_1^i)^2 + (q_1^i)^2 :] (-t\delta q_2^i q_1^i) \\ &= -t\delta q_2^i \exp[(\tau t/2) : (p_1^i)^2 + (q_1^i)^2 :] q_1^i \\ &= -t\delta q_2^i [q_1^i \cos(\tau t) - p_1^i \sin(\tau t)]. \end{aligned} \quad (10.9.35)$$

Consequently, we find for f_1 the explicit result

$$\begin{aligned} f_1 &= \text{ie}x (- : f_2^{(2)} :) f_1^{(2)} = \int_0^1 d\tau \exp(-\tau : f_2^{(2)} :) f_1^{(2)} \\ &= -t\delta q_2^i \int_0^1 d\tau [q_1^i \cos(t\tau) - p_1^i \sin(t\tau)] \\ &= -\delta q_2^i [q_1^i \sin t + p_1^i (\cos t - 1)]. \end{aligned} \quad (10.9.36)$$

Here we have used (8.7.9). Let us now find the effect of \mathcal{M} , with \mathcal{M} given by (9.32), when it acts on q_1^i and p_1^i . We find for $\exp(: f_1 :)$ and $\exp(: f_2 :)$ the results

$$\exp(: f_1 :) q_1^i = q_1^i + \delta q_2^i (\cos t - 1), \quad (10.9.37)$$

$$\exp(: f_1 :) p_1^i = p_1^i - \delta q_2^i \sin t, \quad (10.9.38)$$

$$\exp(: f_2 :) q_1^i = q_1^i \cos t + p_1^i \sin t, \quad (10.9.39)$$

$$\exp(: f_2 :) p_1^i = -q_1^i \sin t + p_1^i \cos t. \quad (10.9.40)$$

Consequently we find for \mathcal{M} the net results

$$q_1(t) = \mathcal{M} q_1^i = \delta q_2^i (\cos t - 1) + q_1^i \cos t + p_1^i \sin t, \quad (10.9.41)$$

$$p_1(t) = \mathcal{M} p_1^i = -\delta q_2^i \sin t - q_1^i \sin t + p_1^i \cos t. \quad (10.9.42)$$

Let us compare these results with those of (9.14) and (9.15), which can be written in the form

$$q_1(t) = \delta(\cos t - 1) + q_1^i \cos t + p_1^i \sin t, \quad (10.9.43)$$

$$p_1(t) = -\delta \sin t - q_1^i \sin t + p_1^i \cos t. \quad (10.9.44)$$

We see that (9.41) and (9.43), and (9.42) and (9.44), agree when we set $q_2^i = 1$. Note that, had we wished, we could have set $q_2^i = 1$ at an earlier stage before carrying out the calculations (9.37) through (9.42).

There is a moral to be learned from this last exercise. Quite generally, we see that once $\hat{\mathcal{M}}$ is computed in the form (9.7) with the aid of the auxiliary variable q_{n+1} , then each \hat{f}_m may be decomposed in the form

$$\hat{f}_m = f_m^m(z) + (q_{n+1}) f_{m-1}^m(z) + \cdots + (q_{n+1})^m f_0^m. \quad (10.9.45)$$

Here, as in Section 9.4, the superscript m on f_ℓ^m indicates that the quantity is associated with \hat{f}_m , and the subscript ℓ indicates that the quantity is homogeneous of degree ℓ in the variables z . After this is done, we may view q_{n+1} as playing the role of ϵ and use the calculus of Section 9.3 to manipulate the $(q_{n+1})^{m-\ell} f_\ell^m$ to produce a map of the form

$$\mathcal{M} = \exp(: f_1 :) \mathcal{R} \exp(: f_3 :) \exp(: f_4 :) \cdots, \quad (10.9.46)$$

and then set $q_{n+1} = 1$ in all the f_m to obtain a final map that involves only the variables z . Better yet, we may simply use the shrinker of Section 9.4 to obtain \mathcal{M} from $\hat{\mathcal{M}}$.

Exercises

10.9.1. Verify that (9.20) is of the form (9.4.46).

10.9.2. The map (9.32), with f_1 and f_2 given by (9.33) and (9.36), is written in reverse factorized form. See Section 7.8. Rewrite the map in forward factorized form. See (9.2.30). Before doing so, set $q_2^i = 1$. Verify that use of the forward factorized map also gives (9.43) and (9.44).

10.10 Wei-Norman and Fer Methods

10.10.1 Wei-Norman Equations

Exercises

10.10.1. Exercise on Wei-Norman equations.

10.10.2 Accelerated Procedure: The Fer Expansion

10.11 Symplectic Integration

Symplectic numerical integration methods, like the numerical integration methods described in Chapter 2, seek to compute trajectories accurately through some order in the time step h or with some prescribed over all accuracy. However, they are special in that they are designed for Hamiltonian systems and seek to satisfy the requirement that the resulting transfer map \mathcal{M} between initial and final conditions be exactly (to machine precision) symplectic.

Many aspects of the construction of symplectic integrators involve map-like methods including Zassenhaus formulas. Moreover, for some symplectic integration methods, in the course of computing a trajectory it is also possible to compute in a natural way the transfer map about this trajectory. Thus, symplectic integration and symplectic maps are closely related, and their discussion could logically form part of this chapter. However, the subject is sufficiently vast to deserve a chapter if its own, and will be treated in Chapter 12.

10.12 Taylor Methods and the Complete Variational Equations

The work of the previous sections dealt for the most part with Hamiltonian systems and the representation of their associated *symplectic* transfer maps \mathcal{M} by Lie transformations. In this section we will work with differential equations that are not necessarily Hamiltonian in form. This may occur if the dynamical system under consideration is intrinsically non-Hamiltonian. Also, there is the possibility that the dynamical system does have a Hamiltonian description, but we do not wish to use it. For example, we may want to consider charged-particle motion but do not wish to be concerned with scalar and vector potentials. Rather, we would

prefer to work only in terms of electric and magnetic fields \mathbf{E} and \mathbf{B} . One such option is to employ the first-order set of equations (1.6.68) and (1.6.69). Another is to employ a first-order set of equations obtained from the second-order set (1.6.74) by the usual means. Finally, the equations may be Hamiltonian in form, but we do not wish to exploit their Hamiltonian structure. However the equations arise, we will seek *Taylor* representations for their associated transfer maps

Let us recapitulate some of the contents of Section 1.3. Consider any set of m first-order differential equations of the form

$$\dot{z}_a = f_a(z_1, \dots, z_m; t; \lambda_1, \dots, \lambda_n), \quad a = 1, \dots, m. \quad (10.12.1)$$

Here t is the independent variable and the z_a are dependent variables. Unlike the Hamiltonian case, the z_a need not be canonical variables and they need not be even in number. See (1.3.4). The λ_b are possible parameters.

Let the quantities z_a^0 be initial conditions specified at some initial time $t = t^0$,

$$z_a(t^0) = z_a^0. \quad (10.12.2)$$

Then, under mild conditions imposed on the functions f_a that appear on the right side of (12.1) and thereby define the set of differential equations, there exists a *unique* solution

$$z_a(t) = g_a(z_1^0, \dots, z_m^0; t^0, t; \lambda_1, \dots, \lambda_n), \quad a = 1, m \quad (10.12.3)$$

of (12.1) with the property

$$z_a(t^0) = g_a(z_1^0, \dots, z_m^0; t^0, t^0; \lambda_1, \dots, \lambda_n) = z_a^0, \quad a = 1, m. \quad (10.12.4)$$

Now assume that the functions f_a are analytic (within some domain) in the quantities z_a , the time t , and the parameters λ_b . Then, according to Poincaré's theorem, the solution given by (12.3) will be analytic (again within some domain) in the initial conditions z_a^0 , the times t^0 and t , and the parameters λ_b .

If the solution $z_a(t)$ is analytic in the initial conditions z_a^0 and the parameters λ_b , then it is possible to expand it in the form of a Taylor series, with time-dependent coefficients, in the variables z_a^0 and λ_b . The aim of this section is to describe how these Taylor coefficients can be found as solutions to what we will call the *complete variational* equations.

To aid further discussion, it is useful to also rephrase our goal in the context of maps. Suppose we rewrite the set of first-order differential equations (12.1) in the more compact vector form

$$\dot{\mathbf{z}} = \mathbf{f}(\mathbf{z}; t; \boldsymbol{\lambda}). \quad (10.12.5)$$

Then, again using vector notation, their solution can be written in the form

$$\mathbf{z}(t) = \mathbf{g}(\mathbf{z}^0; t^0, t; \boldsymbol{\lambda}). \quad (10.12.6)$$

That is, the quantities $\mathbf{z}(t)$ at any time t are uniquely specified by the initial quantities \mathbf{z}^0 given at the initial time t^0 .

We capitalize on this fact by introducing a slightly different notation. First, use t^i instead of t^0 to denote the *initial* time. Similarly use \mathbf{z}^i to denote initial conditions by writing

$$\mathbf{z}^i = \mathbf{z}^0 = \mathbf{z}(t^i). \quad (10.12.7)$$

Next, let t^f be some *final* time, and define final conditions \mathbf{z}^f by writing

$$\mathbf{z}^f = \mathbf{z}(t^f). \quad (10.12.8)$$

Then, with this notation, (12.6) can be rewritten in the form

$$\mathbf{z}^f = \mathbf{g}(\mathbf{z}^i; t^i, t^f; \boldsymbol{\lambda}). \quad (10.12.9)$$

We now view (12.9) as a *map* that sends the initial conditions \mathbf{z}^i to the final conditions \mathbf{z}^f . This map will be called the *transfer map* between the times t^i and t^f , and will be denoted by the symbol \mathcal{M} . What we have emphasized is that a set of first-order differential equations of the form (12.5) can be integrated to produce a transfer map \mathcal{M} . We express the fact that \mathcal{M} sends \mathbf{z}^i to \mathbf{z}^f in symbols by writing

$$\mathbf{z}^f = \mathcal{M}\mathbf{z}^i, \quad (10.12.10)$$

Recall the analogous discussion in Section 1.4. We also note that \mathcal{M} is always invertible: Given \mathbf{z}^f , t^f , and t^i , we can always integrate backward in time from the moment $t = t^f$ to the moment $t = t^i$ and thereby find the initial conditions \mathbf{z}^i .

In the context of maps, our goal is to find a Taylor representation for \mathcal{M} . If parameters are present, we may wish to have an expansion in them as well. Initially, we will seek Taylor expansions of final conditions in terms of initial conditions. Subsequently, we will seek expansions of final conditions in terms of both initial conditions and parameters.

10.12.1 Case of No or Ignored Parameter Dependence

Suppose the equations (12.1) do not depend on any parameters λ_b or we do not wish to make expansions in them. We may then suppress their appearance to rewrite (12.1) in the form

$$\dot{z}_a = f_a(z, t), \quad a = 1, m. \quad (10.12.11)$$

Suppose, as in Section 10.5, that z^d is some given *design* solution to these equations, and we wish to study solutions in the vicinity of this solution. As before, we introduce deviation variables ζ_a by writing

$$z_a = z_a^d + \zeta_a. \quad (10.12.12)$$

Then the equations of motion (12.11) take the form

$$\dot{z}_a^d + \dot{\zeta}_a = f_a(z^d + \zeta, t). \quad (10.12.13)$$

We assume that the right side of (12.11) is analytic about z^d . See Theorem 3.3 of Section 1.3. Then we may write the relation

$$f_a(z^d + \zeta, t) = f_a(z^d, t) + g_a(z^d, t, \zeta) \quad (10.12.14)$$

where each g_a has a Taylor expansion of the form

$$g_a(z^d, t, \zeta) = \sum_r g_a^r(t) G_r(\zeta). \quad (10.12.15)$$

Here the $G_r(\zeta)$ are the various monomials in the variables ζ_b labeled by an *index* r using some convenient labeling scheme. (See, for examples, Tables 12.1 and 12.3. For more detail about monomial labeling schemes, see Section 39.2 and Appendix S.2.1.) And the g_a^r are (generally) time-dependent coefficients that we will call *forcing terms*.¹ By construction, all the monomials occurring in the right side of (12.15) have degree one or greater. We note that the g_a^r are known once $z^d(t)$ is given. By assumption, z^d is a solution of (12.11) and therefore satisfies the relations

$$\dot{z}_a^d = f_a(z^d, t). \quad (10.12.16)$$

It follows that the deviation variables satisfy the equations of motion

$$\dot{\zeta}_a = g_a(z^d, t, \zeta) = \sum_r g_a^r(t) G_r(\zeta). \quad (10.12.17)$$

These equations are evidently generalizations of the usual variational equations (see Exercise 4.6 of Section 1.4), and will be called the *complete variational* equations.

Consider the solution to the complete variational equations with *initial* conditions ζ_b^i specified at some initial time t^i . Under the conditions of Theorem 3.3 in Section 1.3 we expect that this solution will be an analytic function of the initial conditions ζ_b^i . (Here we have used the notation of Section 1.4.) Also, since the right side of (12.17) vanishes when all $\zeta_b = 0$ [all the monomials G_r appearing in (12.17) have degree one or greater], $\zeta(t) = 0$ is a solution to (12.17). It follows that the solution to the complete variational equations has a Taylor expansion of the form

$$\zeta_a(t) = \sum_r h_a^r(t) G_r(\zeta^i) \quad (10.12.18)$$

where the $h_a^r(t)$ are functions to be determined, and again all the monomials that occur have degree one or greater. When the quantities $h_a^r(t)$ are evaluated at some *final* time t^f , (12.18) provides a representation of the transfer map \mathcal{M} about the design orbit in the Taylor form

$$\zeta_a^f = \zeta_a(t^f) = \sum_r h_a^r(t^f) G_r(\zeta^i). \quad (10.12.19)$$

10.12.2 Inclusion of Parameter Dependence

What can be done if we desire to have an expansion in parameters as well? Suppose that there are n such parameters, or that we wish to have expansions in n of them. The work of the previous section can be extended to handle this case by means of a simple trick: View the n parameters as additional *variables*, and “augment” the set of differential equations by additional differential equations that ensure these additional variables remain constant.

In detail, suppose we label the parameters so that those in which we wish to have an expansion are $\lambda_1 \cdots \lambda_n$. Introduce n additional variables z_{m+1}, \dots, z_ℓ where $\ell = m + n$ by making the replacements

$$\lambda_b \rightarrow z_{m+b}, \quad b = 1, n. \quad (10.12.20)$$

¹Here and in what follows the quantities g_a are not to be confused with those appearing in (12.3).

Next augment the equations (12.1) by n more of the form

$$\dot{z}_a = 0, \quad a = m + 1, \ell. \quad (10.12.21)$$

By this device we can rewrite the equations (12.1) in the form

$$\dot{z}_a = f_a(z, t), \quad a = 1, \ell \quad (10.12.22)$$

with the understanding that

$$f_a = f_a(z; t; \lambda^{\text{rem}}) \quad a = 1, m, \quad (10.12.23)$$

where λ^{rem} denotes the other *remaining* parameters, if any, and

$$f_a = 0, \quad a = m + 1, \ell. \quad (10.12.24)$$

For the first m equations we impose, as before, the initial conditions

$$z_a(t^i) = z_a^i, \quad a = 1, m. \quad (10.12.25)$$

For the remaining equations we impose the initial conditions

$$z_a(t^i) = \lambda_{a-m}, \quad a = m + 1, \ell. \quad (10.12.26)$$

Note that the relations (12.21) then ensure that the z_a for $a > m$ retain these values for all t .

To continue, let $z^d(t)$ be some design solution. Then, by construction, we have the result

$$z_a^d(t) = \lambda_{a-m}^d = \lambda_{a-m}, \quad a = m + 1, \ell. \quad (10.12.27)$$

Again introduce deviation variables by writing

$$z_a = z_a^d + \zeta_a \quad a = 1, \ell. \quad (10.12.28)$$

Then the quantities ζ_a for $a > m$ will describe deviations in the parameter values about the values λ_{a-m}^d . Moreover, because we have assumed analyticity in the parameters as well, relations of the forms (12.14) and (12.15) will continue to hold except that the $G_r(\zeta)$ are now the various monomials in the ℓ variables ζ_b . Relations of the forms (12.16) and (12.17) will also hold with the provisos (12.24) and

$$g_a^r(t) = 0, \quad a = m + 1, \ell. \quad (10.12.29)$$

Therefore, we will only need to integrate the equations of the forms (12.16) and (12.17) for $a \leq m$. Finally, relations of the form (12.19) will continue to hold for $a \leq m$ supplemented by the relations

$$\zeta_a^f = \zeta_a^i, \quad a = m + 1, \ell. \quad (10.12.30)$$

Since the $G_r(\zeta^i)$ now involve ℓ variables, the relations of the form (12.19) will provide an expansion of the final quantities ζ_a^f (for $a \leq m$) in terms of the initial quantities ζ_a^i (for $a \leq m$) and also the parameter deviations ζ_a^i with $a = m + 1, \ell$.

10.12.3 Solution of Complete Variational Equations Using Forward Integration

This subsection and Subsection 12.5 describe two methods for the solution of the complete variational equations. This subsection describes the method that employs integration forward in time, and is the conceptually simpler of the two methods.

To determine the functions h_a^r , let us insert the expansion (12.18) into both sides of (12.17). With r'' as a dummy index, the left side becomes the relation

$$\dot{\zeta}_a = \sum_{r''} \dot{h}_a^{r''}(t) G_{r''}(\zeta^i). \quad (10.12.31)$$

For the right side we find the intermediate result

$$\sum_r g_a^r(t) G_r(\zeta) = \sum_r g_a^r(t) G_r \left(\sum_{r'} h_1^{r'}(t) G_{r'}(\zeta^i), \dots, \sum_{r'} h_m^{r'}(t) G_{r'}(\zeta^i) \right). \quad (10.12.32)$$

However, since the G_r are monomials, there are relations of the form

$$G_r \left(\sum_{r'} h_1^{r'}(t) G_{r'}(\zeta^i), \dots, \sum_{r'} h_m^{r'}(t) G_{r'}(\zeta^i) \right) = \sum_{r''} U_r^{r''}(h_n^s) G_{r''}(\zeta^i), \quad (10.12.33)$$

and therefore the right side of (12.17) can be rewritten in the form

$$\sum_r g_a^r(t) G_r(\zeta) = \sum_{r''} \sum_r g_a^r(t) U_r^{r''}(h_n^s) G_{r''}(\zeta^i). \quad (10.12.34)$$

The notation $U_r^{r''}(h_n^s)$ is employed to indicate that these quantities might (at this stage of the argument) depend on all the h_n^s with n ranging from 1 to ℓ , and s ranging over all possible values.

Now, in accord with (12.17), equate the right sides of (12.31) and (12.34) to obtain the relation

$$\sum_{r''} \dot{h}_a^{r''}(t) G_{r''}(\zeta^i) = \sum_{r''} \sum_r g_a^r(t) U_r^{r''}(h_n^s) G_{r''}(\zeta^i). \quad (10.12.35)$$

Since the monomials $G_{r''}(\zeta^i)$ are linearly independent, we must have the result

$$\dot{h}_a^{r''}(t) = \sum_r g_a^r(t) U_r^{r''}(h_n^s). \quad (10.12.36)$$

We have found a set of differential equations that must be satisfied by the h_a^r . Moreover, from (12.18), there is the relation

$$\zeta_a(t^i) = \sum_r h_a^r(t^i) G_r(\zeta^i) = \zeta_a^i. \quad (10.12.37)$$

Thus, all the functions $h_a^r(t)$ have a known value at the initial time t^i , and indeed are mostly initially zero. When the equations (12.36) are integrated *forward* from $t = t^i$ to $t = t^f$ to

obtain the quantities $h_a^r(t^f)$, the result is the transfer map \mathcal{M} about the design orbit in the Taylor form (12.19).

Let us now examine the structure of this set of differential equations. A key observation is that the functions $U_r''(h_n^s)$ are *universal*. That is, as (12.33) indicates, they describe certain *combinatorial* properties of monomials. They depend only on the dimension ℓ of the system under study, and are the *same* for all such systems. As (12.17) shows, what are peculiar to any given system are the forcing terms $g_a^r(t)$.

10.12.4 Application of Forward Integration to the Two-Variable Case

To see what is going on in more detail, it is instructive to work out the first nontrivial case, that with $\ell = 2$. For two variables, all monomials in (ζ_1, ζ_2) are of the form $(\zeta_1)^{j_1}(\zeta_2)^{j_2}$. Here, to simplify notation, we have dropped the superscript i . Table 12.1 below shows a convenient way of labeling such monomials, and for this labeling we write

$$G_r(\zeta) = (\zeta_1)^{j_1}(\zeta_2)^{j_2} \quad (10.12.38)$$

with

$$j_1 = j_1(r) \text{ and } j_2 = j_2(r) \quad (10.12.39)$$

and $D(r)$ denotes the *degree* of each monomial.

Table 10.12.1: A labeling scheme for monomials through degree three in two variables.

r	j_1	j_2	D
1	1	0	1
2	0	1	1
3	2	0	2
4	1	1	2
5	0	2	2
6	3	0	3
7	2	1	3
8	1	2	3
9	0	3	3

Thus, for example,

$$G_1 = \zeta_1, \quad (10.12.40)$$

$$G_2 = \zeta_2, \quad (10.12.41)$$

$$G_3 = \zeta_1^2, \quad (10.12.42)$$

$$G_4 = \zeta_1 \zeta_2, \quad (10.12.43)$$

$$G_5 = \zeta_2^2, \text{ etc.} \quad (10.12.44)$$

Again, for more detail about monomial labeling schemes, see Section 39.2 and Appendix S.2.1.

Let us now compute the first few $U_r''(h_n^s)$. From (12.33) and (12.40) we find the relation

$$G_1 \left(\sum_{r'} h_1^{r'} G_{r'}(\zeta), \sum_{r'} h_2^{r'} G_{r'}(\zeta) \right) = \sum_{r'} h_1^{r'} G_{r'}(\zeta) = \sum_{r''} U_1^{r''} G_{r''}(\zeta). \quad (10.12.45)$$

It follows that there is the result

$$U_1^{r''} = h_1^{r''}. \quad (10.12.46)$$

Similarly, from (12.33) and (12.41), we find the result

$$U_2^{r''} = h_2^{r''}. \quad (10.12.47)$$

From (12.33) and (12.42) we find the relation

$$\begin{aligned} G_3 \left(\sum_{r'} h_1^{r'} G_{r'}(\zeta), \sum_{r'} h_2^{r'} G_{r'}(\zeta) \right) &= \left(\sum_{r'} h_1^{r'} G_{r'}(\zeta) \right)^2 \\ &= \sum_{s,t} h_1^s h_1^t G_s(\zeta) G_t(\zeta) = \sum_{r''} U_3^{r''} G_{r''}(\zeta). \end{aligned} \quad (10.12.48)$$

Use of (12.48) and inspection of (12.40) through (12.44) yield the results

$$U_3^1 = 0, \quad (10.12.49)$$

$$U_3^2 = 0, \quad (10.12.50)$$

$$U_3^3 = (h_1^1)^2, \quad (10.12.51)$$

$$U_3^4 = 2h_1^1 h_1^2, \quad (10.12.52)$$

$$U_3^5 = (h_1^2)^2. \quad (10.12.53)$$

From (12.33) and (12.43) we find the relation

$$\begin{aligned} G_4 \left(\sum_{r'} h_1^{r'} G_{r'}(\zeta), \sum_{r'} h_2^{r'} G_{r'}(\zeta) \right) &= \left(\sum_{r'} h_1^{r'} G_{r'}(\zeta) \right) \left(\sum_{r'} h_2^{r'} G_{r'}(\zeta) \right) \\ &= \sum_{s,t} h_1^s h_2^t G_s(\zeta) G_t(\zeta) = \sum_{r''} U_4^{r''} G_{r''}(\zeta). \end{aligned} \quad (10.12.54)$$

It follows that there are the results

$$U_4^1 = 0, \quad (10.12.55)$$

$$U_4^2 = 0, \quad (10.12.56)$$

$$U_4^3 = h_1^1 h_2^1, \quad (10.12.57)$$

$$U_4^4 = h_1^1 h_2^2 + h_1^2 h_2^1, \quad (10.12.58)$$

$$U_4^5 = h_1^2 h_2^2. \quad (10.12.59)$$

Finally, from (12.33) and (12.44), we find the results

$$U_5^1 = 0, \quad (10.12.60)$$

$$U_5^2 = 0, \quad (10.12.61)$$

$$U_5^3 = (h_2^1)^2, \quad (10.12.62)$$

$$U_5^4 = 2h_2^1 h_2^2, \quad (10.12.63)$$

$$U_5^5 = (h_2^2)^2. \quad (10.12.64)$$

Two features now become apparent. As in Table 12.1, let $D(r)$ be the *degree* of the monomial with label r . Then, from the examples worked out, and quite generally from (12.33), we see that there is the relation

$$U_r^{r''} = 0 \text{ when } D(r) > D(r''). \quad (10.12.65)$$

It follows that the sum on the right side of (12.36) always terminates. Second, for the arguments h_n^s possibly appearing in $U_r^{r''}(h_n^s)$, we see that there is the relation

$$D(s) \leq D(r''). \quad (10.12.66)$$

It follows, again see (12.36), that the right side of the differential equation for any $h_a^{r''}$ involves only the h_n^s for which (12.66) holds. Therefore, to determine the coefficients $h_a^r(t^f)$ of the Taylor expansion (12.19) through terms of some degree D , it is only necessary to integrate a finite number of equations, and the right sides of these equations involve only the coefficients for this degree and lower.

For example, to continue our discussion of the case of two variables, the equations (12.36) take the form

$$\dot{h}_1^1(t) = \sum_{r=1}^2 g_1^r(t) U_r^1 = g_1^1(t) h_1^1(t) + g_1^2(t) h_2^1(t), \quad (10.12.67)$$

$$\dot{h}_2^1(t) = \sum_{r=1}^2 g_2^r(t) U_r^1 = g_2^1(t) h_1^1(t) + g_2^2(t) h_2^1(t), \quad (10.12.68)$$

$$\dot{h}_1^2(t) = \sum_{r=1}^2 g_1^r(t) U_r^2 = g_1^1(t) h_1^2(t) + g_1^2(t) h_2^2(t), \quad (10.12.69)$$

$$\dot{h}_2^2(t) = \sum_{r=1}^2 g_2^r(t) U_r^2 = g_2^1(t) h_1^2(t) + g_2^2(t) h_2^2(t), \quad (10.12.70)$$

$$\begin{aligned} \dot{h}_1^3(t) &= \sum_{r=1}^5 g_1^r(t) U_r^3 \\ &= g_1^1(t) h_1^3(t) + g_1^2(t) h_2^3(t) + g_1^3(t) [h_1^1(t)]^2 \\ &\quad + g_1^4(t) h_1^1(t) h_2^1(t) + g_1^5(t) [h_2^1(t)]^2, \end{aligned} \quad (10.12.71)$$

$$\begin{aligned}
\dot{h}_2^3(t) &= \sum_{r=1}^5 g_2^r(t) U_r^3 \\
&= g_2^1(t) h_1^3(t) + g_2^2(t) h_2^3(t) + g_2^3(t) [h_1^1(t)]^2 \\
&+ g_2^4(t) h_1^1(t) h_2^1(t) + g_2^5(t) [h_2^1(t)]^2,
\end{aligned} \tag{10.12.72}$$

$$\begin{aligned}
\dot{h}_1^4(t) &= \sum_{r=1}^5 g_1^r(t) U_r^4 \\
&= g_1^1(t) h_1^4(t) + g_1^2(t) h_2^4(t) + 2g_1^3(t) h_1^1(t) h_1^2(t) \\
&+ g_1^4(t) [h_1^1(t) h_2^2(t) + h_1^2(t) h_2^1(t)] + 2g_1^5(t) h_1^1(t) h_2^2(t),
\end{aligned} \tag{10.12.73}$$

$$\begin{aligned}
\dot{h}_2^4(t) &= \sum_{r=1}^5 g_2^r(t) U_r^4 \\
&= g_2^1(t) h_1^4(t) + g_2^2(t) h_2^4(t) + 2g_2^3(t) h_1^1(t) h_1^2(t) \\
&+ g_2^4(t) [h_1^1(t) h_2^2(t) + h_1^2(t) h_2^1(t)] + 2g_2^5(t) h_1^1(t) h_2^2(t),
\end{aligned} \tag{10.12.74}$$

$$\begin{aligned}
\dot{h}_1^5(t) &= \sum_{r=1}^5 g_1^r(t) U_r^5 \\
&= g_1^1(t) h_1^5(t) + g_1^2(t) h_2^5(t) + g_1^3(t) [h_1^2(t)]^2 \\
&+ g_1^4(t) h_1^2(t) h_2^2(t) + g_1^5(t) [h_2^2(t)]^2,
\end{aligned} \tag{10.12.75}$$

$$\begin{aligned}
\dot{h}_2^5(t) &= \sum_{r=1}^5 g_2^r(t) U_r^5 \\
&= g_2^1(t) h_1^5(t) + g_2^2(t) h_2^5(t) + g_2^3(t) [h_1^2(t)]^2 \\
&+ g_2^4(t) h_1^2(t) h_2^2(t) + g_2^5(t) [h_2^2(t)]^2, \text{ etc.}
\end{aligned} \tag{10.12.76}$$

From (12.37) we have the initial conditions

$$h_a^r(t^i) = \delta_a^r. \tag{10.12.77}$$

We see that if we desire only the degree one terms in the expansion (12.18), then it is only necessary to integrate the equations (12.67) through (12.70) with the initial conditions (12.77). [A moment's reflection shows that, for the case of two variables and no parameters, these are just the (linear) variational equations for the matrix L in Exercise 4.6 of Section 1.4.] We see that if we desire only the degree one and degree two terms in the expansion (12.18), then it is only necessary to integrate the equations (12.67) through (12.76) with the initial conditions (12.77), etc.

10.12.5 Solution of Complete Variational Equations Using Backward Integration

There is another method of determining the h_a^r that is surprising, ingenious, and in some ways superior to that just described. It involves integrating *backward* in time.²

Let us rewrite (12.19) in the slightly more explicit form

$$\zeta_a^f = \sum_r h_a^r(t^i, t^f) G_r(\zeta^i) \quad (10.12.78)$$

to indicate that there are two times involved, t^i and t^f . From this perspective, (12.36) is a set of differential equations for the quantities $(\partial/\partial t)h_a^r(t^i, t)$ that is to be integrated and evaluated at $t = t^f$. An alternate procedure is to seek a set of differential equations for the quantities $(\partial/\partial \bar{t})h_a^r(\bar{t}, t^f)$ that is to be integrated and evaluated at $\bar{t} = t^i$.

As a first step in considering this alternative, rewrite (12.78) in the form

$$\zeta_a^f = \sum_r h_a^r(\bar{t}, t^f) G_r(\zeta(\bar{t})). \quad (10.12.79)$$

Now reason as follows: If \bar{t} is varied and at the same time the quantities $\zeta(\bar{t})$ are varied (evolve) so as to remain on the solution to (12.17) having final conditions ζ^f , then the quantities ζ^f must remain *unchanged*. Consequently, there is the differential equation result

$$0 = d\zeta_a^f/d\bar{t} = \sum_r [(\partial/\partial \bar{t})h_a^r(\bar{t}, t^f)] G_r(\zeta(\bar{t})) + \sum_r h_a^r(\bar{t}, t^f) (d/d\bar{t})G_r(\zeta(\bar{t})). \quad (10.12.80)$$

Let us introduce the notation $\dot{h}_a^r(\bar{t}, t^f)$ for $(\partial/\partial \bar{t})h_a^r(\bar{t}, t^f)$ so that the first term on the right side of (12.80) can be rewritten in the form

$$\sum_r [(\partial/\partial \bar{t})h_a^r(\bar{t}, t^f)] G_r(\zeta) = \sum_r \dot{h}_a^r G_r(\zeta). \quad (10.12.81)$$

Next, begin working on the second term on the right side of (12.80) by replacing the summation index r by the dummy index r' ,

$$\sum_r h_a^r(\bar{t}, t^f) (d/d\bar{t})G_r(\zeta(\bar{t})) = \sum_{r'} h_a^{r'}(\bar{t}, t^f) (d/d\bar{t})G_{r'}(\zeta(\bar{t})). \quad (10.12.82)$$

Now carry out the indicated differentiation using the chain rule and the relation (12.17) which describes how the quantities ζ vary along a solution,

$$(d/d\bar{t})G_{r'}(\zeta(\bar{t})) = \sum_b (\partial G_{r'}/\partial \zeta_b)(d\zeta_b/d\bar{t}) = \sum_{br''} (\partial G_{r'}/\partial \zeta_b) g_b^{r''}(\bar{t}) G_{r''}(\zeta(\bar{t})). \quad (10.12.83)$$

Watch closely: Since the G_r are simply standard monomials in the ζ , there must be relations of the form

$$[(\partial/\partial \zeta_b)G_{r'}(\zeta)] G_{r''}(\zeta) = \sum_r C_{br'r''}^r G_r(\zeta) \quad (10.12.84)$$

²To integrate backward numerically, simply replace h by $-h$ where h is the step size. See Chapter 2.

where the $C_{br'r''}^r$ are *universal constant coefficients* that describe certain combinatorial properties of monomials. As a result, the second term on the right side of (12.80) can be written in the form

$$\sum_{r'} h_a^{r'}(\bar{t}, t^f)(d/d\bar{t})G_{r'}(\zeta(\bar{t})) = \sum_r G_r(\zeta) \sum_{br'r''} C_{br'r''}^r g_b^{r''}(\bar{t}) h_a^{r'}(\bar{t}, t^f). \quad (10.12.85)$$

Since the monomials G_r are linearly independent, the relations (12.80) through (12.85) imply the result

$$\dot{h}_a^r(\bar{t}, t^f) = - \sum_{br'r''} C_{br'r''}^r g_b^{r''}(\bar{t}) h_a^{r'}(\bar{t}, t^f). \quad (10.12.86)$$

This is a set of differential equations for the h_a^r that are to be integrated from $\bar{t} = t^f$ back to $\bar{t} = t^i$. Also, evaluating (12.79) for $\bar{t} = t^f$ gives the results

$$\zeta_a^f = \sum_r h_a^r(t^f, t^f) G_r(\zeta_a^f), \quad (10.12.87)$$

from which it follows that (with the usual polynomial labeling) the h_a^r satisfy the final conditions

$$h_a^r(t^f, t^f) = \delta_a^r. \quad (10.12.88)$$

Therefore the solution to (12.86) is uniquely defined. Finally, it is evident from the definition (12.84) that the coefficients $C_{br'r''}^r$ satisfy the relation

$$C_{br'r''}^r = 0 \text{ unless } [D(r') - 1] + D(r'') = D(r). \quad (10.12.89)$$

Consequently, since $D(r'') \geq 1$ in (12.86), it follows from (12.89) that the only $h_a^{r'}$ that occur on the right side of (12.86) are those that satisfy

$$D(r') \leq D(r). \quad (10.12.90)$$

Similarly, the only $g_b^{r''}$ that occur are those that satisfy

$$D(r'') \leq D(r). \quad (10.12.91)$$

Therefore, as before, to determine the coefficients h_a^r of the Taylor expansion (12.19) through terms of some degree D , it is only necessary to integrate a finite number of equations, and the right sides of these equations involve only the coefficients for this degree and lower.

Comparison of the differential equation sets (12.36) and (12.86) shows that the latter has the remarkable property of being *linear* in the unknown quantities h_a^r . This feature means that the evaluation of the right side of (12.86) involves only the retrieval of certain universal constants $C_{br'r''}^r$ and straight-forward multiplication and addition. By contrast, the use of (12.36) requires evaluation of the fairly complicated *nonlinear* functions $U_r^{r''}(h_n^s)$. Finally, it is easier to insure that a numerical integration procedure is working properly for a set of linear differential equations than it is for a nonlinear set.

The only complication in the use of (12.86) is that the equations must be integrated backwards in \bar{t} . Correspondingly the equations (12.16) for the design solution must also be integrated backwards since they supply the required quantities g_a^r through use of (12.14) and (12.15). This is no problem if the final quantities $z^d(t^{\text{fin}})$ are known. However if only the initial quantities $z^d(t^{\text{in}})$ are known, then the equations (12.16) for z^d must first be integrated forward in time to find the final quantities $z^d(t^{\text{fin}})$.

10.12.6 The Two-Variable Case Revisited

For clarity, let us also apply this second method to the two-variable case. Table 12.2 shows the nonzero values of $C_{br'r''}^r$ for $r \in [1, 5]$ obtained using (12.40) through (12.44) and (12.84). Note that the rules (12.89) hold. Use of this Table shows that in the two-variable case the equations (12.86) take the form

$$\dot{h}_1^1(\bar{t}, t^f) = -g_1^1(\bar{t})h_1^1(\bar{t}, t^f) - g_2^1(\bar{t})h_1^2(\bar{t}, t^f), \quad (10.12.92)$$

$$\dot{h}_2^1(\bar{t}, t^f) = -g_1^1(\bar{t})h_2^1(\bar{t}, t^f) - g_2^1(\bar{t})h_2^2(\bar{t}, t^f), \quad (10.12.93)$$

$$\dot{h}_1^2(\bar{t}, t^f) = -g_1^2(\bar{t})h_1^1(\bar{t}, t^f) - g_2^2(\bar{t})h_1^2(\bar{t}, t^f), \quad (10.12.94)$$

$$\dot{h}_2^2(\bar{t}, t^f) = -g_1^2(\bar{t})h_2^1(\bar{t}, t^f) - g_2^2(\bar{t})h_2^2(\bar{t}, t^f), \quad (10.12.95)$$

$$\dot{h}_1^3(\bar{t}, t^f) = -2g_1^1(\bar{t})h_1^3(\bar{t}, t^f) - g_1^3(\bar{t})h_1^1(\bar{t}, t^f) - g_2^1(\bar{t})h_1^4(\bar{t}, t^f) - g_2^3(\bar{t})h_1^2(\bar{t}, t^f), \quad (10.12.96)$$

$$\dot{h}_2^3(\bar{t}, t^f) = -2g_1^1(\bar{t})h_2^3(\bar{t}, t^f) - g_1^3(\bar{t})h_2^1(\bar{t}, t^f) - g_2^1(\bar{t})h_2^4(\bar{t}, t^f) - g_2^3(\bar{t})h_2^2(\bar{t}, t^f), \quad (10.12.97)$$

$$\begin{aligned} \dot{h}_1^4(\bar{t}, t^f) &= -g_1^1(\bar{t})h_1^4(\bar{t}, t^f) - 2g_1^2(\bar{t})h_1^3(\bar{t}, t^f) - g_1^4(\bar{t})h_1^1(\bar{t}, t^f) \\ &\quad - 2g_2^1(\bar{t})h_1^5(\bar{t}, t^f) - g_2^2(\bar{t})h_1^4(\bar{t}, t^f) - g_2^4(\bar{t})h_1^2(\bar{t}, t^f), \end{aligned} \quad (10.12.98)$$

$$\begin{aligned} \dot{h}_2^4(\bar{t}, t^f) &= -g_1^1(\bar{t})h_2^4(\bar{t}, t^f) - 2g_1^2(\bar{t})h_2^3(\bar{t}, t^f) - g_1^4(\bar{t})h_2^1(\bar{t}, t^f) \\ &\quad - 2g_2^1(\bar{t})h_2^5(\bar{t}, t^f) - g_2^2(\bar{t})h_2^4(\bar{t}, t^f) - g_2^4(\bar{t})h_2^2(\bar{t}, t^f), \end{aligned} \quad (10.12.99)$$

$$\dot{h}_1^5(\bar{t}, t^f) = -g_1^2(\bar{t})h_1^4(\bar{t}, t^f) - g_1^5(\bar{t})h_1^1(\bar{t}, t^f) - 2g_2^2(\bar{t})h_1^5(\bar{t}, t^f) - g_2^5(\bar{t})h_1^2(\bar{t}, t^f), \quad (10.12.100)$$

$$\dot{h}_2^5(\bar{t}, t^f) = -g_1^2(\bar{t})h_2^4(\bar{t}, t^f) - g_1^5(\bar{t})h_2^1(\bar{t}, t^f) - 2g_2^2(\bar{t})h_2^5(\bar{t}, t^f) - g_2^5(\bar{t})h_2^2(\bar{t}, t^f), \text{ etc.} \quad (10.12.101)$$

As advertised, the right sides of (12.92) through (12.101) are indeed simpler than those of (12.67) through (12.76).

10.12.7 Application to Duffing's Equation

As an extension of our discussion of the case of two variables (and no parameters), let us apply the results obtained so far to Duffing's equation (1.4.27) described earlier in Section 1.4. Recall that by a suitable change of variables this equation can be brought to the form

$$q'' + 2\beta q' + q + q^3 = -\epsilon \sin \omega \tau. \quad (10.12.102)$$

Here, for notational convenience, a prime denotes $d/d\tau$. For our purposes, particularly for the parameter expansion soon to be made in the next subsection, it is useful to make the further change of variables

$$q = \omega Q, \quad (10.12.103)$$

$$\omega = 1/\sigma, \quad (10.12.104)$$

$$\omega \tau = t. \quad (10.12.105)$$

Table 10.12.2: Nonzero values of $C_{br'r''}^r$ for $r \in [1, 5]$ in the two-variable case.

r	b	r'	r''	C
1	1	1	1	1
1	2	2	1	1
2	1	1	2	1
2	2	2	2	1
3	1	1	3	1
3	1	3	1	2
3	2	2	3	1
3	2	4	1	1
4	1	1	4	1
4	1	3	2	2
4	1	4	1	1
4	2	2	4	1
4	2	4	2	1
4	2	5	1	2
5	1	1	5	1
5	1	4	2	1
5	2	2	5	1
5	2	5	2	1

When this is done, there are the relations

$$q' = \omega^2 \dot{Q} \quad (10.12.106)$$

and

$$q'' = \omega^3 \ddot{Q} \quad (10.12.107)$$

where now a dot denotes d/dt . [Note that the variable t here is different from that in (1.4.27).] Correspondingly, Duffing's equation takes the form

$$\ddot{Q} + 2\beta\sigma\dot{Q} + \sigma^2 Q + Q^3 = -\epsilon\sigma^3 \sin t. \quad (10.12.108)$$

Finally, this equation can be converted to a first-order pair of the form (12.11) by writing

$$Q = z_1, \quad (10.12.109)$$

$$\dot{Q} = z_2. \quad (10.12.110)$$

Doing so gives the system

$$\dot{z}_1 = z_2, \quad (10.12.111)$$

$$\dot{z}_2 = -2\beta\sigma z_2 - \sigma^2 z_1 - z_1^3 - \epsilon\sigma^3 \sin t, \quad (10.12.112)$$

and we see that there are the relations

$$f_1(z, t) = z_2, \quad (10.12.113)$$

$$f_2(z, t) = -2\beta\sigma z_2 - \sigma^2 z_1 - z_1^3 - \epsilon\sigma^3 \sin t. \quad (10.12.114)$$

Now we are ready to carry out the expansions (12.14) and (12.15). We find the results

$$f_1(z^d + \zeta, t) = z_2^d + \zeta_2, \quad (10.12.115)$$

$$\begin{aligned} f_2(z^d + \zeta, t) &= -2\beta\sigma(z_2^d + \zeta_2) - \sigma^2(z_1^d + \zeta_1) - (z_1^d + \zeta_1)^3 - \epsilon\sigma^3 \sin t \\ &= -[2\beta\sigma z_2^d + \sigma^2 z_1^d + (z_1^d)^3 + \epsilon\sigma^3 \sin] - \{[\sigma^2 + 3(z_1^d)^2]\zeta_1 + 2\beta\sigma\zeta_2\} \\ &\quad - 3z_1^d\zeta_1^2 - \zeta_1^3. \end{aligned} \quad (10.12.116)$$

Note that the right sides of (12.115) and (12.116) contain only terms of degree 3 and lower in the deviation variables ζ_a . It follows that for Duffing's equation the only nonzero forcing terms are given by the relations

$$g_1^2 = 1, \quad (10.12.117)$$

$$g_2^1 = -\sigma^2 - 3(z_1^d)^2, \quad (10.12.118)$$

$$g_2^2 = -2\beta\sigma, \quad (10.12.119)$$

$$g_2^3 = -3z_1^d, \quad (10.12.120)$$

$$g_2^6 = -1. \quad (10.12.121)$$

And, according to (12.16), the design solution z^d obeys the equations (12.111) and (12.112) with $z = z^d$.

At this point we pause to look particularly at the lowest-degree (linear) variational equations because they have a special simplicity in the Duffing case. Let L be the matrix defined by the relation

$$L = \begin{pmatrix} h_1^1 & h_1^2 \\ h_2^1 & h_2^2 \end{pmatrix} \quad (10.12.122)$$

so that

$$\zeta(t) = L(t)\zeta^i + O[(\zeta^i)^2]. \quad (10.12.123)$$

See (12.18). Then equations (12.67) through (12.70) are equivalent to the matrix equation

$$\dot{L} = AL \quad (10.12.124)$$

where A is the matrix

$$A = \begin{pmatrix} g_1^1 & g_1^2 \\ g_2^1 & g_2^2 \end{pmatrix}. \quad (10.12.125)$$

From (12.125) and (12.117) through (12.121), we find the result

$$\text{tr } A = g_1^1 + g_2^2 = -2\beta\sigma. \quad (10.12.126)$$

Also, in the Duffing case, we may set $t^i = 0$ and require the initial condition

$$L(0) = I. \quad (10.12.127)$$

Based on the results of Exercise 1.4.6, we conclude that there is the relation

$$\det L(t) = \exp(-2\beta\sigma t), \quad (10.12.128)$$

and, in particular, for $t = t^f = 2\pi$ there is the relation

$$\det L(2\pi) = \exp(-4\pi\beta\sigma) = \exp(-4\pi\beta/\omega). \quad (10.12.129)$$

Thus, for the Duffing equation, we are able to find the determinant of the linear part of the transfer map *analytically*. Note also the remarkable feature that for the Duffing equation the determinant of the linear part of the transfer map does not depend on z^d . The determinant is the *same* for any trajectory.

Returning to our main discussion, suppose, for example, that we specify the values of β , ϵ , and ω , and then integrate the system (12.111) and (12.112) from $t = 0$ to $t = 2\pi$. So doing produces an example of the stroboscopic map \mathcal{M} described in Subsection 1.4.3. Suppose, further, that we require that the design solution z^d be periodic (with period 2π) thus yielding a fixed point of \mathcal{M} ,

$$z_a^d(2\pi) = z_a^d(0). \quad (10.12.130)$$

We will see in Chapter 28 that such fixed points exist. Using this $z^d(t)$, we may integrate from $t = t^i = 0$ to $t = t^f = 2\pi$ the equations (12.67) through (12.76), etc., with the g_a^r given by (12.117) through (12.121) and the initial conditions given by (12.77). Alternatively, we may integrate (12.92) though (12.101), etc. from $\bar{t} = t^f = 2\pi$ back to $\bar{t} = t^i = 0$ with the final conditions (12.88). Carrying out either method determines the quantities $h_a^r(0, 2\pi)$, and we see from (12.19) or (12.78) that we have found a Taylor expansion for \mathcal{M} about the *fixed* point $z^d(0)$.

10.12.8 Application to Duffing's Equation Including some Parameter Dependence

Suppose, as described in Subsection 12.2, we wish to include some parameter dependence. Figures 28.8.5 and 28.8.6 in Chapter 28 show, for example, a portion of the Feigenbaum diagram for Duffing's equation as ω is varied. Evidently ω is a parameter and therefore, according to Theorem 3.3 of Section 1.3, it should be possible to Taylor expand the solution to Duffing's equation with respect to ω as well as with respect to the initial conditions. Equivalently, we will seek an expansion of the solution of (12.108) with respect to the parameter σ . See (12.104).

Following the method of Subsection 12.2, we augment the first-order equation set associated with (12.108) by adding the equation

$$\dot{\sigma} = 0. \quad (10.12.131)$$

Then we may view σ as a variable, and (12.131) guarantees that this variable remains a constant. Taken together, (12.108) and (12.131) may be converted to a first-order triplet of the form (12.22) by writing (12.109), (12.110), and

$$\sigma = z_3. \quad (10.12.132)$$

Doing so gives the system

$$\dot{z}_1 = z_2, \quad (10.12.133)$$

$$\dot{z}_2 = -2\beta z_3 z_2 - z_3^2 z_1 - z_1^3 - \epsilon z_3^3 \sin t, \quad (10.12.134)$$

$$\dot{z}_3 = 0, \quad (10.12.135)$$

and we see that there are the relations

$$f_1(z, t) = z_2, \quad (10.12.136)$$

$$f_2(z, t) = -2\beta z_3 z_2 - z_3^2 z_1 - z_1^3 - \epsilon z_3^3 \sin t, \quad (10.12.137)$$

$$f_3(z, t) = 0. \quad (10.12.138)$$

As before, we introduce deviation variables using (12.12) and carry out the steps (12.13) through (12.19). In particular, we write

$$z_3 = z_3^d + \zeta_3 = \sigma^d + \zeta_3. \quad (10.12.139)$$

This time we are working with monomials in the three variables ζ_1 , ζ_2 , and ζ_3 . [That is, a ranges from 1 to 3 in (12.18).] They are conveniently labeled using the indices r given in Table 12.3 below. We see, for example, that if we desire to work with monomials through degree 2, the index r should range from 1 through 9.

Table 10.12.3: A labeling scheme for monomials through degree three in three variables.

r	j_1	j_2	j_3	D
1	1	0	0	1
2	0	1	0	1
3	0	0	1	1
4	2	0	0	2
5	1	1	0	2
6	1	0	1	2
7	0	2	0	2
8	0	1	1	2
9	0	0	2	2
10	3	0	0	3
11	2	1	0	3
12	2	0	1	3
13	1	2	0	3
14	1	1	1	3
15	1	0	2	3
16	0	3	0	3
17	0	2	1	3
18	0	1	2	3
19	0	0	3	3

With regard to the expansions (12.14) and (12.15), we find the results

$$f_1(z^d + \zeta, t) = z_2^d + \zeta_2, \quad (10.12.140)$$

$$\begin{aligned} f_2(z^d + \zeta, t) &= -2\beta(z_3^d + \zeta_3)(z_2^d + \zeta_2) - (z_3^d + \zeta_3)^2(z_1^d + \zeta_1) \\ &\quad - (z_1^d + \zeta_1)^3 - \epsilon(z_3^d + \zeta_3)^3 \sin t \\ &= [-2\beta z_2^d z_3^d - z_1^d(z_3^d)^2 - (z_1^d)^3 - \epsilon(z_3^d)^3 \sin t] \\ &\quad - [3(z_1^d)^2 + (z_3^d)^2]\zeta_1 - 2\beta z_3^d \zeta_2 - [2\beta z_2^d + 2z_1^d z_3^d + 3\epsilon(z_3^d)^2 \sin t]\zeta_3 \\ &\quad - 2\beta \zeta_2 \zeta_3 - (z_1^d + 3\epsilon z_3^d)\zeta_2^2 - z_3^d \zeta_1 \zeta_2 - 3z_1^d \zeta_1^2 \\ &\quad - \zeta_1^3 - \zeta_1 \zeta_3^2 - \epsilon(\sin t)\zeta_3^3. \end{aligned} \quad (10.12.141)$$

$$f_3(z^d + \zeta, t) = 0. \quad (10.12.142)$$

Note the right sides of (12.140) through (12.142) are at most cubic in the deviation variables ζ_a . Therefore, from Table 12.3, we see that the index r for the g_a^r should range from 1 through 19. It follows that for Duffing's equation (with σ parameter expansion) the *only* nonzero forcing terms are given by the relations

$$g_1^2 = 1, \quad (10.12.143)$$

$$g_2^1 = -3(z_1^d)^2 - (z_3^d)^2, \quad (10.12.144)$$

$$g_2^2 = -2\beta z_3^d, \quad (10.12.145)$$

$$g_2^3 = -2\beta z_2^d - 2z_1^d z_3^d - 3\epsilon(z_3^d)^2 \sin t, \quad (10.12.146)$$

$$g_2^4 = -3z_1^d, \quad (10.12.147)$$

$$g_2^6 = -2z_3^d, \quad (10.12.148)$$

$$g_2^8 = -2\beta, \quad (10.12.149)$$

$$g_2^9 = -z_1^d - 3\epsilon z_3^d \sin t, \quad (10.12.150)$$

$$g_2^{10} = -1, \quad (10.12.151)$$

$$g_2^{15} = -1, \quad (10.12.152)$$

$$g_2^{19} = -\epsilon \sin t. \quad (10.12.153)$$

If we choose to use forward integration, and again restrict our attention to monomials through degree 2, the relevant $U_r''(h_n^s)$ are given by the relations

$$U_1^{r''} = h_1^{r''}, \quad (10.12.154)$$

$$U_2^{r''} = h_2^{r''}, \quad (10.12.155)$$

$$U_3^{r''} = h_3^{r''}, \quad (10.12.156)$$

$$U_4^r = 0 \text{ for } r \leq 3, \quad (10.12.157)$$

$$U_4^4 = (h_1^1)^2, \quad (10.12.158)$$

$$U_4^5 = 2h_1^1 h_1^2, \quad (10.12.159)$$

$$U_4^6 = 2h_1^1 h_1^3, \quad (10.12.160)$$

$$U_4^7 = (h_1^2)^2, \quad (10.12.161)$$

$$U_4^8 = 2h_1^2 h_1^3, \quad (10.12.162)$$

$$U_4^9 = (h_1^3)^2, \quad (10.12.163)$$

$$U_5^r = 0 \text{ for } r \leq 3, \quad (10.12.164)$$

$$U_5^4 = h_1^1 h_2^1, \quad (10.12.165)$$

$$U_5^5 = h_1^1 h_2^2 + h_1^2 h_2^1, \quad (10.12.166)$$

$$U_5^6 = h_1^3 h_2^1 + h_1^1 h_2^3, \quad (10.12.167)$$

$$U_5^7 = h_1^2 h_2^2, \quad (10.12.168)$$

$$U_5^8 = h_1^3 h_2^2 + h_1^2 h_2^3, \quad (10.12.169)$$

$$U_5^9 = h_1^3 h_2^3, \quad (10.12.170)$$

$$U_6^r = 0 \text{ for } r \leq 3, \quad (10.12.171)$$

$$U_6^4 = h_1^1 h_3^1, \quad (10.12.172)$$

$$U_6^5 = h_1^1 h_3^2 + h_1^2 h_3^1, \quad (10.12.173)$$

$$U_6^6 = h_1^3 h_3^1 + h_1^1 h_3^3, \quad (10.12.174)$$

$$U_6^7 = h_1^2 h_3^2, \quad (10.12.175)$$

$$U_6^8 = h_1^2 h_3^3 + h_1^3 h_3^2, \quad (10.12.176)$$

$$U_6^9 = h_1^3 h_3^3, \quad (10.12.177)$$

$$U_7^r = 0 \text{ for } r \leq 3, \quad (10.12.178)$$

$$U_7^4 = (h_1^2)^2, \quad (10.12.179)$$

$$U_7^5 = 2h_2^1 h_2^2, \quad (10.12.180)$$

$$U_7^6 = 2h_2^1 h_2^3, \quad (10.12.181)$$

$$U_7^7 = (h_2^2)^2, \quad (10.12.182)$$

$$U_7^8 = 2h_2^2 h_2^3, \quad (10.12.183)$$

$$U_7^9 = (h_2^3)^2, \quad (10.12.184)$$

$$U_8^r = 0 \text{ for } r \leq 3, \quad (10.12.185)$$

$$U_8^4 = h_2^1 h_3^1, \quad (10.12.186)$$

$$U_8^5 = h_2^1 h_3^2 + h_2^2 h_3^1, \quad (10.12.187)$$

$$U_8^6 = h_2^3 h_3^1 + h_2^1 h_3^3, \quad (10.12.188)$$

$$U_8^7 = h_2^2 h_3^2, \quad (10.12.189)$$

$$U_8^8 = h_2^2 h_3^3 + h_2^3 h_3^2, \quad (10.12.190)$$

$$U_8^9 = h_2^3 h_3^3, \quad (10.12.191)$$

$$U_9^r = 0 \text{ for } r \leq 3, \quad (10.12.192)$$

$$U_9^4 = (h_3^1)^2, \quad (10.12.193)$$

$$U_9^5 = 2h_3^1 h_3^2, \quad (10.12.194)$$

$$U_9^6 = 2h_3^1 h_3^3, \quad (10.12.195)$$

$$U_9^7 = (h_3^2)^2, \quad (10.12.196)$$

$$U_9^8 = 2h_3^2 h_3^3, \quad (10.12.197)$$

$$U_9^9 = (h_3^3)^2. \quad (10.12.198)$$

As before, the rules (12.65) and (12.66) hold.

The relevant equations for the h_a^r become

$$\dot{h}_a^{r''} = \sum_{r=1}^3 g_a^r U_r^{r''} \text{ for } r'', a \in [1, 3], \quad (10.12.199)$$

$$\dot{h}_a^{r''} = \sum_{r=1}^9 g_a^r U_r^{r''} \text{ for } r'' \in [4, 9] \text{ and } a \in [1, 3]. \quad (10.12.200)$$

The initial conditions are

$$h_a^r(t^i) = \delta_a^r \text{ for } a \in [1, 3] \text{ and } r \in [1, 9]. \quad (10.12.201)$$

Note that because of (12.135) and (12.201) we can actually restrict a in the differential equations (12.199) and (12.200) for the $h_a^{r''}$ to the values $a = 1$ and $a = 2$ since we have the relations

$$h_3^r(t) = \delta_3^r \text{ for all } t. \quad (10.12.202)$$

For the use of backward integration in the three-variable case, Table 12.4 gives the nonzero values of $C_{br'r''}^r$ for $r \in [1, 9]$. In this method the equations (12.86) are to be integrated from $\bar{t} = t^{\text{fin}}$ back to $\bar{t} = t^{\text{in}}$ with the final conditions (12.88). From these equations, and the information about the $g_b^{r''}$ given by (12.143) through (12.153), it is not immediately obvious that there is the relation

$$h_3^r(\bar{t}, t^f) = \delta_3^r \text{ for all } \bar{t}, \quad (10.12.203)$$

which is consistent with (12.202). (And perhaps this is one minor drawback of this method.) However inspection of Table 12.4 for the cases $b = 1$ and $b = 2$ [the possibility $b = 3$ need not be considered because, as can be checked, $g_3^r = 0$ for all r] reveals there is no nonzero coefficient with $r' = 3$. Therefore, $\dot{h}_3^r(\bar{t}, t^f) = 0$ is the solution to (12.86) and (12.88) with $a = 3$ for the $g_b^{r''}$ in question. Correspondingly, as before, only the equations (12.86) with $a = 1$ and $a = 2$ need be integrated.

For further detail including numerical examples, see Section 29.12.

Table 10.12.4: Nonzero values of $C_{br'r''}^r$ for $r \in [1, 9]$ in the three-variable case.

r	b	r'	r''	C
1	1	1	1	1
1	2	2	1	1
1	3	3	1	1
2	1	1	2	1
2	2	2	2	1
2	3	3	2	1
3	1	1	3	1
3	2	2	3	1
3	3	3	3	1
4	1	1	4	1
4	1	4	1	2
4	2	2	4	1
4	2	5	1	1
4	3	3	4	1
4	3	6	1	1
5	1	1	5	1
5	1	4	2	2
5	1	5	1	1
5	2	2	5	1
5	2	5	2	1
5	2	7	1	2
5	3	3	5	1
5	3	6	2	1
5	3	8	1	1
6	1	1	6	1
6	1	4	3	2
6	1	6	1	1
6	2	2	6	1
6	2	5	3	1
6	2	8	1	1
6	3	3	6	1
6	3	6	3	1
6	3	3	1	2
7	1	1	7	1
7	1	5	2	1
7	2	2	7	1
7	2	7	2	2
7	3	3	7	1
7	3	8	2	1
8	1	1	8	1
8	1	5	3	1
8	1	6	2	1

r	b	r'	r''	C
8	2	2	8	1
8	2	7	3	2
8	2	8	2	1
8	3	3	8	1
8	3	8	3	1
8	3	9	2	2
9	1	1	9	1
9	1	6	3	1
9	2	2	9	1
9	2	8	3	1
9	3	3	9	1
9	3	9	3	2

10.12.9 Taylor Methods for the Hamiltonian Case

Suppose the equations of motion (12.1) arise from some Hamiltonian H . Then we may introduce deviation variables ζ as in Section 10.5, and employ the Hamiltonian $H^{\text{new}}(\zeta, t)$ of (5.4) to find the evolution of the variables ζ . Expand H^{new} in terms of the monomials $G_{r''}$ by writing

$$H^{\text{new}}(\zeta, t) = \sum_{r''} H^{r''}(t) G_{r''}(\zeta). \quad (10.12.204)$$

According to (5.4), all the $G_{r''}$ have degree two or greater. From Hamilton's equations of motion we have the result

$$\dot{\zeta}_a = [\zeta_a, H^{\text{new}}] = \sum_{r''} H^{r''}(t) [\zeta_a, G_{r''}]. \quad (10.12.205)$$

Since the monomials G_r form a basis, there is an expansion of the form

$$[\zeta_a, G_{r''}] = \sum_r E_{ar''}^r G_r \quad (10.12.206)$$

where the $E_{ar''}^r$ are universal coefficients that again describe certain combinatorial properties of monomials. With the aid of these coefficients (12.205) can be rewritten in the form

$$\dot{\zeta}_a = \sum_r \left(\sum_{r''} H^{r''} E_{ar''}^r \right) G_r. \quad (10.12.207)$$

Comparison of (12.17) and (12.207) gives the result

$$g_a^r(t) = \sum_{r''} E_{ar''}^r H^{r''}(t). \quad (10.12.208)$$

This result for the g_a^r may now be used to find the h_a^r in (12.19) by employing either forward or backward integration.

For the method of backward integration there is a further simplification. In its derivation we had to compute the quantities $(d/d\bar{t})G_{r'}(\zeta(\bar{t}))$. See (12.83). In the Hamiltonian case this quantity is given by the rule

$$(d/d\bar{t})G_{r'}(\zeta(\bar{t})) = [G_{r'}, H^{\text{new}}]. \quad (10.12.209)$$

See (1.7.4). Next insert the expansion (12.204) into (12.209) to find the result

$$(d/d\bar{t})G_{r'}(\zeta(\bar{t})) = \sum_{r''} H^{r''}(\bar{t})[G_{r'}, G_{r''}]. \quad (10.12.210)$$

Again, there are universal coefficients $F_{r'r''}^r$ describing certain monomial combinatorial properties such that

$$[G_{r'}, G_{r''}] = \sum_r F_{r'r''}^r G_r. \quad (10.12.211)$$

Note that as a consequence of (7.6.14) there is the restriction

$$F_{r'r''}^r = 0 \text{ unless } D(r') + D(r'') - 2 = D(r). \quad (10.12.212)$$

As a result of (12.210) and (12.211) we may write the relation

$$(d/d\bar{t})G_{r'}(\zeta(\bar{t})) = \sum_r \left(\sum_{r''} F_{r'r''}^r H^{r''}(\bar{t}) \right) G_r. \quad (10.12.213)$$

As a last step combine (12.80) through (12.82) and (12.213) to get the result

$$\dot{h}_a^r(\bar{t}, t^f) = - \sum_{r'r''} F_{r'r''}^r H^{r''}(\bar{t}) h_a^{r'}(\bar{t}, t^f). \quad (10.12.214)$$

These equations are the Hamiltonian analog of the general equations (12.86). As before, they are to be integrated from $\bar{t} = t^f$ back to $\bar{t} = t^i$ with the final conditions (12.88).

Let us compare the method just described for Taylor maps in the Hamiltonian case and that for factored product maps given in Section 10.5. Inspection of the equations (5.60) through (5.66) for the Lie generators show that they become ever more complicated with increasing order. By contrast, the equations (12.214) for the Taylor coefficients can be found easily to any desired order since the coefficients $F_{r'r''}^r$, which describe the monomial Poisson bracket results (12.211), can be obtained easily using Truncated Power Series Algebra. See Section 39.8. The price that must be paid for this ease of programming and computing to any desired order is the need to integrate many more differential equations. The numbers of equations that must be integrated in the Taylor and Lie methods are the same as those required to specify a map in Taylor or factored product Lie form, respectively. See Table 7.10.2.

Of course, no matter how a Taylor map is computed, if it arises from Hamiltonian equations it will be symplectic. And we know from the factorization Theorem of Section 7.6 that once a symplectic map is known in Taylor form, it can be rewritten in the factored product Lie form (7.6.3).

Exercises

10.12.1. Verify the entries in Table 12.2.

10.12.2. Consider the differential equation

$$\dot{x} = tx^2 \quad (10.12.215)$$

with the initial condition x^i at the initial time t^i . Verify that a particular solution is given by $x^d = 0$, and is the solution with $x(t^i) = 0$. Show that it has the general solution

$$x^f = x^i / \{1 - (x^i/2)[(t^f)^2 - (t^i)^2]\}. \quad (10.12.216)$$

Find the first few terms in the Taylor expansion of x^f in powers of x^i . This expansion is, in essence, an expansion about the solution x^d . Also compute the Taylor expansion of x^f in powers of x^i using both the methods of forward and backward integration. Verify that all your results agree.

10.12.3. Verify the derivation of (12.214).

10.12.4. Let $(\eta_1, \eta_2, \dots, \eta_{2n})$ be $2n$ “dummy” variables. Define functions $Z_a(\bar{t}, t^f; \eta)$ by the rule

$$Z_a(\bar{t}, t^f; \eta) = \sum_r h_a^r(\bar{t}, t^f) G_r(\eta), \quad (10.12.217)$$

which can also be written in the equivalent form

$$Z_a(\bar{t}, t^f; \eta) = \sum_{r'} h_a^{r'}(\bar{t}, t^f) G_{r'}(\eta). \quad (10.12.218)$$

Also define a function $H^{\text{new}}(\bar{t}, \eta)$ by the rule

$$H^{\text{new}}(\bar{t}, \eta) = \sum_{r''} H^{r''}(\bar{t}) G_{r''}(\eta). \quad (10.12.219)$$

Using (12.218), (12.219), (12.211), and (12.214) show that

$$[Z_a, H^{\text{new}}]_\eta = \sum_r \left(\sum_{r' r''} h_a^{r'} H^{r''} F_{r' r''}^r \right) G_r = - \sum_r \dot{h}_a^r(\bar{t}, t^f) G_r(\eta). \quad (10.12.220)$$

Hence, show that (12.114) is equivalent to the differential equation

$$\partial Z_a / \partial \bar{t} = -[Z_a, H^{\text{new}}]_\eta. \quad (10.12.221)$$

10.12.5. Suppose the complete variational equations (12.17) happen to be autonomous. That is, the functions g_a^r do not depend on time. In the Hamiltonian case assume, equivalently, that the H^r do not depend on time. This will be the case for idealized beam-line elements. See Chapters 13 and 14. In the case of equation (12.86), define quantities $K_{rr'}$ by the rule

$$K_{rr'} = \sum_{br''} C_{br' r''}^r g_b^{r''}. \quad (10.12.222)$$

They will be time independent since, by hypothesis, the $g_b^{r''}$ are time independent. In the case of equation (12.214), define quantities $K_{rr'}$ by the rule

$$K_{rr'} = \sum_{r''} F_{r'r''}^r H^{r''}. \quad (10.12.223)$$

They too will be time independent if the $H^{r''}$ are. Show that in either case, equation (12.86) or equation (12.214), there is the common result

$$\dot{h}_a^r = - \sum_{r'} K_{rr'} h_a^{r'} \quad (10.12.224)$$

which, when written in vector form, becomes the vector-matrix equation

$$\dot{h}_a = -Kh_a. \quad (10.12.225)$$

Verify that (12.225) has the solution

$$h_a(\bar{t}, t^f) = e^{-(\bar{t}-t^f)K} h_a(t^f, t^f), \quad (10.12.226)$$

and consequently there is the result

$$h_a(t^i, t^f) = e^{-(t^i-t^f)K} h_a(t^f, t^f). \quad (10.12.227)$$

Also, according to (12.88), the vectors $h_a(t^f, t^f)$ are given by the relation

$$h_a^r(t^f, t^f) = \delta_a^r. \quad (10.12.228)$$

Thus, the Taylor map can be found explicitly in the autonomous case in terms of the matrix $\exp[-(t^i - t^f)K]$. This matrix can be computed using the scaling and squaring method of Section 4.1. Compare the results above for the Hamiltonian case with those found in Section 10.7, and in particular that of (7.11). Hint: For simplicity consider the two-variable case and suppose H^{new} consists of only an H_2 and an H_3 .

Bibliography

General References

- [1] E. Forest, *Beam Dynamics: A New Attitude and Framework*, Harwood Academic Press (1998).

- [2] A.J. Dragt and E. Forest, “Computation of nonlinear behavior of Hamiltonian systems using Lie algebraic methods”, *J. Math. Phys.* **24**, 2734 (1983).

Magnus Equations, Wei-Norman Equations, Fer Expansions, and General Lie-Algebraic References

- [3] W. Magnus, “On the exponential solution of differential equations for a linear operator”, *Commun. Pure Appl. Math.* **7**, 649 (1954).

- [4] S. Blanes, F. Casas, J. Oteo, and J. Ros, “The Magnus expansion and some of its applications”, *Physics Reports* **470**, 151-238 (2009).

- [5] A. Arnal, F. Casas, and C. Chiralt, “A general formula for the Magnus expansion in terms of iterated integrals of right-nested commutators”, *Journal of Physics Communications* **2**, 035024 (2018). Or see <https://arxiv.org/abs/1710.10851>.

- [6] F. Fer, “Résolution de l’équation matricielle $\dot{U} = pU$ par produit infini d’exponentielles matricielles”, *Bull. Classe Sci. Acad. Roy. Belg.* **44**, 818 (1958).

- [7] E. Wichmann, “Note on the Algebraic Aspect of the Integration of a System of Ordinary Linear Differential Equations”, *J. Math. Phys.* **6**, 876 (1961).

- [8] P. Moan and J. Niesen, “Convergence of the Magnus Series”, eprint arXiv:math/0609198 (2006).

- [9] J. Wei and E. Norman, “Lie Algebraic Solution of Linear Differential Equations”, *J. Math. Phys.* **4**, 575 (1963).

- [10] J. Wei and E. Norman, “On global representations of the solutions of linear differential equations as products of exponentials”, *Proc. Amer. Math. Soc.* **15**, 327 (1964).

- [11] P.-V. Koseleff, “Formal Calculus for Lie Methods in Hamiltonian Mechanics”, Ph.D. thesis, École Polytechnique (1993).

- [12] M. Torres-Torriti and H. Michalska, “A Software Package for Lie-Algebraic Computations”, *SIAM Review* **47**, 722 (2005).

- [13] F. Casas, J.A. Oteo, and J. Ros, “Lie algebraic approach to Fer’s expansions for classical Hamiltonian systems”, *J. Phys. A: Math. Gen.* **24**, 4037 (1991).
- [14] S. Blanes, F. Casas, J.A. Oteo, and J. Ros, “Magnus and Fer expansions for matrix differential equations: The convergence problem”, *J. Phys. A: Math. Gen.* **31**, 259 (1998).
- [15] S. Blanes, F. Casas, and A. Murua, “Splitting and composition methods in the numerical integration of differential equations”, arXiv:0812.0377v1 [math.NA] 1 Dec 2008.
- [16] A. Iserles and S.P. Nørsett, “On the solution of linear differential equations in Lie groups”, *Phil. Trans. R. Soc. Lond. A* **357**, 983 (1999).

Scaling, Splitting, and Squaring

- [17] A.J. Dragt, “Computation of Maps for Particle and Light Optics by Scaling, Splitting, and Squaring”, *Phys. Rev. Let.* **75**, 1946 (1995).

Taylor Maps

- [18] The method of Backward Integration described in Sections 10.12.5 through 10.12.8 was discovered by F. Neri circa 1986.