

# Chapter 11

## Geometric/Structure-Preserving Integration: Integration on Manifolds

### Overview

The subject of *geometric integration*, also sometimes called *structure-preserving integration*, deals both with the construction of numerical integrators on manifolds and with the construction of numerical integrators that respect various group properties. This chapter provides an introduction to the subject of numerical integration on manifolds, mostly by way of three examples: rigid body motion, spin and qubits, and motion of a charged particle in a static magnetic field. Symplectic integration, the main topic of the next chapter, is a special case of geometric integration that respects a particular group property.<sup>1</sup>

Suppose a set of first-order ordinary differential equations is formulated in terms of coordinates in some ambient space and suppose it is known that, when these equations are integrated exactly, some class of trajectories is confined to a lower dimensional manifold in the ambient space, often a manifold associated with some group. If these same equations are integrated numerically with the aid of some integrator that is locally accurate through terms of order  $h^m$  one may expect, unless the integrator has special properties, that trajectories will deviate, at each integration step, from this lower dimensional manifold by terms of order  $h^{m+1}$ . In this context, a geometric numerical integrator is an integrator that, even though it may make local errors of order  $h^{m+1}$ , is guaranteed to produce trajectories that remain on the desired manifold, perhaps exactly to machine precision, or at least to substantially higher order in  $h$ . The construction and study of such numerical integrators is the subject of numerical integration on manifolds.<sup>2</sup>

Again suppose a set of first-order ordinary differential equations is formulated in terms of coordinates, call them  $z$ , in some ambient space. Let  $t^i$  and  $t^f$  be initial and final times. Now suppose it is known that the relation/map between  $z^i = z(t^i)$  and  $z^f = z(t^f)$  for

---

<sup>1</sup>Although for brevity we will not do so, for precision we should use the terms geometric *numerical integration* and, by association, symplectic *numerical integration*. Strictly speaking, this distinction is necessary because the term *geometric integration* can also refer to the extension of the integration procedures of ordinary calculus to integration over manifolds based on the use of differential forms.

<sup>2</sup>From this perspective numerical integrators that preserve integrals of motion, such as energy and angular momentum, are special instances of geometric integrators.

variable  $z^i$ , arising from exactly integrating initial conditions from  $t = t^i$  to  $t = t^f$ , belongs to some group  $G$  that acts on the ambient space. What happens if these same equations are integrated numerically with the aid of some integrator that is locally accurate through terms of order  $h^m$ ? Let  $t^b$  be the time at the *beginning* of some integration step. We may expect, unless the integrator has special properties, that the relation/map between  $z(t^b)$  and  $z(t^b+h)$  will differ from an element of  $G$  by terms of order  $h^{m+1}$ . In this context, a geometric integrator is an integrator with the property that, even though it may make local errors of order  $h^{m+1}$ , the relation/map between  $z(t^b)$  and  $z(t^b+h)$ , for each integration step, is a map that belongs to  $G$ , perhaps exactly to machine precision, or at least to substantially higher order in  $h$ . Since for such a geometric integrator the maps for all successive integration steps belong to  $G$ , their product will also belong to  $G$ , and thus the map relating  $z(t^i)$  and  $z(t^f)$  will also belong to  $G$ .<sup>3</sup> In the Hamiltonian case, the ambient space is phase space, the group  $G$  is the group of all symplectic maps, and a symplectic numerical integrator is an integrator designed to produce a relation between  $z^i$  and  $z^f$  that is a symplectic map.

## 11.1 Numerical Integration on Manifolds: Rigid-Body Motion

As a first example of integration on manifolds, we will consider the problem of determining the motion of a rigid body with one point fixed and subject to an external torque  $\mathbf{N}$  about the fixed point. We begin with the *kinematics* of rigid-body motion, and then follow with the *dynamics* of rigid-body motion.<sup>4</sup> Subsequent discussion treats various formulations of the equations of motion and various methods for their numerical integration.

### 11.1.1 Angular Velocity and Rigid-Body Kinematics

To describe rigid-body *kinematics*, suppose  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$  is an orthonormal right-hand triad of vectors that is fixed in space, and suppose  $\mathbf{f}_1(t), \mathbf{f}_2(t), \mathbf{f}_3(t)$  is an orthonormal right-hand triad of vectors that is fixed in the body.<sup>5</sup> The  $\mathbf{f}_j(t)$  will generally be time dependent because the body is expected to rotate, and they will be taken to have the initial values

$$\mathbf{f}_j(t^0) = \mathbf{e}_j. \quad (11.1.1)$$

Since the  $\mathbf{e}_j$  and the  $\mathbf{f}_j(t)$  are both orthonormal basis sets, there is a unique *orthogonal* matrix  $R(t)$  such that

$$\mathbf{f}_j(t) = R(t)\mathbf{e}_j. \quad (11.1.2)$$

---

<sup>3</sup>Note that unlike the case of Chapter 10, we do not seek here, at least primarily, to compute the transfer map  $\mathcal{M}$  that relates  $z(t^i)$  and  $z(t^f)$ . Rather, we seek trajectories but require that, were all trajectories to be considered, the relation between  $z^i$  and  $z^f$  for all possible initial conditions  $z^i$  would be a map that belongs to  $G$ .

<sup>4</sup>The kinematic equations are the analog of the single-particle equations  $\mathbf{v} = d\mathbf{r}/dt$  or  $d\mathbf{r}/dt = \mathbf{v}$ , and the dynamic equations are the analog of the equations  $d\mathbf{v}/dt = \mathbf{F}/m$ .

<sup>5</sup>We require that the space-fixed and body-fixed triads have the fixed point of the body as their common origin.

See Section 3.6.3. In view of (1.1) there is the relation

$$R(t^0) = I \quad (11.1.3)$$

and therefore, by continuity,  $R(t) \in SO(3, \mathbb{R})$ .

Let  $(\omega_1^{bf}, \omega_2^{bf}, \omega_3^{bf})$  be the *components* of the *angular velocity* in the *body-fixed frame*. They are *defined* by the rules

$$\omega_j^{bf}(t) = -\mathbf{f}_k(t) \cdot \dot{\mathbf{f}}_\ell(t) \quad (11.1.4)$$

where  $j, k, \ell$  is any cyclic permutation of 1, 2, 3. Some implications of this definition of angular velocity are explored in Exercise 1.1. Here we examine how the  $\omega_j^{bf}(t)$  are related to  $R(t)$ . From (1.2) there are the relations

$$\dot{\mathbf{f}}_\ell(t) = \dot{R}(t)\mathbf{e}_\ell, \quad (11.1.5)$$

and therefore (1.4) can be rewritten in the form

$$\omega_j^{bf}(t) = -[R(t)\mathbf{e}_k] \cdot [\dot{R}(t)\mathbf{e}_\ell] = -\mathbf{e}_k \cdot [R^T(t)\dot{R}(t)\mathbf{e}_\ell] = -\mathbf{e}_k \cdot [R^{-1}(t)\dot{R}(t)\mathbf{e}_\ell]. \quad (11.1.6)$$

Define a matrix  $A(t)$  by the rule

$$A(t) = R^{-1}(t)\dot{R}(t) = R^T(t)\dot{R}(t). \quad (11.1.7)$$

With its use, (1.6) can be written in the yet more compact form

$$\omega_j^{bf}(t) = -\mathbf{e}_k \cdot A(t)\mathbf{e}_\ell. \quad (11.1.8)$$

What can be said about the matrix  $A(t)$ ? From the orthogonality condition

$$R^T(t)R(t) = I \quad (11.1.9)$$

we conclude that

$$R^T(t)R(t) + R^T(t)\dot{R}(t) = 0; \quad (11.1.10)$$

and from (1.7) we conclude that

$$A^T(t) = \dot{R}^T(t)R(t). \quad (11.1.11)$$

See Exercise 1.1. It follows from (1.10) that  $A(t)$  is antisymmetric,

$$A^T(t) + A(t) = 0, \quad (11.1.12)$$

and therefore can be written in the form

$$A(t) = a_1(t)L^1 + a_2(t)L^2 + a_3(t)L^3 = \mathbf{a} \cdot \mathbf{L} \quad (11.1.13)$$

where the coefficients  $a_j(t)$  are to be determined. Here we have used the notation of Exercise 3.7.30. Now insert (1.13) into (1.8). Computation using the properties of the  $L^j$  gives the results

$$a_j(t) = \omega_j^{bf}(t). \quad (11.1.14)$$

For example,

$$\begin{aligned}\omega_1^{bf} &= -\mathbf{e}_2 \cdot A\mathbf{e}_3 = -\mathbf{e}_2 \cdot [(\mathbf{a} \cdot \mathbf{L})\mathbf{e}_3] = -\mathbf{e}_2 \cdot (\mathbf{a} \times \mathbf{e}_3) \\ &= \mathbf{e}_2 \cdot (\mathbf{e}_3 \times \mathbf{a}) = (\mathbf{e}_2 \times \mathbf{e}_3) \cdot \mathbf{a} = \mathbf{e}_1 \cdot \mathbf{a} = a_1.\end{aligned}\quad (11.1.15)$$

To complete the discussion of rigid-body kinematics, rewrite (1.7) in the form

$$\dot{R} = RA \quad (11.1.16)$$

and introduce the notation

$$A = a_1 L^1 + a_2 L^2 + a_3 L^3 = \omega_1^{bf} L^1 + \omega_2^{bf} L^2 + \omega_3^{bf} L^3 = \boldsymbol{\omega}^{bf} \cdot \mathbf{L}. \quad (11.1.17)$$

By substituting (1.17) into (1.16) we find for  $R(t)$  the *kinematic* differential equation of motion

$$\dot{R} = R \boldsymbol{\omega}^{bf} \cdot \mathbf{L}. \quad (11.1.18)$$

Note that, since  $R$  is  $3 \times 3$ , the matrix differential equation (1.18) amounts to nine first-order differential equations. Also note, in passing, that  $\boldsymbol{\omega}^{bf} \cdot \mathbf{L} \in so(3, \mathbb{R})$ .

### 11.1.2 Angular Velocity and Rigid-Body Dynamics

We now turn to the *dynamics* of rigid-body motion. Suppose the body-fixed frame is oriented in the body in such a way that the moment of inertia tensor is diagonal with diagonal entries  $I_1, I_2, I_3$ . (Caution! We will also continue to use, as we have already done, the symbol  $I$  without a subscript to denote the  $3 \times 3$  identity matrix.) Then, from the work of Euler, we know that there are the *dynamic* equations of motion

$$\dot{\omega}_1^{bf} = N_1^{bf}/I_1 + \omega_2^{bf} \omega_3^{bf} (I_2 - I_3)/I_1, \quad (11.1.19)$$

$$\dot{\omega}_2^{bf} = N_2^{bf}/I_2 + \omega_3^{bf} \omega_1^{bf} (I_3 - I_1)/I_2, \quad (11.1.20)$$

$$\dot{\omega}_3^{bf} = N_3^{bf}/I_3 + \omega_1^{bf} \omega_2^{bf} (I_1 - I_2)/I_3. \quad (11.1.21)$$

Here the  $N_j^{bf}(R, \boldsymbol{\omega}^{bf}, t)$  are the components of  $\mathbf{N}$  in the body-fixed frame.<sup>6</sup>

### 11.1.3 Problem of Integrating the Combined Kinematic and Dynamic Equations

Taking both the kinematic and dynamic equations into account, our task is to integrate the nine kinematic differential equations (1.18) and the three dynamic differential equations (1.19) through (1.21) given, at time  $t^i$ , some initial orientation  $R(t^i)$  and some initial angular velocity  $\boldsymbol{\omega}^{bf}(t^i)$ . In this context, the ambient space is  $9 + 3 = 12$  dimensional.

But now we see that there is a computational problem. We know that  $R$  must be an orthogonal matrix. See (1.9). Also, it easily verified that *exact* integration of the matrix

---

<sup>6</sup>Note that we have allowed the possibility that  $\mathbf{N}$  is  $\boldsymbol{\omega}^{bf}$  dependent, and therefore have gone beyond a Lagrangian/Hamiltonian formulation in allowing for the possibility of dissipative forces.

differential equations (1.18), no matter how  $\boldsymbol{\omega}^{bf}$  depends on the time  $t$ , maintains the orthogonality condition (1.9) if the initial matrix  $R(t^i)$  is orthogonal. See Exercise 1.2. Thus we know that, although  $R$  has nine entries, it must lie on the 3-dimensional manifold  $SO(3, \mathbb{R})$ . However, if some *numerical* integration method is used that is locally accurate only through terms of order  $h^m$ , then we expect that the orthogonality condition (1.9) will also be maintained only through terms of order  $h^m$ . Thus, in the course of numerical integration,  $R$  may be expected to move off the manifold  $SO(3, \mathbb{R})$ . Moreover, if  $R$  is not orthogonal, then the quantities  $N_j^{bf}(R, \boldsymbol{\omega}^{bf}, t)$  required in (1.19) through (1.21) are not defined since they are only physically specified when  $R$  is orthogonal.<sup>7</sup>

### 11.1.4 Solution by Projection

What to do? One approach is to orthogonalize the  $R$  provided after or during the course of each integration step, whenever an orthogonal  $R$  is needed to compute the  $N_j^{bf}$  or the actual orientation of the body. The orthogonalization can be done, for example, using any of the methods of Section 3.6.4 and Exercise 4.5.7. This process of orthogonalization is an example of a general problem: Given a submanifold embedded in some larger manifold (or ambient space), and given an element in the larger manifold, how does one find a related element in the submanifold, and what is the optional choice for such an element? In the nomenclature of geometric integration, the process for determining such an element is called *projection*.<sup>8</sup>

### 11.1.5 Solution by Parameterization: Euler Angles

Another approach is to parameterize  $R$  in such a way that it is guaranteed to be orthogonal. For example, suppose we employ the Euler parameterization (3.7.207). Then use of (1.18), rewritten in the form

$$\boldsymbol{\omega}^{bf} \cdot \mathbf{L} = R^{-1} \dot{R}, \quad (11.1.22)$$

gives the relations

$$\omega_1^{bf} = -\dot{\phi} \sin \theta \cos \psi + \dot{\theta} \sin \psi, \quad (11.1.23)$$

$$\omega_2^{bf} = \dot{\phi} \sin \theta \sin \psi + \dot{\theta} \cos \psi, \quad (11.1.24)$$

$$\omega_3^{bf} = \dot{\phi} \cos \theta + \dot{\psi}, \quad (11.1.25)$$

---

<sup>7</sup>Note that a similar problem arises in the numerical integration of (10.4.28) with the initial condition (10.4.29). Although the ambient space is  $(2n)^2$  dimensional, the solution of (10.4.28) is required to originate and remain in the  $n(2n+1)$  dimensional submanifold  $Sp(2n, \mathbb{R})$ .

<sup>8</sup>For the problem at hand we need some projection  $E^9 \rightarrow SO(3, \mathbb{R})$  where  $E^9$  denotes 9-dimensional Euclidean space. Recall also the problem of matrix symplectification, projections  $E^{(2n)^2} \rightarrow Sp(2n, \mathbb{R})$ , treated in Subsection 3.6.5 and Chapter 4. Observe that all the integration methods of Chapter 2 require the evaluation of the right side  $\mathbf{f}$  at intermediate points, and generally these points will not be on the desired manifold, but only nearby. Therefore  $\mathbf{f}$  may not even be defined at these points unless  $\mathbf{f}$  can be extrapolated off the manifold to nearby points in the ambient space. Alternatively, the intermediate points can be projected from the ambient space onto the manifold, and  $\mathbf{f}$  is then computed at these projected intermediate points. In either case one has to ensure that extrapolation or projection does not spoil the desired accuracy of the integration method.

from which it follows that

$$\dot{\phi} = (1/\sin \theta)(\omega_2^{bf} \sin \psi - \omega_1^{bf} \cos \psi), \quad (11.1.26)$$

$$\dot{\theta} = \omega_1^{bf} \sin \psi + \omega_2^{bf} \cos \psi, \quad (11.1.27)$$

$$\dot{\psi} = \omega_3^{bf} + (\cot \theta)(\omega_1^{bf} \cos \psi - \omega_2 \sin \psi). \quad (11.1.28)$$

See Exercise 1.3. In terms of Euler angles, our task is to integrate the equations (1.26) through (1.28) and (1.19) through (1.21) where now

$$N_j^{bf} = N_j^{bf}(\phi, \theta, \psi, \dot{\phi}, \dot{\theta}, \dot{\psi}, t). \quad (11.1.29)$$

### 11.1.6 Problem of Kinematic Singularities

Have we achieved our goal? Only in a fashion. It is true that equations (1.26) through (1.28) and (1.19) through (1.21) constitute a set of six first-order equations of motion, which is what we expect for a system with three degrees of freedom. Also,  $R(\phi, \theta, \psi)$  is guaranteed to be orthogonal no matter how inaccurately the equations of motion are numerically integrated. However, note that the factor  $(1/\sin \theta)$  in (1.26) and the factor  $(\cot \theta)$  in (1.28) become *singular* when  $\theta = 0$  or  $\theta = \pi$ . Therefore these equations are unsuitable for numerical integration whenever  $\theta \simeq 0$  or  $\theta \simeq \pi$ . The singularity at  $\theta = 0$  is particularly alarming because it means that Euler angles do not provide a good coordinate patch in the vicinity  $R \simeq I$ .

Now there is nothing *a priori* to prevent  $\theta \simeq 0$  or  $\theta \simeq \pi$  from happening over the course of a rigid body's motion. (For example, a top is allowed to be vertical or inverted.) These singularities are *kinematic* in the sense that they are an artifact of our choice of coordinate system, and are not intrinsic to the motion of the system being considered. (See Exercise 8.2.11.) However, it can be shown from topological considerations that singularities of this type must arise no matter how  $R$  is parameterized if only three parameters are used. A global three-variable and singularity-free parameterization of  $SO(3, \mathbb{R})$  is impossible. Consequently, if only three variables are used, it is necessary to change coordinate patches whenever a singularity of the coordinate system in current use is approached.<sup>9</sup> This complication might appear to make it difficult to write a robust three-variable numerical integration procedure that would apply for all possible rigid-body motions.<sup>10</sup> However, we will eventually entertain the possibility of changing the coordinate system frequently, and perhaps at every integration step.

---

<sup>9</sup>Moreover, even if a trajectory does not pass directly through a kinematic singularity, the presence of a singularity still affects the integration of nearby trajectories because numerical integration is based on the assumption of analyticity and the use of Taylor series, and nearby singularities affect the convergence of Taylor series.

<sup>10</sup>We also remark that the kind of problem we have encountered here is expected to occur quite generally whenever one seeks to numerically integrate trajectories that are known, and hence required, to lie within some group manifold. Because group manifolds do not generally have the topology of Euclidean space, there is generally no global singularity-free coordinate system that can be used.

### 11.1.7 Quaternions to the Rescue

Our ancestors have discovered that, for rigid-body motion, this troublesome singularity problem can be overcome in an optimum way by the use of (unit) quaternion parameters  $w = (w_0, w_1, w_2, w_3)$ . Again see Exercise 8.2.11. Note that there are now four parameters, rather than three as in the Euler-angle case, and they are related by the constraint

$$w \cdot w = \sum_{j=0}^3 w_j^2 = 1. \quad (11.1.30)$$

Quaternion parameters and  $SO(3, \mathbb{R})$  matrices are connected by (8.2.73). Quaternion and angle-axis parameters are connected by the relations

$$w_0 = \cos(\theta/2), \quad (11.1.31)$$

$$\mathbf{w} = -\mathbf{n} \sin(\theta/2). \quad (11.1.32)$$

Quaternion and Euler-angle parameters are connected by the relations

$$w_0 = \cos(\theta/2) \cos[(1/2)(\phi + \psi)], \quad (11.1.33)$$

$$w_1 = -\sin(\theta/2) \sin[(1/2)(-\phi + \psi)], \quad (11.1.34)$$

$$w_2 = -\sin(\theta/2) \cos[(1/2)(-\phi + \psi)], \quad (11.1.35)$$

$$w_3 = -\cos(\theta/2) \sin[(1/2)(\phi + \psi)]. \quad (11.1.36)$$

See Exercise 1.4.

More to the point for our purposes, the angular velocities are given in terms of quaternion parameters by the relations

$$\omega_1^{bf} = 2(w_1\dot{w}_0 - w_0\dot{w}_1 + w_3\dot{w}_2 - w_2\dot{w}_3), \quad (11.1.37)$$

$$\omega_2^{bf} = 2(w_2\dot{w}_0 - w_0\dot{w}_2 + w_1\dot{w}_3 - w_3\dot{w}_1), \quad (11.1.38)$$

$$\omega_3^{bf} = 2(w_3\dot{w}_0 - w_0\dot{w}_3 + w_2\dot{w}_1 - w_1\dot{w}_2). \quad (11.1.39)$$

See Exercise 1.5. To these relations we add the further relation

$$0 = -\sum_{j=0}^3 w_j \dot{w}_j, \quad (11.1.40)$$

which follows from differentiating (1.30). Together the equations (1.37) through (1.40) can be written in the vector/matrix form

$$\begin{pmatrix} 0 \\ \omega_1^{bf}/2 \\ \omega_2^{bf}/2 \\ \omega_3^{bf}/2 \end{pmatrix} = \begin{pmatrix} -w_0 & -w_1 & -w_2 & -w_3 \\ w_1 & -w_0 & w_3 & -w_2 \\ w_2 & -w_3 & -w_0 & w_1 \\ w_3 & w_2 & -w_1 & -w_0 \end{pmatrix} \begin{pmatrix} \dot{w}_0 \\ \dot{w}_1 \\ \dot{w}_2 \\ \dot{w}_3 \end{pmatrix}. \quad (11.1.41)$$

Remarkably, the  $4 \times 4$  matrix on the right side of (1.41) is orthogonal! See Exercise 1.6. Consequently (1.41) can be inverted easily to give the relation

$$\begin{pmatrix} \dot{w}_0 \\ \dot{w}_1 \\ \dot{w}_2 \\ \dot{w}_3 \end{pmatrix} = \begin{pmatrix} -w_0 & w_1 & w_2 & w_3 \\ -w_1 & -w_0 & -w_3 & w_2 \\ -w_2 & w_3 & -w_0 & -w_1 \\ -w_3 & -w_2 & w_1 & -w_0 \end{pmatrix} \begin{pmatrix} 0 \\ \omega_1^{bf}/2 \\ \omega_2^{bf}/2 \\ \omega_3^{bf}/2 \end{pmatrix}. \quad (11.1.42)$$

Taking components of (1.42) yields the four kinematic equations of motion

$$\dot{w}_0 = (1/2)(\omega_1^{bf}w_1 + \omega_2^{bf}w_2 + \omega_3^{bf}w_3), \quad (11.1.43)$$

$$\dot{w}_1 = (1/2)(-\omega_1^{bf}w_0 - \omega_2^{bf}w_3 + \omega_3^{bf}w_2), \quad (11.1.44)$$

$$\dot{w}_2 = (1/2)(\omega_1^{bf}w_3 - \omega_2^{bf}w_0 - \omega_3^{bf}w_1), \quad (11.1.45)$$

$$\dot{w}_3 = (1/2)(-\omega_1^{bf}w_2 + \omega_2^{bf}w_1 - \omega_3^{bf}w_0). \quad (11.1.46)$$

It is these four kinematic equations, along with the three dynamic equations (1.19) through (1.21), that are to be integrated. Note that equations (1.43) through (1.46) are singularity free. Indeed, they are *linear* in the  $w_j$ . They are therefore ideal for numerical integration. For further elaboration on this point, see the discussion at the end of Exercise 1.13.

It is easily verified that exact integration of the differential equations (1.43) through (1.46) preserves the unit sphere condition (1.30) if this condition is satisfied at some initial time  $t^i$ . See Exercise 1.8. However, if we are integrating the equations of motion numerically by some routine that is only exact through order  $h^m$ , we may expect that the condition (1.30) will be violated by terms of order  $h^{m+1}$  at each integration step. That is, instead of remaining on the unit sphere  $S^3$ ,  $w$  will become a general point in the ambient four-dimensional space  $E^4$ . One simple procedure to overcome this problem is to repeatedly project, by simple scaling,  $w \in E^4$  back onto the unit sphere  $S^3$  any time a unit  $w$  is required to compute  $R(w)$ . So doing requires little computational overhead.

### 11.1.8 Modification of the Quaternion Kinematic Equations of Motion

Another procedure is to modify the kinematic equations of motion. Define an *error* quantity  $\epsilon$  that measures the departure of  $w$  from  $S^3$  by the rule

$$\epsilon = 1 - w \cdot w. \quad (11.1.47)$$

Replace equations (1.43) through (1.46) by the modified equations

$$\dot{w}_0 = (1/2)(\omega_1^{bf}w_1 + \omega_2^{bf}w_2 + \omega_3^{bf}w_3) + k\epsilon w_0, \quad (11.1.48)$$

$$\dot{w}_1 = (1/2)(-\omega_1^{bf}w_0 - \omega_2^{bf}w_3 + \omega_3^{bf}w_2) + k\epsilon w_1, \quad (11.1.49)$$

$$\dot{w}_2 = (1/2)(\omega_1^{bf}w_3 - \omega_2^{bf}w_0 - \omega_3^{bf}w_1) + k\epsilon w_2, \quad (11.1.50)$$

$$\dot{w}_3 = (1/2)(-\omega_1^{bf}w_2 + \omega_2^{bf}w_1 - \omega_3^{bf}w_0) + k\epsilon w_3, \quad (11.1.51)$$

where  $k$  is some constant satisfying  $k > 0$ . Evidently the modified equations have the same solutions as the original equations as long as  $\epsilon = 0$ . But, if for some reason  $\epsilon \neq 0$ , the modified equations *drive*  $\epsilon$  to 0 if  $k$  is positive! Again see Exercise 1.8. Thus, even when integrated numerically, the modified equations (1.48) through (1.51) along with (1.19) through (1.21) provide a satisfactory description of rigid body motion, and are commonly used for this purpose in applications that range from inertial guidance (including space craft orientation/control) through robotics to virtual reality.

### 11.1.9 Local Coordinate Patches

So far, in the case of rigid-body motion, we have been able to finesse the problem of maintaining the manifold condition (1.9) either by some projection  $E^9 \rightarrow SO(3, \mathbb{R})$  or the use of a remarkable coordinate system, namely quaternions, and the simpler projection  $E^3 \rightarrow S^3$  by scaling, or by modification of the quaternion kinematic equations of motion. Are there other approaches, and in particular are there approaches that are also applicable in a more general context?

One procedure is to introduce a local coordinate patch at each integration step. Suppose, for a given integration step, we write

$$R(t) = R^b R^v(t). \quad (11.1.52)$$

Here

$$R^b = R(t^b) \quad (11.1.53)$$

where  $t^b$  is the time at the *beginning* of the integration step, and  $R^v$  is a *variable* rotation matrix near the identity with the property

$$R^v(t^b) = I. \quad (11.1.54)$$

From (1.52) we have the relation

$$\dot{R} = R^b \dot{R}^v, \quad (11.1.55)$$

and substituting this result and (1.52) into (1.18) yields the relation

$$\dot{R}^v = R^v \boldsymbol{\omega}^{bf} \cdot \mathbf{L}. \quad (11.1.56)$$

We wish to integrate (1.56) to find  $R^v(t^b + h)$  starting with the initial condition (1.54). Then, having done so, we have from (1.52) the result

$$R(t^b + h) = R^b R^v(t^b + h). \quad (11.1.57)$$

Note that in view of (1.54) what is needed in any given instance, if parameters are to be employed, is a local coordinate patch in the vicinity of the identity.

### 11.1.10 Canonical Coordinates of the Second Kind: Tait-Bryan Angles

In the case of  $SO(3, \mathbb{R})$ , how can we parameterize  $R^v$  near the identity? One possibility is to employ canonical coordinates of the second kind (see Section 7.9) to write, for example, the Ansatz

$$R^v(\lambda) = \exp(\lambda_1 L^1) \exp(\lambda_2 L^2) \exp(\lambda_3 L^3) \quad (11.1.58)$$

with the quantities  $\lambda_j$  being parameters. [In the context of rigid-body motion, the quantities  $\lambda_j$  in (1.58) are called *Tait-Bryan* or *Cardan* angles.] Unlike the Euler-angle parameterization, application of the BCH formula shows that such parameterizations are well defined for  $R^v$  near  $I$  and the  $\lambda_j$  near 0 since all three  $L^j$  are employed.<sup>11</sup>

For these coordinates use of (1.22) yields the relations

$$\omega_1^{bf} = \dot{\lambda}_1 \cos(\lambda_2) \cos(\lambda_3) + \dot{\lambda}_2 \sin(\lambda_3), \quad (11.1.59)$$

$$\omega_2^{bf} = -\dot{\lambda}_1 \cos(\lambda_2) \sin(\lambda_3) + \dot{\lambda}_2 \cos(\lambda_3), \quad (11.1.60)$$

$$\omega_3^{bf} = \dot{\lambda}_1 \sin(\lambda_2) + \dot{\lambda}_3. \quad (11.1.61)$$

And inverting the relations (1.59) through (1.61) yields the kinematic equations of motion

$$\dot{\lambda}_1 = [1/\cos(\lambda_2)][\omega_1^{bf} \cos(\lambda_3) - \omega_2^{bf} \sin(\lambda_3)], \quad (11.1.62)$$

$$\dot{\lambda}_2 = \omega_1^{bf} \sin(\lambda_3) + \omega_2^{bf} \cos(\lambda_3), \quad (11.1.63)$$

$$\dot{\lambda}_3 = \omega_3^{bf} - \tan(\lambda_2)[\omega_1^{bf} \cos(\lambda_3) - \omega_2^{bf} \sin(\lambda_3)]. \quad (11.1.64)$$

See Exercise 1.9. In terms of these coordinates our task is to integrate the equations (1.62) through (1.64) and (1.19) through (1.21) where now

$$N_j^{bf} = N_j^{bf}(\lambda_1, \lambda_2, \lambda_3, \dot{\lambda}_1, \dot{\lambda}_2, \dot{\lambda}_3, t). \quad (11.1.65)$$

Observe, as anticipated, that the equations of motion are nonsingular for small  $\lambda_j$ . However they are singular when  $\lambda_2 = \pm\pi/2$ . (For the source of these singularities, again see Exercise 1.9.) Thus, like the case for Euler angles, these coordinates cannot be used globally.

### 11.1.11 Canonical Coordinates of the First Kind: Angle-Axis Parameters

Alternatively, inspired by the angle-axis parameterization (3.7.199), another possibility is to introduce parameters  $\lambda_1, \lambda_2, \lambda_3$  and make the Ansatz

$$R^v(\lambda) = \exp(\lambda_1 L^1 + \lambda_2 L^2 + \lambda_3 L^3) = \exp(\boldsymbol{\lambda} \cdot \mathbf{L}). \quad (11.1.66)$$

---

<sup>11</sup>Canonical coordinates of the first and second kind are both well defined near the origin for any finite-dimensional Lie group. However, we know that in some cases canonical coordinates of the first kind cannot be set up globally. See Section 3.8. The global status of canonical coordinates of the second kind is less clear. In general there is a global polar decomposition, which amounts to a hybrid of canonical coordinates of the first and second kinds.

That is, we have expressed  $R^v \in SO(3, \mathbb{R})$  and near the identity in terms of elements in the Lie algebra  $so(3, \mathbb{R})$ ; and we have parameterized the Lie algebra. Since the map between group elements near the identity and Lie elements near the origin is well defined, we know that this parameterization is well defined for  $R^v$  near  $I$  and the  $\lambda_j$  near 0. As described in Section 7.9, this parameterization amounts to using canonical coordinates of the first kind.

This parameterization yields the remarkably symmetric looking kinematic equations of motion

$$\dot{\boldsymbol{\lambda}} = \boldsymbol{\omega}^{bf} + (1/2)(\boldsymbol{\lambda} \times \boldsymbol{\omega}^{bf}) + [(\boldsymbol{\lambda} \cdot \boldsymbol{\omega}^{bf})\boldsymbol{\lambda} - (\boldsymbol{\lambda} \cdot \boldsymbol{\lambda})\boldsymbol{\omega}^{bf}]\{1/|\boldsymbol{\lambda}|^2 - [1/(2|\boldsymbol{\lambda}|)]\cot(|\boldsymbol{\lambda}|/2)\}. \quad (11.1.67)$$

See Exercise 1.10. For the inverse relation that expresses  $\boldsymbol{\omega}^{bf}$  in terms of  $\boldsymbol{\lambda}$  and  $\dot{\boldsymbol{\lambda}}$ , see Exercise 1.11.

Observe, again as anticipated, that the equations of motion are nonsingular for small  $\lambda_j$ , namely when  $|\boldsymbol{\lambda}| < 2\pi$ . But they are singular when  $|\boldsymbol{\lambda}| = 2\pi$ . We expect this singularity to occur because we see from (3.7.188) and (3.7.202) that the individual components of  $\mathbf{n}$  are not defined in terms of  $v$  or  $R$  when  $\theta = |\boldsymbol{\lambda}| = 2\pi$ .

We reiterate that even when coordinates can be set up globally, as in the case of  $SO(3, \mathbb{R})$  using either Euler angles or angle-axis parameters, there may still be singularities in the equations of motion because at some points the inverse map from the group to the Lie algebra may not be well defined. We have seen an example of this problem in the case of Euler angles. Every element of  $SO(3, \mathbb{R})$  can be written in Euler form, but at  $\theta = 0$  and  $\theta = \pi$  the inverse map is not well defined. The same is true for angle-axis parameters. Every element of  $SO(3, \mathbb{R})$  can be written in angle-axis form, but at  $\theta = 2\pi$  the inverse map is not well defined.

### 11.1.12 Cayley Parameters

Quadratic groups, including  $SO(3, \mathbb{R})$  and  $SU(2)$ , can also be parameterized near the identity in terms of Cayley parameters. When Cayley parameters are used for  $SO(3, \mathbb{R})$  or  $SU(2)$ , call them  $\mu_j$ , the task is to find differential equations that specify the  $\dot{\mu}_j$  in terms of the  $\mu_j$  and  $\boldsymbol{\omega}^{bf}$ .

We first consider the case of  $SO(3, \mathbb{R})$  and employ the Cayley parameterization

$$R^v(\boldsymbol{\mu}) = (I + \boldsymbol{\mu} \cdot \mathbf{L})(I - \boldsymbol{\mu} \cdot \mathbf{L})^{-1}. \quad (11.1.68)$$

For this parameterization it can be shown that there are the kinematic equations of motion

$$\dot{\boldsymbol{\mu}} = (1/2)[\boldsymbol{\omega}^{bf} + (\boldsymbol{\mu} \times \boldsymbol{\omega}^{bf}) + (\boldsymbol{\mu} \cdot \boldsymbol{\omega}^{bf})\boldsymbol{\mu}]. \quad (11.1.69)$$

See Exercise 1.13.

Next consider the case of  $SU(2)$  and employ the Cayley parameterization

$$u^v(\boldsymbol{\mu}) = (I + \boldsymbol{\mu} \cdot \mathbf{K})(I - \boldsymbol{\mu} \cdot \mathbf{K})^{-1}. \quad (11.1.70)$$

Here we have again used the notation of Exercise 3.7.30. In this case there are the kinematic equations of motion

$$\dot{\boldsymbol{\mu}} = (1/2)[\boldsymbol{\omega}^{bf} + (\boldsymbol{\mu} \times \boldsymbol{\omega}^{bf}) + (1/2)(\boldsymbol{\mu} \cdot \boldsymbol{\omega}^{bf})\boldsymbol{\mu} - (1/4)(\boldsymbol{\mu} \cdot \boldsymbol{\mu})\boldsymbol{\omega}^{bf}]. \quad (11.1.71)$$

Again see Exercise 1.13.

What can be said about the singularity structure of the kinematic equations of motion (1.69) and (1.71)? Strangely enough, both sets appear to be singularity free! However, this appearance is deceptive because both sets of kinematic equations of motion are singular in  $\mu$  at infinity, and there is the possibility that this singularity can be encountered in *finite* time. Yet again see Exercise 1.13.

### 11.1.13 Summary of Integration Using Local Coordinates

Upon employing (1.58) or (1.66) in (1.56) we can find first-order differential equations that, analogous to the relations (1.26) through (1.28), specify  $\dot{\lambda}_1, \dot{\lambda}_2, \dot{\lambda}_3$  in terms of the  $\lambda_j$  and  $\omega_j^{bf}$ . Moreover, these equations will be singularity free in the neighborhood the origin in  $\lambda$  space. What we have done is to convert the group-space differential equation (1.56) into a set of differential equations for the parameters  $\lambda_j$ . These equations, with the initial conditions

$$\lambda_j(t^b) = 0, \quad (11.1.72)$$

see (1.54), can be integrated numerically for one integration step using any convenient method to find the quantities  $\lambda_j(t^b + h)$ . If  $h$  is sufficiently small, the  $\lambda(t)$  for  $t \in [t^b, t^b + h]$  will remain small, and the differential equations specifying the  $\dot{\lambda}_j$  will remain singularity free over the course of integration. Once the quantities  $\lambda_j(t^b + h)$  have been found,  $R(t^b + h)$  is given, depending on what type of canonical parameterization has been employed, by either (1.58) or (1.66).

As a modification of this procedure, one may integrate for  $k$  steps to find  $\lambda_j(t^b + kh)$  and subsequently  $R(t^b + kh)$ . What is required then is continual checking that the  $\lambda_j$  have not come too close to singular values.

Evidently generalizations of the methods just described can in principle be employed for any Lie group. The difficulties encountered lie only in determining the equations that specify the  $\dot{\lambda}_j$ , see for example Exercises 1.9 and 1.10, and in evaluating the exponentials that occur with the use of (1.58) or (1.66). Assuming these exponentials can be evaluated accurately, the result for  $R^v(t^b + h)$  will lie on the group manifold to high accuracy even though the  $\lambda_j(t^b + h)$  are only exact through terms of order  $h^m$ . Finally, we must assume that the group multiplications involved in (1.57) can also be carried out with high accuracy.

Similarly, with the use of Cayley parameters, the difficulties lie only in determining the equations that specify the  $\dot{\mu}_j$  [see for example (1.69) and (1.71)], and in carrying out the matrix inversions and multiplications that occur with the use of a Cayley representation [see (1.68) and (1.70)]. Assuming these inversions and multiplications can be evaluated accurately, the results for  $R^v(t^b + h)$  or  $u^v(t^b + h)$  will lie on the associated group manifold to high accuracy even though the  $\mu_j(t^b + h)$  are only exact through terms of order  $h^m$ . Finally, we must again assume that the group multiplications involved in (1.57), for example, can also be carried out with high accuracy.

### 11.1.14 Integration in the Lie Algebra: Exponential Representation

With a suitable translation of the origin in time, the differential equation (1.56) with the initial condition (1.54) is a special case of the general differential equation of the form

$$\dot{M}(t) = M(t)A(t) \quad (11.1.73)$$

with the initial condition

$$M(0) = I, \quad (11.1.74)$$

and our goal is to find  $M(h)$ . Here  $M(t)$  is expected to belong to some Lie group  $G$  and is near the identity for small  $t$ ; and  $A(t)$  belongs to the Lie algebra of  $G$ , which we denote by  $\mathcal{L}(G)$ . See Appendix C. In particular, we wish to obtain  $M(h)$  by numerical means with a possible local error of order  $h^{m+1}$  and, despite this possible error, we want to guarantee that  $M(h)$  is in  $G$ .

Since  $M(t)$  is near the identity, we may write

$$M(t) = \exp[B(t)] \quad (11.1.75)$$

where  $B(t)$  is in  $\mathcal{L}(G)$  and is small (near the origin). If we can find  $B(h)$  with a local error of order  $h^{m+1}$  and if, despite this possible error, we can assure that  $B(h)$  is in  $\mathcal{L}(G)$ , then we know that  $M(h) = \exp[B(h)]$  will have the desired local accuracy and is guaranteed to be in  $G$ . We will now see that a suitable  $B(h)$  can be found by converting the differential equation (1.73), a *group* differential equation for the group elements  $M(t)$  in terms of the group elements  $M(t)$  and the Lie elements  $A(t)$ , into a *Lie* differential equation for the Lie elements  $B(t)$  in terms of the Lie elements  $B(t)$  and the Lie elements  $A(t)$ .

Before proceeding further, here is a chance to learn some terminology: Thinking geometrically, we may view  $M(t)$  as a path in  $G$ , and we may view  $B(t)$  as a path in  $\mathcal{L}(G)$ . In this context, the path  $B(t)$  is said to be a *lift* of the path  $M(t)$ . That is, we may view  $\mathcal{L}(G)$ , the Lie algebra of  $G$ , as lying “above” the group  $G$ . Correspondingly, we may say that the path  $B(t)$  is obtained by “lifting” the path  $M(t)$  in the group  $G$  up to a path in  $\mathcal{L}(G)$ . Upon solving (1.75) for  $B$  in terms of  $M$ , we have the relation

$$B(t) = \log[M(t)], \quad (11.1.76)$$

and we see that the lift in question is accomplished by the logarithmic function. Conversely, in view of (1.75), the exponential function “lowers” the path  $B(t)$  in  $\mathcal{L}(G)$  down to the path  $M(t)$  in the group  $G$ .

Let us continue. Given the differential equation (1.73) for the path  $M(t)$ , what is the associated differential equation for the path  $B(t)$ ? The relation (1.75) can be differentiated to yield the result

$$\dot{M}(t) = M(t) \operatorname{dex}[-\#B(t)\#]\dot{B}(t). \quad (11.1.77)$$

See (\*) in Appendix C. Also, in view of (1.73), the relation (1.77) can be rewritten in the form

$$\operatorname{dex}[-\#B(t)\#]\dot{B}(t) = A(t). \quad (11.1.78)$$

Finally, as in Section 10.3, the relation (1.78) can be inverted to become

$$\dot{B}(t) = \{\text{iex}[-\#B(t)\#]\}^{-1}A(t). \quad (11.1.79)$$

We have obtained a (somewhat fearsome looking) differential equation for  $B$  in terms of  $B$  and  $A$ . Our task is to integrate this equation from  $t = 0$  to  $t = h$  with, in view of (1.74), the initial condition

$$B(0) = 0. \quad (11.1.80)$$

What can we make of (1.79)? The function  $[\text{iex}(-w)]^{-1}$  has an expansion of the form

$$[\text{iex}(-w)]^{-1} = \sum_{\ell=0}^{\infty} b_{\ell} w^{\ell} \quad (11.1.81)$$

where the coefficients  $b_{\ell}$  are known. Again see Appendix C. Correspondingly, the differential equation (1.79) is equivalent to the equation

$$\begin{aligned} \dot{B}(t) &= \sum_{\ell=0}^{\infty} b_{\ell} [\#B(t)\#]^{\ell} A(t) \\ &= \{b_0 + b_1 [\#B(t)\#] + b_2 [\#B(t)\#]^2 + \dots\} A(t) \\ &= b_0 A(t) + b_1 \{B(t), A(t)\} + b_2 \{B(t), \{B(t), A(t)\}\} + \dots \end{aligned} \quad (11.1.82)$$

In general, all terms in the expansion (1.82) need to be retained. That is, in general, we need to sum the series (1.82) which, in the general case, can be a formidable task.<sup>12</sup>

However, suppose we only wish to obtain  $B(h)$  through some order in  $h$ . From the Magnus expansion, see Section 10.3, we know that  $B(t)$  is of order  $h$  for  $t \in [0, h]$ . It follows that the term  $b_{\ell} [\#B(t)\#]^{\ell} A(t)$  is of order  $h^{\ell}$  and therefore contributes a term of order  $h^{\ell+1}$  to  $B(h)$ . Suppose we truncate the series (1.82) beyond  $\ell = n$  with  $n$  even. Then the size of the first omitted term will be of order  $h^{n+3}$ .<sup>13</sup> The result is the replacement of the differential equation (1.82) with the truncated equation

$$\begin{aligned} \dot{B}(t) &= \sum_{\ell=0}^{\ell=n} b_{\ell} [\#B(t)\#]^{\ell} A(t) \\ &= b_0 A(t) + b_1 \{B(t), A(t)\} + b_2 \{B(t), \{B(t), A(t)\}\} + \dots \\ &\quad + b_n \{B(t), \{B(t), \{\dots \{B(t), A(t)\} \dots\}\}\}, \end{aligned} \quad (11.1.83)$$

and the understanding is that this truncated equation is to be integrated only over the interval  $t \in [0, h]$ . Then the  $B(h)$  so obtained will be correct through order  $h^m$  with  $m = n + 2$ . We have obtained a tractable problem. We have also been introduced to a new idea: the equation of motion to be integrated may be *modified* depending on the desired local accuracy for the integrated result.

---

<sup>12</sup>Exercise 1.10 shows, in effect, how this summation can be done for the cases of the Lie algebras  $su(2)$  and  $so(3, \mathbb{R})$ .

<sup>13</sup>Examination of the  $b_{\ell}$ , see Appendix C, reveals (save for  $b_1 = 1/2$ ) that they all vanish for odd  $\ell$ .

Still it might appear that, even with all our efforts, not much has been accomplished. We will soon see that progress has indeed been made. First, let us perform a sanity check on the results obtained so far. Suppose that  $B(t) \in \mathcal{L}(G)$ . From the definition

$$\dot{B}(t) = \lim_{\epsilon \rightarrow 0} [B(t + \epsilon) - B(t)]/\epsilon \quad (11.1.84)$$

we see that only vector space operations are involved in the calculation of  $\dot{B}(t)$ , and therefore  $\dot{B}(t)$ , which appears on the left side of (1.83), must also be in  $\mathcal{L}(G)$ . But is the right side of (1.83) in  $\mathcal{L}(G)$ ? We know that  $A(t)$  is in  $\mathcal{L}(G)$ . Also, all the rest of the right side of (1.83) involves sums of commutators of  $B(t)$  with  $A(t)$ . By definition,  $\mathcal{L}(G)$  is closed under addition and commutation. Therefore the right side of (1.83) is also in  $\mathcal{L}(G)$ . We conclude (1.83) is sane at least to the extent that both sides are in  $\mathcal{L}(G)$ .

But now we make a key observation: Suppose (1.83) is integrated by some numerical integrator to find  $B(t)$ . Examination of the various numerical integration schemes (for example all those discussed in Chapter 2) reveals that they all involve just *linear* combinations of the right side of the differential equation in question evaluated at various times and coordinate values. We know, by definition, that  $\mathcal{L}(G)$  is closed under addition. Therefore, if the right side of the differential equation is known to be in  $\mathcal{L}(G)$  at all evaluation points, then the result of numerically integrating such an equation is guaranteed to be in  $\mathcal{L}(G)$ . Since we have verified that the right side of (1.83) is in  $\mathcal{L}(G)$ , it follows that the  $B(t)$  obtained by numerical integration, whatever the accuracy of the method, is guaranteed to be in  $\mathcal{L}(G)$ . Finally, if we wish to obtain  $B(h)$  with an accuracy of some desired order in  $h$ , we may truncate (1.82) to obtain (1.83) with a known accuracy, and then integrate (1.83) using any integrator whose order equals or exceeds the accuracy of (1.83).

In summary, where conveniently feasible, integration in the Lie algebra is advantageous compared to integration in the group because numerical integration schemes, no matter their accuracy, generally preserve Lie algebraic structure.<sup>14</sup> Put another way, suppose the matrices  $A$  and  $B$  are  $k \times k$ . Then (1.83) is a differential equation in an ambient  $k^2$  dimensional space. Numerical integration of (1.83) by any of the standard methods produces a sequence of points in this ambient space corresponding to the times  $t^n$ . We have seen that if the initial point is in the subspace  $\mathcal{L}(G)$ , then all subsequent points will also be in  $\mathcal{L}(G)$ . In our case, the initial point is given by (1.80), and is obviously in  $\mathcal{L}(G)$ . Therefore all subsequent points will also be in  $\mathcal{L}(G)$ .<sup>15</sup>

### 11.1.15 Integration in the Lie Algebra: Cayley Representation

We have used the exponential and logarithmic functions to relate  $G$  and  $\mathcal{L}(G)$ . For quadratic groups, which are often of interest, one may also use a Cayley representation to provide a

---

<sup>14</sup>Note that the work of Subsection 10.4.2 essentially employs a hybrid of canonical coordinates of the first and second kinds for the the nonlinear part of a symplectic map and formulates differential equations in the Lie algebra. The work of Section 10.3 and Subsection 10.4.3 also formulates differential equations in the Lie algebra.

<sup>15</sup>In the context of the exponential representation and in the case of finite-dimensional groups, the strategy of integrating in the Lie algebra and making the truncation (1.83) was pioneered by *Munthe-Kaas*; and the use of Runge-Kutta in this setting is often referred to as *RKMK* integration.

map between group elements near the identity and Lie algebra elements near the origin. That is, lowering and lifting can be done using the Cayley transformation and its inverse. See Section 3.12. Where possible, employing this approach yields a Lie algebraic differential equation analogous to (1.82), but with the advantage that only three terms appear on the right side. Therefore, *no* truncation is required. See Exercise 1.12.

We again seek to integrate equations of the form (1.73) with the initial condition (1.74) where now  $M(t)$  is expected to belong to some quadratic Lie group  $G$ , and  $A(t)$  belongs to its associated Lie algebra. What we again desire is a way of numerically integrating (1.73) that guarantees  $M(t)$  is in  $G$  even though the numerical solution may be locally exact only through terms of order  $h^m$ .

When  $M$  belongs to a quadratic group  $G$ , we may employ the Cayley parameterization

$$M = (I + V)(I - V)^{-1}, \quad (11.1.85)$$

where  $V$  is in the Lie algebra of  $G$ . The challenge now is to find the equation of motion for  $V$ . In Exercise 1.12 you will show that the desired result is the equation of motion

$$\dot{V} = (1/2)(A + \{V, A\} - VAV). \quad (11.1.86)$$

Note that, in contrast to (1.82) whose right side contains an infinite number of terms, the right side of (1.86) contains only three terms. Therefore no truncation is necessary and, unlike the case of  $B(t)$ , the accuracy to which  $V(t)$  is calculated depends only on the method of integration.

Is this result sane? For a quadratic group  $G$  we know that  $V$  is in the Lie algebra. By an argument identical to that made in Subsection 1.14 for the case of  $\dot{B}(t)$ , it follows that  $\dot{V}(t)$  must also be in the Lie algebra of  $G$ . But is the right side of (1.86) in the Lie algebra of  $G$ ? Evidently, since  $A$  is in the Lie algebra of  $G$ , the first two terms on the right side of (1.86) are in the Lie algebra of  $G$ . What about the third term  $VAV$ ? It can be shown that for a quadratic group the quantity  $VAV$  is also in the Lie algebra of  $G$ . Again see Exercise 1.12. Therefore (1.86) is sane at least to the extent that both its sides are in the Lie algebra of  $G$ .

Suppose (1.86) is integrated by some numerical integrator to find  $V(t)$  under the assumption that  $V$  is initially in the Lie algebra of  $G$ . We repeat the key observation of Subsection 1.14: Examination of the usual numerical integration schemes, see Chapter 2, reveals that they all involve just linear combinations of the right side of the differential equation in question evaluated at various times and coordinate values. Therefore, if the right side is known to be in the Lie algebra of  $G$  for all evaluation points, then the result of numerically integrating such an equation is guaranteed to be in the Lie algebra of  $G$ , no matter what the local accuracy of the integrator or the step size employed. Since we have verified that the right side of (1.86) is in the Lie algebra of  $G$ , it follows that  $V(t)$  will be in the Lie algebra of  $G$  if it is initially in the Lie algebra of  $G$ . Finally, since  $V(t)$  is in the Lie algebra of  $G$ , it follows that  $M(t)$  given by (1.85) is in  $G$ .

We have achieved our goal of, in effect, numerically integrating (1.73) in such a way that  $M(t)$  is guaranteed to be in  $G$ . All that is required is accurate matrix multiplication in the calculation of the right side of (1.86) and accurate matrix inversion and multiplication in the evaluation of (1.85). Note, however, that this procedure cannot be carried out globally since

the Cayley parameterization (1.85) cannot be made globally. It may therefore be necessary to change coordinate systems (by left or right group translation) from time to time during the course of a numerical integration in order to stay clear of the singularities associated with any given Cayley parametrization.<sup>16</sup>

### 11.1.16 Parameterization of $G$ and $\mathcal{L}(G)$

To reiterate a point, if the matrices  $M$  or  $B$  are  $k \times k$ , then equations of the kind (1.73) or (1.82) involve  $k^2$  variables. By contrast, the group  $G$ , and correspondingly its Lie algebra  $\mathcal{L}(G)$ , generally have much smaller dimension. Therefore it may be advantageous to parameterize the group or Lie algebra and to convert the differential equations for the group or Lie algebra into (usually) far fewer differential equations for the parameters. This is what was done for the case of  $SO(3, \mathbb{R})$  by the use of quaternion parameters, the use of Tait-Bryan angles, and the use of angle-axis parameters.

By the introduction of a basis it is also possible to parameterize the Lie elements that occur in a Cayley formulation. That is, we parameterize the  $A$  and  $V$  appearing in (1.86). Again, as illustrated in Subsection 1.12, the result is (usually) far fewer equations that need to be integrated. See also Exercise 1.13.

### 11.1.17 Quaternions Revisited

We close this subsection by remarking that, in the case of  $SO(3, \mathbb{R})$  and in the context of local coordinates, there is a still better approach, which again uses quaternions: Namely, suppose we again write (1.56) but now parameterize  $R^v$  in terms of quaternions. In this case the relations (1.43) through (1.46) will continue to hold and, in view of (1.54), there will be the initial conditions

$$w_0(t^b) = 1, \quad (11.1.87)$$

$$w_j(t^b) = 0 \text{ for } j = 1, 2, 3. \quad (11.1.88)$$

Also, if  $h$  is sufficiently small, the  $w_j$  for  $j = 1, 2, 3$  will remain small over the course of a single integration step. We may therefore enforce the condition (1.30) by writing

$$w_0 = [1 - (w_1^2 + w_2^2 + w_3^2)]^{1/2} \quad (11.1.89)$$

and insert this result into (1.44) through (1.46). So doing gives the modified equations of motion

$$\dot{w}_1 = (1/2)\{-\omega_1^{bf}[1 - (w_1^2 + w_2^2 + w_3^2)]^{1/2} - \omega_2^{bf}w_3 + \omega_3^{bf}w_2\}, \quad (11.1.90)$$

$$\dot{w}_2 = (1/2)\{\omega_1^{bf}w_3 - \omega_2^{bf}[1 - (w_1^2 + w_2^2 + w_3^2)]^{1/2} - \omega_3^{bf}w_1\}, \quad (11.1.91)$$

$$\dot{w}_3 = (1/2)\{-\omega_1^{bf}w_2 + \omega_2^{bf}w_1 - \omega_3^{bf}[1 - (w_1^2 + w_2^2 + w_3^2)]^{1/2}\}. \quad (11.1.92)$$

These three kinematic equations, whose right sides depend only on  $(w_1, w_2, w_3)$  and on  $(\omega_1^{bf}, \omega_2^{bf}, \omega_3^{bf})$ , along with the three dynamic equations (1.19) through (1.21), are to be

---

<sup>16</sup>It might appear the the right side of (1.86) is singularity free. However, it is singular at infinity, and this singularity can be reached in finite time. See the discussion in Exercise 1.13.

integrated over the interval  $t \in [t^b, t^b + h]$ . (And any kind of integrator can be used.) Whenever  $w_0$  is needed, it is to be computed from (1.89). Note that in view of (8.2.73), no exponentials need be computed to find  $R^v(t)$ . Also, the kinematic equations are mostly linear and involve, at worst, the calculation of a square root. Moreover, if desired, the equivalent of matrix multiplication in  $SO(3, \mathbb{R})$  can be carried out at the quaternion level where the operations are computationally simpler. See (5.10.24) and Exercises 8.2.10 and 8.2.11. For all these reasons, relatively few floating-point operations are needed to carry out an integration step. Finally, although the quaternion parameters  $(w_1, w_2, w_3)$  may only be computed with a local error of order  $h^{m+1}$ , the resulting matrix  $R_v(t^b + h)$  will be orthogonal to a very high accuracy limited only by roundoff error.

## Exercises

**11.1.1.** The purpose of this exercise is to explore some of the consequences of the definition of angular velocity given in Subsection 1.1. The first task is a bit of housekeeping. In deriving (1.10) through (1.12) it was tacitly assumed that

$$[(d/dt)R]^T = (d/dt)(R^T). \quad (11.1.93)$$

That is, the operations of differentiating and transposing commute. Verify that this is so for any matrix.

Now move on to the main task. The components  $\omega_j^{bf}(t)$  of the angular velocity in the *body-fixed* frame are defined by the rule (1.4). Accordingly, define the angular velocity *vector*  $\boldsymbol{\omega}(t)$  by the rule

$$\boldsymbol{\omega}(t) = \sum_j \omega_j^{bf}(t) \mathbf{f}_j(t). \quad (11.1.94)$$

Let us compute the vector  $\boldsymbol{\omega} \times \mathbf{f}_1$ . Verify the chain of relations

$$\begin{aligned} \boldsymbol{\omega} \times \mathbf{f}_1 &= \omega_2^{bf} \mathbf{f}_2 \times \mathbf{f}_1 + \omega_3^{bf} \mathbf{f}_3 \times \mathbf{f}_1 = -\omega_2^{bf} \mathbf{f}_3 + \omega_3^{bf} \mathbf{f}_2 \\ &= (\mathbf{f}_3 \cdot \dot{\mathbf{f}}_1) \mathbf{f}_3 - (\mathbf{f}_1 \cdot \dot{\mathbf{f}}_2) \mathbf{f}_2 \\ &= (\mathbf{f}_3 \cdot \dot{\mathbf{f}}_1) \mathbf{f}_3 + (\mathbf{f}_2 \cdot \dot{\mathbf{f}}_1) \mathbf{f}_2 + (\mathbf{f}_1 \cdot \dot{\mathbf{f}}_1) \mathbf{f}_1 \\ &= \dot{\mathbf{f}}_1. \end{aligned} \quad (11.1.95)$$

Here we have used the fact

$$\mathbf{f}_1 \cdot \mathbf{f}_2 = 0 \quad (11.1.96)$$

from which it follows that

$$\dot{\mathbf{f}}_1 \cdot \mathbf{f}_2 + \mathbf{f}_1 \cdot \dot{\mathbf{f}}_2 = 0, \quad (11.1.97)$$

and the fact

$$\mathbf{f}_1 \cdot \mathbf{f}_1 = 1 \quad (11.1.98)$$

from which it follows that

$$\mathbf{f}_1 \cdot \dot{\mathbf{f}}_1 = 0. \quad (11.1.99)$$

It is easily checked that there are two more relations like (1.95), and that they together give the general equations of motion

$$\dot{\mathbf{f}}_j(t) = \boldsymbol{\omega}(t) \times \mathbf{f}_j(t). \quad (11.1.100)$$

Verify from (1.100) that

$$\mathbf{f}_j \times \dot{\mathbf{f}}_j = \mathbf{f}_j \times (\boldsymbol{\omega} \times \mathbf{f}_j) = \boldsymbol{\omega} - (\boldsymbol{\omega} \cdot \mathbf{f}_j) \mathbf{f}_j, \quad (11.1.101)$$

and therefore

$$\sum_j \mathbf{f}_j \times \dot{\mathbf{f}}_j = 3\boldsymbol{\omega} - \sum_j (\boldsymbol{\omega} \cdot \mathbf{f}_j) \mathbf{f}_j = 3\boldsymbol{\omega} - \boldsymbol{\omega} = 2\boldsymbol{\omega}. \quad (11.1.102)$$

Conclude that  $\boldsymbol{\omega}$  is also given by the relation

$$\boldsymbol{\omega} = (1/2) \sum_j \mathbf{f}_j \times \dot{\mathbf{f}}_j. \quad (11.1.103)$$

Moreover, suppose  $\mathbf{v}(t)$  is any vector that is “fixed” in the body. That is, suppose  $\mathbf{v}(t)$  has an expansion of the form

$$\mathbf{v}(t) = \sum_j v_j \mathbf{f}_j(t) \quad (11.1.104)$$

where the components  $v_j$  are *constant* numbers. Show that it follows from (1.100) that  $\mathbf{v}$  obeys the equation of motion

$$\dot{\mathbf{v}}(t) = \boldsymbol{\omega}(t) \times \mathbf{v}(t). \quad (11.1.105)$$

Let us compute the  $\omega_j^{bf}(t)$  for a special case. Suppose  $R(t)$  is of the form

$$R(t) = \exp[\theta(t)\mathbf{n} \cdot \mathbf{L}] \quad (11.1.106)$$

where  $\mathbf{n}$  is a constant vector. Verify that

$$\dot{R}(t) = \exp[\theta(t)\mathbf{n} \cdot \mathbf{L}] \dot{\theta}(t) \mathbf{n} \cdot \mathbf{L}. \quad (11.1.107)$$

Show, from (1.22), that in this case

$$\omega_j^{bf}(t) = \dot{\theta}(t)n_j. \quad (11.1.108)$$

This special case illustrates why the name *angular velocity* is appropriate.

What can be said more generally? Suppose  $R(t)$  is a time-dependent matrix in  $SO(3, \mathbb{R})$ . Show that

$$\begin{aligned} R(t + dt) &= R(t) + \dot{R}(t)dt + O[(dt)^2] = R(t)[I + R^{-1}(t)\dot{R}(t)dt] + O[(dt)^2] \\ &= R(t)[I + (dt)\boldsymbol{\omega}^{bf}(t) \cdot \mathbf{L}] + O[(dt)^2] \\ &= R(t) \exp[(dt)\boldsymbol{\omega}^{bf}(t) \cdot \mathbf{L}] + O[(dt)^2]. \end{aligned} \quad (11.1.109)$$

You have shown that, through terms of order  $dt$ ,  $R(t + dt)$  is gotten from  $R(t)$  by translating  $R(t)$  on the right with the near-identity element  $\exp[(dt)\boldsymbol{\omega}^{bf}(t) \cdot \mathbf{L}]$ .

Can  $R(t + dt)$  and  $R(t)$  instead be related by translating  $R(t)$  on the left with a near-identity element? Show that the answer is *yes*. Verify the manipulations

$$\begin{aligned} R(t + dt) &= R(t) + \dot{R}(t)dt + O[(dt)^2] = [I + dt\dot{R}(t)R^{-1}(t)]R(t) + O[(dt)^2] \\ &= \{I + dtR(t)[R^{-1}(t)\dot{R}(t)]R^{-1}(t)\}R(t) + O[(dt)^2] \\ &= \{I + dtR(t)[\boldsymbol{\omega}^{bf}(t) \cdot \mathbf{L}]R^{-1}(t)\}R(t) + O[(dt)^2]. \end{aligned} \quad (11.1.110)$$

Next verify that

$$R(t)[\boldsymbol{\omega}^{bf}(t) \cdot \mathbf{L}]R^{-1}(t) = \boldsymbol{\omega}^{sf}(t) \cdot \mathbf{L} \quad (11.1.111)$$

where  $\boldsymbol{\omega}^{sf}(t)$  is defined by the relation

$$\boldsymbol{\omega}^{sf}(t) = R(t)\boldsymbol{\omega}^{bf}(t). \quad (11.1.112)$$

See (8.2.59). Combine (1.110) through (1.112) to get the result

$$\begin{aligned} R(t + dt) &= \{I + dtR(t)[\boldsymbol{\omega}^{bf}(t) \cdot \mathbf{L}]R^{-1}(t)\}R(t) + O[(dt)^2] \\ &= \{I + dt\boldsymbol{\omega}^{sf}(t) \cdot \mathbf{L}\}R(t) + O[(dt)^2] \\ &= \exp[(dt)\boldsymbol{\omega}^{sf}(t) \cdot \mathbf{L}]R(t) + O[(dt)^2]. \end{aligned} \quad (11.1.113)$$

You have shown that, through terms of order  $dt$ ,  $R(t + dt)$  is gotten from  $R(t)$  by translating  $R(t)$  on the left with the near-identity element  $\exp[(dt)\boldsymbol{\omega}^{sf}(t) \cdot \mathbf{L}]$ . Verify also that (1.113) implies the relation

$$\boldsymbol{\omega}^{sf}(t) \cdot \mathbf{L} = \dot{R}(t)R^{-1}(t), \quad (11.1.114)$$

which is to be compared with (1.22).

According to (1.112), the quantities we have called  $\omega_j^{sf}$  are related to the  $\omega_j^{bf}$  by the rule

$$\omega_j^{sf} = \sum_k R_{jk}\omega_k^{bf}. \quad (11.1.115)$$

What is their significance? Recall the angular velocity vector  $\boldsymbol{\omega}(t)$  defined by (1.94). Rewrite this definition using the dummy index  $k$  to obtain the equivalent definition

$$\boldsymbol{\omega}(t) = \sum_k \omega_k^{bf}(t)\mathbf{f}_k(t). \quad (11.1.116)$$

Since the *space-fixed* vectors  $\mathbf{e}_j$  form a basis, we may make the expansion

$$\boldsymbol{\omega}(t) = \sum_j [\mathbf{e}_j \cdot \boldsymbol{\omega}(t)]\mathbf{e}_j. \quad (11.1.117)$$

Use the representation (1.116) to compute the expansion coefficients  $[\mathbf{e}_j \cdot \boldsymbol{\omega}(t)]$ . Verify that so doing gives the result

$$\begin{aligned} \mathbf{e}_j \cdot \boldsymbol{\omega}(t) &= \sum_k \omega_k^{bf}(t)[\mathbf{e}_j \cdot \mathbf{f}_k(t)] = \sum_k \omega_k^{bf}(t)\{\mathbf{e}_j \cdot [R(t)\mathbf{e}_k(t)]\} \\ &= \sum_k \omega_k^{bf}(t)R_{jk}(t) = \sum_k R_{jk}(t)\omega_k^{bf}(t) = \omega_j^{sf}(t), \end{aligned} \quad (11.1.118)$$

where use has been made of (1.2) and (1.115). By putting everything together, show that (1.117) can be rewritten in the form

$$\boldsymbol{\omega}(t) = \sum_j \omega_j^{sf}(t) \mathbf{e}_j. \quad (11.1.119)$$

We conclude that the quantities  $\omega_j^{sf}(t)$  are the *space-fixed* components of the angular velocity vector  $\boldsymbol{\omega}(t)$ .

Finally show, for the special case (1.106), that

$$\omega_j^{sf}(t) = \omega_j^{bf}(t). \quad (11.1.120)$$

**11.1.2.** The purpose of this exercise is to show that the matrix differential equation (1.18) preserves the orthogonality condition (1.9), and conversely.

Begin with the converse. We already know from the orthogonality assumption (1.9) that the matrix  $(R^{-1}\dot{R})$  must be antisymmetric. See (1.12). Therefore for orthogonal  $3 \times 3$  matrices a relation of the form (1.18) must hold.

Now prove the main assertion. Show, by taking transposes, that (1.18) implies the relation

$$(d/dt)R^T = -\boldsymbol{\omega}^{bf} \cdot \mathbf{L} R^T. \quad (11.1.121)$$

Next show that there is the relation

$$(d/dt)(RR^T) = [(d/dt)R]R^T + R(d/dt)R^T = R(\boldsymbol{\omega}^{bf} \cdot \mathbf{L})R^T - R(\boldsymbol{\omega}^{bf} \cdot \mathbf{L})R^T = 0. \quad (11.1.122)$$

Here use has been made of (1.18) and (1.121). Assume that  $R$  is orthogonal at the initial time  $t^i$ ,

$$R(t^i)R(t^i)^T = I = R^T(t^i)R(t^i). \quad (11.1.123)$$

Verify that the solution to the differential equation (1.122) with the initial condition (1.123) is

$$R(t)R(t)^T = I = R^T(t)R(t). \quad (11.1.124)$$

You have shown that (1.18) preserves orthogonality.

**11.1.3.** The purpose of this exercise is to derive equations (1.23) through (1.28), the expressions for the  $\omega_j^{bf}$  in terms of Euler angles. Recall the Euler-angle parameterization (3.7.207),

$$R(t) = \exp[\phi(t)L^3] \exp[\theta(t)L^2] \exp[\psi(t)L^3]. \quad (11.1.125)$$

According to (1.22), we must compute the quantities  $\omega_j^{bf}$  defined by the relation

$$\boldsymbol{\omega}^{bf} \cdot \mathbf{L} = R^{-1}\dot{R}. \quad (11.1.126)$$

Verify that  $R^{-1}$  is given by the relation

$$R^{-1} = \exp(-\psi L^3) \exp(-\theta L^2) \exp(-\phi L^3). \quad (11.1.127)$$

Show that  $\dot{R}$  is given by the relation

$$\begin{aligned}\dot{R} = & \dot{\phi}L^3 \exp(\phi L^3) \exp(\theta L^2) \exp(\psi L^3) \\ & + \exp(\phi L^3) \dot{\theta}L^2 \exp(\theta L^2) \exp(\psi L^3) \\ & + \exp(\phi L^3) \exp(\theta L^2) \exp(\psi L^3) \dot{\psi}L^3.\end{aligned}\quad (11.1.128)$$

Now begins the fun. Verify that  $R^{-1}\dot{R}$  is given by the seemingly hopeless expression

$$\begin{aligned}R^{-1}\dot{R} = & \exp(-\psi L^3) \exp(-\theta L^2) \exp(-\phi L^3) \dot{\phi}L^3 \exp(\phi L^3) \exp(\theta L^2) \exp(\psi L^3) \\ & + \exp(-\psi L^3) \exp(-\theta L^2) \exp(-\phi L^3) \exp(\phi L^3) \dot{\theta}L^2 \exp(\theta L^2) \exp(\psi L^3) \\ & + \exp(-\psi L^3) \exp(-\theta L^2) \exp(-\phi L^3) \exp(\phi L^3) \exp(\theta L^2) \exp(\psi L^3) \dot{\psi}L^3.\end{aligned}\quad (11.1.129)$$

Simplify each of the three lines in (1.129) so that they become

$$\begin{aligned}\exp(-\psi L^3) \exp(-\theta L^2) \exp(-\phi L^3) \dot{\phi}L^3 \exp(\phi L^3) \exp(\theta L^2) \exp(\psi L^3) \\ = \dot{\phi} \exp(-\psi L^3) \exp(-\theta L^2) L^3 \exp(\theta L^2) \exp(\psi L^3),\end{aligned}\quad (11.1.130)$$

$$\begin{aligned}\exp(-\psi L^3) \exp(-\theta L^2) \exp(-\phi L^3) \exp(\phi L^3) \dot{\theta}L^2 \exp(\theta L^2) \exp(\psi L^3) \\ = \dot{\theta} \exp(-\psi L^3) L^2 \exp(\psi L^3),\end{aligned}\quad (11.1.131)$$

$$\begin{aligned}\exp(-\psi L^3) \exp(-\theta L^2) \exp(-\phi L^3) \exp(\phi L^3) \exp(\theta L^2) \exp(\psi L^3) \dot{\psi}L^3 \\ = \dot{\psi}L^3.\end{aligned}\quad (11.1.132)$$

Line (1.132) is as simple as we could desire. The next more complicated line is (1.131). Show, using the machinery of Exercise 8.2.10, that

$$\exp(-\psi L^3) L^2 \exp(\psi L^3) = \cos(\psi)L^2 + \sin(\psi)L^1. \quad (11.1.133)$$

Thus, the right side of (1.131) becomes

$$\dot{\theta}[\cos(\psi)L^2 + \sin(\psi)L^1]. \quad (11.1.134)$$

Finally, work on line (1.130). Show that

$$\exp(-\theta L^2) L^3 \exp(\theta L^2) = \cos(\theta)L^3 - \sin(\theta)L^1. \quad (11.1.135)$$

Next show that

$$\begin{aligned}\exp(-\psi L^3)[\cos(\theta)L^3 - \sin(\theta)L^1] \exp(\psi L^3) \\ = \cos(\theta)L^3 - \sin(\theta) \exp(-\psi L^3) L^1 \exp(\psi L^3) \\ = \cos(\theta)L^3 - \sin(\theta)[\cos(\psi)L^1 - \sin(\psi)L^2].\end{aligned}\quad (11.1.136)$$

By combining (1.135) and (1.136), verify that

$$\begin{aligned} & \exp(-\psi L^3) \exp(-\theta L^2) L^3 \exp(\theta L^2) \exp(\psi L^3) \\ &= \cos(\theta)L^3 - \sin(\theta)[\cos(\psi)L^1 - \sin(\psi)L^2]. \end{aligned} \quad (11.1.137)$$

Thus, the right side of (1.130) becomes

$$\dot{\phi}\{\cos(\theta)L^3 - \sin(\theta)[\cos(\psi)L^1 - \sin(\psi)L^2]\}. \quad (11.1.138)$$

All the necessary ingredients are at hand. By combining (1.129) through (1.138), show that

$$\begin{aligned} R^{-1}\dot{R} &= \dot{\phi}\{\cos(\theta)L^3 - \sin(\theta)[\cos(\psi)L^1 - \sin(\psi)L^2]\} \\ &+ \dot{\theta}[\cos(\psi)L^2 + \sin(\psi)L^1] \\ &+ \dot{\psi}L^3 \\ &= (-\dot{\phi}\sin\theta\cos\psi + \dot{\theta}\sin\psi)L^1 \\ &+ (\dot{\phi}\sin\theta\sin\psi + \dot{\theta}\cos\psi)L^2 \\ &+ (\dot{\phi}\cos\theta + \dot{\psi})L^3. \end{aligned} \quad (11.1.139)$$

Verify, upon equating coefficients of the  $L^j$  in (1.126) and (1.139), that the relations (1.23) through (1.25) follow. Finally, verify that inverting the relations (1.23) through (1.25) yields the relations (1.26) through (1.28).

**11.1.4.** The purpose of this exercise is to derive equations (1.31) and (1.32), the relation between quaternion and angle-axis parameters, and equations (1.33) through (1.36), the relation between quaternion and Euler-angle parameters. For this purpose, it is convenient to exploit the homomorphism between  $SO(3, \mathbb{R})$  and  $SU(2)$ . Review Exercises 3.7.30, 5.10.13, 8.2.10, and 8.2.11.

Suppose  $u \in SU(2)$  is parameterized in terms of angle-axis parameters by writing

$$u = \exp(\theta \mathbf{n} \cdot \mathbf{K}) = I \cos(\theta/2) + 2(\mathbf{n} \cdot \mathbf{K}) \sin(\theta/2). \quad (11.1.140)$$

See (3.7.186) and (3.7.188). Suppose that  $u$  is also parameterized in terms of unit quaternion matrices by writing

$$u(w) = w_0\sigma^0 + i\mathbf{w} \cdot \boldsymbol{\sigma}. \quad (11.1.141)$$

Equate (1.140) and (1.141) and employ the relations

$$I = \sigma^0, \quad (11.1.142)$$

$$\mathbf{K} = (-i/2)\boldsymbol{\sigma}. \quad (11.1.143)$$

Use the linear independence of the matrices  $\sigma^0$  through  $\sigma^3$  to verify (1.31) and (1.32).

Next suppose  $u \in SU(2)$  is parameterized in terms of Euler angles as in (3.7.195) in Exercise 3.7.30. Verify that there is the decomposition

$$\begin{aligned} u(\phi, \theta, \psi) &= \begin{pmatrix} \cos(\theta/2) \exp[-(i/2)(\phi + \psi)] & -\sin(\theta/2) \exp[(i/2)(-\phi + \psi)] \\ \sin(\theta/2) \exp[-(i/2)(-\phi + \psi)] & \cos(\theta/2) \exp[(i/2)(\phi + \psi)] \end{pmatrix} \\ &= \cos(\theta/2) \cos[(1/2)(\phi + \psi)] \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - i \sin(\theta/2) \sin[(1/2)(-\phi + \psi)] \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ &\quad - i \sin(\theta/2) \cos[(1/2)(-\phi + \psi)] \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} - i \cos(\theta/2) \sin[(1/2)(\phi + \psi)] \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned} \tag{11.1.144}$$

Thus, using the definition of the Pauli matrices, verify that

$$\begin{aligned} u &= \sigma^0 \{\cos(\theta/2) \cos[(1/2)(\phi + \psi)]\} \\ &\quad - i\sigma^1 \{\sin(\theta/2) \sin[(1/2)(-\phi + \psi)]\} \\ &\quad - i\sigma^2 \{\sin(\theta/2) \cos[(1/2)(-\phi + \psi)]\} \\ &\quad - i\sigma^3 \{\cos(\theta/2) \sin[(1/2)(\phi + \psi)]\}. \end{aligned} \tag{11.1.145}$$

Finally, verify that comparison of (1.141) and (1.145), and use of the linear independence of the matrices  $\sigma^0$  through  $\sigma^3$ , yields (1.33) through (1.36).

**11.1.5.** The purpose of this exercise is to derive equations (1.37) through (1.39), the expressions for the angular velocities  $\omega_j^{bf}$  in terms of quaternion parameters. Surprisingly, this task turns out to be computationally simpler than the Euler-angle case of Exercise 1.3. Review Exercise 1.4 above for notation.

The relation (1.22) specifies the  $\omega_j^{bf}$  in terms of  $SO(3, \mathbb{R})$  quantities. Based on the homomorphism between  $SO(3, \mathbb{R})$  and  $SU(2)$  (see Exercises 3.7.30, 5.10.13, 8.2.10, and 8.2.11), verify that there is an associated  $SU(2)$  relation given by the formula

$$u^{-1}\dot{u} = \omega_1^{bf} K^1 + \omega_2^{bf} K^2 + \omega_3^{bf} K^3 = \boldsymbol{\omega}^{bf} \cdot \mathbf{K}. \tag{11.1.146}$$

Indeed, suppose  $u$  is parameterized in terms of Euler angles by writing

$$u(t) = \exp[\phi(t)K^3] \exp[\theta(t)K^2] \exp[\psi(t)K^3]. \tag{11.1.147}$$

See (3.7.194). Show that then use of (1.147) in (1.146) reproduces the results of Exercise 1.3.

Instead, we will parameterize  $u$  in terms of unit quaternion matrices by writing (1.141). Then we have the results

$$u^{-1} = u^\dagger = w_0 \sigma^0 - i\mathbf{w} \cdot \boldsymbol{\sigma} \tag{11.1.148}$$

and

$$\dot{u}(w) = \dot{w}_0 \sigma^0 + i\dot{\mathbf{w}} \cdot \boldsymbol{\sigma}. \tag{11.1.149}$$

Verify that carrying out the required multiplication  $u^{-1}\dot{u}$  yields the intermediate result

$$\begin{aligned} u^{-1}\dot{u} &= [w_0\sigma^0 - i\mathbf{w} \cdot \boldsymbol{\sigma}][\dot{w}_0\sigma^0 + i\dot{\mathbf{w}} \cdot \boldsymbol{\sigma}] \\ &= w_0\dot{w}_0\sigma^0 + iw_0\dot{\mathbf{w}} \cdot \boldsymbol{\sigma} - i\dot{w}_0\mathbf{w} \cdot \boldsymbol{\sigma} + (\mathbf{w} \cdot \boldsymbol{\sigma})(\dot{\mathbf{w}} \cdot \boldsymbol{\sigma}). \end{aligned} \quad (11.1.150)$$

Now use (5.7.44) to write

$$(\mathbf{w} \cdot \boldsymbol{\sigma})(\dot{\mathbf{w}} \cdot \boldsymbol{\sigma}) = (\mathbf{w} \cdot \dot{\mathbf{w}})\sigma^0 + i(\mathbf{w} \times \dot{\mathbf{w}}) \cdot \boldsymbol{\sigma}. \quad (11.1.151)$$

Substitute (1.151) into (1.150) to yield the next intermediate result

$$u^{-1}\dot{u} = (w_0\dot{w}_0 + \mathbf{w} \cdot \dot{\mathbf{w}})\sigma^0 + i(w_0\dot{\mathbf{w}} - \dot{w}_0\mathbf{w} + \mathbf{w} \times \dot{\mathbf{w}}) \cdot \boldsymbol{\sigma}. \quad (11.1.152)$$

But in view of (1.40), which follows from the requirement that  $u$  be a unit quaternion, the first term on the right of (1.152) vanishes, and (1.152) therefore becomes

$$u^{-1}\dot{u} = i(w_0\dot{\mathbf{w}} - \dot{w}_0\mathbf{w} + \mathbf{w} \times \dot{\mathbf{w}}) \cdot \boldsymbol{\sigma}. \quad (11.1.153)$$

There is also the result (1.143), and therefore (1.153) can be written as

$$u^{-1}\dot{u} = -2(w_0\dot{\mathbf{w}} - \dot{w}_0\mathbf{w} + \mathbf{w} \times \dot{\mathbf{w}}) \cdot \mathbf{K}. \quad (11.1.154)$$

Upon comparing (1.146) and (1.154), show that

$$\omega_j^{bf} = -2(w_0\dot{\mathbf{w}} - \dot{w}_0\mathbf{w} + \mathbf{w} \times \dot{\mathbf{w}}) \cdot \mathbf{e}_j. \quad (11.1.155)$$

Verify that (1.37) through (1.39) are equivalent to (1.155).

**11.1.6.** The purpose of this exercise is to verify that the  $4 \times 4$  matrices appearing in (1.41) and (1.42) are orthogonal. Suppose  $(w_0, w_1, w_2, w_3)^T$  is a unit four vector. That is, suppose (1.30) is satisfied. Consider the mapping of this four vector into the space of  $4 \times 4$  matrices given by the rule

$$\begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{pmatrix} \rightarrow M(w) = \begin{pmatrix} w_0 & -w_1 & -w_2 & -w_3 \\ w_1 & w_0 & w_3 & -w_2 \\ w_2 & -w_3 & w_0 & w_1 \\ w_3 & w_2 & -w_1 & w_0 \end{pmatrix}. \quad (11.1.156)$$

Verify that all the columns of  $M$  are unit vectors. Verify that all the columns of  $M$  are mutually orthogonal. It follows that  $M$  is an orthogonal matrix,

$$M^T M = I. \quad (11.1.157)$$

Verify that the matrices appearing in (1.41) and (1.42) are  $-M^T$  and  $-M$ , and hence both are also orthogonal.

**11.1.7.** Intrigued by the remarkable mapping (1.156), this exercise is devoted to a further exploration of what is going on. We will learn that what is involved is a mapping of a pair of unit quaternions into  $SO(4, \mathbb{R})$ .

Show that since  $M$  is orthogonal, it must satisfy

$$\det M = \pm 1. \quad (11.1.158)$$

Let  $e_0$  be the unit vector

$$e_0 = (1, 0, 0, 0)^T. \quad (11.1.159)$$

Verify that

$$M(e_0) = I, \quad (11.1.160)$$

and therefore

$$\det M(e_0) = +1. \quad (11.1.161)$$

Show that any unit vector  $w \in S^3$  is connected to  $e_0$  by a continuous path in  $S^3$ . [Hint: Show, for example, that  $SO(4, \mathbb{R})$  acts transitively on  $S^3$ .] Verify that the mapping (1.156) is continuous. Show that it follows, by continuity, that

$$\det M(w) = +1, \quad (11.1.162)$$

and therefore  $M(w) \in SO(4, \mathbb{R})$ . That is, (1.156) produces a map

$$S^3 \rightarrow SO(4, \mathbb{R}). \quad (11.1.163)$$

Can all elements of  $SO(4, \mathbb{R})$  be written in the form  $M(w)$ ? No, because  $w \in S^3$  involves three parameters, and we know that  $SO(4, \mathbb{R})$  is six dimensional. What elements in  $SO(4, \mathbb{R})$  can be written in the form  $M(w)$ ? Let's see. Define three matrices  $D^1, D^2, D^3$  by the rules

$$D^1 = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \quad (11.1.164)$$

$$D^2 = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad (11.1.165)$$

$$D^3 = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \quad (11.1.166)$$

Verify the relation

$$M(w) = w_0 I + w_1 D^1 + w_2 D^2 + w_3 D^3. \quad (11.1.167)$$

Show that the matrices  $D^j$  satisfy the relations

$$(D^j)^2 = -I, \quad (11.1.168)$$

$$D^1 D^2 = -D^2 D^1 = -D^3, \quad (11.1.169)$$

$$D^2 D^3 = -D^3 D^2 = -D^1, \quad (11.1.170)$$

$$D^3 D^1 = -D^1 D^3 = -D^2. \quad (11.1.171)$$

By employing these relations, show that

$$M(w)M(w') = M(w'') \quad (11.1.172)$$

with

$$\begin{aligned} w''_0 &= w_0 w'_0 - w_1 w'_1 - w_2 w'_2 - w_3 w'_3, \\ w''_1 &= w_0 w'_1 + w'_0 w_1 - w_2 w'_3 + w_3 w'_2, \\ w''_2 &= w_0 w'_2 + w'_0 w_2 - w_3 w'_1 + w_1 w'_3, \\ w''_3 &= w_0 w'_3 + w'_0 w_3 - w_1 w'_2 + w_2 w'_1, \end{aligned} \quad (11.1.173)$$

which, with the notation  $\mathbf{w} = (w_1, w_2, w_3)$ , can be written more compactly in the form

$$w''_0 = w_0 w'_0 - \mathbf{w} \cdot \mathbf{w}', \quad (11.1.174)$$

$$\mathbf{w}'' = w_0 \mathbf{w}' + w'_0 \mathbf{w} - \mathbf{w} \times \mathbf{w}'. \quad (11.1.175)$$

Show that if  $w \in S^3$  and  $w' \in S^3$ , then  $w'' \in S^3$ . Show also that

$$M(w + w') = M(w) + M(w'). \quad (11.1.176)$$

Observe that the relations (1.174) and (1.175) are exactly those for quaternion matrix multiplication. See Section 5.10.4 and Exercise 5.10.15. Verify that there is the correspondence

$$\begin{aligned} I &\leftrightarrow e, \\ -D^1 &\leftrightarrow j, \\ -D^2 &\leftrightarrow k \\ -D^3 &\leftrightarrow \ell. \end{aligned} \quad (11.1.177)$$

That is, the  $4 \times 4$  matrices  $I, -D^1, -D^2, -D^3$  provide a representation for quaternion algebra. Observe that, unlike the  $2 \times 2$  representation given by the complex matrices in (5.10.64), these  $4 \times 4$  matrices are all real.

Verify that the  $D^j$  are antisymmetric,

$$(D^j)^T = -D^j. \quad (11.1.178)$$

Verify that

$$[M(w)]^T = M(w^*) \quad (11.1.179)$$

with the definitions

$$w_0^* = w_0, \quad w_1^* = -w_1, \quad w_2^* = -w_2, \quad w_3^* = -w_3. \quad (11.1.180)$$

Verify that

$$(w^*)^* = w. \quad (11.1.181)$$

Verify that  $w^* \in S^3$  if  $w \in S^3$ , and verify the relation

$$M(w^*) = [M(w)]^{-1}. \quad (11.1.182)$$

Thus, verify that the  $M(w)$  for  $w \in S^3$  form a group.

Finally, what elements in  $SO(4, \mathbb{R})$  can be written in the form  $M(w)$ ? We know from Exercise 5.10.13 that unit quaternions form a group that is isomorphic to  $SU(2)$ . We also know from Exercises 4.4.19 and 4.3.20 that  $so(4, \mathbb{R})$  is the direct sum of two commuting  $su(2)$  Lie algebras. From these same exercises we know that the  $H^j$  generate the  $SU(2)$  associated with one of these  $su(2)$  Lie algebras. Verify from (4.3.152) and (1.161) through (1.163) that there is the relation

$$D^j = -2H^j. \quad (11.1.183)$$

Show that it follows that the elements in  $SO(4, \mathbb{R})$  that can be written in the form  $M(w)$  are those that belong to the  $SU(2)$  subgroup generated by the  $H^j$ . Verify, in the terminology of Exercises 4.3.19 and 4.3.20, that these are all elements of the form  $\exp(\mathbf{t} \cdot \mathbf{H})$ . Find the relation between  $w \in S^3$  and  $\mathbf{t}$ . Lastly verify that, as expected from the work of Exercise 4.3.19, that all  $M(w)$  with  $w \in S^3$  are also symplectic with respect to the  $J$  of (4.3.65).

But wait, as they say in infomercials, there's more! As just stated, we know from Exercises 4.3.19 and 4.3.20 that  $SO(4, \mathbb{R})$  is the direct product of two commuting  $SU(2)$  subgroups, and we know from Exercise 5.10.13 that any  $SU(2)$  has an associated unit quaternion equivalent. Therefore, there should be *two* unit quaternions associated with  $SO(4, \mathbb{R})$ . How can the second unit quaternion be found/employed?

Motivated by (1.183), we might define matrices  $E^j$  by the rule

$$E^j \stackrel{?}{=} -2G^j \quad (11.1.184)$$

with the  $G^j$  given by (4.3.151). This is a workable possibility. However, a choice that yields more aesthetic results is to make the definitions

$$\begin{aligned} E^1 &= 2G^3, \\ E^2 &= -2G^2, \\ E^3 &= -2G^1. \end{aligned} \quad (11.1.185)$$

The  $E^j$  have the explicit form

$$E^1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \quad (11.1.186)$$

$$E^2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad (11.1.187)$$

$$E^3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}. \quad (11.1.188)$$

Form the matrix  $N(w)$  by the rule

$$N(w) = w_0 I + w_1 E^1 + w_2 E^2 + w_3 E^3. \quad (11.1.189)$$

Verify that  $N(w)$  has the explicit form

$$N(w) = \begin{pmatrix} w_0 & w_1 & w_2 & w_3 \\ -w_1 & w_0 & w_3 & -w_2 \\ -w_2 & -w_3 & w_0 & w_1 \\ -w_3 & w_2 & -w_1 & w_0 \end{pmatrix}. \quad (11.1.190)$$

Check that the columns of  $N(w)$  are mutually orthogonal and, if  $w \in S^3$ , are also unit vectors. It follows that  $N(w)$  is an orthogonal matrix if  $w \in S^3$ . Check that

$$N(e_0) = I, \quad (11.1.191)$$

and therefore show, as was done for  $M(w)$ , that  $N(w) \in SO(4, \mathbb{R})$ . Thus the correspondence

$$\begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{pmatrix} \rightarrow N(w) = \begin{pmatrix} w_0 & w_1 & w_2 & w_3 \\ -w_1 & w_0 & w_3 & -w_2 \\ -w_2 & -w_3 & w_0 & w_1 \\ -w_3 & w_2 & -w_1 & w_0 \end{pmatrix} \quad (11.1.192)$$

also provides a mapping

$$S^3 \rightarrow SO(4, \mathbb{R}). \quad (11.1.193)$$

Review the relations (1.168) through (1.182). Show that there are completely analogous relations for the  $E^j$  and  $N(w)$ . Among other things, the  $4 \times 4$  matrices  $I, -E^1, -E^2, -E^3$  also provide a representation for quaternion algebra. Also verify that  $M(w')$  and  $N(w)$  commute.

Finally, in analogy with (4.3.166) and (4.3.174), define a matrix  $O(w, w')$  by the rule

$$O(w, w') = N(w)M(w'). \quad (11.1.194)$$

Call the pair  $(w, w') = S^3 \times S^3$  a double three-sphere. The relation (1.194) provides a two-to-one mapping of the double three-sphere into  $SO(4, \mathbb{R})$ ,

$$S^3 \times S^3 \rightarrow SO(4, \mathbb{R}). \quad (11.1.195)$$

**11.1.8.** The purpose of this exercise is to determine the behavior of  $w \cdot w$  when  $w$  evolves as in equations (1.42) or in equations (1.48) through (1.51), and to explore the nature of these equations. Let  $\Omega$  be the vector

$$\Omega = \begin{pmatrix} 0 \\ \omega_1^{bf}/2 \\ \omega_2^{bf}/2 \\ \omega_3^{bf}/2 \end{pmatrix}. \quad (11.1.196)$$

Verify that the equations (1.48) through (1.51) can be written in the compact form

$$\dot{w} = -M(w)\Omega + k\epsilon w \quad (11.1.197)$$

with  $M(w)$  defined by (1.156). It follows that

$$(w, \dot{w}) = -(w, M(w)\Omega) + k\epsilon(w, w). \quad (11.1.198)$$

Verify that

$$M^T(w)w = e_0 \quad (11.1.199)$$

with  $e_0$  given by (1.159). Next verify that

$$(w, M(w)\Omega) = (M^T(w)w, \Omega) = (e_0, \Omega) = 0. \quad (11.1.200)$$

It follows that

$$(1/2)(d/dt)(w, w) = (w, \dot{w}) = k\epsilon(w, w), \quad (11.1.201)$$

and therefore  $w \cdot w$  is conserved if  $k = 0$ , which is the case for the equations of motion (1.42). In particular, if  $w \cdot w = 1$  initially, it will remain so.

What about the evolution of  $w \cdot w$  when  $k \neq 0$ ? Verify that

$$\dot{\epsilon} = -2(w, \dot{w}). \quad (11.1.202)$$

Show that, consequently, (1.201) and (1.202) together yield the relation

$$\dot{\epsilon} = -2k\epsilon(1 - \epsilon). \quad (11.1.203)$$

Verify that (1.203) has the implicit solution

$$\epsilon(t)/[1 - \epsilon(t)] = [\epsilon_0/(1 - \epsilon_0)] \exp[-2k(t - t_0)] \quad (11.1.204)$$

where

$$\epsilon_0 = \epsilon(t_0). \quad (11.1.205)$$

To solve (1.204) explicitly for  $\epsilon(t)$ , let

$$r(t) = [\epsilon_0/(1 - \epsilon_0)] \exp[-2k(t - t_0)]. \quad (11.1.206)$$

Show that

$$\epsilon(t) = r(t)/[1 + r(t)]. \quad (11.1.207)$$

Evidently, for  $k > 0$ ,  $\epsilon(t) \rightarrow 0$  as  $t \rightarrow +\infty$ , essentially exponentially. Correspondingly,  $w \cdot w \rightarrow 1$  as  $t \rightarrow +\infty$ .

There is an alternate way of looking at the equations (1.43) through (1.46) or (1.47) through (1.51) that emphasizes linearity in  $w$ . Let  $A(\omega^{bf})$  be the matrix defined by the rule

$$A(\omega^{bf}) = (1/2) \begin{pmatrix} 0 & \omega_1^{bf} & \omega_2^{bf} & \omega_3^{bf} \\ -\omega_1^{bf} & 0 & \omega_3^{bf} & -\omega_2^{bf} \\ -\omega_2^{bf} & -\omega_3^{bf} & 0 & \omega_1^{bf} \\ -\omega_3^{bf} & \omega_2^{bf} & -\omega_1^{bf} & 0 \end{pmatrix}. \quad (11.1.208)$$

Show that (1.43) through (1.46) can be written in the vector/matrix form

$$\dot{w} = A(\boldsymbol{\omega}^{bf})w. \quad (11.1.209)$$

Verify that  $A$  is antisymmetric, and therefore it immediately follows from (1.209) that

$$(w, \dot{w}) = (w, Aw) = 0. \quad (11.1.210)$$

Similarly, show that (1.48) through (1.51) can be written in the form

$$\dot{w} = A(\boldsymbol{\omega}^{bf})w + k\epsilon w. \quad (11.1.211)$$

In this case, it follows directly from (1.211) and the antisymmetry of  $A$  that

$$(w, \dot{w}) = (w, Aw) + k\epsilon(w, w) = k\epsilon(w, w), \quad (11.1.212)$$

as before.

**11.1.9.** The aim of the this exercise is to find the  $\omega_j^{bf}$  corresponding to the Tait-Bryan angle parameterization (1.58). The tools for this purpose will be very similar to those employed in Exercise 1.3, which you should review. Our discussion begins with the key relation

$$(R^v)^{-1}\dot{R}^v = \boldsymbol{\omega}^{bf} \cdot \mathbf{L}. \quad (11.1.213)$$

Show that this relation follows from (1.56).

For  $R^v$  given by (1.58), verify that

$$(R^v)^{-1} = \exp(-\lambda_3 L^3) \exp(-\lambda_2 L^2) \exp(-\lambda_1 L^1) \quad (11.1.214)$$

and

$$\begin{aligned} \dot{R}^v &= \dot{\lambda}_1 L^1 \exp(\lambda_1 L^1) \exp(\lambda_2 L^2) \exp(\lambda_3 L^3) \\ &+ \exp(\lambda_1 L^1) \dot{\lambda}_2 L^2 \exp(\lambda_2 L^2) \exp(\lambda_3 L^3) \\ &+ \exp(\lambda_1 L^1) \exp(\lambda_2 L^2) \exp(\lambda_3 L^3) \dot{\lambda}_3 L^3. \end{aligned} \quad (11.1.215)$$

Next verify that  $(R^v)^{-1}\dot{R}^v$  is given by the expression

$$\begin{aligned} (R^v)^{-1}\dot{R}^v &= \exp(-\lambda_3 L^3) \exp(-\lambda_2 L^2) \exp(-\lambda_1 L^1) \dot{\lambda}_1 L^1 \exp(\lambda_1 L^1) \exp(\lambda_2 L^2) \exp(\lambda_3 L^3) \\ &+ \exp(-\lambda_3 L^3) \exp(-\lambda_2 L^2) \exp(-\lambda_1 L^1) \exp(\lambda_1 L^1) \dot{\lambda}_2 L^2 \exp(\lambda_2 L^2) \exp(\lambda_3 L^3) \\ &+ \exp(-\lambda_3 L^3) \exp(-\lambda_2 L^2) \exp(-\lambda_1 L^1) \exp(\lambda_1 L^1) \exp(\lambda_2 L^2) \exp(\lambda_3 L^3) \dot{\lambda}_3 L^3. \end{aligned} \quad (11.1.216)$$

Simplify each of the three lines in (1.216) so that they become

$$\begin{aligned} &\exp(-\lambda_3 L^3) \exp(-\lambda_2 L^2) \exp(-\lambda_1 L^1) \dot{\lambda}_1 L^1 \exp(\lambda_1 L^1) \exp(\lambda_2 L^2) \exp(\lambda_3 L^3) \\ &= \dot{\lambda}_1 \exp(-\lambda_3 L^3) \exp(-\lambda_2 L^2) L^1 \exp(\lambda_2 L^2) \exp(\lambda_3 L^3), \end{aligned} \quad (11.1.217)$$

$$\begin{aligned} \exp(-\lambda_3 L^3) \exp(-\lambda_2 L^2) \exp(-\lambda_1 L^1) \exp(\lambda_1 L^1) \dot{\lambda}_2 L^2 \exp(\lambda_2 L^2) \exp(\lambda_3 L^3) \\ = \dot{\lambda}_2 \exp(-\lambda_3 L^3) L^2 \exp(\lambda_3 L^3), \end{aligned} \quad (11.1.218)$$

$$\begin{aligned} \exp(-\lambda_3 L^3) \exp(-\lambda_2 L^2) \exp(-\lambda_1 L^1) \exp(\lambda_1 L^1) \exp(\lambda_2 L^2) \exp(\lambda_3 L^3) \dot{\lambda}_3 L^3 \\ = \dot{\lambda}_3 L^3. \end{aligned} \quad (11.1.219)$$

Line (1.219) is as simple as we could desire. The next more complicated line is (1.218). Show, using the machinery of Exercise 8.2.10, that

$$\exp(-\lambda_3 L^3) L^2 \exp(\lambda_3 L^3) = \cos(\lambda_3) L^2 + \sin(\lambda_3) L^1. \quad (11.1.220)$$

Thus, the right side of (1.218) becomes

$$\dot{\lambda}_2 [\cos(\lambda_3) L^2 + \sin(\lambda_3) L^1]. \quad (11.1.221)$$

Finally, work on line (1.217). Show that

$$\exp(-\lambda_2 L^2) L^1 \exp(\lambda_2 L^2) = \cos(\lambda_2) L^1 + \sin(\lambda_2) L^3. \quad (11.1.222)$$

Next show that

$$\begin{aligned} & \exp(-\lambda_3 L^3) [\cos(\lambda_2) L^1 + \sin(\lambda_2) L^3] \exp(\lambda_3 L^3) \\ &= \sin(\lambda_2) L^3 + \cos(\lambda_2) \exp(-\lambda_3 L^3) L^1 \exp(\lambda_3 L^3) \\ &= \sin(\lambda_2) L^3 + \cos(\lambda_2) [\cos(\lambda_3) L^1 - \sin(\lambda_3) L^2]. \end{aligned} \quad (11.1.223)$$

By combining (1.222) and (1.223), verify that

$$\begin{aligned} & \exp(-\lambda_3 L^3) \exp(-\lambda_2 L^2) L^1 \exp(\lambda_2 L^2) \exp(\lambda_3 L^3) \\ &= \sin(\lambda_2) L^3 + \cos(\lambda_2) [\cos(\lambda_3) L^1 - \sin(\lambda_3) L^2]. \end{aligned} \quad (11.1.224)$$

Thus, the right side of (1.217) becomes

$$\dot{\lambda}_1 \{ \sin(\lambda_2) L^3 + \cos(\lambda_2) [\cos(\lambda_3) L^1 - \sin(\lambda_3) L^2] \}. \quad (11.1.225)$$

Now we have everything we need. By combining (1.216) through (1.225), show that

$$\begin{aligned} (R^v)^{-1} \dot{R}^v &= \dot{\lambda}_1 \{ \sin(\lambda_2) L^3 + \cos(\lambda_2) [\cos(\lambda_3) L^1 - \sin(\lambda_3) L^2] \} \\ &+ \dot{\lambda}_2 [\cos(\lambda_3) L^2 + \sin(\lambda_3) L^1] \\ &+ \dot{\lambda}_3 L^3 \\ &= [\dot{\lambda}_1 \cos(\lambda_2) \cos(\lambda_3) + \dot{\lambda}_2 \sin(\lambda_3)] L^1 \\ &+ [-\dot{\lambda}_1 \cos(\lambda_2) \sin(\lambda_3) + \dot{\lambda}_2 \cos(\lambda_3)] L^2 \\ &+ [\dot{\lambda}_1 \sin(\lambda_2) + \dot{\lambda}_3] L^3. \end{aligned} \quad (11.1.226)$$

Verify, upon equating coefficients of the  $L^j$  in (1.213) and (1.226), that there are the relations

$$\omega_1^{bf} = \dot{\lambda}_1 \cos(\lambda_2) \cos(\lambda_3) + \dot{\lambda}_2 \sin(\lambda_3), \quad (11.1.227)$$

$$\omega_2^{bf} = -\dot{\lambda}_1 \cos(\lambda_2) \sin(\lambda_3) + \dot{\lambda}_2 \cos(\lambda_3), \quad (11.1.228)$$

$$\omega_3^{bf} = \dot{\lambda}_1 \sin(\lambda_2) + \dot{\lambda}_3. \quad (11.1.229)$$

Finally, verify that inverting the relations (1.227) through (1.229) yields the equations of motion

$$\dot{\lambda}_1 = [1/\cos(\lambda_2)][\omega_1^{bf} \cos(\lambda_3) - \omega_2^{bf} \sin(\lambda_3)], \quad (11.1.230)$$

$$\dot{\lambda}_2 = \omega_1^{bf} \sin(\lambda_3) + \omega_2^{bf} \cos(\lambda_3), \quad (11.1.231)$$

$$\dot{\lambda}_3 = \omega_3^{bf} - \tan(\lambda_2)[\omega_1^{bf} \cos(\lambda_3) - \omega_2^{bf} \sin(\lambda_3)]. \quad (11.1.232)$$

Verify that these equations of motion are nonsingular for small  $\lambda_j$ . Note, however, that there are singularities when  $\lambda_2 = \pm\pi/2$ . What causes these singularities? Show that

$$\begin{aligned} & \exp(\lambda_1 L^1) \exp[(\pi/2)L^2] \exp(\lambda_3 L^3) \\ &= \exp(\lambda_1 L^1) \exp[(\pi/2)L^2] \exp(\lambda_3 L^3) \exp[-(\pi/2)L^2] \exp[(\pi/2)L^2] \\ &= \exp(\lambda_1 L^1) \exp(\lambda_3 L^1) \exp[(\pi/2)L^2] \\ &= \exp[(\lambda_1 + \lambda_3)L^1] \exp[(\pi/2)L^2], \end{aligned} \quad (11.1.233)$$

$$\begin{aligned} & \exp(\lambda_1 L^1) \exp[-(\pi/2)L^2] \exp(\lambda_3 L^3) \\ &= \exp(\lambda_1 L^1) \exp[-(\pi/2)L^2] \exp(\lambda_3 L^3) \exp[(\pi/2)L^2] \exp[-(\pi/2)L^2] \\ &= \exp(\lambda_1 L^1) \exp(-\lambda_3 L^1) \exp[-(\pi/2)L^2] \\ &= \exp[(\lambda_1 - \lambda_3)L^1] \exp[-(\pi/2)L^2]. \end{aligned} \quad (11.1.234)$$

Verify that only the combination  $(\lambda_1 + \lambda_3)$  is well defined when  $\lambda_2 = \pi/2$ , and only the combination  $(\lambda_1 - \lambda_3)$  is well defined when  $\lambda_2 = -\pi/2$ . Therefore the parameterization (1.58) fails when  $\lambda_2 = \pm\pi/2$ .

**11.1.10.** The aim of this exercise, which is not easy, is to find the  $\dot{\lambda}_j$  in terms of the  $\omega_j^{bf}$  for the angle-axis parameterization (1.66). Although not essential, it is convenient for this purpose to use the  $SU(2)$  version (1.146) of the  $SO(3, \mathbb{R})$  relation (1.22). In analogy to (1.66), begin by writing

$$u(\boldsymbol{\lambda}) = \exp(\boldsymbol{\lambda} \cdot \mathbf{K}). \quad (11.1.235)$$

From the rule for differentiating the exponential function there is the result

$$(d/dt)u = u \operatorname{iex}(-\#\boldsymbol{\lambda} \cdot \mathbf{K}\#)(d/dt)(\boldsymbol{\lambda} \cdot \mathbf{K}), \quad (11.1.236)$$

which can be rewritten in the form

$$\dot{u} = u \operatorname{iex}(-\#\boldsymbol{\lambda} \cdot \mathbf{K}\#)(\dot{\boldsymbol{\lambda}} \cdot \mathbf{K}). \quad (11.1.237)$$

See Appendix C. Combine (1.146) and (1.237) to show that

$$\boldsymbol{\omega}^{bf} \cdot \mathbf{K} = \operatorname{iex}(-\#\boldsymbol{\lambda} \cdot \mathbf{K}\#)(\dot{\boldsymbol{\lambda}} \cdot \mathbf{K}). \quad (11.1.238)$$

Now for some daring steps. Show that, as in Section 10.3, the relation (1.238) can be inverted to become

$$\begin{aligned}\dot{\lambda} \cdot K &= [\text{iex}(-\#\lambda \cdot K\#)]^{-1}(\omega^{bf} \cdot K) \\ &= [I + (1/2)\#\lambda \cdot K\# + (1/12)(\#\lambda \cdot K\#)^2 + \dots](\omega^{bf} \cdot K) \\ &= (\omega^{bf} \cdot K) + (1/2)\{(\lambda \cdot K), (\omega^{bf} \cdot K)\} \\ &\quad + (1/12)\{(\lambda \cdot K), \{(\lambda \cdot K), (\omega^{bf} \cdot K)\}\} + \dots\end{aligned}\quad (11.1.239)$$

The result (1.239) is quite general in the sense that analogous results hold for any Lie group. For the specific case of  $SU(2)$  we will seek to explicitly sum the series (1.239).

From Appendix C we know that there is the expansion

$$[\text{iex}(-\#\lambda \cdot K\#)]^{-1} = \sum_{m=0}^{\infty} b_m (\#\lambda \cdot K\#)^m. \quad (11.1.240)$$

Verify that insertion of (1.240) into (1.239) gives the result

$$\begin{aligned}\dot{\lambda} \cdot K &= \left[ \sum_{m=0}^{\infty} b_m (\#\lambda \cdot K\#)^m \right] (\omega^{bf} \cdot K) \\ &= b_0(\omega^{bf} \cdot K) + b_1(\#\lambda \cdot K\#)(\omega^{bf} \cdot K) \\ &\quad + b_2(\#\lambda \cdot K\#)^2(\omega^{bf} \cdot K) + b_3(\#\lambda \cdot K\#)^3(\omega^{bf} \cdot K) \\ &\quad + b_4(\#\lambda \cdot K\#)^4(\omega^{bf} \cdot K) + b_5(\#\lambda \cdot K\#)^5(\omega^{bf} \cdot K) + \dots\end{aligned}\quad (11.1.241)$$

Now evaluate the terms appearing in (1.241). Show that there are the results

$$(\#\lambda \cdot K\#)(\omega^{bf} \cdot K) = \{\lambda \cdot K, \omega^{bf} \cdot K\} = (\lambda \times \omega^{bf}) \cdot K, \quad (11.1.242)$$

$$\begin{aligned}(\#\lambda \cdot K\#)^2(\omega^{bf} \cdot K) &= \{\lambda \cdot K, (\lambda \times \omega^{bf}) \cdot K\} = [\lambda \times (\lambda \times \omega^{bf})] \cdot K \\ &= (\lambda \cdot \omega^{bf})(\lambda \cdot K) - (\lambda \cdot \lambda)(\omega^{bf} \cdot K),\end{aligned}\quad (11.1.243)$$

$$\begin{aligned}(\#\lambda \cdot K\#)^3(\omega^{bf} \cdot K) &= (\#\lambda \cdot K\#)(\#\lambda \cdot K\#)^2(\omega^{bf} \cdot K) \\ &= -(\lambda \cdot \lambda)\{\lambda \cdot K, \omega^{bf} \cdot K\} \\ &= -(\lambda \cdot \lambda)(\lambda \times \omega^{bf}) \cdot K \\ &= (i\lambda \cdot i\lambda)(\#\lambda \cdot K\#)(\omega^{bf} \cdot K),\end{aligned}\quad (11.1.244)$$

$$\begin{aligned}(\#\lambda \cdot K\#)^4(\omega^{bf} \cdot K) &= (\#\lambda \cdot K\#)(\#\lambda \cdot K\#)^3(\omega^{bf} \cdot K) \\ &= (\#\lambda \cdot K\#)(-1)(\lambda \cdot \lambda)(\#\lambda \cdot K\#)(\omega^{bf} \cdot K) \\ &= (i\lambda \cdot i\lambda)(\#\lambda \cdot K\#)^2(\omega^{bf} \cdot K), \text{ etc.}\end{aligned}\quad (11.1.245)$$

Note the similarity of (3.7.201) and (1.244).

Verify that putting everything together gives the net result

$$\begin{aligned}
& \left[ \sum_{m=0}^{\infty} b_m (\#\boldsymbol{\lambda} \cdot \mathbf{K}\#)^m \right] (\boldsymbol{\omega}^{bf} \cdot \mathbf{K}) = b_0 (\boldsymbol{\omega}^{bf} \cdot \mathbf{K}) \\
& + b_1 (\#\boldsymbol{\lambda} \cdot \mathbf{K}\#) (\boldsymbol{\omega}^{bf} \cdot \mathbf{K}) + b_3 (\#\boldsymbol{\lambda} \cdot \mathbf{K}\#)^3 (\boldsymbol{\omega}^{bf} \cdot \mathbf{K}) + \dots \\
& + b_2 (\#\boldsymbol{\lambda} \cdot \mathbf{K}\#)^2 (\boldsymbol{\omega}^{bf} \cdot \mathbf{K}) + b_4 (\#\boldsymbol{\lambda} \cdot \mathbf{K}\#)^4 (\boldsymbol{\omega}^{bf} \cdot \mathbf{K}) + \dots \\
& \quad = b_0 (\boldsymbol{\omega}^{bf} \cdot \mathbf{K}) \\
& + [(\#\boldsymbol{\lambda} \cdot \mathbf{K}\#)(\boldsymbol{\omega}^{bf} \cdot \mathbf{K})][b_1 + b_3(i\boldsymbol{\lambda} \cdot i\boldsymbol{\lambda}) + b_5(i\boldsymbol{\lambda} \cdot i\boldsymbol{\lambda})^2 \dots] \\
& + [(\#\boldsymbol{\lambda} \cdot \mathbf{K}\#)^2(\boldsymbol{\omega}^{bf} \cdot \mathbf{K})][b_2 + b_4(i\boldsymbol{\lambda} \cdot i\boldsymbol{\lambda}) + b_6(i\boldsymbol{\lambda} \cdot i\boldsymbol{\lambda})^2 \dots] \\
& \quad = b_0 (\boldsymbol{\omega}^{bf} \cdot \mathbf{K}) \\
& + [(\boldsymbol{\lambda} \times \boldsymbol{\omega}^{bf}) \cdot \mathbf{K}][b_1 + b_3(i\boldsymbol{\lambda} \cdot i\boldsymbol{\lambda}) + b_5(i\boldsymbol{\lambda} \cdot i\boldsymbol{\lambda})^2 \dots] \\
& + [(\boldsymbol{\lambda} \cdot \boldsymbol{\omega}^{bf})(\boldsymbol{\lambda} \cdot \mathbf{K}) - (\boldsymbol{\lambda} \cdot \boldsymbol{\lambda})(\boldsymbol{\omega}^{bf} \cdot \mathbf{K})][b_2 + b_4(i\boldsymbol{\lambda} \cdot i\boldsymbol{\lambda}) + b_6(i\boldsymbol{\lambda} \cdot i\boldsymbol{\lambda})^2 \dots]. \tag{11.1.246}
\end{aligned}$$

Also verify that combining (1.241) and the last result in (1.246) gives the relation

$$\begin{aligned}
& \dot{\boldsymbol{\lambda}} = b_0 \boldsymbol{\omega}^{bf} \\
& + (\boldsymbol{\lambda} \times \boldsymbol{\omega}^{bf})[b_1 + b_3(i\boldsymbol{\lambda} \cdot i\boldsymbol{\lambda}) + b_5(i\boldsymbol{\lambda} \cdot i\boldsymbol{\lambda})^2 \dots] \\
& + [(\boldsymbol{\lambda} \cdot \boldsymbol{\omega}^{bf})\boldsymbol{\lambda} - (\boldsymbol{\lambda} \cdot \boldsymbol{\lambda})\boldsymbol{\omega}^{bf}][b_2 + b_4(i\boldsymbol{\lambda} \cdot i\boldsymbol{\lambda}) + b_6(i\boldsymbol{\lambda} \cdot i\boldsymbol{\lambda})^2 \dots]. \tag{11.1.247}
\end{aligned}$$

What remains to be done is to sum the series in (1.247). Begin by defining a quantity  $w$  by the rule

$$w = \sqrt{(i\boldsymbol{\lambda} \cdot i\boldsymbol{\lambda})} = i|\boldsymbol{\lambda}|. \tag{11.1.248}$$

With this definition we have the relations

$$[b_1 + b_3(i\boldsymbol{\lambda} \cdot i\boldsymbol{\lambda}) + b_5(i\boldsymbol{\lambda} \cdot i\boldsymbol{\lambda})^2 \dots] = (1/w) \sum_{\text{odd } m} b_m w^m, \tag{11.1.249}$$

$$[b_2 + b_4(i\boldsymbol{\lambda} \cdot i\boldsymbol{\lambda}) + b_6(i\boldsymbol{\lambda} \cdot i\boldsymbol{\lambda})^2 \dots] = (1/w^2) \sum_{\text{even } m > 0} b_m w^m. \tag{11.1.250}$$

We also know, see Appendix C, that

$$\sum_{m=0}^{\infty} b_m w^m = w/[1 - \exp(-w)], \tag{11.1.251}$$

from which it follows that

$$\sum_{m>0} b_m w^m = w/[1 - \exp(-w)] - 1. \tag{11.1.252}$$

(Here we have used the result  $b_0 = 1$ , which you should check.) Verify the identity

$$\begin{aligned}
w/[1 - \exp(-w)] &= w \exp(w/2)/[\exp(w/2) - \exp(-w/2)] \\
&= (w/2)[\cosh(w/2) + \sinh(w/2)]/\sinh(w/2) \\
&= w/2 + (w/2)\coth(w/2). \tag{11.1.253}
\end{aligned}$$

Use this identity to show that

$$\sum_{m>0} b_m w^m = [w/2] + [(w/2) \coth(w/2) - 1]. \quad (11.1.254)$$

From (1.254), by equating odd and even parts, show that

$$\sum_{\text{odd } m} b_m w^m = w/2, \quad (11.1.255)$$

$$\sum_{\text{even } m>0} b_m w^m = (w/2) \coth(w/2) - 1. \quad (11.1.256)$$

Use these results to show that

$$(1/w) \sum_{\text{odd } m} b_m w^m = 1/2, \quad (11.1.257)$$

$$(1/w^2) \sum_{\text{even } m>0} b_m w^m = [1/(2w)] \coth(w/2) - 1/w^2. \quad (11.1.258)$$

Verify that employing (1.248) in (1.258) yields the result

$$(1/w^2) \sum_{\text{even } m>0} b_m w^m = 1/|\boldsymbol{\lambda}|^2 - [1/(2|\boldsymbol{\lambda}|)] \cot(|\boldsymbol{\lambda}|/2). \quad (11.1.259)$$

At last, verify the final (and amazing) result

$$\dot{\boldsymbol{\lambda}} = \boldsymbol{\omega}^{bf} + (1/2)(\boldsymbol{\lambda} \times \boldsymbol{\omega}^{bf}) + [(\boldsymbol{\lambda} \cdot \boldsymbol{\omega}^{bf})\boldsymbol{\lambda} - (\boldsymbol{\lambda} \cdot \boldsymbol{\lambda})\boldsymbol{\omega}^{bf}] \{1/|\boldsymbol{\lambda}|^2 - [1/(2|\boldsymbol{\lambda}|)] \cot(|\boldsymbol{\lambda}|/2)\}. \quad (11.1.260)$$

Check that wherever  $|\boldsymbol{\lambda}|$  appears in (1.260), it appears as an *even* power. Therefore, there is no overall ambiguity in (1.260) despite the sign ambiguity present in the definition (1.248). Show, moreover, that the right side of (1.260) is *analytic* in the components of  $\boldsymbol{\lambda}$  for  $|\boldsymbol{\lambda}| < 2\pi$ , but is singular when  $|\boldsymbol{\lambda}| = 2\pi$ . As stated earlier, we expect this singularity to occur because we see from (3.7.188) and (3.7.202) that the individual components of  $\mathbf{n}$  are not defined in terms of  $v$  or  $R$  when  $\theta = |\boldsymbol{\lambda}| = 2\pi$ .

There is another way of writing (1.260) that is of interest. Define a function  $f(\boldsymbol{\lambda})$  by the rule

$$f(\boldsymbol{\lambda}) = 1/|\boldsymbol{\lambda}|^2 - [1/(2|\boldsymbol{\lambda}|)] \cot(|\boldsymbol{\lambda}|/2). \quad (11.1.261)$$

Verify that  $f(\boldsymbol{\lambda})$  is even in  $|\boldsymbol{\lambda}|$ , is analytic in the components of  $\boldsymbol{\lambda}$  for  $|\boldsymbol{\lambda}| < 2\pi$ , and is singular when  $|\boldsymbol{\lambda}| = 2\pi$ . Define a  $3 \times 3$  matrix  $M(\boldsymbol{\lambda})$  by the rule

$$M(\boldsymbol{\lambda}) = I + (1/2)(\boldsymbol{\lambda} \cdot \mathbf{L}) + f(\boldsymbol{\lambda})(\boldsymbol{\lambda} \cdot \mathbf{L})^2. \quad (11.1.262)$$

Verify that (1.260) can also be written in the form

$$\dot{\boldsymbol{\lambda}} = M(\boldsymbol{\lambda})\boldsymbol{\omega}^{bf}, \quad (11.1.263)$$

which highlights linearity in  $\boldsymbol{\omega}^{bf}$  and the role of the matrix  $\boldsymbol{\lambda} \cdot \mathbf{L}$ .

**11.1.11.** Review Exercise 1.10. It found the  $\dot{\lambda}_j$  in terms of the  $\omega_j^{bf}$  for the parameterization (1.66). For some purposes it is useful to also find the  $\omega_j^{bf}$  in terms of the  $\dot{\lambda}_j$ . That is the aim of this exercise.

There are at least two ways to proceed. The first begins with the relation (1.238) and makes an expansion of the form

$$\text{lex}(-\#\boldsymbol{\lambda} \cdot \mathbf{K}\#) = \sum_{m=0}^{\infty} d_m (\#\boldsymbol{\lambda} \cdot \mathbf{K}\#)^m. \quad (11.1.264)$$

This expansion is then manipulated, in the spirit of Exercise 1.10, to find and sum expansions that specify the  $\omega_j^{bf}$  in terms of the  $\dot{\lambda}_j$ .

A second way exploits more of what we already know from Exercise 1.10. Suppose we could invert the matrix  $M(\boldsymbol{\lambda})$  given in (1.262). Then we could rewrite (1.263) in the form

$$\boldsymbol{\omega}^{bf} = [M(\boldsymbol{\lambda})]^{-1} \dot{\boldsymbol{\lambda}}, \quad (11.1.265)$$

and we would have found the  $\omega_j^{bf}$  in terms of the  $\dot{\lambda}_j$ .

We now proceed to construct  $[M(\boldsymbol{\lambda})]^{-1}$ . Let  $N(\boldsymbol{\lambda})$  be a  $3 \times 3$  matrix of the form

$$N(\boldsymbol{\lambda}) = I + a(\boldsymbol{\lambda})(\boldsymbol{\lambda} \cdot \mathbf{L}) + b(\boldsymbol{\lambda})(\boldsymbol{\lambda} \cdot \mathbf{L})^2 \quad (11.1.266)$$

where  $a(\boldsymbol{\lambda})$  and  $b(\boldsymbol{\lambda})$  are coefficients to be determined. Next form the product of  $M(\boldsymbol{\lambda})$  and  $N(\boldsymbol{\lambda})$ . Verify that so doing yields the result

$$\begin{aligned} MN &= I + (1/2)(\boldsymbol{\lambda} \cdot \mathbf{L}) + f(\boldsymbol{\lambda} \cdot \mathbf{L})^2 \\ &+ a(\boldsymbol{\lambda} \cdot \mathbf{L}) + (1/2)a(\boldsymbol{\lambda} \cdot \mathbf{L})^2 + af(\boldsymbol{\lambda} \cdot \mathbf{L})^3 \\ &+ b(\boldsymbol{\lambda} \cdot \mathbf{L})^2 + (1/2)b(\boldsymbol{\lambda} \cdot \mathbf{L})^3 + bf(\boldsymbol{\lambda} \cdot \mathbf{L})^4. \end{aligned} \quad (11.1.267)$$

Verify the property

$$(\boldsymbol{\lambda} \cdot \mathbf{L})^3 = -|\boldsymbol{\lambda}|^2 \boldsymbol{\lambda} \cdot \mathbf{L}, \quad (11.1.268)$$

and employ it in (1.267) to achieve the net result

$$\begin{aligned} MN &= I + (1/2)(\boldsymbol{\lambda} \cdot \mathbf{L}) + f(\boldsymbol{\lambda} \cdot \mathbf{L})^2 \\ &+ a(\boldsymbol{\lambda} \cdot \mathbf{L}) + (1/2)a(\boldsymbol{\lambda} \cdot \mathbf{L})^2 - |\boldsymbol{\lambda}|^2 af(\boldsymbol{\lambda} \cdot \mathbf{L}) \\ &+ b(\boldsymbol{\lambda} \cdot \mathbf{L})^2 - |\boldsymbol{\lambda}|^2 (1/2)b(\boldsymbol{\lambda} \cdot \mathbf{L}) - |\boldsymbol{\lambda}|^2 bf(\boldsymbol{\lambda} \cdot \mathbf{L})^2 \\ &= I + [(1/2) + a - |\boldsymbol{\lambda}|^2 af - |\boldsymbol{\lambda}|^2 (1/2)b](\boldsymbol{\lambda} \cdot \mathbf{L}) \\ &+ [f + (1/2)a + b - |\boldsymbol{\lambda}|^2 bf](\boldsymbol{\lambda} \cdot \mathbf{L})^2. \end{aligned} \quad (11.1.269)$$

Suppose we can arrange that the coefficients of  $(\boldsymbol{\lambda} \cdot \mathbf{L})$  and  $(\boldsymbol{\lambda} \cdot \mathbf{L})^2$  in the net result (1.269) vanish. So doing requires the conditions

$$(1/2) + a - |\boldsymbol{\lambda}|^2 af - |\boldsymbol{\lambda}|^2 (1/2)b = 0 \quad (11.1.270)$$

and

$$f + (1/2)a + b - |\boldsymbol{\lambda}|^2bf = 0. \quad (11.1.271)$$

If these conditions are met, it follows that  $MN = I$ , and therefore

$$N(\boldsymbol{\lambda}) = [M(\boldsymbol{\lambda})]^{-1}. \quad (11.1.272)$$

Verify that the linear equations (1.270) and (1.271) can be written in the standard form

$$(1 - |\boldsymbol{\lambda}|^2 f)a - (|\boldsymbol{\lambda}|^2/2)b = -1/2, \quad (11.1.273)$$

$$(1/2)a + (1 - |\boldsymbol{\lambda}|^2 f)b = -f. \quad (11.1.274)$$

Show from the definition (1.261) of  $f$  that

$$(1 - |\boldsymbol{\lambda}|^2 f) = (|\boldsymbol{\lambda}|/2) \cot(|\boldsymbol{\lambda}|/2). \quad (11.1.275)$$

Show that (1.273) and (1.274) have the solution

$$a = -(2/|\boldsymbol{\lambda}|^2) \sin^2(|\boldsymbol{\lambda}|/2), \quad (11.1.276)$$

$$b = (1/|\boldsymbol{\lambda}|^2)[1 - (1/|\boldsymbol{\lambda}|) \sin(|\boldsymbol{\lambda}|)]. \quad (11.1.277)$$

Evaluate  $a(0)$  and  $b(0)$  and verify that both  $a(\boldsymbol{\lambda})$  and  $b(\boldsymbol{\lambda})$  are analytic in the components of  $\boldsymbol{\lambda}$  everywhere except at  $\infty$ .

**11.1.12.** The aim of this exercise is to explore the use of Cayley transforms and parameterizations for the purpose of integration on manifolds. Review Section 3.12 and Exercises 3.12.1, 3.12.5, and 3.12.6 to recall the subject of Cayley transforms for quadratic groups  $G$ . We will begin with the general case. Then in a following exercise, we will study in more detail the cases of  $SO(3, \mathbb{R})$  and  $SU(2)$ , which are more tractable. Specifically, in this exercise we will study equations of motion of the form

$$\dot{M}(t) = M(t)A(t) \quad (11.1.278)$$

where  $M(t)$  is expected to belong to some quadratic Lie group  $G$  and  $A(t)$  belongs to its associated Lie algebra. What we seek is a way of numerically integrating (1.278) that guarantees  $M(t)$  is in  $G$  even though the numerical solution may be locally exact only through terms of order  $h^m$ .

We could also study the related equations of motion of the form

$$\dot{N}(t) = \bar{A}(t)N(t) \quad (11.1.279)$$

but, as will be seen from some of the work of Exercise 2.7, this case is equivalent to the case (1.278) under the substitutions  $N = M^{-1}$  and  $\bar{A} = -A$ .

From Appendix C we know in general that exact integration of (1.278) assures that  $M(t)$  is in  $G$ . Here, before going further, your first task is to provide a simple proof of this fact in the case of quadratic groups.

Begin with the converse claim. Let  $M(t)$  be some path in matrix space. By Taylor's theorem there is the result

$$M(t + dt) = M(t) + dt\dot{M}(t) + O[(dt)^2]. \quad (11.1.280)$$

Define a matrix  $A(t)$  by the rule

$$A(t) = M^{-1}(t)\dot{M}(t). \quad (11.1.281)$$

Verify that (1.278) follows from (1.281). What remains is to determine the properties of  $A(t)$ .

For  $M(t)$  to belong to a quadratic group  $G$  it must satisfy a relation of the form

$$M^T(t)L M(t) = L. \quad (11.1.282)$$

See Exercise 3.12.5. From (1.282) verify that

$$M^T(t + dt)L M(t + dt) = L. \quad (11.1.283)$$

Since  $M(t)$  and  $M(t + dt)$  are nearby matrices, it follows that the product  $M^{-1}(t)M(t + dt)$  must be near the identity. Indeed, verify that we may write

$$\begin{aligned} M^{-1}(t)M(t + dt) &= M^{-1}(t)[M(t) + dt\dot{M}(t)] + O[(dt)^2] \\ &= I + dtM^{-1}(t)\dot{M}(t) + O[(dt)^2] \\ &= I + dtA(t) + O[(dt)^2] \\ &= \exp[dtA(t)] + O[(dt)^2]. \end{aligned} \quad (11.1.284)$$

Rewrite (1.284) in the form

$$M(t + dt) = M(t)\exp[dtA(t)] + O[(dt)^2]. \quad (11.1.285)$$

Show that employing this relation in (1.283) and equating powers of  $dt$  yields the condition

$$A^T(t)L + LA(t) = 0, \quad (11.1.286)$$

which demonstrates that  $A(t)$  belongs to the Lie algebra of  $G$ .

Next consider the direct claim. Suppose that  $M(t)$  satisfies (1.278) and that  $A(t)$  satisfies (1.286). Assume also that at some time  $t^0$  there is the relation

$$M^T(t^0)L M(t^0) = L. \quad (11.1.287)$$

Such will be the case, in particular, if  $M(t^0) = I$ . Your task is to show that then (1.282) must hold for all  $t$ .

Begin by showing that (1.287) can be rewritten in the form

$$[M^{-1}(t^0)]^T L [M^{-1}(t^0)] = L. \quad (11.1.288)$$

Next, from the identity

$$M^{-1}(t)M(t) = I \quad (11.1.289)$$

and (1.278), show that

$$(d/dt)[M^{-1}(t)] = -A(t)M^{-1}(t). \quad (11.1.290)$$

As a further step show that

$$\begin{aligned} (d/dt)[(M^{-1})^T LM^{-1}] &= \{(d/dt)[(M^{-1})^T]\}LM^{-1} + (M^{-1})^T L(d/dt)(M^{-1}) \\ &= [(d/dt)(M^{-1})]^T LM^{-1} + (M^{-1})^T L(d/dt)(M^{-1}) \\ &= [-AM^{-1}]^T LM^{-1} + (M^{-1})^T L[-AM^{-1}] \\ &= -(M^{-1})^T [A^T L + LA]M^{-1} = 0. \end{aligned} \quad (11.1.291)$$

Verify that the unique solution to the differential equation (1.291) with the initial condition (1.288) is the relation

$$[M^{-1}(t)]^T L[M^{-1}(t)] = L. \quad (11.1.292)$$

Finally, show that (1.282) follows from (1.292).

With this background work out of the way, the main task of this exercise is to consider use of the Cayley parameterization. Specifically, for  $M$  we employ the Cayley parameterization

$$M = (I + V)(I - V)^{-1}, \quad (11.1.293)$$

see (3.12.36), and your task is to find the equation of motion for  $V$ . In particular, you are to show how this may be done starting from (1.278) rewritten as (1.281).

Begin by writing  $M$  in the form

$$M = BC \quad (11.1.294)$$

where

$$B = I + V \quad (11.1.295)$$

and

$$C = D^{-1} \quad (11.1.296)$$

with

$$D = I - V. \quad (11.1.297)$$

Show that the product differentiation rule yields the result

$$\dot{M} = \dot{B}C + B\dot{C}. \quad (11.1.298)$$

Simple calculation with (1.295) yields the result

$$\dot{B} = \dot{V}. \quad (11.1.299)$$

The calculation of  $\dot{C}$  is more involved. Show from (1.296) and the product differentiation rule that

$$CD = I, \quad (11.1.300)$$

$$\dot{C}D + C\dot{D} = 0, \quad (11.1.301)$$

and therefore

$$\dot{C} = -C\dot{D}C. \quad (11.1.302)$$

Show that use of (1.297) gives the result

$$\dot{D} = -\dot{V}, \quad (11.1.303)$$

and therefore

$$\dot{C} = C\dot{V}C. \quad (11.1.304)$$

Verify that combining the results obtained so far gives the next conclusion

$$\dot{M} = \dot{V}C + BC\dot{V}C. \quad (11.1.305)$$

Manipulate some more. Verify the steps

$$\begin{aligned} \dot{M} &= \dot{V}C + BC\dot{V}C = C^{-1}\dot{C}V + BC\dot{V}C = (C^{-1} + B)\dot{C}V \\ &= (D + B)\dot{C}V = [(I - V) + (I + V)]\dot{C}V = 2\dot{C}V. \end{aligned} \quad (11.1.306)$$

According to (1.281) what is needed is the quantity  $M^{-1}\dot{M}$ . Show that

$$M^{-1} = C^{-1}B^{-1} \quad (11.1.307)$$

and therefore

$$M^{-1}\dot{M} = 2C^{-1}B^{-1}\dot{C}V. \quad (11.1.308)$$

Verify that  $B$  and  $C$  commute and therefore  $B^{-1}$  and  $C^{-1}$  commute. It follows that

$$C^{-1}B^{-1}C = B^{-1}C^{-1}C = B^{-1} \quad (11.1.309)$$

so that

$$M^{-1}\dot{M} = 2C^{-1}B^{-1}\dot{C}V = 2B^{-1}\dot{V}C, \quad (11.1.310)$$

from which we conclude, with the aid of (1.281), that

$$A = 2B^{-1}\dot{V}C. \quad (11.1.311)$$

Solve (1.311) for  $\dot{V}$  to yield the result

$$\dot{V} = (1/2)BAC^{-1} = (1/2)(I + V)A(I - V) = (1/2)(A + VA - AV - VAV), \quad (11.1.312)$$

which can be written as

$$\dot{V} = (1/2)(A + \{V, A\} - VAV). \quad (11.1.313)$$

You have found  $\dot{V}$  in terms of  $A$ . Note that, in contrast to (1.82) whose right side contains an infinite number of terms, the right side of (1.313) contains only three terms.

Is this result sane? For a quadratic group  $G$  we know that  $V$  is in the Lie algebra. From the definition

$$\dot{V}(t) = \lim_{\epsilon \rightarrow 0}[V(t + \epsilon) - V(t)]/\epsilon \quad (11.1.314)$$

we see that only vector space operations are involved in the calculation of  $\dot{V}(t)$ , and therefore  $\dot{V}(t)$  must also be in the Lie algebra of  $G$ . But is the right side of (1.313) in the Lie algebra

of  $G$ ? Evidently, since  $A$  is in the Lie algebra of  $G$ , the first two terms on the right side of (1.313) are in the Lie algebra of  $G$ . What about the third term  $VAV$ ? According to Exercise 3.12.4 the condition for  $A$  and  $V$  to be in the Lie algebra of  $G$  is that

$$L^{-1}A^T L = -A, \quad (11.1.315)$$

$$L^{-1}V^T L = -V. \quad (11.1.316)$$

See (3.12.34). Verify it follows by matrix manipulation that

$$L^{-1}(VAV)^T L = -VAV, \quad (11.1.317)$$

and therefore  $VAV$  is also in the Lie algebra of  $G$ . Therefore (1.313) is sane at least to the extent that both its sides are in the Lie algebra of  $G$ .

To return to the main discussion, suppose (1.313) is integrated by some numerical integrator to find  $V(t)$  under the assumption that  $V$  is initially in the Lie algebra of  $G$ . We repeat the key observation of Subsection 1.14: Examination of the usual numerical integration schemes, see Chapter 2, reveals that they all involve just linear combinations of the right side of the differential equation in question evaluated at various times and coordinate values. Therefore, if the right side is known to be in the Lie algebra of  $G$  for all evaluation points, then the result of numerically integrating such an equation is guaranteed to be in the Lie algebra of  $G$ , no matter what the local accuracy of the integrator or the step size employed. Since we have verified that the right side of (1.313) is in the Lie algebra of  $G$ , it follows that  $V(t)$  will be in the Lie algebra of  $G$  if it is initially in the Lie algebra of  $G$ . Finally, since  $V(t)$  is in the Lie algebra of  $G$ , it follows that  $M(t)$  given by (1.293) is in  $G$ .

We have achieved our goal of, in effect, numerically integrating (1.278) in such a way that  $M(t)$  is guaranteed to be in  $G$ . Note, however, that this procedure cannot be carried out globally since the Cayley parameterization (1.293) cannot be made globally. It may therefore be necessary to change coordinate systems (by left or right group translation) from time to time during the course of a numerical integration in order to stay clear of the singularities associated with any given Cayley parametrization.

Although we have achieved our goal, there is still one undesirable feature of our procedure. Namely, if  $A$  and  $V$  are  $k \times k$  matrices, then the integration of (1.313) involves the integration of  $k^2$  equations. Generally the group  $G$  has dimension considerably less than  $k^2$ . What we would like is a way of parameterizing the Lie algebra of  $G$  and a procedure that only involves the integration of differential equations for these parameters. In the next exercise we will illustrate such a procedure for the cases of  $SO(3, \mathbb{R})$  and  $SU(2)$  where the necessary operations can be carried out relatively easily.

We close this exercise with a small variation. Suppose the relations (1.293) through (1.307) remain in force, but the task is to find

$$\bar{A} = \dot{M} M^{-1} \quad (11.1.318)$$

rather than (1.281). Verify, using (1.306) and (1.307), that

$$\dot{M} M^{-1} = 2C\dot{V}CC^{-1}B^{-1} = 2C\dot{V}B^{-1}. \quad (11.1.319)$$

Show that solving (1.319) for  $\dot{V}$  with the aid of (1.318) yields the result

$$\begin{aligned}\dot{V} &= (1/2)C^{-1}\bar{A}B = (1/2)D\bar{A}B = (1/2)I - V)\bar{A}(I + V) \\ &= (1/2)(\bar{A} - \{V, \bar{A}\} - V\bar{A}V).\end{aligned}\quad (11.1.320)$$

This result will be of use in Subsection 2.9.

**11.1.13.** The aim of this exercise is to apply the methods of the previous exercise to the cases of  $SO(3, \mathbb{R})$  and  $SU(2)$  including parameterization of the associated Lie algebras. In the case of  $SO(3, \mathbb{R})$  we seek to integrate the equation

$$\dot{R} = R \boldsymbol{\omega}^{bf} \cdot \mathbf{L}. \quad (11.1.321)$$

Recall (1.18). And in the case of  $SU(2)$  we seek to integrate (1.146) rewritten in the form

$$\dot{u} = u \boldsymbol{\omega}^{bf} \cdot \mathbf{K}. \quad (11.1.322)$$

In these two cases we write the Cayley parameterizations

$$R(\boldsymbol{\mu}) = (I + \boldsymbol{\mu} \cdot \mathbf{L})(I - \boldsymbol{\mu} \cdot \mathbf{L})^{-1} \quad (11.1.323)$$

and

$$u(\boldsymbol{\mu}) = (I + \boldsymbol{\mu} \cdot \mathbf{K})(I - \boldsymbol{\mu} \cdot \mathbf{K})^{-1}. \quad (11.1.324)$$

What you are to find is the relation between  $\dot{\boldsymbol{\mu}}$  and  $\boldsymbol{\omega}^{bf}$  for these two cases. At this point you should review Exercise 1.12 if you have not previously studied it.

For the case of  $SO(3, \mathbb{R})$  compare (1.321) with (1.278) and compare (1.323) with (1.293). Show that in this case there are the correspondences

$$A = \boldsymbol{\omega}^{bf} \cdot \mathbf{L}, \quad (11.1.325)$$

$$V = \boldsymbol{\mu} \cdot \mathbf{L}, \quad \dot{V} = \dot{\boldsymbol{\mu}} \cdot \mathbf{L}. \quad (11.1.326)$$

For the case of  $SU(2)$  compare (1.322) with (1.278) and compare (1.324) with (1.293). Show that in this case there are the correspondences

$$A = \boldsymbol{\omega}^{bf} \cdot \mathbf{K}, \quad (11.1.327)$$

$$V = \boldsymbol{\mu} \cdot \mathbf{K}, \quad \dot{V} = \dot{\boldsymbol{\mu}} \cdot \mathbf{K}. \quad (11.1.328)$$

What remains is to insert these results into (1.313) and to work out the consequences. Begin with the case of  $SO(3, \mathbb{R})$ . Show that use of (1.325) and (1.326) in (1.313) yields the result

$$\dot{\boldsymbol{\mu}} \cdot \mathbf{L} = (1/2)[\boldsymbol{\omega}^{bf} \cdot \mathbf{L} + \{\boldsymbol{\mu} \cdot \mathbf{L}, \boldsymbol{\omega}^{bf} \cdot \mathbf{L}\} - (\boldsymbol{\mu} \cdot \mathbf{L})(\boldsymbol{\omega}^{bf} \cdot \mathbf{L})(\boldsymbol{\mu} \cdot \mathbf{L})]. \quad (11.1.329)$$

Next manipulate the terms appearing on the right side of this equation. The commutator term is easy. It has the value

$$\{\boldsymbol{\mu} \cdot \mathbf{L}, \boldsymbol{\omega}^{bf} \cdot \mathbf{L}\} = (\boldsymbol{\mu} \times \boldsymbol{\omega}^{bf}) \cdot \mathbf{L}. \quad (11.1.330)$$

See (3.7.183). The evaluation of

$$(\boldsymbol{\mu} \cdot \mathbf{L})(\boldsymbol{\omega}^{bf} \cdot \mathbf{L})(\boldsymbol{\mu} \cdot \mathbf{L}) = ? \quad (11.1.331)$$

is more tedious. By using the  $3 \times 3$  matrix form for each of the factors in (1.331), multiplying out the matrices, and rewriting the result in the form  $\mathbf{c} \cdot \mathbf{L}$ , show that

$$(\boldsymbol{\mu} \cdot \mathbf{L})(\boldsymbol{\omega}^{bf} \cdot \mathbf{L})(\boldsymbol{\mu} \cdot \mathbf{L}) = -(\boldsymbol{\mu} \cdot \boldsymbol{\omega}^{bf})(\boldsymbol{\mu} \cdot \mathbf{L}). \quad (11.1.332)$$

Verify that combining (1.329) through (1.332) gives the result

$$\dot{\boldsymbol{\mu}} \cdot \mathbf{L} = (1/2)[\boldsymbol{\omega}^{bf} + (\boldsymbol{\mu} \times \boldsymbol{\omega}^{bf}) + (\boldsymbol{\mu} \cdot \boldsymbol{\omega}^{bf})\boldsymbol{\mu}] \cdot \mathbf{L}. \quad (11.1.333)$$

From this result it follows that there are the equations of motion

$$\dot{\boldsymbol{\mu}} = (1/2)[\boldsymbol{\omega}^{bf} + (\boldsymbol{\mu} \times \boldsymbol{\omega}^{bf}) + (\boldsymbol{\mu} \cdot \boldsymbol{\omega}^{bf})\boldsymbol{\mu}]. \quad (11.1.334)$$

We have achieved the desired goal for the case of  $SO(3, \mathbb{R})$ . Note that, since  $R$  is  $3 \times 3$  real, the integration of (1.321), and its Cayley counterpart (1.313), involves 9 real differential equations. By contrast, the integration of (1.334) involves only 3 real differential equations. Although (1.313) and its integration preserves lie algebraic structure, it does not exploit this structure. By contrast, based on the introduction of a basis, (1.334) exploits Lie algebraic structure. And, of course, if local errors of order  $h^{m+1}$  arise in the numerical integration of (1.334), the resulting  $R(\boldsymbol{\mu})$  is still guaranteed to be in  $SO(3, \mathbb{R})$  because of the Ansatz (1.323).

The case of  $SU(2)$  requires more calculations, but these calculations involve only results already known. Verify that inserting (1.327) and (1.328) into (1.313) yields the relation

$$\dot{\boldsymbol{\mu}} \cdot \mathbf{K} = (1/2)[\boldsymbol{\omega}^{bf} \cdot \mathbf{K} + \{\boldsymbol{\mu} \cdot \mathbf{K}, \boldsymbol{\omega}^{bf} \cdot \mathbf{K}\} - (\boldsymbol{\mu} \cdot \mathbf{K})(\boldsymbol{\omega}^{bf} \cdot \mathbf{K})(\boldsymbol{\mu} \cdot \mathbf{K})]. \quad (11.1.335)$$

Now manipulate the terms in (1.335) using known results. Recall that

$$\{\boldsymbol{\mu} \cdot \mathbf{K}, \boldsymbol{\omega}^{bf} \cdot \mathbf{K}\} = (\boldsymbol{\mu} \times \boldsymbol{\omega}^{bf}) \cdot \mathbf{K}. \quad (11.1.336)$$

See (3.7.182). Next recall that

$$(\boldsymbol{\mu} \cdot \mathbf{K})(\boldsymbol{\omega}^{bf} \cdot \mathbf{K}) = -(1/4)(\boldsymbol{\mu} \cdot \boldsymbol{\omega}^{bf})I + (1/2)(\boldsymbol{\mu} \times \boldsymbol{\omega}^{bf}) \cdot \mathbf{K}. \quad (11.1.337)$$

See (3.7.176). It follows that

$$(\boldsymbol{\mu} \cdot \mathbf{K})(\boldsymbol{\omega}^{bf} \cdot \mathbf{K})(\boldsymbol{\mu} \cdot \mathbf{K}) = -(1/4)(\boldsymbol{\mu} \cdot \boldsymbol{\omega}^{bf})(\boldsymbol{\mu} \cdot \mathbf{K}) + (1/2)[(\boldsymbol{\mu} \times \boldsymbol{\omega}^{bf}) \cdot \mathbf{K}](\boldsymbol{\mu} \cdot \mathbf{K}). \quad (11.1.338)$$

Verify that

$$\begin{aligned} [(\boldsymbol{\mu} \times \boldsymbol{\omega}^{bf}) \cdot \mathbf{K}](\boldsymbol{\mu} \cdot \mathbf{K}) &= -(1/4)[(\boldsymbol{\mu} \times \boldsymbol{\omega}^{bf}) \cdot \boldsymbol{\mu}]I + (1/2)[(\boldsymbol{\mu} \times \boldsymbol{\omega}^{bf}) \times \boldsymbol{\mu}] \cdot \mathbf{K} \\ &= 0 - (1/2)[\boldsymbol{\mu} \times (\boldsymbol{\mu} \times \boldsymbol{\omega}^{bf})] \cdot \mathbf{K} = -(1/2)[(\boldsymbol{\mu} \cdot \boldsymbol{\omega}^{bf})\boldsymbol{\mu} - (\boldsymbol{\mu} \cdot \boldsymbol{\mu})\boldsymbol{\omega}^{bf}] \cdot \mathbf{K}, \end{aligned} \quad (11.1.339)$$

and that, consequently,

$$\begin{aligned}
 & (\boldsymbol{\mu} \cdot \mathbf{K})(\boldsymbol{\omega}^{bf} \cdot \mathbf{K})(\boldsymbol{\mu} \cdot \mathbf{K}) \\
 &= -(1/4)(\boldsymbol{\mu} \cdot \boldsymbol{\omega}^{bf})(\boldsymbol{\mu} \cdot \mathbf{K}) - (1/4)[(\boldsymbol{\mu} \cdot \boldsymbol{\omega}^{bf})\boldsymbol{\mu} - (\boldsymbol{\mu} \cdot \boldsymbol{\mu})\boldsymbol{\omega}^{bf}] \cdot \mathbf{K} \\
 &= [-(1/2)(\boldsymbol{\mu} \cdot \boldsymbol{\omega}^{bf})\boldsymbol{\mu} + (1/4)(\boldsymbol{\mu} \cdot \boldsymbol{\mu})\boldsymbol{\omega}^{bf}] \cdot \mathbf{K}.
 \end{aligned} \tag{11.1.340}$$

Verify that combining (1.335) through (1.340) gives the result

$$\dot{\boldsymbol{\mu}} \cdot \mathbf{K} = (1/2)[\boldsymbol{\omega}^{bf} + (\boldsymbol{\mu} \times \boldsymbol{\omega}^{bf}) + (1/2)(\boldsymbol{\mu} \cdot \boldsymbol{\omega}^{bf})\boldsymbol{\mu} - (1/4)(\boldsymbol{\mu} \cdot \boldsymbol{\mu})\boldsymbol{\omega}^{bf}] \cdot \mathbf{K}. \tag{11.1.341}$$

From this result it follows that there are the equations of motion

$$\dot{\boldsymbol{\mu}} = (1/2)[\boldsymbol{\omega}^{bf} + (\boldsymbol{\mu} \times \boldsymbol{\omega}^{bf}) + (1/2)(\boldsymbol{\mu} \cdot \boldsymbol{\omega}^{bf})\boldsymbol{\mu} - (1/4)(\boldsymbol{\mu} \cdot \boldsymbol{\mu})\boldsymbol{\omega}^{bf}]. \tag{11.1.342}$$

We have achieved our goal for the case of  $SU(2)$ . Note that, since  $\boldsymbol{u}$  is  $2 \times 2$  complex, the integration of (1.322), and its Cayley counterpart, involves 8 real differential equations. By contrast, the integration of (1.342) again involves only 3 real differential equations. And, again, if local errors of order  $h^{m+1}$  arise in the numerical integration of (1.342), the resulting  $\boldsymbol{u}(\boldsymbol{\mu})$  is still guaranteed to be in  $SU(2)$  because of the Ansatz (1.324).

What can be said about the singularity structure of Cayley parameterization? Evidently (1.293) is singular when

$$\det(I - V) = 0, \tag{11.1.343}$$

that is, when  $V$  has  $+1$  as an eigenvalue. Also, (1.311) can be rewritten in the form

$$A = 2(I + V)^{-1}\dot{V}(I - V)^{-1} \tag{11.1.344}$$

so that the relation between  $A$  and  $\dot{V}$  is singular when

$$\det(I + V) = 0 \text{ and } \det(I - V) = 0. \tag{11.1.345}$$

Strangely enough, the general equation of motion (1.313), and the specific  $SO(3, \mathbb{R})$  and  $SU(2)$  equations of motion (1.334) and (1.342), appear to be singularity free. This appearance is deceptive, because, for example, (1.334) and (1.342) are singular in  $\boldsymbol{\mu}$  at infinity, and there is the possibility that this singularity can be encountered in *finite* time.

To realize this possible singularity in the case of  $SO(3, \mathbb{R})$ , suppose that

$$\omega_j^{bf} = \Omega\delta_{j3}, \tag{11.1.346}$$

where  $\Omega$  is a constant, and make the Ansatz

$$\mu_j = f(t)\delta_{j3} \tag{11.1.347}$$

where  $f$  satisfies the initial condition

$$f(0) = 0, \tag{11.1.348}$$

but is otherwise to be determined. Show that putting this Ansatz into the equation of motion (1.334) yields the result

$$\dot{f} = (1/2)\Omega(1 + f^2). \quad (11.1.349)$$

Show that the solution to (1.349) with the initial condition (1.348) is

$$f = \tan(\Omega t/2). \quad (11.1.350)$$

Review Exercise 3.12.6. Show that the corresponding  $\lambda$  is given by

$$\lambda_j = \Omega t \delta_{j3} \quad (11.1.351)$$

and therefore  $R$  is given by

$$R = \exp(\Omega t L^3). \quad (11.1.352)$$

See (3.12.61). Observe that (1.350), and hence (1.347), are singular when  $\Omega t = \pi$  and therefore when  $|\lambda| = \pi$  which, according to Exercise 3.12.6, is the expected condition for singularity in the case of  $SO(3, \mathbb{R})$ .

To realize this possible singularity in the case of  $SU(2)$ , suppose that (1.346) through (1.348) continue to hold. Show that putting this Ansatz into (1.342) yields the differential equation

$$\dot{f} = (\Omega/2)(1 + f^2/4), \quad (11.1.353)$$

and that this equation has the solution

$$f = 2 \tan(\Omega t/4). \quad (11.1.354)$$

Show that the corresponding  $\lambda$  is again given by

$$\lambda_j = \Omega t \delta_{j3} \quad (11.1.355)$$

and therefore  $u$  is given by

$$u = \exp(\Omega t K^3). \quad (11.1.356)$$

See (3.12.73). Observe that (1.354), and hence (1.347), are singular when  $\Omega t = 2\pi$  and therefore when  $|\lambda| = 2\pi$  which, according to Exercise 3.12.6, is the expected condition for singularity in the case of  $SU(2)$ .

At this point you, the observant reader, might object that the quaternion parameter equations of motion (1.43) through (1.46), which were lauded as being wonderful, are also singular in the components of  $w$  at infinity. They are indeed singular at infinity, but this singularity cannot be reached in *real* time because these equations preserve the condition  $w \cdot w = 1$ . This singularity can be reached in *complex* time, but because the equations of motion are *linear* in  $w$ , the time must be *infinite* complex. Therefore, for quaternion parameterization, there are no singularities in finite time, real or complex.

## 11.2 Numerical Integration on Manifolds: Spin and Qubits

As a second example of integration on manifolds, we consider an equation that occurs in several contexts. Let  $\mathbf{s}(t)$  be a time-dependent 3-dimensional vector that evolves according to the rule

$$d\mathbf{s}/dt = \bar{\boldsymbol{\omega}}(t) \times \mathbf{s} \quad (11.2.1)$$

where  $\bar{\boldsymbol{\omega}}(t)$  is some other specified, possibly time dependent, 3-dimensional vector. This equation is called the *Bloch* equation in the context of nuclear magnetic resonance (NMR or MRI) and electron spin resonance (ESR), and the *Bargmann-Michel-Telegdi* (BMT or *Thomas*-BMT or *Thomas-Frenkel*-BMT) equation in the context of determining the evolution of a particle's spin polarization vector as it traverses some accelerator or beam line. It also occurs in the context of rigid-body motion. See (1.105) in Exercise 1.1. Finally, it is relevant to the general quantum mechanical treatment of two level systems (qubits), and therefore plays a prominent role in quantum information theory and quantum computation. See Exercises 2.15 and 2.16. Note also that the equation of motion (1.6.112) can be written as

$$d\mathbf{v}/dt = [-(q/m^*)\mathbf{B}] \times \mathbf{v}, \quad (11.2.2)$$

which also appears to be of the form (2.1). See Section 3 for further discussion of this observation.

Suppose our task is to find  $\mathbf{s}(t)$  given  $\bar{\boldsymbol{\omega}}(t)$  and the initial vector

$$\mathbf{s}^0 = \mathbf{s}(t^0) \quad (11.2.3)$$

at time  $t^0$ .<sup>17</sup> It is easily verified that the equations of motion (2.1) preserve  $\mathbf{s} \cdot \mathbf{s}$ . Define a quantity  $s^*$  by the rule

$$s^* = \sqrt{\mathbf{s}^0 \cdot \mathbf{s}^0}, \quad (11.2.4)$$

and let  $S^{2*}$  denote the two-sphere of radius  $s^*$  imbedded in the ambient space  $E^3$ . With this notation, we may say that the equations of motion (2.1) preserve  $S^{2*}$ , and are equations of motion on the manifold  $S^{2*}$  embedded in the ambient space  $E^3$ . Note also that the equations of motion (2.1) are *linear*. Therefore, for many applications, there is no loss in generality in taking  $\mathbf{s}^0$  to be a unit vector:  $\mathbf{s}^0 \in S^2$  where, as usual,  $S^2$  denotes the unit two-sphere ( $s^* = 1$ ). Solutions corresponding to initial conditions that are not unit vectors can be obtained by simple scaling of the solutions corresponding to unit-vector initial conditions. Therefore, unless specifically specified otherwise, we will work with the  $S^2$  case. However, where useful, we will treat explicitly the general  $S^{2*}$  case.

Our task is to integrate (2.1) numerically in such a way that, even if local errors of order  $h^{m+1}$  are made, the solution is guaranteed to be in  $S^2$ . One procedure for so doing is to employ any of the standard methods of Chapter 2 for one step at a time (thereby making local errors of order  $h^{m+1}$ ) and to then project after each step the resulting  $\mathbf{s}$  back onto  $S^2$  by simple scaling. Alternatively, we may consider other approaches that parameterize  $S^2$  or exploit other features of the problem. The purpose of this section is to describe several such approaches.

---

<sup>17</sup>For a discussion of the inverse problem, namely given  $\mathbf{s}(t)$  find  $\bar{\boldsymbol{\omega}}(t)$ , see Exercise 2.1.

### 11.2.1 Constrained Cartesian Coordinates Are Not Global

In Cartesian coordinates (2.1) yields the three coupled linear equations

$$\dot{s}_1 = \bar{\omega}_2 s_3 - \bar{\omega}_3 s_2, \quad (11.2.5)$$

$$\dot{s}_2 = \bar{\omega}_3 s_1 - \bar{\omega}_1 s_3, \quad (11.2.6)$$

$$\dot{s}_3 = \bar{\omega}_1 s_2 - \bar{\omega}_2 s_1. \quad (11.2.7)$$

To recapitulate, if they are integrated numerically by a method that makes local errors of order  $h^{m+1}$ , the quantity  $\mathbf{s} \cdot \mathbf{s}$  will generally be locally preserved only through terms of order  $h^m$ . If we wish to preserve the condition  $\mathbf{s} \in S^2$  to machine precision, one simple procedure is to project  $\mathbf{s} \in E^3$  back onto  $S^2$  by simple scaling after each integration step.

Another procedure, assuming  $\mathbf{s}^0 \in S^2$ , is to enforce the condition  $\mathbf{s} \in S^2$  by making the definition

$$s_1 = +(1 - s_2^2 - s_3^2)^{1/2} \quad (11.2.8)$$

and inserting this definition/constraint into the equations (2.6) and (2.7) to yield the equations of motion

$$\dot{s}_2 = \bar{\omega}_3(1 - s_2^2 - s_3^2)^{1/2} - \bar{\omega}_1 s_3, \quad (11.2.9)$$

$$\dot{s}_3 = \bar{\omega}_1 s_2 - \bar{\omega}_2(1 - s_2^2 - s_3^2)^{1/2}. \quad (11.2.10)$$

In essence, we have parameterized  $S^2$  by the coordinates  $s_2$  and  $s_3$ . If  $\mathbf{s}$  is initially in the front hemisphere,  $s_1 > 0$ , these equations for  $s_2$  and  $s_3$  can be integrated as long as  $(s_2^2 + s_3^2) < 1$ . However, they become singular if  $\mathbf{s}$  crosses the plane  $s_1 = 0$  (which divides the front and rear hemispheres), as is certainly mathematically/physically possible, and they therefore cannot be generally used to produce a global solution.

### 11.2.2 Polar-Angle Coordinates Are Not Global

Yet another possibility is to parameterize  $\mathbf{s} \in S^2$  by the use of polar-angle coordinates  $\theta$  and  $\phi$ . Make the Ansatz

$$s_1 = \sin(\theta) \cos(\phi), \quad (11.2.11)$$

$$s_2 = \sin(\theta) \sin(\phi), \quad (11.2.12)$$

$$s_3 = \cos(\theta). \quad (11.2.13)$$

This Ansatz guarantees  $\mathbf{s} \in S^2$ . From (2.11) through (2.13) we find the relations

$$\dot{s}_1 = \dot{\theta} \cos(\theta) \cos(\phi) - \dot{\phi} \sin(\theta) \sin(\phi), \quad (11.2.14)$$

$$\dot{s}_2 = \dot{\theta} \cos(\theta) \sin(\phi) + \dot{\phi} \sin(\theta) \cos(\phi), \quad (11.2.15)$$

$$\dot{s}_3 = -\dot{\theta} \sin(\theta). \quad (11.2.16)$$

Solving these relations for  $\dot{\theta}$  and  $\dot{\phi}$  yields the results

$$\dot{\theta} = -[1/\sin(\theta)]\dot{s}_3, \quad (11.2.17)$$

$$\dot{\phi} = -[1/\sin(\theta)][\dot{s}_1 \sin(\phi) - \dot{s}_2 \cos(\phi)]. \quad (11.2.18)$$

Finally, combining (2.5) through (2.7) and (2.11) through (2.13) with (2.17) and (2.18) yields the equations of motion

$$\dot{\theta} = -\bar{\omega}_1 \sin(\phi) + \bar{\omega}_2 \cos(\phi), \quad (11.2.19)$$

$$\dot{\phi} = \bar{\omega}_3 - [\cos(\theta)/\sin(\theta)][\bar{\omega}_1 \cos(\phi) + \bar{\omega}_2 \sin(\phi)]. \quad (11.2.20)$$

Observe that (2.20) is singular at the poles  $\theta = 0$  and  $\theta = \pi$ . (Note that  $\phi$  is ill defined at the poles, and consequently these singularities are to be expected). Therefore these equations are also not suitable for finding global solutions.

### 11.2.3 Local Tangent-Space Coordinates

One way to insure that a numerical trajectory will remain on an invariant manifold is to introduce local coordinates in the ambient space at some point on the manifold, locally parameterize the manifold, formulate differential equations for these parameters, and finally numerically integrate the differential equations for the parameters. By so doing, even if single-step errors of order  $h^{m+1}$  occur in the parameters over the process of integration, the resulting trajectory is guaranteed to remain on the manifold. We will illustrate this process for the equation of motion (2.1) and, for future use, we will explicitly treat the general case  $S^{2*}$ .

Let  $\mathbf{s}^b$  be some point on the manifold  $S^{2*}$  at the *beginning* of an integration step to be initiated at time  $t^b$ . Parameterize points in the ambient space and in the vicinity of  $\mathbf{s}^b$  by writing

$$\mathbf{s}(t) = \mathbf{s}^b + \mathbf{s}^v(t) \quad (11.2.21)$$

where  $\mathbf{s}^v(t)$  is a *variable* vector with the property

$$\mathbf{s}^v(t^b) = 0. \quad (11.2.22)$$

Next work to insure that  $\mathbf{s}(t)$  remains on  $S^{2*}$  as  $t$  varies. Let  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$  be a fixed right-hand triad of orthonormal vectors. Construct a second right-hand triad of orthonormal vectors  $\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3$  associated with  $\mathbf{s}^b$  as follows: Begin by defining

$$\mathbf{f}_1 = \mathbf{s}^b/s^*. \quad (11.2.23)$$

Next examine the quantities  $\mathbf{e}_j \cdot \mathbf{f}_1$  and select the  $\mathbf{e}_j$  for which  $\mathbf{e}_j \cdot \mathbf{f}_1$  has the least magnitude. This will be the  $\mathbf{e}_j$  that is most nearly perpendicular to  $\mathbf{s}^b$ . Use this  $\mathbf{e}_j$ , call it  $\mathbf{e}_k$ , to define the unit vector

$$\mathbf{f}_2 = (\mathbf{e}_k \times \mathbf{f}_1)/|\mathbf{e}_k \times \mathbf{f}_1|. \quad (11.2.24)$$

By construction  $\mathbf{f}_2$  is perpendicular to  $\mathbf{f}_1$ ,

$$\mathbf{f}_1 \cdot \mathbf{f}_2 = 0. \quad (11.2.25)$$

Finally, define the unit vector  $\mathbf{f}_3$  by the rule

$$\mathbf{f}_3 = \mathbf{f}_1 \times \mathbf{f}_2. \quad (11.2.26)$$

Continue on by expressing  $\mathbf{s}^v(t)$  in the form

$$\mathbf{s}^v(t) = s_1^{vf}(t)\mathbf{f}_1 + s_2^{vf}(t)\mathbf{f}_2 + s_3^{vf}(t)\mathbf{f}_3. \quad (11.2.27)$$

Here the superscript  $f$  reminds us that the  $\mathbf{f}_j$  have been employed as a basis. That is, the  $s_j^{vf}$  are the components of  $\mathbf{s}^v$  with respect to the  $\mathbf{f}_j$  basis. By construction, vectors of the form

$$\mathbf{s}^{tan} = \mathbf{s}^b + s_2^{vf}\mathbf{f}_2 + s_3^{vf}\mathbf{f}_3 \quad (11.2.28)$$

comprise the *tangent* space to  $S^{2*}$  at  $\mathbf{s}^b$ ; and we may view  $s_2^{vf}$  and  $s_3^{vf}$  as being tangent-space coordinates. Combining (2.21) and (2.27) gives the result

$$\mathbf{s}(t) = [s^* + s_1^{vf}(t)]\mathbf{f}_1 + s_2^{vf}(t)\mathbf{f}_2 + s_3^{vf}(t)\mathbf{f}_3. \quad (11.2.29)$$

Now enforce the condition that  $\mathbf{s}(t)$  lie in  $S^{2*}$ . So doing gives the relation

$$[s^* + s_1^{vf}(t)]^2 + [s_2^{vf}(t)]^2 + [s_3^{vf}(t)]^2 = (s^*)^2, \quad (11.2.30)$$

from which it follows that

$$s_1^{vf}(t) = \{(s^*)^2 - [s_2^{vf}(t)]^2 - [s_3^{vf}(t)]^2\}^{1/2} - s^*. \quad (11.2.31)$$

We see that in the vicinity of  $\mathbf{s}^b$  the manifold  $S^{2*}$  can be parameterized by  $s_2^{vf}$  and  $s_3^{vf}$  providing (2.31) is used to specify  $s_1^{vf}$ .

What remains is to find equations of motion for  $s_2^{vf}$  and  $s_3^{vf}$ . The first step is to expand  $\bar{\omega}(t)$  in terms of the  $\mathbf{f}_j$  by writing

$$\bar{\omega}(t) = \sum_j [\bar{\omega}(t) \cdot \mathbf{f}_j] \mathbf{f}_j = \sum_j \bar{\omega}_j^f(t) \mathbf{f}_j \quad (11.2.32)$$

where we have made the definition

$$\bar{\omega}_j^f(t) = \bar{\omega}(t) \cdot \mathbf{f}_j \quad (11.2.33)$$

and again the superscript  $f$  reminds us that the  $\mathbf{f}_j$  has been employed as a basis. Use of (2.29) gives for the left side of (2.1) the result

$$d\mathbf{s}(t)/dt = \dot{s}_1^{vf}(t)\mathbf{f}_1 + \dot{s}_2^{vf}(t)\mathbf{f}_2 + \dot{s}_3^{vf}(t)\mathbf{f}_3. \quad (11.2.34)$$

Use of (2.29) and (2.32) gives for the right side of (2.1) the result

$$\begin{aligned} \bar{\omega}(t) \times \mathbf{s} &= [\bar{\omega}_2^f(t)s_3^{vf}(t) - \bar{\omega}_3^f(t)s_2^{vf}(t)]\mathbf{f}_1 \\ &+ \{\bar{\omega}_3^f(t)[s^* + s_1^{vf}(t)] - \bar{\omega}_1^f(t)s_3^{vf}(t)\}\mathbf{f}_2 \\ &+ \{\bar{\omega}_1^f(t)s_2^{vf}(t) - \bar{\omega}_2^f(t)[s^* + s_1^{vf}(t)]\}\mathbf{f}_3. \end{aligned} \quad (11.2.35)$$

Now equate the second and third components of (2.34) and (2.35) to find the relations

$$\dot{s}_2^{vf}(t) = \bar{\omega}_3^f(t)[s^* + s_1^{vf}(t)] - \bar{\omega}_1^f(t)s_3^{vf}(t), \quad (11.2.36)$$

$$\dot{s}_3^{vf}(t) = \bar{\omega}_1^f(t)s_2^{vf}(t) - \bar{\omega}_2^f(t)[s^* + s_1^{vf}(t)]. \quad (11.2.37)$$

Finally, employing (2.31) in (2.36) and (2.37) yields the equations of motion

$$\dot{s}_2^{vf}(t) = \bar{\omega}_3^f(t)\{(s^*)^2 - [s_2^{vf}(t)]^2 - [s_3^{vf}(t)]^2\}^{1/2} - \bar{\omega}_1^f(t)s_3^{vf}(t), \quad (11.2.38)$$

$$\dot{s}_3^{vf}(t) = \bar{\omega}_1^f(t)s_2^{vf}(t) - \bar{\omega}_2^f(t)\{(s^*)^2 - [s_2^{vf}(t)]^2 - [s_3^{vf}(t)]^2\}^{1/2}. \quad (11.2.39)$$

It is these equations that are to be numerically integrated from the time  $t^b$  to the time  $t^b + h$  (or perhaps  $t^b + kh$ ) starting with the initial conditions  $s_2^{vf}(t^b) = s_3^{vf}(t^b) = 0$ .<sup>18</sup> Then, once  $\mathbf{s}^v(t^b + h)$  [or perhaps  $\mathbf{s}^v(t^b + kh)$ ] has been obtained,  $\mathbf{s}(t^b + h)$  [or perhaps  $\mathbf{s}(t^b + kh)$ ] is given by (2.21).<sup>19</sup> At this point, the whole process just described is repeated as often as desired. That is, the vectors  $\mathbf{f}_j$  are reconstructed based on the most recently obtained  $\mathbf{s}$ , etc.

We close this subsection by noting that, as was the case with constrained Cartesian coordinates and polar-angle coordinates, there are only *two* equations to be integrated, namely (2.38) and (2.39), whereas working in the ambient space  $E^3$  as in (2.5) through (2.7) required the integration of *three* equations.

### 11.2.4 Exploiting Connection with Rigid-Body Kinematics

We next consider approaches related to those used in the rigid-body case. Suppose we seek the general solution of (2.1). Observe, with the aid of the matrices  $L^j$ , that (2.1) can be written in the form

$$d\mathbf{s}/dt = (\bar{\boldsymbol{\omega}} \cdot \mathbf{L})\mathbf{s}. \quad (11.2.40)$$

Recall (3.7.200). Also, since (2.1) is linear, we may make the general Ansatz

$$\mathbf{s}(t) = S(t)\mathbf{s}^0 \quad (11.2.41)$$

where  $S$  is a  $3 \times 3$  matrix to be determined. Now insert (2.41) into (2.40) to find the relation

$$\dot{S}(t)\mathbf{s}^0 = (\bar{\boldsymbol{\omega}} \cdot \mathbf{L})S(t)\mathbf{s}^0. \quad (11.2.42)$$

Since we wish  $\mathbf{s}^0$  to be an arbitrary unit vector and (2.42) is linear, the relation (2.42) is equivalent to the matrix differential equation

$$\dot{S}(t) = (\bar{\boldsymbol{\omega}} \cdot \mathbf{L})S(t), \quad (11.2.43)$$

and from (2.3) and (2.41) we find the initial condition

$$S(t^0) = I. \quad (11.2.44)$$

In summary, solving (2.43) with the initial condition (2.44) provides the general solution to (2.1). This approach has the advantage that once  $S(t)$  has been found,  $\mathbf{s}(t)$  can be found

---

<sup>18</sup>Observe that the equations of motion (2.38) and (2.39) agree with the equations of motion (2.9) and (2.10) in the case where the  $\mathbf{f}_j$  agree with the  $\mathbf{e}_j$  and  $s^* = 1$ , which is a nice check of our work.

<sup>19</sup>If  $k > 1$  is attempted, one must monitor  $[(s_2^v)^2 + (s_3^v)^2]$  to ensure that the square root singularity in (2.31) is not approached too closely.

for *all* initial conditions  $\mathbf{s}^0$  by the easy computation (2.41). In essence,  $S(t)$  is the transfer map associated with the differential equation (2.1). By contrast, if (2.5) through (2.7), or (2.9) and (2.10), or (2.19) and (2.20), or (2.38) and (2.39) are employed, these differential equations must be integrated afresh for each different initial  $\mathbf{s}^0$ .

At this point we observe that (2.43) and (1.18) are quite similar. Indeed, suppose we pass back and forth between matrices  $R$  and  $S$  by the rule

$$R = S^{-1} \text{ or } S = R^{-1}, \quad (11.2.45)$$

from which it follows that

$$R(t^0) = I. \quad (11.2.46)$$

Then it can be shown from (2.43) that there is the relation

$$\dot{R} = R(\boldsymbol{\omega}^{bf} \cdot \mathbf{L}) \quad (11.2.47)$$

with

$$\boldsymbol{\omega}^{bf} = -\bar{\boldsymbol{\omega}}. \quad (11.2.48)$$

See Exercise 2.7.

We know that  $R$  is orthogonal and therefore, from (2.45), we conclude that  $S$  is also orthogonal. We also observe that in the orthogonal case the relation (2.45) is equivalent to the computationally simpler relation

$$R = S^T \text{ or } S = R^T. \quad (11.2.49)$$

We see that all the machinery developed for and the conclusions drawn about rigid body motion in Section 1 are also applicable here.

### 11.2.5 What Just Happened? Generalizations

In the last subsection we saw that the problem of determining the path  $\mathbf{s}(t)$  in the manifold  $S^2$  and satisfying the *manifold* differential equation (2.1) [or, equivalently (2.40)] with the initial condition (2.3) was converted into finding a path  $S(t)$  in the group  $SO(3, \mathbb{R})$  that satisfied the *group* differential equation (2.43) with the initial condition (2.44). We also observe that the group  $SO(3, \mathbb{R})$  acts *transitively* on the manifold  $S^2$ . (Evidently any point in  $S^2$  can be rotated into any other point in  $S^2$ .) Therefore  $S^2$  is a homogeneous space with respect to the group  $SO(3, \mathbb{R})$ , and is in fact isomorphic to the coset space  $SO(3, \mathbb{R})/SO(2, \mathbb{R})$ . Recall the discussion of homogeneous spaces in Subsections 5.12.3 through 5.12.5.

Observe moreover that (2.41) is a relation that sends the path  $S(t)$  in the group  $SO(3, \mathbb{R})$  to the path  $\mathbf{s}(t)$  in the manifold  $S^2$ . The path  $S(t)$  begins at the identity, see (2.44), and the path  $\mathbf{s}(t)$  begins at  $\mathbf{s}^0$ . In the language of manifold theory, the relation (2.41) is said to *push forward* the path  $S(t) \in SO(3, \mathbb{R})$  to produce the path  $\mathbf{s}(t) \in S^2$ .<sup>20</sup> Correspondingly, the group differential equation (2.43) is the *pullback* to  $SO(3, \mathbb{R})$  of the  $S^2$  manifold differential equation (2.1). Finally, since  $SO(3, \mathbb{R})$  acts on and *preserves*  $S^2$ , we are guaranteed that  $\mathbf{s}(t)$  will be in  $S^2$  if  $S(t)$  is in  $SO(3, \mathbb{R})$ . Therefore a numerical integrator that preserves

---

<sup>20</sup>Put another way,  $\mathbf{s}(t)$  is the orbit of  $\mathbf{s}^0$  under the action of the  $SO(3, \mathbb{R})$  group elements  $S(t)$ .

the  $SO(3, \mathbb{R})$  group manifold for any group differential equation automatically produces a pushed-forward path that is guaranteed to lie in the  $S^2$  manifold.

It should now be evident that this strategy has some general applicability. Suppose we are given a manifold, call it  $\Gamma$ , perhaps embedded in some larger ambient space, and a first-order differential equation for a path, call it  $\gamma(t)$ , that from the differential equation and the initial condition  $\gamma^0$  can be shown to lie in  $\Gamma$ . Suppose we can find a group, call it  $\mathcal{G}$ , such that  $\mathcal{G}$  acts transitively on  $\Gamma$  so that  $\Gamma$  is a homogeneous space with respect to  $\mathcal{G}$ . Introduce the notation  $G(t)$  to denote a path in  $\mathcal{G}$  with the beginning point

$$G(t^0) = I. \quad (11.2.50)$$

Then this path in  $\mathcal{G}$  may be pushed forward to produce a path  $\gamma(t)$  in  $\Gamma$ . In a notation analogous to that of Subsection 5.12.4, we write

$$\gamma(t) = T_{G(t)}(\gamma^0) \quad (11.2.51)$$

where  $T_G$  describes the action of  $\mathcal{G}$  on  $\Gamma$ . With some suitable parameterization of  $\mathcal{G}$ , perhaps involving an embedding in an ambient space of its own, the differential equation for  $\gamma(t)$  can be pulled back to produce a group differential equation for  $G(t)$ . And if this group differential equation can be integrated numerically in such a way that the group manifold is preserved, say either by parameterizing  $\mathcal{G}$  or its Lie algebra or by integrating in its Lie algebra, then the  $\gamma(t)$  given by (2.51) will be (locally) accurate through terms of order  $h^m$  and is guaranteed to lie in  $\Gamma$ . Thus, whatever means can be found to integrate group differential equations numerically in such a way that the group manifold is preserved, by the same means one has found a procedure to numerically integrate in a manifold preserving way all differential equations defined on the homogeneous spaces associated with  $\mathcal{G}$ .

### 11.2.6 Exploiting an Important Simplification: Lie Taylor Factorization and Lie Taylor Runge Kutta

We observe that there is one way in which the context for (2.43) is simpler than that for (1.18). Namely, in (2.43)  $\bar{\omega}$  is assumed to be a *given* function of  $t$  *independent* of  $S$  whereas in (1.18)  $\omega^{bf}$  must be determined dynamically from the Euler equations (1.19) through (1.21) which themselves may depend on  $R$ . This simplification can be used to good advantage in integrating (2.43) numerically. Of course, since  $S(t)$  is orthogonal when computed exactly, we will want a numerical integrator that guarantees this property for the numerical solution. In this subsection we will see how this simplification can be exploited to perform what we call *Lie Taylor factorization*, and in so doing we will produce in effect a special kind of Runge Kutta that we will call *Lie Taylor Runge Kutta*.

Subsequently, in the next two subsections, we will explore how this requirement that  $\bar{\omega}(t)$  be known in advance can be relaxed in the context of two other special forms of Runge Kutta that we will call *factored Lie Runge Kutta* and *Magnus Lie Runge Kutta*. A final subsection revisits the use of integration in the Lie algebra.

For the purposes of this subsection it is convenient to capitalize on the homomorphism between  $SO(3, \mathbb{R})$  and  $SU(2)$ . Let  $u$  be the  $2 \times 2$  matrix that satisfies the differential equation

$$\dot{u}(t) = (\bar{\omega} \cdot \mathbf{K})u(t) \quad (11.2.52)$$

with the initial condition

$$u(t^0) = I. \quad (11.2.53)$$

By the results of Section 1.3,  $u(t)$  is uniquely defined. Also, it can be shown that  $u(t)$  is in  $SU(2)$  for all  $t$ . See Exercise 2.4. Next, use the homomorphism between  $SO(3, \mathbb{R})$  and  $SU(2)$  to define  $S(t)$  by the rule

$$S_{\alpha\beta}(t) = S_{\alpha\beta}[u(t)] = (1/2)\text{tr}[u^\dagger(t)\sigma^\alpha u(t)\sigma^\beta]. \quad (11.2.54)$$

See (8.2.54). Then it can be shown that  $S(t)$  satisfies (2.43) with the initial condition (2.44). See Exercise 2.5.<sup>21</sup>

Since  $u$  is a  $2 \times 2$  complex matrix, it lives in the ambient space  $\mathbb{C}^4 = E^8$ . However, we know that  $u$  is also in  $SU(2)$  for all  $t$ . Therefore (2.52) is an equation of motion on the three-dimensional manifold  $SU(2)$  imbedded in the ambient space  $E^8$ . Based on the assumption that  $\bar{\omega}$  is a given function of  $t$ , what we seek is a method, beyond those already described, for numerically integrating (2.52) in such a way that  $u$  is guaranteed to remain in  $SU(2)$ .<sup>22</sup> Such a method is presented below.

In the language of Section 2.6, let  $H$  be the duration of a meso integration step, and suppose  $H$  is divided into  $M$  micro steps each of duration  $h = H/M$ . Let  $t^b$  be the time at which a meso step is to be initiated so that we wish to integrate from  $t^b$  to  $t^b + H$ . We also suppose that

$$u^b = u(t^b) \quad (11.2.55)$$

is known and is an element of  $SU(2)$ .

Introduce a relative time  $\tau$  by the rule

$$t = t^b + \tau \quad (11.2.56)$$

so that, in terms of  $\tau$ , we wish to integrate from  $\tau = \tau^0 = 0$  to  $\tau = \tau^M = H$ . Also, define a quantity  $\hat{\omega}(\tau)$  by the rule

$$\hat{\omega}(\tau) = \bar{\omega}(t^b + \tau). \quad (11.2.57)$$

In the spirit of (1.52), write

$$u(t) = u^v(t)u^b \quad (11.2.58)$$

with  $u^v$  being a variable matrix near the identity satisfying

$$u^v(t^b) = I. \quad (11.2.59)$$

---

<sup>21</sup>Why can we move from  $SO(3, \mathbb{R})$  matrices to  $SU(2)$  matrices? It can be shown that the solution to any matrix differential equation of the forms  $dM/dt = AM$  or  $dM/dt = MA$  is governed by the Lie algebra generated by the  $A(t)$  at different times. See the paper by E. Wichmann cited in the Bibliography for Chapter 10. See also Section 10.3 and Appendix C. We may therefore use any set of matrices that obey the Lie algebra in question. In our case this Lie algebra is  $so(3, \mathbb{R})$  or, equivalently,  $su(2)$ . Why should we move from  $SO(3, \mathbb{R})$  to  $SU(2)$ ? It is computationally advantageous to use the matrices that have the lowest dimension. In this case, the matrices with lowest dimension that satisfy  $su(2)$  commutation rules are the  $K^j$ .

<sup>22</sup>By methods “already described” we mean that  $u$ , like  $R$ , can be parameterized in terms of Euler or Tait-Bryan angles or angle-axis or quaternion or Cayley parameters, or can be integrated in its Lie algebra; and so doing produces equations of motion analogous to those for rigid-body motion.

Finally, define a variable matrix  $\hat{u}^v(\tau)$  by the rule

$$\hat{u}^v(\tau) = u^v(t^b + \tau). \quad (11.2.60)$$

It then follows from (2.52), (2.53), and (2.55) through (2.60) that  $\hat{u}^v(\tau)$  obeys the equation of motion

$$d\hat{u}^v(\tau)/d\tau = [\hat{\omega}(\tau) \cdot \mathbf{K}] \hat{u}^v(\tau) \quad (11.2.61)$$

with the initial condition

$$\hat{u}^v(0) = I. \quad (11.2.62)$$

Suppose the  $M + 1$  vectors  $\hat{\omega}^n$  are given with

$$\hat{\omega}^n = \hat{\omega}(\tau^{(n)}) \quad (11.2.63)$$

and

$$\tau^{(n)} = nh \text{ for } n = 0, 1, \dots, M. \quad (11.2.64)$$

We want to use this information to find a numerical approximation to  $\hat{u}^v(H)$  that is both accurate and exactly in  $SU(2)$ . How to proceed? First use the  $M + 1$  vectors  $\hat{\omega}^n$  to produce a polynomial *fit* to  $\hat{\omega}(\tau)$  of the form

$$\hat{\omega}^{\text{fit}}(\tau) = \sum_{m=0}^M \mathbf{c}_m \tau^m. \quad (11.2.65)$$

Here we have introduced the notation  $\tau^{(n)}$  to denote the  $n^{\text{th}}$  sampling point and the notation  $\tau^m$  to denote the  $m^{\text{th}}$  power of  $\tau$ . We also remark that some sampling procedure other than equal spacing could be used to obtain the expansion (2.65). All we need for present purposes are the expansion coefficients  $\mathbf{c}_m$  in (2.65), and need not specify how they are to be obtained.

Next, assume that  $\hat{u}^v(\tau)$  has a *factorized Taylor* approximation of the form

$$\hat{u}^v(\tau) \simeq \hat{u}^{v\text{fac}}(\tau) = \exp(\tau^{M+1} \mathbf{d}_M \cdot \mathbf{K}) \exp(\tau^M \mathbf{d}_{M-1} \cdot \mathbf{K}) \cdots \exp(\tau \mathbf{d}_0 \cdot \mathbf{K}). \quad (11.2.66)$$

Note that this Ansatz satisfies the relation

$$\hat{u}^{v\text{fac}}(0) = I, \quad (11.2.67)$$

as is desirable in view of (2.62).<sup>23</sup> Finally, insert (2.65) and (2.66) into (2.61) to yield the approximate relation

$$d\hat{u}^{v\text{fac}}(\tau)/d\tau = [\hat{\omega}^{\text{fit}}(\tau) \cdot \mathbf{K}] \hat{u}^{v\text{fac}}(\tau), \quad (11.2.68)$$

which can also be written in the form

$$[d\hat{u}^{v\text{fac}}(\tau)/d\tau][\hat{u}^{v\text{fac}}(\tau)]^{-1} = \hat{\omega}^{\text{fit}}(\tau) \cdot \mathbf{K}. \quad (11.2.69)$$

---

<sup>23</sup>The justification for the Ansatz (2.66) is as follows: Under the assumption that  $\hat{\omega}(\tau)$  is analytic in  $\tau$ , the solution  $\hat{u}^v(\tau)$  to (2.61) will be analytic in  $\tau$ . See Section 1.3. It follows that the logarithm of  $\hat{u}^v(\tau)$ , the  $su(2)$  element corresponding to  $\hat{u}^v(\tau)$ , can be expanded as a power series in  $\tau$  assuming  $\hat{u}^v(\tau)$  is near the origin, which it is for small  $\tau$ . See Subsection 3.7.1. Finally, we may pass from a power series in the exponent to a product of exponentials with the aid of the BCH series, thereby yielding the factorization (2.66).

The strategy now is to equate powers of  $\tau$  on both sides of (2.69) to determine the vectors  $\mathbf{d}_n$  in terms of the vectors  $\mathbf{c}_m$ .

As an example, let us see how this procedure plays out for the case  $M = 3$ . Then there are the results

$$\hat{\boldsymbol{\omega}}^{\text{fit}}(\tau) = \mathbf{c}_0 + \mathbf{c}_1\tau + \mathbf{c}_2\tau^2 + \mathbf{c}_3\tau^3, \quad (11.2.70)$$

$$\hat{u}^{v\text{fac}}(\tau) = \exp(\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) \exp(\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(\tau \mathbf{d}_0 \cdot \mathbf{K}), \quad (11.2.71)$$

$$[\hat{u}^{v\text{fac}}(\tau)]^{-1} = \exp(-\tau \mathbf{d}_0 \cdot \mathbf{K}) \exp(-\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(-\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(-\tau^4 \mathbf{d}_3 \cdot \mathbf{K}), \quad (11.2.72)$$

$$\begin{aligned} d\hat{u}^{v\text{fac}}(\tau)/d\tau &= \exp(\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) \exp(\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) (\mathbf{d}_0 \cdot \mathbf{K}) \exp(\tau \mathbf{d}_0 \cdot \mathbf{K}) \\ &\quad + \exp(\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) \exp(\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) (2\tau \mathbf{d}_1 \cdot \mathbf{K}) \exp(\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(\tau \mathbf{d}_0 \cdot \mathbf{K}) \\ &\quad + \exp(\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) (3\tau^2 \mathbf{d}_2 \cdot \mathbf{K}) \exp(\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(\tau \mathbf{d}_0 \cdot \mathbf{K}) \\ &\quad + (4\tau^3 \mathbf{d}_3 \cdot \mathbf{K}) \exp(\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) \exp(\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(\tau \mathbf{d}_0 \cdot \mathbf{K}). \end{aligned} \quad (11.2.73)$$

Next, in view of (2.69), combine (2.72) and (2.73) to yield the result

$$\begin{aligned} &[d\hat{u}^{v\text{fac}}(\tau)/d\tau][\hat{u}^{v\text{fac}}(\tau)]^{-1} = \\ &\exp(\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) \exp(\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) (\mathbf{d}_0 \cdot \mathbf{K}) \exp(\tau \mathbf{d}_0 \cdot \mathbf{K}) \times \\ &\quad \exp(-\tau \mathbf{d}_0 \cdot \mathbf{K}) \exp(-\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(-\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(-\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) \\ &+ \exp(\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) \exp(\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) (2\tau \mathbf{d}_1 \cdot \mathbf{K}) \exp(\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(\tau \mathbf{d}_0 \cdot \mathbf{K}) \times \\ &\quad \exp(-\tau \mathbf{d}_0 \cdot \mathbf{K}) \exp(-\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(-\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(-\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) \\ &+ \exp(\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) (3\tau^2 \mathbf{d}_2 \cdot \mathbf{K}) \exp(\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(\tau \mathbf{d}_0 \cdot \mathbf{K}) \times \\ &\quad \exp(-\tau \mathbf{d}_0 \cdot \mathbf{K}) \exp(-\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(-\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(-\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) \\ &+ (4\tau^3 \mathbf{d}_3 \cdot \mathbf{K}) \exp(\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) \exp(\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(\tau \mathbf{d}_0 \cdot \mathbf{K}) \times \\ &\quad \exp(-\tau \mathbf{d}_0 \cdot \mathbf{K}) \exp(-\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(-\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(-\tau^4 \mathbf{d}_3 \cdot \mathbf{K}). \end{aligned} \quad (11.2.74)$$

After cancellations of various factors in (2.74) against their inverses, (2.74) simplifies to become

$$\begin{aligned} &[d\hat{u}^{v\text{fac}}(\tau)/d\tau][\hat{u}^{v\text{fac}}(\tau)]^{-1} = \\ &\exp(\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) \exp(\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) (\mathbf{d}_0 \cdot \mathbf{K}) \times \\ &\quad \exp(-\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(-\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(-\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) \\ &+ \exp(\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) \exp(\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) (2\tau \mathbf{d}_1 \cdot \mathbf{K}) \exp(-\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(-\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) \\ &\quad + \exp(\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) (3\tau^2 \mathbf{d}_2 \cdot \mathbf{K}) \exp(-\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) \\ &\quad + (4\tau^3 \mathbf{d}_3 \cdot \mathbf{K}). \end{aligned} \quad (11.2.75)$$

Now expand the various terms on the right side of (2.75) as power series in  $\tau$  through terms of order  $\tau^3$ . The first term becomes

$$\begin{aligned} &\exp(\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) \exp(\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) (\mathbf{d}_0 \cdot \mathbf{K}) \times \\ &\quad \exp(-\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(-\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(-\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) = \\ &\exp(\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) (\mathbf{d}_0 \cdot \mathbf{K}) \exp(-\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(-\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) + O(\tau^4). \end{aligned} \quad (11.2.76)$$

Observe that

$$\begin{aligned}\exp(\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) (\mathbf{d}_0 \cdot \mathbf{K}) \exp(-\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) &= \mathbf{d}_0 \cdot \mathbf{K} + \tau^2 \{\mathbf{d}_1 \cdot \mathbf{K}, \mathbf{d}_0 \cdot \mathbf{K}\} + O(\tau^4) \\ &= \mathbf{d}_0 \cdot \mathbf{K} + \tau^2 (\mathbf{d}_1 \times \mathbf{d}_0) \cdot \mathbf{K} + O(\tau^4).\end{aligned}\tag{11.2.77}$$

See (3.7.182) and (8.2.5). It follows that

$$\begin{aligned}\exp(\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) (\mathbf{d}_0 \cdot \mathbf{K}) \exp(-\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(-\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) &= \\ \exp(\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) (\mathbf{d}_0 \cdot \mathbf{K}) \exp(-\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) + \tau^2 (\mathbf{d}_1 \times \mathbf{d}_0) \cdot \mathbf{K} + O(\tau^4) &= \\ \mathbf{d}_0 \cdot \mathbf{K} + \tau^3 \{\mathbf{d}_2 \cdot \mathbf{K}, \mathbf{d}_0 \cdot \mathbf{K}\} + \tau^2 (\mathbf{d}_1 \times \mathbf{d}_0) \cdot \mathbf{K} + O(\tau^4) &= \\ \mathbf{d}_0 \cdot \mathbf{K} + \tau^3 (\mathbf{d}_2 \times \mathbf{d}_0) \cdot \mathbf{K} + \tau^2 (\mathbf{d}_1 \times \mathbf{d}_0) \cdot \mathbf{K} + O(\tau^4).\end{aligned}\tag{11.2.78}$$

The net result is that the first term has the expansion

$$\begin{aligned}\exp(\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) \exp(\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) (\mathbf{d}_0 \cdot \mathbf{K}) \times \\ \exp(-\tau^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(-\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(-\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) = \\ \mathbf{d}_0 \cdot \mathbf{K} + \tau^2 (\mathbf{d}_1 \times \mathbf{d}_0) \cdot \mathbf{K} + \tau^3 (\mathbf{d}_2 \times \mathbf{d}_0) \cdot \mathbf{K} + O(\tau^4).\end{aligned}\tag{11.2.79}$$

What remains is to expand the second and third terms. The second term has the expansion

$$\begin{aligned}\exp(\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) \exp(\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) (2\tau \mathbf{d}_1 \cdot \mathbf{K}) \exp(-\tau^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(-\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) = \\ 2\tau \mathbf{d}_1 \cdot \mathbf{K} + O(\tau^4).\end{aligned}\tag{11.2.80}$$

The third term has the expansion

$$\exp(\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) (3\tau^2 \mathbf{d}_2 \cdot \mathbf{K}) \exp(-\tau^4 \mathbf{d}_3 \cdot \mathbf{K}) = 3\tau^2 \mathbf{d}_2 \cdot \mathbf{K} + O(\tau^4).\tag{11.2.81}$$

The fourth term,  $4\tau^3 \mathbf{d}_3 \cdot \mathbf{K}$ , is already as simple as possible.

Now gather all the terms together. The result, as a power series in  $\tau$ , is that (2.75) becomes

$$\begin{aligned}[d\hat{u}^{vfac}(\tau)/d\tau][\hat{u}^{vfac}(\tau)]^{-1} = \\ \mathbf{d}_0 \cdot \mathbf{K} + 2\tau \mathbf{d}_1 \cdot \mathbf{K} + \tau^2 [3\mathbf{d}_2 + (\mathbf{d}_1 \times \mathbf{d}_0)] \cdot \mathbf{K} + \tau^3 [4\mathbf{d}_3 + (\mathbf{d}_2 \times \mathbf{d}_0)] \cdot \mathbf{K} + O(\tau^4).\end{aligned}\tag{11.2.82}$$

We are ready to equate powers of  $\tau$  on both sides of (2.69). So doing yields the relations

$$\mathbf{d}_0 = \mathbf{c}_0,\tag{11.2.83}$$

$$2\mathbf{d}_1 = \mathbf{c}_1,\tag{11.2.84}$$

$$3\mathbf{d}_2 + (\mathbf{d}_1 \times \mathbf{d}_0) = \mathbf{c}_2, \quad (11.2.85)$$

$$4\mathbf{d}_3 + (\mathbf{d}_2 \times \mathbf{d}_0) = \mathbf{c}_3; \quad (11.2.86)$$

and these relations have the solution

$$\mathbf{d}_0 = \mathbf{c}_0, \quad (11.2.87)$$

$$\mathbf{d}_1 = \mathbf{c}_1/2, \quad (11.2.88)$$

$$\mathbf{d}_2 = (1/3)\mathbf{c}_2 + (1/6)(\mathbf{c}_0 \times \mathbf{c}_1), \quad (11.2.89)$$

$$\mathbf{d}_3 = (1/4)\mathbf{c}_3 + (1/12)(\mathbf{c}_0 \times \mathbf{c}_2) + (1/24)[\mathbf{c}_0 \times (\mathbf{c}_0 \times \mathbf{c}_1)]. \quad (11.2.90)$$

In the case  $M = 3$  we have found the approximation

$$\begin{aligned} \hat{u}^v(H) &\simeq \hat{u}^{v\text{fac}}(H) = \\ \exp(H^4 \mathbf{d}_3 \cdot \mathbf{K}) \exp(H^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(H^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(H \mathbf{d}_0 \cdot \mathbf{K}) \end{aligned} \quad (11.2.91)$$

with the coefficients  $\mathbf{d}_0$  through  $\mathbf{d}_3$  given by (2.87) through (2.90). And in general we have the result

$$\begin{aligned} \hat{u}^v(H) &\simeq \hat{u}^{v\text{fac}}(H) = \\ \cdots \exp(H^{n+1} \mathbf{d}_n \cdot \mathbf{K}) \exp(H^n \mathbf{d}_{n-1} \cdot \mathbf{K}) \cdots \exp(H \mathbf{d}_0 \cdot \mathbf{K}). \end{aligned} \quad (11.2.92)$$

What can be said about the error in this approximation? We begin by observing that any given  $\mathbf{d}_n$  depends only on the  $\mathbf{c}_m$  with  $m \leq n$ , and is independent of  $M$  as long as  $M \geq n$ . Also if we use  $(M+1)$  sampling points to find  $(M+1)$  values of  $\hat{\omega}$ , and use these values to compute  $\mathbf{c}_0$  through  $\mathbf{c}_M$ , then we can find  $\mathbf{d}_0$  through  $\mathbf{d}_M$  using relations of the kind (2.87) through (2.90). With this information we can compute  $\hat{u}^v(H)$  given by

$$\hat{u}^v(H) \simeq \hat{u}^{v\text{fac}}(H) = \exp(H^{M+1} \mathbf{d}_M \cdot \mathbf{K}) \exp(H^M \mathbf{d}_{M-1} \cdot \mathbf{K}) \cdots \exp(H \mathbf{d}_0 \cdot \mathbf{K}), \quad (11.2.93)$$

which is locally accurate through terms of order  $H^{M+1}$ , and exactly in  $SU(2)$ . In effect, we have produced a special kind of Runge Kutta that we will call *Lie Taylor Runge Kutta*. Indeed, in the terminology of Runge-Kutta integration, we may think of  $(M+1)$  as being the number of stages. Thus, the local accuracy of  $\hat{u}^v(H)$  equals the number of stages. Comparison of this result with the entries of Table 2.3.1 shows that this performance of Lie Taylor Runge Kutta equals or exceeds that of ordinary explicit Runge-Kutta; and reference to (T.166) and (T.169) shows that Lie Taylor Runge Kutta has one order lower accuracy than Newton Cotes for odd values of  $M+1$ , and equal accuracy for even values of  $M+1$ .<sup>24</sup>

There is one other observation that is worth consideration. In determining the  $\mathbf{c}_0$  through  $\mathbf{c}_M$  there is no need for the sampling points to lie within the interval of meso-step integration as was done in (2.64). Suppose, for example, that we wish to integrate from the time  $t = t^0$  to the time  $t = t^0 + T$  using  $N$  meso steps each of duration

$$H = T/N. \quad (11.2.94)$$

---

<sup>24</sup>Note that here we are demanding more than the integration of an equation of the form (T.157) since in this instance the right side of (2.52) depends on  $u$  as well as  $t$ .

Also suppose that over the *full* interval  $t \in [t^0, t^0 + T]$  the vector function  $\bar{\omega}(t)$  can be well fit by a polynomial of degree  $M$  in  $t$  with vector coefficients. Under this assumption, the coefficients of this polynomial can be obtained by evaluating  $\bar{\omega}(t)$  at  $(M+1)$  sampling points. Use this global polynomial to form the local expansion (2.65) for each meso-step integration, and use a relation of the form (2.93) to find the result for each meso-step integration. Then the local error for each meso step is of order  $H^{M+2}$ , and the global error in integrating from the time  $t = t^0$  to the time  $t = t^0 + T$  is given by

$$\text{global error} \approx NH^{M+2} = N(T/N)^{M+2} = T(T/N)^{M+1}. \quad (11.2.95)$$

We see that the global *integration* error can be made arbitrarily small by increasing  $N$  (and, correspondingly, decreasing  $H$ ) *without* changing the number of sampling points ( $M+1$ ). Put another way, for sufficiently large  $N$ , the full global error is *only* the error associated with making the global polynomial fit to  $\bar{\omega}(t)$  in the full interval  $t \in [t^0, t^0 + T]$ . The goodness of this polynomial fit in turn depends only on the analytic properties of  $\bar{\omega}(t)$ . In particular, a good fit is easiest to achieve when  $\bar{\omega}(t)$  does not vary too rapidly over the interval  $[t^0, t^0 + T]$ .<sup>25</sup> Finally, we may truncate (2.93) at  $M'$  with  $M' < M$  and still achieve convergence, but at the slower rate of

$$\text{global error} \approx NH^{M'+2} = N(T/N)^{M'+2} = T(T/N)^{M'+1}. \quad (11.2.96)$$

In this case we only need to work out formulas for the  $\mathbf{d}_n$  with  $n \leq M'$

Where applicable, employing the ideas just described should result in a substantial savings in computer time because only  $(M+1)$  evaluations of  $\bar{\omega}(t)$  are required for the full integration run.

### 11.2.7 Factored Lie Runge Kutta

#### Purpose, Motivation, and Plan

The purpose of this subsection is to describe a special form of Runge Kutta designed to preserve group properties. We will see that for integrating equations of the form (2.43), unlike Lie Taylor Runge Kutta, it does not require the values of  $\bar{\omega}(t)$  in advance.

By way of motivation, suppose we seek to integrate (2.43) by the simplest Runge-Kutta method, namely the crude Euler method of Section 2.2. Doing so yields the stepping formula

$$S^{n+1} = S^n + h\dot{S}^n = S^n + h[\bar{\omega}(t^n) \cdot \mathbf{L}]S^n = \{I + h[\bar{\omega}(t^n) \cdot \mathbf{L}]\}S^n. \quad (11.2.97)$$

Observe that  $\{I + h[\bar{\omega}(t^n) \cdot \mathbf{L}]\}$  is generally not an orthogonal matrix. Consequently,  $S^{n+1}$  will generally not be orthogonal even if  $S^n$  is. Consider, instead, the modified stepping formula

$$S^{n+1} = \exp[h\bar{\omega}(t^n) \cdot \mathbf{L}]S^n. \quad (11.2.98)$$

As can be seen by expanding  $\exp[h\bar{\omega}(t^n) \cdot \mathbf{L}]$  in powers of  $h$ , (2.98) and (2.97) agree through terms of order  $h$ . Therefore, use of (2.98) provides an integration algorithm that is of the

---

<sup>25</sup>More precisely,  $\bar{\omega}(t)$  needs to be analytic in the complex  $t$  plane in a disk of radius  $T/2$  and centered on  $t = t^0 + T/2$ .

same order as the Euler method (2.97). However the algorithm (2.98), even though (like Euler) it makes local errors of order  $h^2$ , preserves  $SO(3, \mathbb{R})$  exactly because  $\exp[h\bar{\omega}(t^n) \cdot \mathbf{L}]$  is orthogonal.

The relation (2.98) provides an example of what we call a *factored Lie Runge-Kutta* algorithm designed to preserve some group, in this case  $SO(3, \mathbb{R})$ .<sup>26</sup> We now discuss the possibility of finding such algorithms that are of order  $h^m$  with  $m > 1$ .

The equation (2.43) is a special case of a more general equation, and some results are known about factored Lie Runge-Kutta for this more general equation. Our plan is to discuss this more general equation, and then apply the known results for the more general equation to the special case (2.43).

### Factored Lie Runge Kutta

Let  $G$  be some Lie group of  $n \times n$  matrices, and let  $Y$  denote matrices in  $G$ . Next assume that there is some  $n \times n$  matrix function  $A(Y, t)$  such that  $A(Y, t)$  is in the Lie algebra of  $G$  for all  $Y \in G$  and all  $t$ . Let  $t^0$  be some initial time and let  $Y^0$  be some initial matrix in  $G$ . Consider the matrix differential equation

$$\dot{Y}(t) = A(Y, t)Y(t). \quad (11.2.99)$$

Then it can be shown that the solution to (2.99) lies in  $G$  for all time.<sup>27</sup> See Exercise 2.8. Comparison of (2.43) and (2.99) shows that (2.43) is a special case of (2.99) with  $S$  playing the role of  $Y$  and  $[\bar{\omega}(t) \cdot \mathbf{L}]$  playing the role of  $A(Y, t)$ :

$$S \leftrightarrow Y, \quad (11.2.100)$$

$$[\bar{\omega}(t) \cdot \mathbf{L}] \leftrightarrow A(Y, t). \quad (11.2.101)$$

Note that for the special case (2.43) the matrices  $A(Y, t)$  are in fact *independent* of  $Y$ .

*Crouch, Grossman*, and others have developed factored Lie Runge-Kutta algorithms for the numerical integration of (2.99). These algorithms are constructed in such a way that, although they may make local errors of order  $h^{m+1}$ ,  $Y(t)$  is guaranteed to lie in  $G$  to machine precision and evaluations of  $A(Y, t)$  are required only for matrices  $Y$  in  $G$ .

Applying crude Euler to (2.99) produces the stepping rule

$$Y^{n+1} = Y^n + h\dot{Y}^n = [I + hA(Y^n, t^n)]Y^n \quad (11.2.102)$$

which, to the same order in  $h$ , can be rewritten in the exponential form

$$Y^{n+1} = \{\exp[hA(Y^n, t^n)]\}Y^n. \quad (11.2.103)$$

Suppose  $Y^n \in G$ . Then, by assumption,  $A(Y^n, t^n)$  is in the Lie algebra of  $G$ , from which it follows that  $\{\exp[hA(Y^n, t^n)]\} \in G$ , and therefore  $Y^{n+1} \in G$ . The stepping rule (2.103) preserves  $G$ .

---

<sup>26</sup>The significance of the adjective *factored* will become apparent subsequently. See (2.113).

<sup>27</sup>Note that there is a consistency consideration here. If the solution  $Y(t)$  were to leave  $G$  even though the initial matrix  $Y^0$  is in  $G$ , then  $A(Y, t)$  could become undefined.

Consider a single-stage Butcher tableau of the form

$$\begin{array}{c|c} c_1 & a_{11} \\ \hline & b_1 \end{array}. \quad (11.2.104)$$

It describes crude Euler when  $c_1 = a_{11} = 0$  and  $b_1 = 1$ . See Exercise 2.3.3. Define quantities  $Y_1^n$  and  $K_1$  by the rules

$$Y_1^n = Y^n, \quad (11.2.105)$$

$$K_1 = A(Y_1^n, t^n + hc_1) = A(Y_1^n, t^n) = A(Y^n, t^n), \quad (11.2.106)$$

Then we see that (2.103) can be written in the form

$$Y^{n+1} = \exp(hb_1 K_1) Y^n. \quad (11.2.107)$$

Thus, there is a correspondence between (2.103) and the Butcher tableau for crude Euler.

Next consider a two-stage Butcher tableau of the form

$$\begin{array}{c|cc} c_1 & 0 & 0 \\ \hline c_2 & a_{21} & 0 \\ \hline & b_1 & b_2 \end{array}. \quad (11.2.108)$$

Note that the matrix  $a$  is strictly lower triangular, and therefore the associated Runge-Kutta method is explicit, as is the single-stage method specified by (2.104) when  $a_{11} = 0$ . Also, we continue to enforce the consistency condition (2.3.16) so that, in fact,  $c_1 = 0$ . Corresponding to the Butcher tableau (2.108), consider the following rule for stepping from  $Y^n$  to  $Y^{n+1}$ :

$$Y_1^n = Y^n, \quad (11.2.109)$$

$$K_1 = A(Y_1^n, t^n + hc_1) = A(Y_1^n, t^n) = A(Y^n, t^n), \quad (11.2.110)$$

$$Y_2^n = [\exp(ha_{21} K_1)] Y^n, \quad (11.2.111)$$

$$K_2 = A(Y_2^n, t^n + hc_2), \quad (11.2.112)$$

$$Y^{n+1} = [\exp(hb_2 K_2) \exp(hb_1 K_1)] Y^n. \quad (11.2.113)$$

Observe that the term appearing in square brackets on the right side of (2.113) is *faktored* into a product of group elements, each in exponential form.

Let us examine the ingredients in this rule. Assuming  $Y^n \in G$ , we see from (2.109) that  $Y_1^n \in G$ . Next, we see from (2.110) that  $K_1$  is in the Lie algebra of  $G$ . Now look at (2.111). Since  $K_1$  is in the Lie algebra,  $[\exp(ha_{21} K_1)]$  is in  $G$ , and therefore  $Y_2^n \in G$ . With regard to (2.112), we see that the arguments of  $A$  are in its domain of definition, and therefore  $K_2$  is well defined and in the Lie algebra of  $G$ . Finally, examination of (2.113) shows that  $Y^{n+1} \in G$ . Our goal has at least been partially achieved: Starting from  $Y^0 \in G$ , we have produced a sequence of matrices  $Y^1, Y^2, \dots$ , all of which are in  $G$  to machine precision.

What about the error associated with this rule? It can be shown that this algorithm is locally correct through terms of order  $h^2$  (local error of order  $h^3$ ) if the coefficients  $a$ ,  $b$ , and  $c$  satisfy the consistency condition (2.3.16) and the order conditions (2.3.42), and (2.3.43). Thus, in the case of explicit one and two-stage methods, the order conditions on the Butcher

tableau and the local accuracy for factored Lie Runge-Kutta are the same as those for the ordinary Runge-Kutta methods of Chapter 2.

To continue our discussion, consider a three-stage explicit Butcher tableau of the form

$$\begin{array}{c|ccc} c_1 & 0 & 0 & 0 \\ c_2 & a_{21} & 0 & 0 \\ c_3 & a_{31} & a_{32} & 0 \\ \hline b_1 & b_2 & b_3 \end{array} \quad (11.2.114)$$

Corresponding to the Butcher tableau (2.114), we make the following rule for stepping from  $Y^n$  to  $Y^{n+1}$ :

$$Y_1^n = Y^n, \quad (11.2.115)$$

$$K_1 = A(Y_1^n, t^n + hc_1) = A(Y_1^n, t^n) = A(Y^n, t^n), \quad (11.2.116)$$

$$Y_2^n = [\exp(ha_{21}K_1)]Y^n, \quad (11.2.117)$$

$$K_2 = A(Y_2^n, t^n + hc_2), \quad (11.2.118)$$

$$Y_3^n = [\exp(ha_{32}K_2)][\exp(ha_{31}K_1)]Y^n, \quad (11.2.119)$$

$$K_3 = A(Y_3^n, t^n + hc_3), \quad (11.2.120)$$

$$Y^{n+1} = \exp(hb_3K_3) \exp(hb_2K_2) \exp(hb_1K_1)Y^n. \quad (11.2.121)$$

Again we see that the  $Y_j^n$  are in  $G$ , the arguments of  $A$  are in its domain of definition, and consequently the  $K_j$  are well defined and in the Lie algebra of  $G$ . And examination of (2.121) shows that therefore  $Y^{n+1} \in G$ . We conclude that our goal has again at least been partially achieved.

What about the error associated with this rule? It can be shown that this algorithm is locally correct through terms of order  $h^3$  (local error of order  $h^4$ ) if the Butcher tableau satisfies the consistency condition (2.3.16), the order conditions (2.3.42) through (2.3.45), and the *additional* order condition

### Additional Order 3:

$$\sum_i b_i^2 c_i + 2 \sum_{i < j} b_i c_i b_j = 1/3. \quad (11.2.122)$$

We know that the consistency condition (2.3.16) and the order conditions (2.3.42) through (2.3.45) are necessary and sufficient for the three-stage explicit ordinary Runge-Kutta methods of Chapter 2 to be locally accurate through terms of order  $h^3$ . Because of the additional order condition (2.122), more is required for the case of factored Lie Runge-Kutta to achieve a local accuracy through terms of order  $h^3$ . Fortunately, there are three-stage explicit Butcher tableaux that meet the requirements (2.3.16), (2.3.42) through (2.3.45), and (2.122). Two such Butcher tableaux, found by Crouch and Grossman, are given below:

$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ -1/24 & -1/24 & 0 & 0 \\ 17/24 & 161/24 & -6 & 0 \\ \hline 1 & -2/3 & 2/3 \end{array}, \quad (11.2.123)$$

0	0	0	0	
3/4	3/4	0	0	
17/24	119/216	17/108	0	
	13/51	-2/3	24/17	

(11.2.124)

Note the curious feature that use of Butcher tableau (2.123) entails the evaluation of  $A$  outside the temporal interval  $[t^n, t^n + h]$ , the interval over which integration is being performed, because for this tableau  $c_2 < 0$ .

At this point we are prepared to state the general recipe for factored Lie Runge-Kutta methods. We have already discussed the cases of  $s = 1$  or  $s = 2$  or  $s = 3$  stages. Consider a Butcher tableau with  $s$  stages with  $s > 3$  and again suppose that the matrix  $a$  is strictly lower triangular. Use this tableau to make the following stepping rule: For  $1 \leq j \leq 3$  define, as before,  $Y_j^n$  and  $K_j$  by the rules (2.115) through (2.120). And for  $4 \leq j \leq s$  make the definitions

$$Y_j^n = \exp(ha_{j,j-1}K_{j-1}) \exp(ha_{j,j-2}K_{j-2}) \cdot \dots \cdot \exp(ha_{j,1}K_1)Y^n, \quad (11.2.125)$$

$$K_j = A(Y_j^n, t^n + hc_j). \quad (11.2.126)$$

Finally, step from  $Y^n$  to  $Y^{n+1}$  using the rule

$$Y^{n+1} = \exp(hb_s K_s) \exp(hb_{s-1} K_{s-1}) \cdot \dots \cdot \exp(hb_1 K_1) Y^n. \quad (11.2.127)$$

What can be said about the error in this case? For an optimum choice of coefficients, how many stages are required to achieve order  $m$ ? That is, what is the analog of Table 2.3.1 for the case of factored Lie Runge-Kutta methods? This is a difficult question. For factored Lie Runge-Kutta and for  $m \geq 3$ , compared to the ordinary Runge-Kutta methods of Chapter 2, there are many more conditions that the entries in the Butcher tableau must meet to achieve order  $m$ . For example, to achieve  $m = 4$  there are 5 more order conditions for factored Lie Runge-Kutta compared to ordinary Runge-Kutta. And, unlike ordinary Runge-Kutta, it is impossible with only 4 stages to satisfy all the factored Lie Runge-Kutta order conditions required to achieve  $m = 4$ . At least  $s = 5$  stages are required for factored Lie Runge-Kutta to achieve order  $m = 4$ , some 5-stage Butcher tableaux with this property have been obtained by *Owren and Marthinsen*, and they have published one of them. Finally, It is believed that the minimum number of stages required for factored Lie Runge-Kutta to achieve an order  $m$  with  $m > 4$  grows rapidly with increasing  $m$ .

### Application of Factored Lie Runge-Kutta

As described at the beginning of this subsection, our goal is to find higher-order versions of (2.98). This is now easily done based on what we have learned of factored Lie Runge-Kutta. From (2.101) and (2.126) we see for the case of (2.43) that we may write

$$K_j \leftrightarrow \bar{\omega}(t^n + hc_j) \cdot \mathbf{L}. \quad (11.2.128)$$

And, using (2.127), we see that there is the stepping rule

$$\begin{aligned} S^{n+1} = & \exp[hb_s \bar{\omega}(t^n + hc_s) \cdot \mathbf{L}] \exp[hb_{s-1} \bar{\omega}(t^n + hc_{s-1}) \cdot \mathbf{L}] \times \\ & \dots \times \exp[hb_1 \bar{\omega}(t^n + hc_1) \cdot \mathbf{L}] S^n. \end{aligned} \quad (11.2.129)$$

We have achieved our objective. Given an  $s$ -stage factored Lie Runge-Kutta method of order  $m$ , (2.129) provides a stepping rule that may make local errors of order  $h^{m+1}$ , but is guaranteed to preserve  $SO(3, \mathbb{R})$  to machine precision.

For example, based on the Butcher tableau (2.124), there is the third-order (but exactly orthogonality-preserving) stepping rule

$$S^{n+1} = \exp\{(24/17)h\bar{\omega}[t^n + (17/24)h] \cdot \mathbf{L}\} \exp\{-(2/3)h\bar{\omega}[t^n + (3/4)h] \cdot \mathbf{L}\} \times \exp\{(13/51)h\bar{\omega}[t^n] \cdot \mathbf{L}\} S^n. \quad (11.2.130)$$

We close this subsection with two comments. The first is based on the observation that the Butcher tableau for a factored Lie Runge-Kutta method can also be used as a Butcher tableau for an ordinary Runge-Kutta method. This result follows because the order conditions for factored Lie Runge-Kutta contain as a subset all the order conditions for ordinary Runge-Kutta. Therefore if a factored Lie Runge-Kutta method is used to track particle spin, which amounts to use of the rule (2.129) to track particle spin, then the same algorithm (the same Butcher tableau) could be used for an ordinary Runge-Kutta routine to compute the particle trajectory. In this way, the same times ( $t^n + hc_j$ ) would occur in both the spin and particle trajectory routines, thus facilitating the computation of the required quantities  $\bar{\omega}(t^n + hc_j)$ .

There is a corollary to this observation. In the previous subsection  $\bar{\omega}(t)$  was assumed to be a *given* function of  $t$ . With the use of factored Lie Runge-Kutta this assumption is no longer necessary since there is no need with this method for an explicit fit of the form (2.65) with known coefficients. All that is required at each step are the sampling-point values  $\bar{\omega}(t^n + hc_j)$ . This is true even if the  $\bar{\omega}(t^n + hc_j)$  need be determined dynamically.

At this point it is tempting to imagine that this approach could be applied to the case of spin if there were spin-orbit coupling (Stern-Gerlach effect) so that the equations of motion for the particle trajectory could be visualized as depending on  $S$  as well as the particle's position and momentum. However, there are quantum-mechanical reasons why this approach is not applicable.

The Stern-Gerlach effect is an example of *quantum entanglement*. Conceptually, and in a fully quantum treatment, when entering a Stern-Gerlach apparatus a single particle (assumed to have spin 1/2) is described by an initial state vector that is the tensor product of a spin state eigenvector (an eigenvector with eigenvalue +1 for  $\mathbf{n} \cdot \boldsymbol{\sigma}$  for some specified unit vector  $\mathbf{n}$ ) and an orbital state vector for a wave packet well localized (consistent with the uncertainty principle) both in position and momentum about some initial point  $z^i$  in phase space. After passing through the Stern-Gerlach apparatus the particle is no longer described by a product state vector. That is, the outgoing final quantum state vector *cannot* be written in tensor product form. Rather, it is described by a superposition of two vectors, each expressible in tensor product form. The first vector has a spin state that is an eigenvector with eigenvalue +1 for  $\mathbf{m} \cdot \boldsymbol{\sigma}$  for some specified unit vector  $\mathbf{m}$  that is determined by the orientation of the Stern-Gerlach apparatus, and an orbital part consisting of a well localized packet about some final point  $z^{f+}$ . The second vector has a spin state that is an eigenvector of  $\mathbf{m} \cdot \boldsymbol{\sigma}$  with eigenvalue -1 and an orbital part consisting of a well localized packet about some final point  $z^{f-}$ . Moreover, if the Stern-Gerlach experiment is a success, the points  $z^{f+}$  and  $z^{f-}$  are sufficiently separated and the associated wave packets

sufficiently localized so that there is little overlap between them. Finally, there is a definite phase relation between the two vectors. This state of affairs has no classical analog, and therefore cannot be treated classically.

We also remark that the Stern-Gerlach experiment, which is intended to separate a (single) beam of particles (with different particles characterized by different initial vectors  $\mathbf{n}$ ) into two well-separated beams with the first consisting of particles in spin eigenstates characterized by  $\mathbf{m}$  and the second consisting of particles in spin eigenstates characterized by  $-\mathbf{m}$ , may not necessarily be possible for all kinds of particles. The original Stern-Gerlach experiment was performed with silver atoms that are both *heavy* and were *neutral* (not ionized). There is an argument, due to Bohr and generalized by others, to the effect that it is not possible to magnetically separate a beam, consisting of particles with various spin orientations, into two beams with specified polarizations if the particles are *charged* and too light, or have too small a magnetic moment. For example, it is argued that it is not possible to achieve a Stern-Gerlach effect with electrons.

The essence of the argument, which is semiclassical, is as follows: Particles that are both charged and have a magnetic moment experience two forces when they move in a magnetic field, a Lorentz force due to the field itself and a Stern-Gerlach force due to field inhomogeneities. Since the magnetic field must be divergence free, desirable inhomogeneity (inhomogeneity transverse to the beam direction and in the direction of the main field) employed to produce a Stern-Gerlach force must lead to undesirable inhomogeneity in the other transverse direction. This undesirable field inhomogeneity leads to an undesirable Lorentz force that tends to spread the beam. Therefore the beam must be made small so that only small field variations are actually encountered by particles in the beam. But then, by the quantum uncertainty principle, there will be a corresponding spread in velocity space. This spread in velocity space again leads to an uncertainty in the Lorentz force. It may happen, if the particle mass is too small (thereby leading to a very large spread in velocity space and a corresponding large uncertainty in the Lorentz force) or if the magnetic moment is too small, that the uncertainty in the Lorentz force exceeds the Stern-Gerlach force. The net effect then (if the argument is to be believed) is that only a single final beam is produced whose spread due to the quantum-related uncertainty in the Lorentz force exceeds the splitting expected from the Stern-Gerlach force, thereby washing out any anticipated Stern-Gerlach effect.

The moral to be drawn from these considerations is that a full quantum treatment is required to find reliably the complete effect of an inhomogeneous magnetic field on a beam of particles that are charged and possibly too light or have too small a magnetic moment. See the references to the Stern-Gerlach effect at the end of this chapter.<sup>28</sup>

Although the Stern-Gerlach effect cannot be treated classically, there is a classical problem that is somewhat analogous, namely that of rigid-body motion, for which factored Lie Runge Kutta can be used to good advantage. As described earlier, in the case of rigid-body motion there is the complication that the  $\omega^{bf}$  must be determined dynamically from the Euler equations (1.19) through (1.21), which themselves may depend on  $R$ . This complication

---

<sup>28</sup>In fact, the semiclassical arguments that are used to describe the Stern-Gerlach effect even in the case of a neutral beam are suspect because they do not include beam spreading due to the uncertainty principle, the Stern-Gerlach force in the other transverse direction due to the undesirable field inhomogeneity, and the precession of the magnetic moment in the main field. See Exercise 2.17.

causes no problem if both the kinematic equations (1.18) and the dynamic equations (1.19) through (1.21) are integrated using the same factored Lie Runge-Kutta Butcher tableau. In fact, all that is really required is that both the kinematic equations (1.18) and the dynamic equations (1.19) through (1.21) be integrated using Runge-Kutta Butcher tableaux that have the same sampling times, i.e. the same vector  $c$ . However, it is important to recognize that the integration method(s) must be capable, as factored Lie Runge-Kutta is, of producing the intermediate values  $Y_j^n$  so that values of  $R$  are available when needed in the various Runge-Kutta stages.

The second comment is that the  $s = 5$  and  $m = 4$  factored Lie Runge-Kutta Butcher tableau published by Owren and Marthinsen has for its  $b$  and  $c$  entries the values

$$b_1 = (1 + \kappa + \kappa^2)/(2\kappa + 2\kappa^2), \quad (11.2.131)$$

$$b_2 = 0, \quad (11.2.132)$$

$$b_3 = -(1 + 2\kappa + \kappa^2)/(12 + 6\kappa + 6\kappa^2), \quad (11.2.133)$$

$$b_4 = -1/(2\kappa + 2\kappa^2), \quad (11.2.134)$$

$$b_5 = b_1 = (1 + \kappa + \kappa^2)/(2\kappa + 2\kappa^2), \quad (11.2.135)$$

$$c_1 = 0, \quad (11.2.136)$$

$$c_2 = 3/2, \quad (11.2.137)$$

$$c_3 = 2/3 + \kappa/3 + \kappa^2/6 \simeq * \dots, \quad (11.2.138)$$

$$c_4 = 1/3 - \kappa/3 - \kappa^2/6 \simeq - * \dots, \quad (11.2.139)$$

$$c_5 = 1, \quad (11.2.140)$$

where

$$\kappa = 2^{1/3}. \quad (11.2.141)$$

Note that  $b_2 = 0$ . Thus, when (2.129) is used in this case as a stand-alone formula, it is effectively a *four* stage ( $s = 4$ )  $m = 4$  formula. Observe, however, that  $c_3 > 1$ . Therefore use of this Butcher tableau involves an evaluation of  $\bar{\omega}(t)$  outside the interval  $[t^n, t^n + h]$ . One might also worry that  $c_2 > 1$ , which also leads to  $t$  values outside  $[t^n, t^n + h]$ . But, since  $b_2 = 0$ , this is not a concern. Finally, observe that  $c_4 < 0$ , so a second evaluation of  $\bar{\omega}(t)$  outside the interval  $[t^n, t^n + h]$  is also required.

### 11.2.8 Magnus Lie Runge Kutta

The use of factored Lie Runge Kutta is not particularly attractive for our purposes because only relatively low-order results are available and because sometimes some evaluation points lie outside the interval  $[t^n, t^{n+1}]$ . But factored Lie Runge Kutta is designed to handle the case (2.99) for which  $A$  is allowed to depend on  $Y$ . What happens if we relax this requirement, and consider *only* equations of the simpler form

$$\dot{Y}(t) = A(t)Y(t)? \quad (11.2.142)$$

Note that, in view of (2.101), our problem of particular interest, namely (2.61), is of this simpler form.<sup>29</sup> We will learn that for the case (2.142) there are far more attractive results.

Our approach will be to *combine* all the exponents in the Lie Taylor factorization (2.93) to write  $\hat{u}^{v^{\text{fac}}}(H)$  in *single-exponent* form. For example, in the case  $M = 3$ , suppose we write

$$\begin{aligned}\hat{u}^v(H) &\simeq \hat{u}^{v^{\text{fac}}}(H) = \exp(H^4 \mathbf{d}_3 \cdot \mathbf{K}) \exp(H^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(H^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(H \mathbf{d}_0 \cdot \mathbf{K}) \\ &\simeq \exp[G(H)]\end{aligned}\tag{11.2.143}$$

where

$$G(H) = H^4 \mathbf{e}_3 \cdot \mathbf{K} + H^3 \mathbf{e}_2 \cdot \mathbf{K} + H^2 \mathbf{e}_1 \cdot \mathbf{K} + H \mathbf{e}_0 \cdot \mathbf{K}.\tag{11.2.144}$$

Since the use of a single-exponent representation is in the spirit of the Magnus equations and our results will eventually be cast in Runge-Kutta form, we will refer to this procedure as *Magnus Lie Runge Kutta*. Indeed, the quantities  $\mathbf{e}_n$  could be found in terms of the  $\mathbf{c}_m$  by integrating the Magnus equations. See Section 10.3. Equivalently, they could be found by making temporal Taylor expansions of equations of the forms (1.67) or (1.83) and equating like powers of  $t$ .

Alternatively, since we already know the  $\mathbf{d}_n$ , as a first step we can convert the left side of (2.143) to the right side of (2.143) using the BCH formula (3.7.41). So doing will provide the  $\mathbf{e}_n$  in terms of the  $\mathbf{d}_m$ . Also, according to (2.87) through (2.90), the  $\mathbf{d}_m$  are already known in terms of the  $\mathbf{c}_\ell$ . Therefore, in a second step, we can find the  $\mathbf{e}_n$  in terms of the  $\mathbf{c}_m$  by simple algebraic substitution. We will now carry out this task.

Begin by observing that

$$\exp(H^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(H \mathbf{d}_0 \cdot \mathbf{K}) \simeq \exp(E)\tag{11.2.145}$$

with

$$\begin{aligned}E &= H^2 \mathbf{d}_1 \cdot \mathbf{K} + H \mathbf{d}_0 \cdot \mathbf{K} + (1/2)H^3 \{\mathbf{d}_1 \cdot \mathbf{K}, \mathbf{d}_0 \cdot \mathbf{K}\} \\ &\quad + (1/12)H^4 \{\mathbf{d}_0 \cdot \mathbf{K}, \{\mathbf{d}_0 \cdot \mathbf{K}, \mathbf{d}_1 \cdot \mathbf{K}\}\} \\ &= H \mathbf{d}_0 \cdot \mathbf{K} + H^2 \mathbf{d}_1 \cdot \mathbf{K} + (1/2)H^3 \{\mathbf{d}_1 \cdot \mathbf{K}, \mathbf{d}_0 \cdot \mathbf{K}\} \\ &\quad + (1/12)H^4 \{\mathbf{d}_0 \cdot \mathbf{K}, \{\mathbf{d}_0 \cdot \mathbf{K}, \mathbf{d}_1 \cdot \mathbf{K}\}\} \\ &= H \mathbf{d}_0 \cdot \mathbf{K} + H^2 \mathbf{d}_1 \cdot \mathbf{K} + (1/2)H^3 (\mathbf{d}_1 \times \mathbf{d}_0) \cdot \mathbf{K} \\ &\quad + (1/12)H^4 [\mathbf{d}_0 \times (\mathbf{d}_0 \times \mathbf{d}_1)] \cdot \mathbf{K}.\end{aligned}\tag{11.2.146}$$

Next we find that

$$\begin{aligned}\exp(H^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(H^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(H \mathbf{d}_0 \cdot \mathbf{K}) &\simeq \\ \exp(H^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(E) &= \exp(F)\end{aligned}\tag{11.2.147}$$

---

<sup>29</sup>Note also that (10.4.28) is of this form.

with

$$\begin{aligned}
 F &\simeq H^3 \mathbf{d}_2 \cdot \mathbf{K} + E + (1/2)\{H^3 \mathbf{d}_2 \cdot \mathbf{K}, E\} \simeq \\
 &H^3 \mathbf{d}_2 \cdot \mathbf{K} + E + (1/2)H^4 \{\mathbf{d}_2 \cdot \mathbf{K}, \mathbf{d}_0 \cdot \mathbf{K}\} = \\
 &H^3 \mathbf{d}_2 \cdot \mathbf{K} + E + (1/2)H^4 (\mathbf{d}_2 \times \mathbf{d}_0) \cdot \mathbf{K} = \\
 &H \mathbf{d}_0 \cdot \mathbf{K} + H^2 \mathbf{d}_1 \cdot \mathbf{K} + H^3 [\mathbf{d}_2 \cdot \mathbf{K} + (1/2)(\mathbf{d}_1 \times \mathbf{d}_0) \cdot \mathbf{K}] \\
 &+ H^4 \{(1/2)(\mathbf{d}_2 \times \mathbf{d}_0) \cdot \mathbf{K} + (1/12)[\mathbf{d}_0 \times (\mathbf{d}_0 \times \mathbf{d}_1)] \cdot \mathbf{K}\}.
 \end{aligned} \tag{11.2.148}$$

Finally, we see that

$$\begin{aligned}
 \exp(H^4 \mathbf{d}_3 \cdot \mathbf{K}) \exp(H^3 \mathbf{d}_2 \cdot \mathbf{K}) \exp(H^2 \mathbf{d}_1 \cdot \mathbf{K}) \exp(H \mathbf{d}_0 \cdot \mathbf{K}) \simeq \\
 \exp(H^4 \mathbf{d}_3 \cdot \mathbf{K}) \exp(F) = \exp(G)
 \end{aligned} \tag{11.2.149}$$

with

$$G \simeq H^4 \mathbf{d}_3 \cdot \mathbf{K} + F. \tag{11.2.150}$$

The net result, through terms of order  $H^4$ , is that

$$\begin{aligned}
 G &= H \mathbf{d}_0 \cdot \mathbf{K} + H^2 \mathbf{d}_1 \cdot \mathbf{K} + H^3 [\mathbf{d}_2 \cdot \mathbf{K} + (1/2)(\mathbf{d}_1 \times \mathbf{d}_0) \cdot \mathbf{K}] \\
 &+ H^4 \{\mathbf{d}_3 \cdot \mathbf{K} + (1/2)(\mathbf{d}_2 \times \mathbf{d}_0) \cdot \mathbf{K} + (1/12)[\mathbf{d}_0 \times (\mathbf{d}_0 \times \mathbf{d}_1)] \cdot \mathbf{K}\}.
 \end{aligned} \tag{11.2.151}$$

Upon comparing (2.144) and (2.151), we conclude that there are the relations

$$\mathbf{e}_0 = \mathbf{d}_0, \tag{11.2.152}$$

$$\mathbf{e}_1 = \mathbf{d}_1, \tag{11.2.153}$$

$$\mathbf{e}_2 = \mathbf{d}_2 + (1/2)(\mathbf{d}_1 \times \mathbf{d}_0), \tag{11.2.154}$$

$$\mathbf{e}_3 = \mathbf{d}_3 + (1/2)(\mathbf{d}_2 \times \mathbf{d}_0) + (1/12)[\mathbf{d}_0 \times (\mathbf{d}_0 \times \mathbf{d}_1)]. \tag{11.2.155}$$

The first step is complete. To finish our task, we employ the relations (2.87) through (2.90) in the relations (2.152) through (2.155) to find the results

$$\mathbf{e}_0 = \mathbf{c}_0, \tag{11.2.156}$$

$$\mathbf{e}_1 = \mathbf{c}_1/2, \tag{11.2.157}$$

$$\mathbf{e}_2 = (1/3)\mathbf{c}_2 - (1/12)(\mathbf{c}_0 \times \mathbf{c}_1), \tag{11.2.158}$$

$$\mathbf{e}_3 = (1/4)\mathbf{c}_3 - (1/12)(\mathbf{c}_0 \times \mathbf{c}_2). \tag{11.2.159}$$

Remarkably, although there are double cross products in the intermediate results (2.90) and (2.155), there is (due to cancelations) no double cross product in the final results (2.156) through (2.159).

And there is a further remarkable result. Define a vector  $\boldsymbol{\Omega}$  by the rule

$$\boldsymbol{\Omega}(H) = H \mathbf{e}_0 + H^2 \mathbf{e}_1 + H^3 \mathbf{e}_2 + H^4 \mathbf{e}_3 \tag{11.2.160}$$

so that  $G$  can be written the form

$$G(H) = \boldsymbol{\Omega}(H) \cdot \mathbf{K}. \tag{11.2.161}$$

By (2.156) through (2.159) we may also write

$$\begin{aligned}\Omega(H) = & H\mathbf{c}_0 + (1/2)H^2\mathbf{c}_1 + (1/3)H^3\mathbf{c}_2 + (1/4)H^4\mathbf{c}_3 \\ & -(1/12)H^3(\mathbf{c}_0 \times \mathbf{c}_1) - (1/12)H^4(\mathbf{c}_0 \times \mathbf{c}_2).\end{aligned}\quad (11.2.162)$$

Next, in accord with the stipulation that  $M = 3$ , define a vector  $\hat{\boldsymbol{\omega}}^{\text{fit}3}(\tau)$  by the rule

$$\hat{\boldsymbol{\omega}}^{\text{fit}3}(\tau) = \sum_{m=0}^3 \mathbf{c}_m \tau^m = \mathbf{c}_0 + \mathbf{c}_1 \tau + \mathbf{c}_2 \tau^2 + \mathbf{c}_3 \tau^3,\quad (11.2.163)$$

which is the  $M = 3$  version of (2.65). Observe that

$$\int_0^H \hat{\boldsymbol{\omega}}^{\text{fit}3}(\tau) d\tau = H\mathbf{c}_0 + (1/2)H^2\mathbf{c}_1 + (1/3)H^3\mathbf{c}_2 + (1/4)H^4\mathbf{c}_3,\quad (11.2.164)$$

and that the right side of (2.164) also appears on the right side of right (2.162). Therefore we many rewrite (2.162) in the form

$$\begin{aligned}\Omega(H) = & \int_0^H \hat{\boldsymbol{\omega}}^{\text{fit}3}(\tau) d\tau - (1/12)H^3(\mathbf{c}_0 \times \mathbf{c}_1) - (1/12)H^4(\mathbf{c}_0 \times \mathbf{c}_2).\end{aligned}\quad (11.2.165)$$

### Enter Legendre Gauss

The occurrence of the integral (2.164) suggests the application of quadrature formulas. Suppose we define two sampling times  $\tau_i$  by the rule

$$(\tau_1, \tau_2) = (Hx_1, Hx_2)\quad (11.2.166)$$

where  $x_1$  and  $x_2$  are the  $k = 2$  Legendre-Gauss sampling points. See (T.1.29). Then, since  $k = 2$  Legendre Gauss has  $\ell_{\max} = 3$ , see (T.1.11), there is the relation

$$\int_0^H \hat{\boldsymbol{\omega}}^{\text{fit}3}(\tau) d\tau = (H/2)[\hat{\boldsymbol{\omega}}^{\text{fit}3}(\tau_1) + \hat{\boldsymbol{\omega}}^{\text{fit}3}(\tau_2)].\quad (11.2.167)$$

See (T.1.72).

Also, it can be verified that there is the result

$$\hat{\boldsymbol{\omega}}^{\text{fit}3}(\tau_1) \times \hat{\boldsymbol{\omega}}^{\text{fit}3}(\tau_2) = (1/\sqrt{3})[H(\mathbf{c}_0 \times \mathbf{c}_1) + H^2(\mathbf{c}_0 \times \mathbf{c}_2)] + O(H^3).\quad (11.2.168)$$

See Exercise 2.12. It follows that

$$\begin{aligned}- & (\sqrt{3}/12)H^2[\hat{\boldsymbol{\omega}}^{\text{fit}3}(\tau_1) \times \hat{\boldsymbol{\omega}}^{\text{fit}3}(\tau_2)] = \\ - & (1/12)[H^3(\mathbf{c}_0 \times \mathbf{c}_1) + H^4(\mathbf{c}_0 \times \mathbf{c}_2)] + O(H^5).\end{aligned}\quad (11.2.169)$$

Comparison of (2.165) with (2.167) and (2.169) now reveals that, through terms of order  $H^4$ , there is the even more remarkable result

$$\Omega(H) = (H/2)[\hat{\boldsymbol{\omega}}^{\text{fit}3}(\tau_1) + \hat{\boldsymbol{\omega}}^{\text{fit}3}(\tau_2)] - (\sqrt{3}/12)H^2[\hat{\boldsymbol{\omega}}^{\text{fit}3}(\tau_1) \times \hat{\boldsymbol{\omega}}^{\text{fit}3}(\tau_2)].\quad (11.2.170)$$

The ingredients for computing  $\Omega(H)$  through terms of order  $H^4$  can be obtained by computing the value of  $\hat{\omega}^{\text{fit3}}(\tau)$  at just *two* sampling points! By contrast, the utilization of  $M = 3$  Lie Taylor factorization requires the evaluation of  $\hat{\omega}(\tau)$  at  $M + 1 = 4$  sampling points.

Let us summarize what has been accomplished. From (2.143) and (2.161) we have the result

$$\hat{u}^v(H) = \exp[G(H)] = \exp[\Omega(H) \cdot \mathbf{K}] \quad (11.2.171)$$

where, through terms of order  $H^4$ ,  $\Omega(H)$  is given by (2.170). Upon making in (2.170) the substitution

$$\hat{\omega}^{\text{fit3}}(\tau_i) \simeq \hat{\omega}(\tau_i) \quad (11.2.172)$$

we obtain, in effect, a *two-stage explicit fourth-order* Runge Kutta method for computing  $\hat{u}^v(H)$  with the *guarantee* that  $\hat{u}^v(H)$  is *exactly* in  $SU(2)$ . Note that order 4 is the highest order that can be obtained even for the simplest problem of  $k = 2$  quadrature of ordinary functions. Again see (T.1.72).

At this point we make two remarks. The first is that (2.164), the first term in (2.165), is what might be expected if the quantities  $\hat{\omega}(\tau) \cdot \mathbf{K}$  at various times  $\tau$  all commuted. It is the analog of the term (10.3.17) in the Magnus expansion. Correspondingly (2.169), the remaining terms in (2.165), takes into account the possibility that the quantities  $\hat{\omega}(\tau) \cdot \mathbf{K}$  at various times may not all commute. It is the analog of the term (10.3.19) in the Magnus expansion.

The second remark is that if (2.170) through (2.172) are to be used to track particle spin, then Gauss4 could be used to compute the particle trajectory. See the Butcher tableau (2.3.19). This is possible because both algorithms would then use the common sampling times  $t^n + \tau_1$  and  $t^n + \tau_2$ , and Gauss4 is a *collocation* method so that the results at each stage produce accurate values for the orbit and hence accurate values for  $\hat{\omega}$  at the sampling times. See Exercise 2.3.12. Moreover Gauss4, when applied to Hamiltonian systems, has the further important property of being symplectic. See Section \*.

### Enter Newton Cotes

With the idea of quadrature still in mind, we recall from Table T.1.1 that  $k = 3$  Newton Cotes also has  $\ell_{\max} = 3$ . Suppose we now define three sampling times  $\tau_i$  by the rule

$$(\tau_1, \tau_2, \tau_3) = (Hx_1, Hx_2, Hx_3) \quad (11.2.173)$$

where  $x_1$ ,  $x_2$ , and  $x_3$  are the  $k = 3$  Newton-Cotes sampling points. See (T.1.15). Then we have the relation

$$\int_0^H \hat{\omega}^{\text{fit3}}(\tau) d\tau = (H/6)[\hat{\omega}^{\text{fit3}}(\tau_1) + 4\hat{\omega}^{\text{fit3}}(\tau_2) + \hat{\omega}^{\text{fit3}}(\tau_3)]. \quad (11.2.174)$$

See (T.1.16) and (T.1.66). Also, there is the result

$$\begin{aligned} \hat{\omega}^{\text{fit3}}(\tau_1) \times \hat{\omega}^{\text{fit3}}(\tau_3) &= \hat{\omega}^{\text{fit3}}(0) \times \hat{\omega}^{\text{fit3}}(H) \\ &= \mathbf{c}_0 \times [\mathbf{c}_0 + \mathbf{c}_1 H + \mathbf{c}_2 H^2 + \mathbf{c}_3 H^3] \\ &= H(\mathbf{c}_0 \times \mathbf{c}_1) + H^2(\mathbf{c}_0 \times \mathbf{c}_2) + O(H^3). \end{aligned} \quad (11.2.175)$$

It follows that

$$-(1/12)H^2[\hat{\omega}^{\text{fit3}}(\tau_1) \times \hat{\omega}^{\text{fit3}}(\tau_3)] = -(1/12)[H^3(\mathbf{c}_0 \times \mathbf{c}_1) + H^4(\mathbf{c}_0 \times \mathbf{c}_2)] + O(H^5). \quad (11.2.176)$$

Comparison of (2.165) with (2.174) and (2.176) reveals that, through terms of order  $H^4$ , there is also the result

$$\begin{aligned} \Omega(H) &= (H/6)[\hat{\omega}^{\text{fit3}}(\tau_1) + 4\hat{\omega}^{\text{fit3}}(\tau_2) + \hat{\omega}^{\text{fit3}}(\tau_3)] \\ &\quad - (1/12)H^2[\hat{\omega}^{\text{fit3}}(\tau_1) \times \hat{\omega}^{\text{fit3}}(\tau_3)]. \end{aligned} \quad (11.2.177)$$

The ingredients for computing  $\Omega(H)$  through terms of order  $H^4$  can be obtained by computing the value of  $\hat{\omega}^{\text{fit3}}(\tau)$  at the three sampling times (2.173). Moreover, we observe that these sampling times are the same as those for classic RK4. See the Butcher tableau (2.3.14). It follows that use of (2.171), (2.172), and (2.177) are ideal for tracking particle spin when the particle trajectory is computed using classic RK4 equipped with *dense output*. See Section 2.3.4. The dense output feature would be used to provide values for the coordinates and hence the  $\hat{\omega}$  at the times  $\tau_i$ , and since classic RK4 (although not a collocation method) employs the same sampling times, these interpolated values are expected to be especially accurate.

One can also evaluate the integral on the left side of (2.174) using  $k = 4$  Newton Cotes, which also has  $\ell_{\max} = 3$ . See (T.1.20) through (T.1.22) and Table T.1.1. So doing produces a  $k = 4$  formula for  $\Omega(H)$  that involves four sampling times  $\tau_i$  given by the rule

$$(\tau_1, \tau_2, \tau_3, \tau_4) = (Hx_1, Hx_2, Hx_3, Hx_4) \quad (11.2.178)$$

where the  $x_i$  are the  $k = 4$  Newton-Cotes sampling points. See (T.1.20). Use of these sampling points gives, through terms of order  $H^4$ , the relation

$$\begin{aligned} \Omega(H) &= (H/8)[\hat{\omega}^{\text{fit3}}(\tau_1) + 3\hat{\omega}^{\text{fit3}}(\tau_2) + 3\hat{\omega}^{\text{fit3}}(\tau_3) + \hat{\omega}^{\text{fit3}}(\tau_4)] \\ &\quad - (1/12)H^2[\hat{\omega}^{\text{fit3}}(\tau_1) \times \hat{\omega}^{\text{fit3}}(\tau_4)]. \end{aligned} \quad (11.2.179)$$

And, in this case, one can use for trajectory integration the explicit RK4 version given by the Butcher tableau (2.3.15), again equipped with dense output, since this RK4 (although again not a collocation method) has the same sampling times  $\tau_1$  through  $\tau_4$ .

We close this subsection with two final remarks. The first remark is that the results we have just found for spin readily generalize to all equations of the form (2.142). Suppose we seek to integrate (2.142) using a stepping rule of the form

$$Y^{n+1} = Y(t^n + H) = \exp(G_n)Y^n. \quad (11.2.180)$$

Then it can be shown, for example and in analogy with (2.170), that through terms of order  $H^4$  there is the rule

$$G_n = (H/2)[A(t^n + \tau_1) + A(t^n + \tau_2)] - (\sqrt{3}/12)H^2\{A(t^n + \tau_1), A(t^n + \tau_2)\}. \quad (11.2.181)$$

There are also rules analogous to (2.177) and (2.179).

The second remark is that there is a systematic procedure for finding the coefficients of the expansion of  $G(H)$  in powers of  $H$  in terms of the Taylor expansion of  $A(t)$  in powers of  $t$ , which we have just done for  $M \leq 3$ , and some specific results are known for the next few orders including the cases  $M \leq 6$ . See the paper of *Blanes, Casas, and Ros* listed under the references to Magnus Lie Runge Kutta in the Bibliography at the end of this chapter. When translated to the context of spin tracking, they provide higher-order generalizations of (2.170), (2.177), and (2.179). For example, a three-stage sixth-order generalization of (2.170) could be constructed that could be used in conjunction with Gauss6.

### 11.2.9 Integration in the Lie Algebra Revisited

In the previous Subsection 2.8 we studied the integration of the equation of motion

$$\dot{Y}(t) = A(t)Y(t). \quad (11.2.182)$$

We also mentioned earlier, and you will prove in Exercise 2.7, that one can work equally well with an equation of motion having the form

$$\dot{M}(t) = M(t)A(t). \quad (11.2.183)$$

Moreover, from the work of Subsections 1.14 and 1.15 , we know how to integrate (2.183) in its Lie algebra. It follows that we also know how to integrate (2.182) in its Lie algebra. The purpose of this subsection is to study how a generalization of (2.182), namely an equation of motion of the form

$$\dot{Y}(t) = A(Y, t)Y(t), \quad (11.2.184)$$

can be integrated in its Lie algebra.<sup>30</sup> In so doing we will describe an alternative to factored Lie Runge Kutta. Recall Subsection 2.7 and (2.99).

For the purposes of Runge Kutta it is sufficient to describe how to take one step. In the spirit of Subsection 2.6, let  $t^b$  be the time at which an integration step is to be initiated so that we wish to integrate from  $t^b$  to  $t^b + h$ . We also suppose that

$$Y^b = Y(t^b) \quad (11.2.185)$$

is known and is an element of the group in question. Write

$$Y(t) = Y^v(t)Y^b \quad (11.2.186)$$

with  $Y^v$  being a variable matrix near the identity satisfying

$$Y^v(t^b) = I. \quad (11.2.187)$$

Then it follows from (2.184) and (2.186) that  $Y^v(t)$  obeys the equation of motion

$$\dot{Y}^v(t) = A(Y^vY^b, t)Y^v(t) \quad (11.2.188)$$

with the initial condition (2.187).

---

<sup>30</sup>See Exercise 2.14 for the treatment of a related problem equivalent to the generalization of (2.183).

Next introduce a relative time  $\tau$  by the rule

$$t = t^b + \tau. \quad (11.2.189)$$

Also, define quantities  $\hat{Y}^v$  and  $\hat{A}$  by the rules

$$\hat{Y}^v(\tau) = Y^v(t^b + \tau), \quad (11.2.190)$$

$$\hat{A}(\hat{Y}^v, \tau) = A[Y^v(t^b + \tau)Y^b, t^b + \tau] = A[\hat{Y}^v(\tau)Y^b, t^b + \tau]. \quad (11.2.191)$$

It follows from these definitions and (2.188) that  $\hat{Y}^v(\tau)$  obeys the equation of motion

$$d\hat{Y}^v(\tau)/d\tau = \hat{A}(\hat{Y}^v, \tau)\hat{Y}^v(\tau) \quad (11.2.192)$$

with the initial condition

$$\hat{Y}^v(0) = I. \quad (11.2.193)$$

Our task is to integrate (2.192) from  $\tau = 0$  to  $\tau = h$  in such a way that  $\hat{Y}^v(h)$  is accurate through terms of order  $h^m$ , has possible errors of order  $h^{m+1}$ , but is still exactly in the group in question, or at least is in the group through terms of substantially higher order than  $h^m$ .

### Integration in the Lie Algebra: Exponential Representation

Begin by making the exponential Ansatz

$$\hat{Y}^v(\tau) = \exp[\Omega(\tau)]. \quad (11.2.194)$$

The relation (2.194) can be differentiated and manipulated to yield the result

$$[d\hat{Y}^v(\tau)/d\tau][\hat{Y}^v(\tau)]^{-1} = \text{iex}(\#\Omega\#)(d\Omega/d\tau). \quad (11.2.195)$$

See (\*) in Appendix C. Also, the equation of motion (2.192) can be rewritten in the form

$$[d\hat{Y}^v(\tau)/d\tau][\hat{Y}^v(\tau)]^{-1} = \hat{A}(\hat{Y}^v, \tau). \quad (11.2.196)$$

Comparison of (2.195) and (2.196) and use of (2.194) give the result

$$\text{iex}(\#\Omega\#)(d\Omega/d\tau) = \hat{A}[\exp(\Omega), \tau]. \quad (11.2.197)$$

Finally, solving (2.197) for  $d\Omega/d\tau$  yields the equation of motion

$$d\Omega/d\tau = [\text{iex}(\#\Omega\#)]^{-1}\hat{A}[\exp(\Omega), \tau]; \quad (11.2.198)$$

and use of (2.193) and (2.194) gives the initial condition

$$\Omega(0) = 0. \quad (11.2.199)$$

The function  $[\text{iex}(w)]^{-1}$  has an expansion of the form

$$[\text{iex}(w)]^{-1} = \sum_{\ell=0}^{\infty} c_{\ell} w^{\ell} \quad (11.2.200)$$

with  $c_0 = 1$ ,  $c_1 = -1/2$ ,  $c_2 = 1/12$ , and  $c_\ell = b_\ell$  for  $\ell > 1$ . Again see Appendix C. With the aid of this expansion, (2.198) becomes

$$d\Omega/d\tau = \sum_{\ell=0}^{\infty} c_\ell [\#\Omega(\tau)\#]^\ell \hat{A}[\exp(\Omega), \tau]. \quad (11.2.201)$$

As in the work of Subsection 1.14, the sum in (2.201) can be truncated beyond an even number  $n$  to obtain a result that is correct through order  $h^m$  with  $m = n + 2$ . Thus, to this accuracy, the equation to be solved is

$$\begin{aligned} d\Omega/d\tau &= \sum_{\ell=0}^{\ell=n} c_\ell [\#\Omega(\tau)\#]^\ell \hat{A}[\exp(\Omega), \tau] \\ &= c_0 \hat{A}[\exp(\Omega), \tau] + c_1 \{\Omega(\tau), \hat{A}[\exp(\Omega), \tau]\} \\ &\quad + c_2 \{\Omega(\tau), \{\Omega(\tau), \hat{A}[\exp(\Omega), \tau]\}\} + \dots \\ &\quad + c_n \{\Omega(\tau), \{\dots \{\Omega(\tau), \hat{A}[\exp(\Omega), \tau]\} \dots\}\}, \end{aligned} \quad (11.2.202)$$

and the understanding is that this truncated equation is to be integrated only over the interval  $\tau \in [0, h]$ .

For the sake of pedagogy, let us work out a specific case in detail. Suppose we take  $n = 2$ , in which case we expect a local accuracy through terms of order  $h^4$ . Then (2.202) becomes

$$\begin{aligned} d\Omega/d\tau &= \hat{A}[\exp(\Omega), \tau] - (1/2)\{\Omega(\tau), \hat{A}[\exp(\Omega), \tau]\} \\ &\quad + (1/12)\{\Omega(\tau), \{\Omega(\tau), \hat{A}[\exp(\Omega), \tau]\}\}. \end{aligned} \quad (11.2.203)$$

Correspondingly, we will use classic RK4 for integration. To do so, and to conform to previous notation, it is convenient to write (2.203) in the form

$$d\Omega/d\tau = f(\Omega, \tau) \quad (11.2.204)$$

where

$$\begin{aligned} f(\Omega, \tau) &= \hat{A}[\exp(\Omega), \tau] - (1/2)\{\Omega, \hat{A}[\exp(\Omega), \tau]\} \\ &\quad + (1/12)\{\Omega, \{\Omega, \hat{A}[\exp(\Omega), \tau]\}\}. \end{aligned} \quad (11.2.205)$$

For the first stage of classic RK4, in this instance, and adopting a notation analogous to that of (2.3.8) through (2.3.10), we have the results

$$\tau_1 = 0, \quad (11.2.206)$$

$$\Omega_1 = 0, \quad (11.2.207)$$

$$K_1 = f(\Omega_1, \tau_1) = f(0, 0) = \hat{A}[I, 0]. \quad (11.2.208)$$

See (2.201) and (2.3.9), and note that according to (2.3.14) the first row in the Butcher tableau for classic RK4 vanishes. For the second stage there are the results

$$\tau_2 = h/2, \quad (11.2.209)$$

$$\Omega_2 = (h/2)K_1, \quad (11.2.210)$$

$$\begin{aligned} K_2 &= f(\Omega_2, \tau_2) \\ &= \hat{A}[\exp(\Omega_2), h/2] - (1/2)\{\Omega_2, \hat{A}[\exp(\Omega_2), h/2]\} \\ &\quad + (1/12)\{\Omega_2, \{\Omega_2, \hat{A}[\exp(\Omega_2), h/2]\}\}. \end{aligned} \quad (11.2.211)$$

For the third stage there are the results

$$\tau_3 = h/2, \quad (11.2.212)$$

$$\Omega_3 = (h/2)K_2, \quad (11.2.213)$$

$$\begin{aligned} K_3 &= f(\Omega_3, \tau_3) \\ &= \hat{A}[\exp(\Omega_3), h/2] - (1/2)\{\Omega_3, \hat{A}[\exp(\Omega_3), h/2]\} \\ &\quad + (1/12)\{\Omega_3, \{\Omega_3, \hat{A}[\exp(\Omega_3), h/2]\}\}. \end{aligned} \quad (11.2.214)$$

For the fourth stage there are the results

$$\tau_4 = h, \quad (11.2.215)$$

$$\Omega_4 = hK_3, \quad (11.2.216)$$

$$\begin{aligned} K_4 &= f(\Omega_4, \tau_4) \\ &= \hat{A}[\exp(\Omega_4), h] - (1/2)\{\Omega_4, \hat{A}[\exp(\Omega_4), h]\} \\ &\quad + (1/12)\{\Omega_4, \{\Omega_4, \hat{A}[\exp(\Omega_4), h]\}\}. \end{aligned} \quad (11.2.217)$$

The net result of the full RK4 step is that

$$\Omega(h) = (h/6)K_1 + (h/3)K_2 + (h/3)K_3 + (h/6)K_4. \quad (11.2.218)$$

See (2.3.11).

Finally, from (2.194), we find that, through terms of order  $h^4$ ,

$$\hat{Y}^v(h) = \exp[\Omega(h)]. \quad (11.2.219)$$

Also, from the definitions (2.186), (2.189), and (2.190), we have the relation

$$Y(t^b + h) = Y^v(t^b + h)Y^b = \hat{Y}^v(h)Y^b. \quad (11.2.220)$$

Therefore, upon setting  $t^b = t^n$  and making the identification

$$Y^b = Y^n, \quad (11.2.221)$$

we obtain the stepping rule

$$Y^{n+1} = \exp[\Omega(h)]Y^n. \quad (11.2.222)$$

Let us reflect on what has been accomplished. Let  $G$  be the group in question and let  $\mathcal{L}(G)$  be its Lie algebra. From (2.207) we see that  $\Omega_1 \in \mathcal{L}(G)$ . Also, since  $I \in G$ , we see from (2.208) that  $\hat{A}$  is evaluated in its domain, and therefore  $K_1 \in \mathcal{L}(G)$ . Next look at the second stage results. By (2.210),  $\Omega_2 \in \mathcal{L}(G)$  since  $K_1 \in \mathcal{L}(G)$ . Also, since  $\Omega_2 \in \mathcal{L}(G)$ , there is the result that  $\exp(\Omega_2) \in G$  and therefore again  $\hat{A}$  is evaluated in its domain. See the right side of (2.211). It follows that all the ingredients in the right side of (2.211) are in  $\mathcal{L}(G)$ . Moreover they are combined there in such a way that the full right side of (2.211) is in  $\mathcal{L}(G)$ , and therefore  $K_2 \in \mathcal{L}(G)$ . Evidently analogous results hold for all the stages so that  $\hat{A}$  is always evaluated in its domain and all the  $K_j$  are in  $\mathcal{L}(G)$ . Therefore, by (2.218),  $\Omega(h) \in \mathcal{L}(G)$  from which it follows that  $\exp[\Omega(h)] \in G$ . And from (2.222) we conclude that  $Y^{n+1} \in G$  if  $Y^n \in G$ . Despite generally having made local errors of order  $h^5$ , arising both from the truncation of (2.201) and the use of RK4, we have preserved  $G$  to machine precision.

Finally we note that, contrary to appearances, the relation (2.222) between  $Y^{n+1}$  and  $Y^n$  is generally *not* linear. This is the case because generally  $\Omega(h)$  depends on  $Y^n$ . See the far right side of (2.191) and recall (2.221).

### Integration in the Lie Algebra: Cayley Representation

Suppose  $G$  is a quadratic group. The results (2.185) through (2.193), since they are general, continue to hold. But if  $G$  is a quadratic group, then we may also make, in place of (2.194), the Ansatz

$$\hat{Y}^v(\tau) = \text{cay}[V(\tau)], \quad (11.2.223)$$

where

$$\text{cay}(V) = (I + V)(I - V)^{-1}. \quad (11.2.224)$$

Note that here the definition of  $\text{cay}(V)$  given by (2.224) differs by a sign from that given in (3.12.45). Again  $V$  will be in  $\mathcal{L}(G)$  if  $\hat{Y}^v$  is in  $G$ , and conversely.

From (2.223) and (2.224) it follows that

$$[d\hat{Y}^v(\tau)/d\tau][\hat{Y}^v(\tau)]^{-1} = 2C(dV/d\tau)B^{-1}, \quad (11.2.225)$$

where  $B$  and  $C$  are given in terms of  $V$  by the relations (1.295) through (1.297). See the calculations at the end of Exercise 1.12. Recall also that (2.192) can be rewritten in the form (2.196). Combining (2.196) and (2.225) gives the relation

$$2C(dV/d\tau)B^{-1} = \hat{A}[\text{cay}(V), \tau]. \quad (11.2.226)$$

Finally, solving (2.226) for  $dV/d\tau$  yields the equation of motion

$$dV/d\tau = (1/2)\hat{A}[\text{cay}(V), \tau] - (1/2)\{V, \hat{A}[\text{cay}(V), \tau]\} - (1/2)V\hat{A}[\text{cay}(V), \tau]V. \quad (11.2.227)$$

See (1.320) at the end of Exercise 1.12. And use of (2.193) and (2.223) gives the initial condition

$$V(0) = 0. \quad (11.2.228)$$

Note that the right side of (2.227), unlike (2.201), involves only a finite number of terms, namely three, and therefore in this case no truncation is required.

Because no truncation of the equation of motion has occurred, the only source of error in this case is that associated with numerical integration. Suppose we write (2.227) in the form

$$dV/d\tau = f(V, \tau) \quad (11.2.229)$$

where

$$f(V, \tau) = (1/2)\hat{A}[\text{cay}(V), \tau] - (1/2)\{V, \hat{A}[\text{cay}(V), \tau]\} - (1/2)V\hat{A}[\text{cay}(V), \tau]V. \quad (11.2.230)$$

And suppose for purposes of illustration that we again use classic RK4 for integration. For the first stage of classic RK4, and again adopting a notation analogous to that of (2.3.8) through (2.3.10), we now have the results

$$\tau_1 = 0, \quad (11.2.231)$$

$$V_1 = 0, \quad (11.2.232)$$

$$K_1 = f(V_1, \tau_1) = f(0, 0) = (1/2)\hat{A}[I, 0]. \quad (11.2.233)$$

See (2.228). For the second stage there are the results

$$\tau_2 = h/2, \quad (11.2.234)$$

$$V_2 = (h/2)K_1, \quad (11.2.235)$$

$$\begin{aligned} K_2 &= f(V_2, \tau_2) \\ &= (1/2)\hat{A}[\text{cay}(V_2), h/2] - (1/2)\{V_2, \hat{A}[\text{cay}(V_2), h/2]\} \\ &\quad - (1/2)V_2\hat{A}[\text{cay}(V_2), h/2]V_2. \end{aligned} \quad (11.2.236)$$

For the third stage there are the results

$$\tau_3 = h/2, \quad (11.2.237)$$

$$V_3 = (h/2)K_2, \quad (11.2.238)$$

$$\begin{aligned} K_3 &= f(V_3, \tau_3) \\ &= (1/2)\hat{A}[\text{cay}(V_3), h/2] - (1/2)\{V_3, \hat{A}[\text{cay}(V_3), h/2]\} \\ &\quad - (1/2)V_3\hat{A}[\text{cay}(V_3), h/2]V_3. \end{aligned} \quad (11.2.239)$$

For the fourth stage there are the results

$$\tau_4 = h, \quad (11.2.240)$$

$$V_4 = hK_3, \quad (11.2.241)$$

$$\begin{aligned} K_4 &= f(V_4, \tau_4) \\ &= (1/2)\hat{A}[\text{cay}(V_4), h] - (1/2)\{V_4, \hat{A}[\text{cay}(V_4), h]\} \\ &\quad - (1/2)V_4\hat{A}[\text{cay}(V_4), h]V_4. \end{aligned} \quad (11.2.242)$$

The net result of the full RK4 step is that

$$V(h) = (h/6)K_1 + (h/3)K_2 + (h/3)K_3 + (h/6)K_4. \quad (11.2.243)$$

Finally, from (2.223), we find that, through terms of order  $h^4$ ,

$$\hat{Y}^v(h) = \text{cay}[V(h)]. \quad (11.2.244)$$

Look again at the relations (2.220) and (2.221). Combining them with (2.244) yields the stepping rule

$$Y^{n+1} = \text{cay}[V(h)]Y^n. \quad (11.2.245)$$

Let us again reflect on what has been accomplished. Let  $G$  be the group in question and let  $\mathcal{L}(G)$  be its Lie algebra. From (2.232) we see that  $V_1 \in \mathcal{L}(G)$ . Also, since  $I \in G$ , we see from (2.233) that  $\hat{A}$  is evaluated in its domain, and therefore  $K_1 \in \mathcal{L}(G)$ . Next look at the second stage results. By (2.235),  $V_2 \in \mathcal{L}(G)$  since  $K_1 \in \mathcal{L}(G)$ . Also, since  $V_2 \in \mathcal{L}(G)$ , there is the result that  $\text{cay}(V_2) \in G$  and therefore again  $\hat{A}$  is evaluated in its domain. See the right side of (2.236). It follows that all the ingredients in the right side of (2.236) are in  $\mathcal{L}(G)$ . Moreover they are combined there in such a way that the full right side of (2.236) is in  $\mathcal{L}(G)$ , and therefore  $K_2 \in \mathcal{L}(G)$ . Recall the discussion in Exercise 1.12. Evidently analogous results hold for all the stages so that  $\hat{A}$  is always evaluated in its domain and all the  $K_j$  are in  $\mathcal{L}(G)$ . Therefore, by (2.243),  $V(h) \in \mathcal{L}(G)$  from which it follows that  $\text{cay}[V(h)] \in G$ . And from (2.245) we conclude that  $Y^{n+1} \in G$  if  $Y^n \in G$ . Despite generally having made local errors of order  $h^5$  arising from the use of RK4, we have preserved  $G$  to machine precision.

Finally we note again that, contrary to appearances, the relation (2.245) between  $Y^{n+1}$  and  $Y^n$  is generally *not* linear. This is the case because generally  $V(h)$  depends on  $Y^n$ . Again see the far right side of (2.191) and recall (2.221).

## Exercises

**11.2.1.** Much of Section 2 discussed the problem of finding  $\mathbf{s}(t)$  given  $\bar{\omega}(t)$ , the equation of motion (2.1), and the initial condition (2.3). The aim of this exercise is to treat the inverse *control theory* problem: given  $\mathbf{s}(t)$ , can one find an  $\bar{\omega}(t)$  such that solving (2.1) yields  $\mathbf{s}(t)$ ? Stated geometrically, given a path in  $S^2$ , is there an  $\bar{\omega}(t)$  such that solving (2.1) yields this path? You are to show that the answer to this question is *yes*, and that there are in fact many such  $\bar{\omega}(t)$ . Then you are to show that a knowledge of *two* paths suffices to define  $\bar{\omega}(t)$  uniquely.<sup>31</sup> Treatment of the two-path case employs some of the machinery of Exercise 1.1, which you should review.

Begin by showing that, since by assumption  $\mathbf{s}(t) \in S^2$ , there is the condition

$$\mathbf{s}(t) \cdot \dot{\mathbf{s}}(t) = 0. \quad (11.2.246)$$

Next show that (2.1) can be rewritten in the forms

$$d\mathbf{s}/dt = \bar{\omega}(t) \times \mathbf{s} = -\mathbf{s} \times \bar{\omega}(t) = -[\mathbf{s}(t) \cdot \mathbf{L}]\bar{\omega}(t), \quad (11.2.247)$$

---

<sup>31</sup>The idea of considering two paths, as well as the relation (1.103), were suggested by Sateesh Mane.

$$[\mathbf{s}(t) \cdot \mathbf{L}] \bar{\boldsymbol{\omega}}(t) = -\dot{\mathbf{s}}(t). \quad (11.2.248)$$

According to (2.248), given  $\mathbf{s}(t)$  and  $\dot{\mathbf{s}}(t)$ , we seek to find a vector  $\bar{\boldsymbol{\omega}}(t)$  such that the vector/matrix equation (2.248) is satisfied.

Evidently, at any moment  $t'$ , there are three possibilities:

$$\mathbf{s}(t') \in S^2 \text{ and } \mathbf{s}(t') = +\mathbf{e}_3, \quad (11.2.249)$$

$$\mathbf{s}(t') \in S^2 \text{ and } \mathbf{s}(t') = -\mathbf{e}_3, \quad (11.2.250)$$

$$\mathbf{s}(t') \in S^2 \text{ and } \mathbf{s}(t') \neq \pm\mathbf{e}_3. \quad (11.2.251)$$

First consider the nongeneric possibilities (2.249) and (2.250). Verify that then the relation (2.248) becomes

$$\pm L^3 \bar{\boldsymbol{\omega}}(t') = -\dot{\mathbf{s}}(t'). \quad (11.2.252)$$

Write out  $\bar{\boldsymbol{\omega}}(t')$  and  $\dot{\mathbf{s}}(t')$  in component form,

$$\bar{\boldsymbol{\omega}}(t') = \begin{pmatrix} \bar{\omega}_1(t') \\ \bar{\omega}_2(t') \\ \bar{\omega}_3(t') \end{pmatrix}, \quad (11.2.253)$$

$$\dot{\mathbf{s}}(t') = \begin{pmatrix} \dot{s}_1(t') \\ \dot{s}_2(t') \\ 0 \end{pmatrix}. \quad (11.2.254)$$

Here, in writing (2.254), we have used the fact  $\dot{s}_3(t') = 0$  which follows from (2.249), (2.250), and the condition (2.246). Show that

$$L^3 \bar{\boldsymbol{\omega}}(t') = \begin{pmatrix} -\bar{\omega}_2(t') \\ \bar{\omega}_1(t') \\ 0 \end{pmatrix}. \quad (11.2.255)$$

Therefore verify (2.252) is satisfied in the possibilities (2.249) and (2.250) providing that (respectively)

$$\bar{\omega}_1(t') = \mp \dot{s}_2(t'), \quad (11.2.256)$$

$$\bar{\omega}_2(t') = \pm \dot{s}_1(t'), \quad (11.2.257)$$

$$\bar{\omega}_3(t') = \text{anything}. \quad (11.2.258)$$

What remains is the generic possibility (2.251). For this possibility parameterize  $\mathbf{s}(t') \in S^2$  in terms of polar angle coordinates as in (2.11) through (2.13), and verify that both  $\theta(t')$  and  $\phi(t')$  are well defined. Show that, in terms of the Euler angle parameterization (3.7.207), the rotation  $R(\phi, \theta, 0)$  relates  $\mathbf{s}(t') \in S^2$  and  $\mathbf{e}_3$  by the equation

$$\mathbf{s}(t') = R[\phi(t'), \theta(t'), 0] \mathbf{e}_3. \quad (11.2.259)$$

See (3.7.208). Next verify that

$$RL^3R^{-1} = R(\mathbf{e}_3 \cdot \mathbf{L})R^{-1} = [R(\mathbf{e}_3) \cdot \mathbf{L}] = \mathbf{s}(t') \cdot \mathbf{L}. \quad (11.2.260)$$

Recall (8.2.59). Show that inserting this result into (2.248) yields, at the generic moment  $t = t'$ , the relation

$$[RL^3R^{-1}]\bar{\omega}(t') = -\dot{s}(t') \quad (11.2.261)$$

from which it follows that

$$L^3[R^{-1}\bar{\omega}(t')] = -R^{-1}\dot{s}(t'). \quad (11.2.262)$$

Introduce the notation

$$\mathbf{u} = R^{-1}\bar{\omega}(t'), \quad (11.2.263)$$

$$\mathbf{v} = R^{-1}\dot{s}(t'), \quad (11.2.264)$$

so that (2.262) becomes

$$L^3\mathbf{u} = -\mathbf{v}. \quad (11.2.265)$$

Verify also that

$$\mathbf{e}_3 \cdot \mathbf{v} = [R^{-1}\mathbf{s}] \cdot [R^{-1}\dot{\mathbf{s}}] = \mathbf{s} \cdot \dot{\mathbf{s}} = 0. \quad (11.2.266)$$

Here we have used the orthogonality of  $R$ . Consequently, verify that we may write

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}, \quad (11.2.267)$$

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ 0 \end{pmatrix}. \quad (11.2.268)$$

We are back to the first case we considered, and conclude that

$$u_1 = -v_2, \quad (11.2.269)$$

$$u_2 = v_1, \quad (11.2.270)$$

$$u_3 = \text{anything}. \quad (11.2.271)$$

Verify that (2.269) through (2.271) can be written in the vector/matrix form

$$\mathbf{u}(t') = L^3\mathbf{v}(t') + g(t')\mathbf{e}_3 \quad (11.2.272)$$

where  $g$  is any function. We are almost done. Show that inserting (2.263) and (2.264) into (2.272) yields the result

$$R^{-1}\bar{\omega}(t') = L^3R^{-1}\dot{s}(t') + g(t')\mathbf{e}_3, \quad (11.2.273)$$

from which it follows, for any generic time  $t = t'$ , that

$$\begin{aligned} \bar{\omega}(t) &= RL^3R^{-1}\dot{s}(t) + g(t)R\mathbf{e}_3 \\ &= [\mathbf{s}(t) \cdot \mathbf{L}]\dot{s}(t) + g(t)\mathbf{s}(t) \\ &= \mathbf{s}(t) \times \dot{s}(t) + g(t)\mathbf{s}(t). \end{aligned} \quad (11.2.274)$$

If all has been done correctly, (2.274) provides the general solution to (2.248). Verify that this general solution also covers the nongeneric possibilities (2.249) and (2.250) for which the

polar angles were not well defined. Verify, by direct computation, that the general solution satisfies (2.248). Hint: Use (3.7.200). In practice we might require that  $g(t)$  be continuous, and for convenience we might set  $g(t) = 0$ .

You have shown that there is a choice of  $\bar{\omega}(t)$  that will produce any desired path  $\mathbf{s}(t) \in S^2$ , and that this choice is not unique. What if there are two different paths  $\mathbf{r}(t) \in S^2$  and  $\mathbf{s}(t) \in S^2$  that are supposed to be produced by the same  $\bar{\omega}(t)$ . That is,  $\mathbf{s}(t)$  satisfies (2.1) and  $\mathbf{r}(t)$  satisfies the analogous relation

$$d\mathbf{r}/dt = \bar{\omega}(t) \times \mathbf{r}. \quad (11.2.275)$$

Is  $\bar{\omega}(t)$  then uniquely determined? You are to show that the answer is *yes*.

Can  $\mathbf{r}(t) \in S^2$  and  $\mathbf{s}(t) \in S^2$  be specified arbitrarily and independently? The answer is *no*. There are some mutual restrictions on  $\mathbf{r}(t) \in S^2$  and  $\mathbf{s}(t) \in S^2$  since they are supposed to be produced by the same differential equation, namely (2.1) and (2.275). We will call a path on  $S^2$  that satisfies this differential equation an *allowable path*.

First, we must assume that  $\mathbf{r}^0 \neq \pm \mathbf{s}^0$ . In the first case the initial conditions would be the same and therefore  $\mathbf{r}(t)$  and  $\mathbf{s}(t)$  would be identical. In the second case, since by linearity  $-\mathbf{s}(t)$  is an allowable path if  $\mathbf{s}(t)$  is, we have again gained no new information. Therefore we may assume that  $\mathbf{r}^0$  and  $\mathbf{s}^0$  are linearly independent. Show that it follows that

$$-1 < \mathbf{r}^0 \cdot \mathbf{s}^0 < 1 \quad (11.2.276)$$

and

$$|\mathbf{r}^0 \times \mathbf{s}^0| \neq 0. \quad (11.2.277)$$

Consider the quantity  $\mathbf{r}(t) \cdot \mathbf{s}(t)$ . Verify that

$$\begin{aligned} (d/dt)[\mathbf{r}(t) \cdot \mathbf{s}(t)] &= \dot{\mathbf{r}}(t) \cdot \mathbf{s}(t) + \mathbf{r}(t) \cdot \dot{\mathbf{s}}(t) \\ &= [\bar{\omega}(t) \times \mathbf{r}(t)] \cdot \mathbf{s}(t) + \mathbf{r}(t) \cdot [\bar{\omega}(t) \times \mathbf{s}(t)] \\ &= [\bar{\omega}(t) \times \mathbf{r}(t)] \cdot \mathbf{s}(t) + [\mathbf{r}(t) \times \bar{\omega}(t)] \cdot \mathbf{s}(t) = 0. \end{aligned} \quad (11.2.278)$$

Conclude that

$$\mathbf{r}(t) \cdot \mathbf{s}(t) = \mathbf{r}^0 \cdot \mathbf{s}^0. \quad (11.2.279)$$

The dot product of the two time-dependent vectors associated with two allowable paths (and also the dot product of such a vector with itself) remains constant as the paths are traversed.

Next consider the vector  $\mathbf{r}(t) \times \mathbf{s}(t)$ . Verify that

$$\begin{aligned} (d/dt)[\mathbf{r}(t) \times \mathbf{s}(t)] &= \dot{\mathbf{r}}(t) \times \mathbf{s}(t) + \mathbf{r}(t) \times \dot{\mathbf{s}}(t) \\ &= [\bar{\omega}(t) \times \mathbf{r}(t)] \times \mathbf{s}(t) + \mathbf{r}(t) \times [\bar{\omega}(t) \times \mathbf{s}(t)] \\ &= -\{\mathbf{s}(t) \times [\bar{\omega}(t) \times \mathbf{r}(t)]\} + \{\mathbf{r}(t) \times [\bar{\omega}(t) \times \mathbf{s}(t)]\} \\ &= -\{\bar{\omega}(t)[\mathbf{r}(t) \cdot \mathbf{s}(t)] - \mathbf{r}(t)[\bar{\omega}(t) \cdot \mathbf{s}(t)]\} \\ &\quad + \{\bar{\omega}(t)[\mathbf{r}(t) \cdot \mathbf{s}(t)] - \mathbf{s}(t)[\bar{\omega}(t) \cdot \mathbf{r}(t)]\} \\ &= \mathbf{r}(t)[\bar{\omega}(t) \cdot \mathbf{s}(t)] - \mathbf{s}(t)[\bar{\omega}(t) \cdot \mathbf{r}(t)] \\ &= \bar{\omega}(t) \times [\mathbf{r}(t) \times \mathbf{s}(t)]. \end{aligned} \quad (11.2.280)$$

You have shown, apart from normalization, that  $\mathbf{r}(t) \times \mathbf{s}(t)$  is an allowable path if  $\mathbf{r}(t)$  and  $\mathbf{s}(t)$  are allowable paths. Argue, from the results of the previous paragraph, that there is also the result

$$|\mathbf{r}(t) \times \mathbf{s}(t)| = |\mathbf{r}^0 \times \mathbf{s}^0|. \quad (11.2.281)$$

Define three *fixed* vectors by the rules

$$\mathbf{e}_1 = \mathbf{r}^0, \quad (11.2.282)$$

$$\mathbf{e}_2 = (\mathbf{r}^0 \times \mathbf{s}^0)/|(\mathbf{r}^0 \times \mathbf{s}^0)|, \quad (11.2.283)$$

$$\mathbf{e}_3 = \mathbf{e}_1 \times \mathbf{e}_2. \quad (11.2.284)$$

Verify that together the  $\mathbf{e}_j$  form a right-hand triad of orthonormal vectors. That is,

$$\mathbf{e}_j \cdot \mathbf{e}_k = \delta_{jk}, \quad (11.2.285)$$

$$\mathbf{e}_j \times \mathbf{e}_k = \mathbf{e}_\ell, \quad (11.2.286)$$

where  $j, k, \ell$  is any cyclic permutation of 1, 2, 3.

Define three *time-dependent* vectors by the rules

$$\mathbf{f}_1(t) = \mathbf{r}(t), \quad (11.2.287)$$

$$\mathbf{f}_2(t) = [\mathbf{r}(t) \times \mathbf{s}(t)]/|(\mathbf{r}^0 \times \mathbf{s}^0)|, \quad (11.2.288)$$

$$\mathbf{f}_3(t) = \mathbf{f}_1(t) \times \mathbf{f}_2(t). \quad (11.2.289)$$

Based on the work above, show that the  $\mathbf{f}_j(t)$  have the initial conditions

$$\mathbf{f}_j(t^0) = \mathbf{e}_j \quad (11.2.290)$$

and are all allowable paths,

$$d\mathbf{f}_j/dt = \bar{\boldsymbol{\omega}}(t) \times \mathbf{f}_j. \quad (11.2.291)$$

Show that they also form a right-hand triad of orthonormal vectors at each instant  $t$ ,

$$\mathbf{f}_j \cdot \mathbf{f}_k = \delta_{jk}, \quad (11.2.292)$$

$$\mathbf{f}_j \times \mathbf{f}_k = \mathbf{f}_\ell, \quad (11.2.293)$$

where  $j, k, \ell$  is any cyclic permutation of 1, 2, 3.

Next observe that since the  $\mathbf{e}_j$  and  $\mathbf{f}_j(t)$  both form triads of orthonormal vectors, they must be related by an orthogonal transformation  $R(t)$ ,

$$\mathbf{f}_j(t) = R(t)\mathbf{e}_j, \quad (11.2.294)$$

with  $R(t^0) = I$ . Moreover, (2.294) specifies  $R(t)$  uniquely. See Section 3.6.3. By continuity  $R(t)$  must satisfy  $\det R(t) = 1$ , and therefore  $R(t) \in SO(3, \mathbb{R})$ . Thus, from a knowledge of two allowable paths, we are able to define a rigid body motion.

Also, from the work of Exercise 1.1, we know that there is the relation

$$\dot{R}(t)R^{-1}(t) = \boldsymbol{\omega}^{sf}(t) \cdot \mathbf{L}. \quad (11.2.295)$$

See (1.114). Are  $\bar{\omega}(t)$ ,  $\omega^{sf}(t)$ ,  $\omega^{bf}(t)$ , and  $\omega(t)$  related? And, if so, how? From (2.294) and (2.295) show that

$$\dot{\mathbf{f}}_j = \dot{R}\mathbf{e}_j = \dot{R}R^{-1}R\mathbf{e}_j = [\omega^{sf} \cdot \mathbf{L}]\mathbf{f}_j = \omega^{sf} \times \mathbf{f}_j. \quad (11.2.296)$$

Upon comparing (2.291) and (2.296) we are led to make the conjecture

$$\bar{\omega}(t) \stackrel{?}{=} \omega^{sf}(t). \quad (11.2.297)$$

Note that in writing the far right side of (2.296) some explanation must be given as to what is meant by  $\omega^{sf}(t)$  when it is employed in a cross product. We will see that it is the vector with the components  $\omega_j^{sf}(t)$  when the  $\mathbf{e}_j$  are used as a basis. We also observe that, according to (1.103),  $\omega(t)$  is specified once the  $\mathbf{f}_j(t)$  are known.

Because there are two basis sets involved, namely the  $\mathbf{e}_j$  and the  $\mathbf{f}_j(t)$ , let us make this conjecture precise. From the definition (1.4), and the fact that the  $\mathbf{f}_j(t)$  are all allowable paths, verify that

$$\begin{aligned} \omega_j^{bf}(t) &= -\mathbf{f}_k(t) \cdot \dot{\mathbf{f}}_\ell(t) = -\mathbf{f}_k(t) \cdot [\bar{\omega}(t) \times \mathbf{f}_\ell(t)] \\ &= \mathbf{f}_k(t) \cdot [\mathbf{f}_\ell(t) \times \bar{\omega}(t)] = [\mathbf{f}_k(t) \times \mathbf{f}_\ell(t)] \cdot \bar{\omega}(t) \\ &= \mathbf{f}_j(t) \cdot \bar{\omega}(t) \end{aligned} \quad (11.2.298)$$

where  $j, k, \ell$  is any cyclic permutation of 1, 2, 3. Show, since the  $\mathbf{f}_j(t)$  form an orthonormal basis, it follows that

$$\bar{\omega}(t) = \sum_j [\mathbf{f}_j(t) \cdot \bar{\omega}(t)] \mathbf{f}_j(t) = \sum_j \omega_j^{bf}(t) \mathbf{f}_j(t) = \sum_j \omega_j^{sf}(t) \mathbf{e}_j = \omega(t). \quad (11.2.299)$$

Here we have again used the results and notation of Exercise 1.1. We conclude that a knowledge of two allowable paths completely specifies  $\bar{\omega}(t)$ .

As an interesting side calculation, verify that there are also the relations

$$\begin{aligned} -\mathbf{f}_2(t) \cdot \{\mathbf{f}_1(t) \times \dot{\mathbf{f}}_2(t)\} &= -\mathbf{f}_2(t) \cdot \{\mathbf{f}_1(t) \times [\bar{\omega} \times \mathbf{f}_2(t)]\} \\ &= -\mathbf{f}_2(t) \cdot \{-\mathbf{f}_2(t)[\mathbf{f}_1(t) \cdot \bar{\omega}(t)]\} \\ &= \mathbf{f}_1(t) \cdot \bar{\omega}(t), \end{aligned} \quad (11.2.300)$$

$$\begin{aligned} -\mathbf{f}_3(t) \cdot \{\mathbf{f}_2(t) \times \dot{\mathbf{f}}_3(t)\} &= -\mathbf{f}_3(t) \cdot \{\mathbf{f}_2(t) \times [\bar{\omega} \times \mathbf{f}_3(t)]\} \\ &= -\mathbf{f}_3(t) \cdot \{-\mathbf{f}_3(t)[\mathbf{f}_2(t) \cdot \bar{\omega}(t)]\} \\ &= \mathbf{f}_2(t) \cdot \bar{\omega}(t), \end{aligned} \quad (11.2.301)$$

$$\begin{aligned} -\mathbf{f}_1(t) \cdot \{\mathbf{f}_3(t) \times \dot{\mathbf{f}}_1(t)\} &= -\mathbf{f}_1(t) \cdot \{\mathbf{f}_3(t) \times [\bar{\omega} \times \mathbf{f}_1(t)]\} \\ &= -\mathbf{f}_1(t) \cdot \{-\mathbf{f}_1(t)[\mathbf{f}_3(t) \cdot \bar{\omega}(t)]\} \\ &= \mathbf{f}_3(t) \cdot \bar{\omega}(t). \end{aligned} \quad (11.2.302)$$

There is one last observation. Suppose  $\mathbf{q}^0$  is any point in  $S^2$ . Since the  $\mathbf{e}_j$  form a basis in  $E^3$ , any such  $\mathbf{q}^0$  can be written as a linear combination of the  $\mathbf{e}_j$ ,

$$\mathbf{q}^0 = \sum_{j=1}^3 \alpha_j \mathbf{e}_j \quad (11.2.303)$$

with

$$\alpha_j = \mathbf{e}_j \cdot \mathbf{q}^0 \quad (11.2.304)$$

and

$$\sum_{j=1}^3 \alpha_j^2 = 1. \quad (11.2.305)$$

It follows that  $\mathbf{q}(t)$  given by

$$\mathbf{q}(t) = \sum_{j=1}^3 \alpha_j \mathbf{f}_j(t) \quad (11.2.306)$$

will be an allowable path on  $S^2$  with initial condition  $\mathbf{q}^0$ . Alternatively, we may write

$$\mathbf{q}(t) = R(t)\mathbf{q}^0. \quad (11.2.307)$$

We conclude that at  $t = t^0$  every point on  $S^2$  can be viewed as the starting point for a unique allowable path. Conversely, since  $R(t)$  is in  $SO(3, \mathbb{R})$  and therefore invertible at each time  $t$ , every point on  $S^2$  at time  $t$  may be viewed as lying on a unique allowable path. Or, put another way, all allowable paths may be viewed as arising from a time-dependent rotation of  $S^2$ . Finally, we have learned that  $R(t)$ , and all allowable paths  $\mathbf{q}(t)$ , can be built from the two allowable paths  $\mathbf{r}(t)$  and  $\mathbf{s}(t)$ .

**11.2.2.** Verify that the equations of motion (2.5) through (2.7) follow from assuming the validity of (2.8) through (2.10). That is, if (2.8) through (2.10) are satisfied, then (2.5) through (2.7) are satisfied.

**11.2.3.** Verify (2.11) through (2.20).

**11.2.4.** The aim of this exercise is to verify that (2.52) preserves  $SU(2)$ . Begin by assuming that  $u(t)$  belongs to  $SU(2)$  at some time  $t = t^0$ :

$$u^\dagger(t^0)u(t^0) = I \quad (11.2.308)$$

and

$$\det[u(t^0)] = 1. \quad (11.2.309)$$

First show that (2.52) preserves the determinant of  $u$ . Verify from (2.52) that

$$\begin{aligned} u(t + dt) &= u(t) + \dot{u}(t)dt + O[(dt)^2] \\ &= u(t) + (\bar{\omega} \cdot \mathbf{K})u(t)dt + O[(dt)^2] \\ &= \exp[(dt)\bar{\omega} \cdot \mathbf{K}]u(t) + O[(dt)^2]. \end{aligned} \quad (11.2.310)$$

Show that taking the determinant of both sides of (2.310) yields the result

$$\begin{aligned}\det[u(t + dt)] &= \det\{\exp[(dt)\bar{\boldsymbol{\omega}} \cdot \mathbf{K}]u(t)\} + O[(dt)^2] \\ &= \det\{\exp[(dt)\bar{\boldsymbol{\omega}} \cdot \mathbf{K}]\} \det[u(t)] + O[(dt)^2] \\ &= \exp[(dt) \operatorname{tr}(\bar{\boldsymbol{\omega}} \cdot \mathbf{K})] \det[u(t)] + O[(dt)^2] \\ &= \det[u(t)] + O[(dt)^2].\end{aligned}\quad (11.2.311)$$

(Recall that the  $K^j$  are traceless.) Show from (2.311) that

$$\{\det[u(t + dt)] - \det[u(t)]\}/dt = O(dt) \quad (11.2.312)$$

and therefore

$$d\{\det[u(t)]\}/dt = 0. \quad (11.2.313)$$

Verify that the solution to the differential equation (2.313) with the initial condition (2.309) is the relation

$$\det[u(t)] = 1. \quad (11.2.314)$$

What remains to be verified is that (2.52) preserves unitarity. Show from (2.52) that

$$(d/dt)u^\dagger(t) = -u^\dagger(t)(\bar{\boldsymbol{\omega}} \cdot \mathbf{K}). \quad (11.2.315)$$

Next verify from (2.52) and (2.315) that

$$\begin{aligned}(d/dt)[u^\dagger(t)u(t)] &= [(d/dt)u^\dagger(t)u(t)] + u^\dagger(t)(d/dt)u(t) \\ &= -u^\dagger(t)(\bar{\boldsymbol{\omega}} \cdot \mathbf{K})u(t) + u^\dagger(t)(\bar{\boldsymbol{\omega}} \cdot \mathbf{K})u(t) = 0.\end{aligned}\quad (11.2.316)$$

Finally, verify that the solution to the differential equation (2.316) with the initial condition (2.308) is the relation

$$u^\dagger(t)u(t) = I. \quad (11.2.317)$$

**11.2.5.** The purpose of this exercise is to prove that  $S$  defined by (2.54) satisfies (2.43) and (2.44) providing  $u$  satisfies (2.52) and (2.53). You will need some of the ingredients of Exercise 2.4 above. Begin by verifying that (2.44) is satisfied because

$$(1/2)\operatorname{tr}[u^\dagger(t^0)\sigma^\alpha u(t^0)\sigma^\beta] = (1/2)\operatorname{tr}[\sigma^\alpha\sigma^\beta] = \delta_{\alpha\beta} = S_{\alpha\beta}(t^0). \quad (11.2.318)$$

Next work on proving that  $S$  satisfies (2.43). Show that

$$\dot{S}_{\alpha\beta}(t) = (1/2)\operatorname{tr}(\dot{u}^\dagger\sigma^\alpha u\sigma^\beta) + (1/2)\operatorname{tr}(u^\dagger\sigma^\alpha\dot{u}\sigma^\beta). \quad (11.2.319)$$

Verify that employing (2.52) and (2.315) in (2.319) yields the result

$$\begin{aligned}\dot{S}_{\alpha\delta}(t) &= -(1/2)\operatorname{tr}[u^\dagger(\bar{\boldsymbol{\omega}} \cdot \mathbf{K})\sigma^\alpha u\sigma^\delta] + (1/2)\operatorname{tr}[u^\dagger\sigma^\alpha(\bar{\boldsymbol{\omega}} \cdot \mathbf{K})u\sigma^\delta] \\ &= (1/2)\operatorname{tr}[u^\dagger\{\sigma^\alpha, (\bar{\boldsymbol{\omega}} \cdot \mathbf{K})\}u\sigma^\delta].\end{aligned}\quad (11.2.320)$$

Verify that

$$\{\sigma^\alpha, (\bar{\boldsymbol{\omega}} \cdot \mathbf{K})\} = \sum_\beta \bar{\omega}_\beta \{\sigma^\alpha, \sigma^\beta\} = \sum_{\beta\gamma} \bar{\omega}_\beta \epsilon_{\alpha\beta\gamma} \sigma^\gamma. \quad (11.2.321)$$

Recall (1.143) and (5.7.39). Insert (2.321) into (2.320) and use (3.7.181) to show that

$$\begin{aligned}\dot{S}_{\alpha\delta}(t) &= \sum_{\beta\gamma} \bar{\omega}_\beta \epsilon_{\alpha\beta\gamma} (1/2) \text{tr}[u^\dagger \sigma^\gamma u \sigma^\delta] \\ &= \sum_{\beta\gamma} \bar{\omega}_\beta \epsilon_{\beta\alpha\gamma} S_{\gamma\delta} = \sum_{\beta\gamma} \bar{\omega}_\beta \epsilon_{\alpha\beta\gamma} S_{\gamma\delta} \\ &= \sum_{\beta\gamma} \bar{\omega}_\beta L_{\alpha\gamma}^\beta S_{\gamma\delta} = \sum_\gamma (\bar{\omega} \cdot \mathbf{L})_{\alpha\gamma} S_{\gamma\delta},\end{aligned}\quad (11.2.322)$$

or, in more compact matrix form,

$$\dot{S} = (\bar{\omega} \cdot \mathbf{L}) S. \quad (11.2.323)$$

**11.2.6.** This exercise is in part a continuation of Exercises 1.7 and 1.8, which you should review. Our aim here is to explore the analogy between (2.40), which describes rotations in  $E^3$ , and (1.209), which will be seen to be related to rotations in  $E^4$ .

Verify that  $A(\boldsymbol{\omega}^{bf})$  as given by (1.208) is  $4 \times 4$  and antisymmetric, and therefore belongs to the Lie algebra  $so(4, \mathbb{R})$ . Recall that  $(\bar{\omega} \cdot \mathbf{L})$  is  $3 \times 3$  and antisymmetric, and therefore belongs to the Lie algebra  $so(3, \mathbb{R})$ . Thus, (1.209) appears to be a four-dimensional analog of (2.40). Moreover because  $A(\boldsymbol{\omega}^{bf}) \in so(4, \mathbb{R})$ , (1.209) describes, as asserted, (at least some) rotations in  $E^4$ . Therefore it is no wonder that (1.209) preserves  $w \cdot w$  (preserves  $S^3$ ) just as (2.40) preserves  $S^2$ .

Verify, however, that  $A(\boldsymbol{\omega}^{bf})$  is given by

$$A(\boldsymbol{\omega}^{bf}) = (1/2)(\omega_1^{bf} E^1 + \omega_2^{bf} E^2 + \omega_3^{bf} E^3) = (1/2)\boldsymbol{\omega}^{bf} \cdot \mathbf{E}, \quad (11.2.324)$$

and therefore  $A$  belongs to one of the  $su(2)$  Lie algebras in  $so(4, \mathbb{R})$ . Thus the matrices  $A$  given by (1.208) do not span all of  $so(4, \mathbb{R})$ . [Recall that  $so(4, \mathbb{R})$  is six dimensional.]

We have seen that the vector equation (2.40) has the corresponding matrix equation (2.43). Similarly, the vector equation (1.209) has the  $4 \times 4$  matrix equation counterpart

$$\dot{T} = A(\boldsymbol{\omega}^{bf}) T \quad (11.2.325)$$

with the initial condition

$$T(t^0) = I. \quad (11.2.326)$$

Therefore part of our task is also to explore the analogy between (2.43) and (2.325).

Verify that  $T(t) \in SO(4, \mathbb{R})$ , but that not all  $SO(4, \mathbb{R})$  matrices can be produced in this fashion because the matrices  $A$  given by (1.208) do not span all of  $so(4, \mathbb{R})$ . Therefore the analogy between (2.40) and (1.209), and between (2.43) and (2.325), is not complete.<sup>32</sup>

The observation that  $A$  belongs to an  $su(2)$  Lie algebra begs further exploration because we know that what counts, when solving differential equations, is not the matrix size of  $A$  but rather the Lie algebra to which it belongs. (The same reasoning led to the replacement

---

<sup>32</sup>Although the analogy between (2.40) and (1.209) is not complete, it can be shown that any path in  $S^3$  can be produced by solution of (1.209) for some choice of  $\boldsymbol{\omega}^{bf}(t)$  when employed in  $A[\boldsymbol{\omega}^{bf}(t)]$ . This is possible for two reasons: First, as manifolds,  $SU(2)$  and  $S^3$  are the same. Recall Exercise 5.10.3. Second,  $SU(2)$  acts transitively on itself (as does any group) by both left and right multiplication.

of the  $L_j$  by the  $K_j$  in Exercise 1.5 and in Subsection 2.6.) Thus, we expect that there must be an  $SU(2)$  formulation of the equations of motion (2.325). There is. Verify the commutation rules

$$\{(-E^j/2), (-E^k/2)\} = (-E^\ell/2) \quad (11.2.327)$$

where  $j, k, \ell$  are any cyclic permutation of 1,2,3. Evidently these commutation rules are the same as those for the  $K^j$ , see (3.7.172), and therefore we may make the correspondence

$$(-E^j/2) \leftrightarrow K^j. \quad (11.2.328)$$

Consequently, in view of (2.324), there is also the correspondence

$$A(\boldsymbol{\omega}^{bf}) \leftrightarrow -\boldsymbol{\omega}^{bf} \cdot \mathbf{K} = \bar{\boldsymbol{\omega}} \cdot \mathbf{K} \quad (11.2.329)$$

with

$$\bar{\boldsymbol{\omega}} = -\boldsymbol{\omega}^{bf}. \quad (11.2.330)$$

It follows that the  $SU(2)$  analog of (2.325) is the differential equation

$$\dot{u}(t) = (\bar{\boldsymbol{\omega}} \cdot \mathbf{K})u(t) \quad (11.2.331)$$

with the initial conditions

$$u(t^0) = I \quad (11.2.332)$$

where  $u$  is a complex  $2 \times 2$  matrix.

We know from Exercise 2.4 that  $u(t)$ , the solution to (2.331) with the initial condition (2.332), will be in  $SU(2)$  for all  $t$ . Moreover from Exercise 2.5 we know that, because of the homomorphism between  $SU(2)$  and  $SO(3, \mathbb{R})$ , the  $S$  associated with  $u$  by the rule (2.54) satisfies the equation of motion (2.43) with the initial condition (2.44). Thus, we have come full circle back to (2.43).

There are other interesting paths we can take. For example, Exercise 2.15 below shows that there is a connection between quaternion parameters and solutions to the general Schroedinger equation in a two-dimensional complex Hilbert space.

**11.2.7.** The aim of this exercise is to explore the relations between matrix differential equations of the form

$$\dot{M}(t) = M(t)A(t) \quad (11.2.333)$$

with the initial condition

$$M(0) = I, \quad (11.2.334)$$

and matrix differential equations of the form

$$\dot{N}(t) = \bar{A}(t)N(t) \quad (11.2.335)$$

with the initial condition

$$N(0) = I. \quad (11.2.336)$$

Begin with  $N(t)$ , the solution to (2.335) with the initial condition (2.336). Define a matrix  $M(t)$  by the rule

$$M = N^{-1}, \quad (11.2.337)$$

from which it follows that

$$M(0) = I \quad (11.2.338)$$

and

$$MN = I. \quad (11.2.339)$$

Verify that differentiating both sides of (2.339) gives the result

$$\dot{M}N + M\dot{N} = 0, \quad (11.2.340)$$

and therefore

$$\dot{M} = -M\dot{N}N^{-1} = -M\bar{A}NN^{-1} = -M\bar{A}. \quad (11.2.341)$$

Here we have also used (2.335). Evidently (2.341) agrees with (2.333) providing we make the identification

$$A = -\bar{A}. \quad (11.2.342)$$

Let us pause for a side check on consistency with previous results. For the case of  $SO(3, \mathbb{R})$ , verify that what we have just found is consistent with the relations (1.18) and (1.121). For the Cayley parameterization (1.293) in the case of quadratic groups, verify that the substitution

$$M \leftrightarrow M^{-1} \quad (11.2.343)$$

is equivalent to the substitution

$$V \leftrightarrow -V. \quad (11.2.344)$$

Verify that, under the substitution (2.344) and the substitution

$$A \leftrightarrow -\bar{A}, \quad (11.2.345)$$

the relation (1.313) is transformed into the relation (1.320).

Continue on. The only possible difficulty in making the transition between the two cases (2.333) and (2.335) is the inversion (2.337). It can be shown that matrices that satisfy differential equations of the form (2.333) and (2.335) must generally be invertible, and therefore the transition is generally possible. Your next task is to prove this result.

Review Exercise 1.4.6. Using the methods of that exercise show, starting with (2.335), that there is the relation

$$\det N(t) = [\det N(t^0)] \exp\left[\int_{t^0}^t dr' \operatorname{tr} \bar{A}(t')\right]. \quad (11.2.346)$$

Assume that  $\bar{A}(t')$  is finite for all  $t'$ . Show it follows that the factor  $\exp\left[\int_{t^0}^t dr' \operatorname{tr} \bar{A}(t')\right]$  is finite and nonzero. Verify it follows that if  $N(t^0)$  is invertible at any time  $t^0$ ,  $\det N(t^0) \neq 0$ , then  $N(t)$  is invertible at all times  $t$ . Show, starting with (2.333), that an analogous result holds for  $M(t)$ .

Although matrix inversion is generally computationally intensive, it can be carried out easily if  $M$  (and consequently also  $N$ ) belong to various groups. Let us recall some groups for which inverse elements are easily computed. The unitary group comes first to mind. If  $M$  is unitary, then

$$M^{-1} = M^\dagger. \quad (11.2.347)$$

And, if  $M$  is orthogonal,

$$M^{-1} = M^T. \quad (11.2.348)$$

Finally, if  $M$  belongs to a quadratic group, then

$$M^{-1} = L^{-1}M^TL. \quad (11.2.349)$$

See (3.12.32).

**11.2.8.** The purpose of this exercise is to prove that the exact solution to (2.99) preserves  $G$ . Verify that (2.103) solves, over the interval  $[t^n, t^{n+1}]$ , the equation (2.99) through terms of order  $h$ . It may make local errors of order  $h^2$ , but verify that it preserves  $G$ . In the terminology of Sections 2.1 and 2.2, let  $h \rightarrow 0$  and correspondingly  $N \rightarrow \infty$  to find  $Y(t^0 + T)$ . Verify that this result is the exact solution of (2.99) evaluated at  $t = t^0 + T$  and that this result is in  $G$ .

**11.2.9.** We have claimed that the two-stage Lie Runge Kutta routine (2.109) through (2.113) has order  $m = 2$ . Verify, for example, that this is true in the case of (2.43), when (2.100) and (2.101) hold, and the Butcher tableau (2.3.36) is employed.

**11.2.10.** Verify that the entries in the Butcher tableau (2.124) satisfy the consistency relation (2.3.16), the order conditions (2.3.42) through (2.3.45), and the additional order 3 condition (2.122).

**11.2.11.** Verify that (2.156) through (2.159) follow from (2.87) through (2.90) and (2.152) through (2.155).

**11.2.12.** Verify (2.168) using (2.163), (2.166), and (T.1.29).

**11.2.13.** Emboldened by the remarkable results in Subsection 2.8, achieved by combining everything into one Lie element using the exponential map, the purpose to the exercise is to explore what happens when everything is combined into one Lie element using the Cayley map. We will call the result *Cayley Lie Runge Kutta*. Again we will concentrate our efforts on  $SU(2)$  for ease of computation.

Begin by assuming that  $\hat{u}^v(\tau)$  has a *Cayley Taylor* approximation of the form

$$\hat{u}^v(\tau) \simeq \hat{u}^{vcay}(\tau) = (I + \boldsymbol{\mu} \cdot \mathbf{K})/(I - \boldsymbol{\mu} \cdot \mathbf{K}) \quad (11.2.350)$$

where

$$\boldsymbol{\mu}(\tau) = \mathbf{f}_0\tau + \mathbf{f}_1\tau^2 + \cdots + \mathbf{f}_M\tau^{M+1}, \quad (11.2.351)$$

$$\boldsymbol{\mu}(H) = \mathbf{f}_0H + \mathbf{f}_1H^2 + \cdots + \mathbf{f}_MH^{M+1}, \quad (11.2.352)$$

and the coefficients  $\mathbf{f}_n$  are to be determined. See (3.12.71).

The coefficients  $\mathbf{f}_n$  could be determined in terms of the  $\mathbf{c}_m$  by integrating equations of the form (11.1.71). [.] Alternatively, we may assume that we already know  $\hat{u}^v(H)$  in exponential form and therefore can use a relation of the form (3.12.73). Specifically, show that we may write

$$\begin{aligned} \boldsymbol{\mu} &= 2(\Omega/|\Omega|) \tan(|\Omega|/4) \\ &= 2(\Omega/|\Omega|)[(|\Omega|/4) + (1/3)(|\Omega|/4)^3 + (2/15)(|\Omega|/4)^5 + \cdots] \\ &= (1/2)(\Omega)[1 + *|\Omega|^2 + *|\Omega|^4 + \cdots] \end{aligned} \quad (11.2.353)$$

with  $\Omega$  given, through terms of order  $H^4$ , by (2.162). Show , through terms of order  $H^4$ , that

$$|\Omega|^2 =, \quad (11.2.354)$$

$$|\Omega|^4 =, \quad (11.2.355)$$

Use these relations and (2.312) to obtain the results

$$\mathbf{f}_0 =, \quad (11.2.356)$$

$$\mathbf{f}_1 =, \quad (11.2.357)$$

$$\mathbf{f}_2 =, \quad (11.2.358)$$

$$\mathbf{f}_3 = . \quad (11.2.359)$$

**11.2.14.** Subsection 2.9 treated the integration of (2.183) in its Lie algebra using both exponential and Cayley representations. Carry out the analogous tasks for the equation of motion

$$\dot{Y}(t) = Y(t)A(Y,t). \quad (11.2.360)$$

Hint: Use (\*) in Appendix C and (1.312) in Exercise 1.12.

**11.2.15.** The Schroedinger equation reads

$$d\psi/dt = (-iH/\hbar)\psi. \quad (11.2.361)$$

Here the state vector  $\psi$  is supposed to belong to some Hilbert space (a vector space equipped with an inner product), and  $(-iH/\hbar)$  is supposed to be an anti-Hermitian operator with respect to this inner product. Consider the case for which the Hilbert space is two dimensional. Then  $\psi$  is a two-component vector, and we may take the inner product  $\langle *, * \rangle$  to be the usual complex inner product for finite-dimensional vector spaces over the complex field. Also,  $(-iH/\hbar)$  is then a  $2 \times 2$  anti-Hermitian matrix. Show that the most general such matrix is of the form

$$-iH/\hbar = i\delta(t)I + \check{\omega}(t) \cdot \mathbf{K} \quad (11.2.362)$$

where  $\delta$  is any real possibly time-dependent number and  $\check{\omega}$  is any real possibly time-dependent three-dimensional vector. Thus, for a two-dimensional Hilbert space, the most general Schroedinger equation reads

$$d\psi/dt = [i\delta(t)I + \check{\omega}(t) \cdot \mathbf{K}]\psi. \quad (11.2.363)$$

Our task is to find  $\psi(t)$  given

$$\psi^0 = \psi(t^0). \quad (11.2.364)$$

Since (2.172) is linear, we may write  $\psi(t)$  in the form

$$\psi(t) = v(t)\psi^0 \quad (11.2.365)$$

where  $v$  is a  $2 \times 2$  matrix to be determined. Show that the general solution to (2.172) is of the form (2.174) providing  $v$  satisfies the equation

$$\dot{v} = [i\delta(t)I + \check{\omega}(t) \cdot \mathbf{K}]v \quad (11.2.366)$$

with the initial condition

$$v(t^0) = I. \quad (11.2.367)$$

Verify that  $v$  defined by (2.175) and (2.176) is an element in  $U(2)$ , and therefore  $\langle \psi, \psi \rangle$  is preserved as is necessary for a probability interpretation of the state vector.<sup>33</sup>

Suppose that the  $\delta$  term in (2.172) and (2.175) is omitted to yield differential equations of the form

$$\dot{\psi}' = \check{\omega} \cdot \mathbf{K} \psi', \quad (11.2.368)$$

with the same initial condition

$$\psi'(t^0) = \psi^0, \quad (11.2.369)$$

and

$$\dot{v}' = \check{\omega}(t) \cdot \mathbf{K} v', \quad (11.2.370)$$

with the same initial condition

$$v'(t^0) = I. \quad (11.2.371)$$

We already know from Exercise 2.3 that  $v'$  defined by (2.179) and (2.180) is an element in  $SU(2)$ . Verify that

$$\psi'(t) = v'(t)\psi^0. \quad (11.2.372)$$

How are  $\psi$  and  $\psi'$ , and  $v$  and  $v'$ , related? Define a quantity  $\Delta(t)$  by the rule

$$\Delta(t) = \int_{t^0}^t dt' \delta(t'). \quad (11.2.373)$$

Verify the relations

$$\psi(t) = \exp[i\Delta(t)]\psi'(t) \quad (11.2.374)$$

and

$$v(t) = \exp[i\Delta(t)]v'(t). \quad (11.2.375)$$

We see that  $\psi$  and  $\psi'$ , and also  $v$  and  $v'$ , differ only by a phase factor. Since overall phase factors are supposed to be unobservable in quantum mechanics, we conclude there is no loss in generality in omitting the  $\delta$  term. Thus, without loss of quantum-mechanical generality, we may drop primes and require that  $v$  satisfies the differential equation

$$\dot{v} = \check{\omega}(t) \cdot \mathbf{K} v \quad (11.2.376)$$

with the initial condition (2.176). Correspondingly, the Schrödinger equation now reads

$$\dot{\psi} = \check{\omega} \cdot \mathbf{K} \psi, \quad (11.2.377)$$

and has the general solution (2.174) with  $v \in SU(2)$ .

By construction, the relations (2.185) and (2.186) entail each other. But we now observe that the relations (2.185) and (2.176) have the same form as (2.52) and (2.53). Therefore, relations of the form (2.52) and (2.53) may also be viewed as arising from the most general Schrödinger equation in the case of a two-dimensional Hilbert space.

---

<sup>33</sup>That is why  $(-iH/\hbar)$  is required to be anti-Hermitian.

What can be said about the nature of two-dimensional Hilbert space? Let  $\psi^\uparrow$  and  $\psi^\downarrow$  be the orthonormal basis vectors

$$\psi^\uparrow = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad (11.2.378)$$

$$\psi^\downarrow = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (11.2.379)$$

Write the most general  $\psi$  in the form

$$\psi = \alpha\psi^\uparrow + \beta\psi^\downarrow. \quad (11.2.380)$$

Since both  $\alpha$  and  $\beta$  are complex, they each have the topology of  $\mathbb{C} = E^2$ , and together they have the topology of  $\mathbb{C} \times \mathbb{C} = E^4$ . Show that requiring that  $\psi$  be a unit vector, which is necessary for a probability interpretation of quantum mechanics, yields the restriction

$$|\alpha|^2 + |\beta|^2 = (\Re\alpha)^2 + (\Im\alpha)^2 + (\Re\beta)^2 + (\Im\beta)^2 = 1. \quad (11.2.381)$$

Observe that (2.190) is the equation for  $S^3$ , the three-dimensional surface of a sphere in  $E^4$ . Thus the topology of unit vectors in two-dimensional Hilbert space is that of  $S^3$ .

Show that the general solution to the restriction (2.190) is given by

$$\alpha = \exp(i\gamma) \cos(\theta/2), \quad (11.2.382)$$

$$\beta = \exp(i\gamma) \exp(i\phi) \sin(\theta/2), \quad (11.2.383)$$

where  $\gamma$ ,  $\theta$ , and  $\phi$  are real, but otherwise arbitrary. Actually, not all values of  $\gamma$ ,  $\theta$ , and  $\phi$  are required for full generality. Show that all possibilities are covered by making the restrictions  $\gamma \in [0, 2\pi]$ ,  $\theta \in [0, \pi]$ ,  $\phi \in [0, 2\pi]$ . Moreover, since overall phase factors are unobservable, we may set  $\gamma = 0$  if we wish, and without loss of quantum-mechanical generality. Thus, the set (collection of equivalence classes) of all *unit rays* can be written in the form

$$\psi = \cos(\theta/2)\psi^\uparrow + \exp(i\phi) \sin(\theta/2)\psi^\downarrow \quad (11.2.384)$$

with

$$\theta \in [0, \pi] \text{ and } \phi \in [0, 2\pi]. \quad (11.2.385)$$

(Recall that a unit ray is the equivalence class of a unit vector multiplied by an arbitrary phase factor.) The quantities  $\theta$  and  $\phi$  with the restrictions (2.194) may be regarded as the polar angles for points on  $S^2$ , the unit sphere in  $E^3$ . That is, in terms of Cartesian coordinates for  $E^3$ , we may write

$$x_1 = \sin(\theta) \cos(\phi), \quad (11.2.386)$$

$$x_2 = \sin(\theta) \sin(\phi), \quad (11.2.387)$$

$$x_3 = \cos(\theta). \quad (11.2.388)$$

Thus, the topology of unit rays in two-dimensional Hilbert space is that of  $S^2$ . In the context of quantum mechanics, this  $S^2$  is frequently called the *Bloch sphere*. (It is the Poincaré sphere in the context of describing polarized light.) Show that the north pole  $(0, 0, 1)$  of the

Bloch sphere corresponds to the state  $\psi^\uparrow$  and the south pole  $(0, 0, -1)$  corresponds (up to an irrelevant phase factor) to the state  $\psi^\downarrow$ . Show that points on the equator correspond to the states  $(1/\sqrt{2})(\psi^\uparrow + \exp(i\phi)\psi^\downarrow)$ . Show that the vectors  $\psi$  corresponding to diametrically opposite points on the Bloch sphere are orthogonal.

Let  $\psi(\theta, \phi)$  denote the vector given by (2.384). Show that, in terms of the Euler-angle parameterization given by (3.7.194) and (3.7.195), there is the relation

$$\psi(\theta, \phi) = v(\phi, \theta, -\phi)\psi^\uparrow. \quad (11.2.389)$$

Let  $\mathbf{n}$  denote the unit vector defined by the relation

$$\mathbf{n} = R(\phi, \theta, -\phi)\mathbf{e}_3 = \cos\phi\sin\theta\mathbf{e}_1 + \sin\phi\sin\theta\mathbf{e}_2 + \cos\theta\mathbf{e}_3. \quad (11.2.390)$$

See (3.7.208) and note the resemblance between (2.390) and (2.386) through (2.388). Show that there is the relation

$$(\mathbf{n} \cdot \boldsymbol{\sigma})\psi(\theta, \phi) = \psi(\theta, \phi). \quad (11.2.391)$$

That is,  $\psi(\theta, \phi)$  is an eigenvector of  $\mathbf{n} \cdot \boldsymbol{\sigma}$  with eigenvalue +1.

Verify the coset relation

$$SU(2)/U(1) = S^2. \quad (11.2.392)$$

See Subsection 5.12..4. Show that the correspondence between unit rays in a two-dimensional Hilbert space and points in  $S^2$  is a consequence of this coset relation.

We are ready for a parenthetical remark about the field of quantum computing and quantum information. Let  $|0\rangle$  and  $|1\rangle$  be the *qubit* (quantum bit) states corresponding to the states 0 and 1 of a classical bit.<sup>34</sup> In this field it is conventional to define the vectors  $|0\rangle$  and  $|1\rangle$  by the relations

$$|0\rangle = \psi^\uparrow, \quad (11.2.393)$$

$$|1\rangle = \psi^\downarrow; \quad (11.2.394)$$

and a general qubit state  $|\psi\rangle$  takes the superposition form

$$|\psi\rangle = \cos(\theta/2)|0\rangle + \exp(i\phi)\sin(\theta/2)|1\rangle. \quad (11.2.395)$$

Back to the main thread. There is still more to be said. Define a three-component vector  $\mathbf{s}$  by the rule

$$\mathbf{s} = 2i\langle\psi, \mathbf{K}\psi\rangle = \langle\psi, \boldsymbol{\sigma}\psi\rangle. \quad (11.2.396)$$

Show that  $\mathbf{s}$  is real, and does not depend on the phase of  $\psi$ . Show that using (2.193) for  $\psi$  and evaluating (2.201) gives the results (2.10) through (2.12). Thus,  $\mathbf{s}$  is a unit vector with polar angles  $\theta$  and  $\phi$ . We see that  $\mathbf{s}$  is a point on the Bloch sphere, and can be any point on the Bloch sphere. Conversely, given any point on the Bloch sphere, we can determine its polar angles  $\theta$  and  $\phi$ , and from these angles we can determine  $\psi$  up to an overall phase factor.

---

<sup>34</sup>A qubit is a quantum-mechanical system that is well described by a two-dimensional Hilbert space. Examples include spin 1/2 particles and polarized light, plus more complicated systems, including atoms and superconducting devices, that can be effectively regarded as two-state systems consisting of a ground state and first excited state, with higher excited states ignorable because they are well separated in energy from these states.

Therefore, for a two-dimensional Hilbert space, knowledge of a unit ray and knowledge of a unit vector  $\mathbf{s}$  are equivalent.

Finally, suppose  $\psi$  evolves according to (2.186). How does the corresponding  $\mathbf{s}$  evolve? Show, using (2.186) and (2.201), that in terms of components

$$\begin{aligned}\dot{s}_\alpha &= 2i[\langle \dot{\psi}, K^\alpha \psi \rangle + \langle \psi, K^\alpha \dot{\psi} \rangle] \\ &= 2i[\langle \check{\omega} \cdot \mathbf{K} \psi, K^\alpha \psi \rangle + \langle \psi, K^\alpha \check{\omega} \cdot \mathbf{K} \psi \rangle] \\ &= 2i[\langle \psi, (\check{\omega} \cdot \mathbf{K})^\dagger K^\alpha \psi \rangle + \langle \psi, K^\alpha \check{\omega} \cdot \mathbf{K} \psi \rangle] \\ &= 2i[\langle \psi, (-\check{\omega} \cdot \mathbf{K}) K^\alpha \psi \rangle + \langle \psi, K^\alpha \check{\omega} \cdot \mathbf{K} \psi \rangle] \\ &= 2i\langle \psi, \{K^\alpha, \check{\omega} \cdot \mathbf{K}\} \psi \rangle.\end{aligned}\tag{11.2.397}$$

Next verify that

$$\{K^\alpha, \check{\omega} \cdot \mathbf{K}\} = \sum_\beta \{K^\alpha, K^\beta\} \check{\omega}_\beta = \sum_{\beta\gamma} \epsilon_{\alpha\beta\gamma} \check{\omega}_\beta K^\gamma.\tag{11.2.398}$$

Now combine (2.202) and (2.203) to show that

$$\dot{s}_\alpha = 2i \sum_{\beta\gamma} \epsilon_{\alpha\beta\gamma} \check{\omega}_\beta \langle \psi, K^\gamma \psi \rangle = \sum_{\beta\gamma} \epsilon_{\alpha\beta\gamma} \check{\omega}_\beta s_\gamma = (\check{\omega} \times \mathbf{s})_\alpha.\tag{11.2.399}$$

In vector notation, you have demonstrated that

$$\dot{\mathbf{s}} = \check{\omega} \times \mathbf{s},\tag{11.2.400}$$

a relation of the form (2.1). Moreover, since a knowledge of  $\mathbf{s}$  determines  $\psi$  up to a phase, given the equation (2.205) we may view it as arising from the Schroedinger equation (2.186). That is, given the  $\check{\omega}$  appearing in (2.205), we may insert it into the Schroedinger equation (2.186) to find  $\psi(t)$ , and this  $\psi(t)$  will yield  $\mathbf{s}(t)$  by way of (2.201). Thus, we may also view (2.186) and (2.205) as entailing each other.

Your last task in this exercise is to verify a relation between components of  $\psi$  and quaternion parameters  $w$ . Let  $\psi^u(t)$  and  $\psi^d(t)$  be solutions to the Schroedinger equation (2.186) with the initial conditions

$$\psi^u(t^0) = \psi^\uparrow,\tag{11.2.401}$$

$$\psi^d(t^0) = \psi^\downarrow.\tag{11.2.402}$$

Here the superscript mnemonics  $u$  and  $d$  stand for *up* and *down*. Suppose  $v$  is parameterized in terms of quaternions in analogy to (1.132). Verify, using (1.174), (2.187), (2.188), and (5.10.29), that there are the relations

$$\langle \psi^\uparrow, \psi^u \rangle = v_{11} = w_0 + iw_3,\tag{11.2.403}$$

$$\langle \psi^\uparrow, \psi^d \rangle = v_{12} = iw_1 + w_2,\tag{11.2.404}$$

$$\langle \psi^\downarrow, \psi^u \rangle = v_{21} = iw_1 - w_2,\tag{11.2.405}$$

$$\langle \psi^\downarrow, \psi^d \rangle = v_{22} = w_0 - iw_3.\tag{11.2.406}$$

Solve (2.208) through (2.211) to yield the relations

$$w_0(t) = (1/2)(\langle \psi^\uparrow, \psi^u \rangle + \langle \psi^\downarrow, \psi^d \rangle), \quad (11.2.407)$$

$$w_1(t) = (-i/2)(\langle \psi^\uparrow, \psi^d \rangle + \langle \psi^\downarrow, \psi^u \rangle), \quad (11.2.408)$$

$$w_2(t) = (1/2)(\langle \psi^\uparrow, \psi^d \rangle - \langle \psi^\downarrow, \psi^u \rangle) \quad (11.2.409)$$

$$w_3(t) = (-i/2)(\langle \psi^\uparrow, \psi^u \rangle - \langle \psi^\downarrow, \psi^d \rangle). \quad (11.2.410)$$

### 11.2.16. Exact solutions to (2.52).

**11.2.17.** Consider a Stern-Gerlach apparatus in which the beam propagates in the  $y$  direction, the main magnetic field is in the  $z$  direction, and also has a gradient in the  $z$  direction. Such a field, when expanded about the beam axis (taken to be  $x = z = 0$ ) and near the beam axis, consists essentially of a (skew) quadrupole field superimposed on a dipole field. That is, ignoring end effects, the magnetic field has, to the lowest nontrivial order, the expansion

$$\mathbf{B}(\mathbf{r}) = B^d \mathbf{e}_z + Q^q(z \mathbf{e}_z - x \mathbf{e}_x). \quad (11.2.411)$$

Here  $B^d$  is the strength of the main (dipole) field, and  $Q^q$  is the strength (field gradient) of the quadrupole field. Verify that this field is divergence and curl free as is desired. Note that there is a desired field gradient in the  $z$  direction, intended to produce a Stern-Gerlach force along the  $z$  direction, as well as an “undesirable” gradient along the  $x$  direction that will produce a Stern-Gerlach force along the  $x$  direction.

$$\Psi = f^\uparrow(\mathbf{r}, t)\psi^\uparrow + f^\downarrow(\mathbf{r}, t)\psi^\downarrow. \quad (11.2.412)$$

The Schrödinger equation reads

$$\partial\Psi/\partial t = (-i/\hbar)H\Psi \quad (11.2.413)$$

where

$$H = \mathbf{p} \cdot \mathbf{p}/(2m) + \mu \mathbf{B}(\mathbf{r}) \cdot \boldsymbol{\sigma}. \quad (11.2.414)$$

Here  $m$  is the particle mass and  $\mu$  is a measure of its magnetic moment. Indeed, suppose we ignore the kinetic energy term in (2.414). Then we find that

$$-iH/\hbar \approx -i(\mu/\hbar)\mathbf{B}(\mathbf{r}) \cdot \boldsymbol{\sigma} = (2\mu/\hbar)\mathbf{B}(\mathbf{r}) \cdot \mathbf{K}. \quad (11.2.415)$$

Upon comparing (2.362) with (2.415) we see that we should make the identification

$$\check{\boldsymbol{\omega}} = (2\mu/\hbar)\mathbf{B}. \quad (11.2.416)$$

$$\partial f^\uparrow(\mathbf{r}, t)/\partial t = (-i/\hbar)\{[\hbar^2/(2m)]\nabla^2 f^\uparrow(\mathbf{r}, t) + \mu(B^d + Q^q z)f^\uparrow(\mathbf{r}, t) + \mu Q^q x f^\downarrow(\mathbf{r}, t)\}, \quad (11.2.417)$$

$$\partial f^\downarrow(\mathbf{r}, t)/\partial t = (-i/\hbar)\{[\hbar^2/(2m)]\nabla^2 f^\downarrow(\mathbf{r}, t) - \mu(B^d + Q^q z)f^\downarrow(\mathbf{r}, t) + \mu Q^q x f^\uparrow(\mathbf{r}, t)\}. \quad (11.2.418)$$

## 11.3 Numerical Integration on Manifolds: Charged Particle Motion in a Static Magnetic Field

### Overview

Sections 1.1 and 1.2 treated the problem of integrating on manifolds largely either by projection or by integration in the Lie algebra or by parameterizing the manifold in question, finding the associated differential equation for the parameters, and integrating these equations using standard integration algorithms such as those described in Chapter 2. The only exception to this approach was the work of Subsections 2.6 through 2.8. The purpose of the section is to describe how manifold-preserving integration methods may be applied to the problem of charged-particle motion in a static magnetic field.

The reader may have been somewhat puzzled by the assertion (made at the beginning of Section 1.2) that (1.80) was like (1.79). The equation of motion (1.6.112) is equivalent to the pair

$$d\mathbf{r}/dt = \mathbf{v}, \quad (11.3.1)$$

$$d\mathbf{v}/dt = \bar{\boldsymbol{\omega}}(\mathbf{r}) \times \mathbf{v}, \quad (11.3.2)$$

with

$$\bar{\boldsymbol{\omega}}(\mathbf{r}) = -(q/m^*)\mathbf{B}(\mathbf{r}). \quad (11.3.3)$$

Since  $\bar{\boldsymbol{\omega}}$  depends on  $\mathbf{r}$  and, according to (1.148),  $\mathbf{r}$  in turn depends on  $\mathbf{v}$ , the  $\bar{\boldsymbol{\omega}}$  appearing in (1.149) is *not* a given function of  $t$  independent of everything else including  $\mathbf{v}$ . However, it is the case that the pair (1.148) and (1.149) does preserve the quantity  $\mathbf{v} \cdot \mathbf{v}$ . That is, the pair preserves the manifold

$$\Gamma = E^3 \times S^{2*} \quad (11.3.4)$$

with  $\mathbf{r} \in E^3$  and  $\mathbf{v} \in S^{2*}$ . Here  $S^{2*}$  denotes a 2-sphere whose radius is given by  $v^* = |\mathbf{v}^0|$ , or equivalently, is determined by  $\gamma = m^*/m$  with

$$\gamma = 1/\sqrt{1 - |\mathbf{v}^0|^2/c^2}. \quad (11.3.5)$$

See (1.6.113) and (1.6.114).

In Subsection 3.1 we will describe how the methods already developed in Sections 1 and 2 can be exploited to provide numerical integrators that preserve  $\Gamma$ . In Subsection 3.2 we will describe splitting methods that also preserve  $\Gamma$  but are more akin to some of the methods for symplectic integration to be described in Chapter 12.

### 11.3.1 Exploitation of Previous Results

The purpose of this subsection is to describe how the methods of Sections 1 and 2 can be applied to the computation of charged-particle motion in static magnetic fields. We will begin with the use of local tangent-space coordinates as illustrated in Subsection 2.3.

### Use of Local Tangent-Space Coordinates

In analogy with the work of Subsection 2.3, let  $\mathbf{v}^b$  be the velocity at the *beginning* of an integration step, and write

$$\mathbf{v}(t) = \mathbf{v}^b + \mathbf{v}^v(t) \quad (11.3.6)$$

with

$$\mathbf{v}^b = \mathbf{v}(t^b) \quad (11.3.7)$$

and

$$\mathbf{v}^v(t^b) = 0. \quad (11.3.8)$$

(Here we apologize for our notation: As before, when  $v$  appears as a superscript, it stands for *variable*. Elsewhere it denotes vector or scalar *velocity*.) It follows that we may also write

$$\mathbf{v}^v(t) = v_1^{vf}(t)\mathbf{f}_1 + v_2^{vf}(t)\mathbf{f}_2 + v_3^{vf}(t)\mathbf{f}_3 \quad (11.3.9)$$

and

$$\mathbf{v}(t) = [v^* + v_1^{vf}(t)]\mathbf{f}_1 + v_2^{vf}(t)\mathbf{f}_2 + v_3^{vf}(t)\mathbf{f}_3 \quad (11.3.10)$$

with

$$v_1^{vf}(t) = \{(v^*)^2 - [v_2^{vf}(t)]^2 - [v_3^{vf}(t)]^2\}^{1/2} - v^*. \quad (11.3.11)$$

Here, *mutatis mutandis*, the vectors  $\mathbf{f}_j$  are constructed as in Subsection 2.3.

We have (locally) parameterized  $S^{2*}$  in terms of  $v_2^{vf}$  and  $v_3^{vf}$ . Put another way, we have changed variables from  $\mathbf{v}$  to  $v_2^{vf}$  and  $v_3^{vf}$  in such a way that the  $S^{2*}$  manifold condition is built in. Corresponding to this change of variables, the  $\mathbf{v}$  equations of motion (3.2) becomes the pair

$$\dot{v}_2^{vf}(t) = \bar{\omega}_3^f(\mathbf{r})\{(v^*)^2 - [v_2^{vf}(t)]^2 - [v_3^{vf}(t)]^2\}^{1/2} - \bar{\omega}_1^f(\mathbf{r})v_3^{vf}(t), \quad (11.3.12)$$

$$\dot{v}_3^{vf}(t) = \bar{\omega}_1^f(\mathbf{r})v_2^{vf}(t) - \bar{\omega}_2^f(\mathbf{r})\{(v^*)^2 - [v_2^{vf}(t)]^2 - [v_3^{vf}(t)]^2\}^{1/2}. \quad (11.3.13)$$

For the  $\mathbf{r}$  variables make the decomposition

$$\mathbf{r}(t) = \mathbf{r}^b + \mathbf{r}^v(t) \quad (11.3.14)$$

with

$$\mathbf{r}^b = \mathbf{r}(t^b) \quad (11.3.15)$$

and

$$\mathbf{r}^v(t^b) = 0. \quad (11.3.16)$$

For  $\mathbf{r}^v$  make the expansion

$$\mathbf{r}^v = r_1^{vf}\mathbf{f}_1 + r_2^{vf}\mathbf{f}_2 + r_3^{vf}\mathbf{f}_3. \quad (11.3.17)$$

Then, in view of (3.13) and (3.14), the equations of motion (3.1) for  $\mathbf{r}$  become the set

$$\dot{r}_1^{vf} = v_1^{vf} = \{(v^*)^2 - [v_2^{vf}(t)]^2 - [v_3^{vf}(t)]^2\}^{1/2} - v^*, \quad (11.3.18)$$

$$\dot{r}_2^{vf} = v_2^{vf}, \quad (11.3.19)$$

$$\dot{r}_3^{vf} = v_3^{vf}. \quad (11.3.20)$$

It is these five equations that are to be numerically integrated from the time  $t^b$  to the time  $t^b + h$  (or perhaps  $t^b + kh$ ) starting with the initial conditions  $v_2^{vf}(t^b) = v_3^{vf}(t^b) = 0$  and  $r_1^{vf}(t^b) = r_2^{vf}(t^b) = r_3^{vf}(t^b) = 0$ . Then, once  $\mathbf{v}^v(t^b + h)$  and  $\mathbf{r}^v(t^b + h)$  [or perhaps  $\mathbf{v}^v(t^b + kh)$  and  $\mathbf{r}^v(t^b + kh)$ ] have been obtained,  $\mathbf{v}(t^b + h)$  and  $\mathbf{r}(t^b + h)$  [or perhaps  $\mathbf{v}(t^b + kh)$  and  $\mathbf{r}(t^b + kh)$ ] are given by (2.20) and \*<sup>35</sup>. At this point, the whole process just described is repeated as often as desired. That is, the vectors  $\mathbf{f}_j$  are reconstructed based on the most recently obtained  $\mathbf{v}$ , etc. Note that any numerical integration method may be used to carry out the desired integration. By construction, although the numerical integration may make local errors of order  $h^{m+1}$  in finding  $\mathbf{v}(t^b + h)$  and  $\mathbf{r}(t^b + h)$ , these quantities are guaranteed to be in the manifold  $\Gamma = E^3 \times S^{2*}$  to machine precision.

### Use of Connection with Rigid-Body Kinematics

An alternative to the use of local tangent-space coordinates to ensure that  $\mathbf{v}$  and  $\mathbf{r}$  lie in  $\Gamma$  is to make an Ansatz for  $\mathbf{v}$  that involves rotations. We will seek to use the results of Subsections 2.4 and 2.5.

Again let  $\mathbf{v}^b$  be the velocity at the *beginning* of an integration step,

$$\mathbf{v}^b = \mathbf{v}(t^b), \quad (11.3.21)$$

and write

$$\mathbf{v}(t) = R(t; \mathbf{v}^b, \mathbf{r}^b) \mathbf{v}^b \quad (11.3.22)$$

with and  $R(t; \mathbf{v}^b, \mathbf{r}^b)$  being a rotation to be determined subject to the initial condition

$$R(t^b; \mathbf{v}^b, \mathbf{r}^b) = I. \quad (11.3.23)$$

At this point, explanations are in order both about the notation  $R(t; \mathbf{v}^b, \mathbf{r}^b)$  and the generality of the Ansatz (3.27). The notation is meant to indicate that the relation (3.27) between  $\mathbf{v}(t)$  and  $\mathbf{v}^b$  need not (and generally will not) be linear because  $R(t; \mathbf{v}^b, \mathbf{r}^b)$  can depend on  $\mathbf{v}^b$  and  $\mathbf{r}^b$ .

#### 11.3.2 Splitting: Exploitation of Future Results

### Exercises

However, before doing so, let us consider the cost of implementing (1.156). The evaluation of  $\exp[h\bar{\omega}(\mathbf{r}^n) \cdot \mathbf{L}]$  can be performed using the Rodrigues formula (3.7.202). This evaluation is fairly expensive because it involves, among other things, the evaluation of trigonometric functions. Alternatively, (1.153) could be rewritten in the form

$$\mathbf{v}^{n+1} = [I + h\bar{\omega}(\mathbf{r}^n) \cdot \mathbf{L}] \mathbf{v}^n. \quad (11.3.24)$$

To the same accuracy, one could orthogonalize the matrix  $[I + h\bar{\omega}(\mathbf{r}^n) \cdot \mathbf{L}]$  using one of the methods of Subsection 3.6.4, and then apply this matrix to  $\mathbf{v}^n$  to obtain  $\mathbf{v}^{n+1}$ . These

---

<sup>35</sup>If  $k > 1$  is attempted, one must monitor  $[(v_2^{vf})^2 + (v_3^{vf})^2]$  to ensure that the square root singularity in (2.30) is not approached too closely.

methods involve the computation of square roots. Finally, as already mentioned, at each step one could simply renormalize the  $\mathbf{v}^{n+1}$  in (1.153) by simple scaling to project it back onto  $S^{2*}$ . So doing requires the evaluation of a square root.

**11.3.1.** Exercise on the growth of  $\mathbf{v}$  dot  $\mathbf{v}$  based on Euler result (1.154).



# Bibliography

General References. See also the references for Chapter 12

- [1] E. Celledoni and B. Owren, “Lie group methods for rigid body dynamics and time integration on manifolds”, *Computer Methods in Applied Mechanics and Engineering* **192**, 421–438 (2003).
- [2] E. Celledoni, H. Marthinsen, and B. Owren, “An introduction to Lie group integrators—basics, new developments, and applications”, <http://arxiv.org/pdf/1207.0069.pdf>.
- [3] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure Preserving Algorithms for Ordinary Differential Equations*, Second Edition, Springer (2006).
- [4] H. Munthe-Kaas, A. Zanna, “Numerical integration of differential equations on homogeneous manifolds”, *Foundations of Computational Mathematics*, Springer Verlag (1997).
- [5] H. Munthe-Kaas and B. Owren, “Computations in a free Lie algebra”, *Phil. Trans. R. Soc. Lond. A* **357**, 957–981 (1999).
- [6] A. Iserles, H. Munthe-Kaas, S. Nørsett, and A. Zanna, “Lie-group methods”, *Acta Numerica* **14**, 1–148 (2005). Also available on the Web at [http://www.damtp.cam.ac.uk/user/na/NA\\_papers/NA2000\\_03.pdf](http://www.damtp.cam.ac.uk/user/na/NA_papers/NA2000_03.pdf).

Factored Lie Runge Kutta

- [7] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure Preserving Algorithms for Ordinary Differential Equations*, Second Edition, Springer (2006). See Section IV.8.
- [8] P.E. Crouch and R. Grossman, “Numerical integration of ordinary differential equations on manifolds”, *J. Nonlinear Sci.* **3**, 1 (1993).
- [9] B. Owren and A Marthinsen, “Runge-Kutta methods adapted to manifolds and based on rigid frames ”, *BIT* **39**, 116 (1999).

Magnus Lie Runge Kutta

- [10] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure Preserving Algorithms for Ordinary Differential Equations*, Second Edition, Springer (2006). See Section IV.7.
- [11] A. Zanna, “Collocation and relaxed collocation for the Fer and the Magnus expansions”, *SIAM J. Numer. Anal.* **36**, 1145 (1999).
- [12] S. Blanes, F. Casas, and J. Ros, “Improved high order integrators based on the Magnus expansion”, *BIT* **40**, 434 (2000).

#### Spin and Qubits

- [13] S. Mane, Yu. Shatunov, and K. Yokoya, “Spin-polarized charged particle beams in high-energy accelerators”, *Rep. Prog. Phys.* **68**, 1997 (2005).
- [14] D.P. Barber, References to and lectures on spin: <http://www.desy.de/~mpybar/>, [http://www.desy.de/~mpybar/psdump/Combined\\_CI\\_2012\\_16.pdf](http://www.desy.de/~mpybar/psdump/Combined_CI_2012_16.pdf).
- [15] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*, Cambridge (2000).

#### Stern-Gerlach Effect

- [16] N. Mott and H. Massey, *The Theory of Atomic Collisions*, Third Edition, Oxford (1965).
- [17] J. Kessler, *Polarized Electrons*, Second Edition, Springer-Verlag (1985).
- [18] H. Dehmelt, “Continuous Stern-Gerlach effect: Principle and idealized apparatus”, *Proc. Natl. Acad. Sci. USA*, Vol. 83, pp. 2291-2294 (1986).
- [19] H. Batelaan, T. Gay, and J. Schwendiman, “Stern-Gerlach Effect for Electron Beams”, *Phys. Rev. Let.* **79**, pp. 4517-4521 (1997).
- [20] G. Gallup, H. Batelaan, and T. Gay, “Quantum-Mechanical Analysis of a Longitudinal Stern-Gerlach Effect”, *Phys. Rev. Let.* **86**, pp. 4508-4511 (2001).
- [21] G. Werth, H. Häffner, and W. Quint, “Continuous Stern-Gerlach Effect on Atomic Ions”, *Advances in Atomic, Molecular, and Optical Physics*, Vol. 48, pp. 191-217 (2002).
- [22] W. Quint, J. Alonso, S. Djekić, H.-J. Kluge, S. Stahl, T. Valenzuela, J. Verdú, M. Vogel, and G. Werth, “Continuous Stern-Gerlach effect and the magnetic moment of the antiproton”, *Nucl. Instru. Meth. Phys. Res. B*, Vol. 214, pp. 207-210 (2004).
- [23] M. Kohda, S. Nakamura, Y. Nishihara, K. Lobayashi, T. Ono, J. Ohe, Y. Tokuram, T. Mineno, and J. Nitta, “Spin-orbit induced electronic spin separation in semiconductor nanostructures”, *Nature Communications* **3** (2012).

# Chapter 12

## Geometric/Structure-Preserving Integration: Symplectic Integration

### Overview

Imagine we wish to compute, by numerical integration, a trajectory governed by some Hamiltonian. Suppose  $z^i$  is an initial condition and  $z^f$  is an associated final condition. A numerical integrator is called a *symplectic integrator* for this Hamiltonian if the relation between  $z^i$  and  $z^f$  produced by use of this integrator is (to within roundoff errors) a symplectic map. Sometimes we will want an integrator to be a symplectic integrator for general Hamiltonians. Often, as we will see, it suffices to have a symplectic integrator for some class of Hamiltonians.

There are many things that might be said about symplectic integrators. Indeed, books have been and are being written on the subject. However, we must limit our discussion to a single chapter. Are symplectic integrators important, and if so, why? The answers to these questions are not fully known. As we have seen in Chapter 6, the production of symplectic maps is a key feature of Hamiltonian systems, and the preservation of this feature by any approximation scheme, including numerical integration, would appear to be highly desirable. To date, much experience with symplectic integrators, particularly when one is concerned with studying the long-term behavior of trajectories, seems to confirm this belief. A particular aspect of this subject is discussed in Chapter 34. For a broader discussion, see the references listed at the end of the present chapter. However, as might be feared, satisfying the symplectic condition comes at a cost. We will see that in general  $m$ th-order but exactly symplectic integrators require much more work (many more function evaluations) than the  $m$ th-order methods of Chapter 2, and are therefore considerably slower.<sup>1</sup>

What can be said about the numerical integration methods of Chapter 2 when applied to Hamiltonian systems? In general, they are not symplectic. Typically, at each step, they violate the symplectic condition by an amount of order  $h^{m+1}$  if the integration method is

---

<sup>1</sup>It is sometimes argued that this cost is compensated by the possibility of using larger step sizes in symplectic integration. Although the solution thereby obtained may not be particularly accurate, it is at least qualitatively correct whereas solutions obtained by other integration methods may exhibit spurious damping or spurious growth. Whether this trade-off can be realized or is acceptable depends on the problem being considered.

locally correct through terms of order  $h^m$ . Consequently, they are not exactly symplectic for any finite value of  $h$ .<sup>2</sup>

Are there modifications of the integration methods of Chapter 2 that make them symplectic? It is known that, for *general* Hamiltonians, there are no *explicit* Runge-Kutta methods that are symplectic. See Exercise 3.1. However, we will learn that there are *implicit* Runge-Kutta methods that are symplectic. Like corrector methods, at each step implicit Runge-Kutta methods require iteration, which may be slow, in order to solve a set of implicit equations.

What about the usual predictor-corrector finite-difference methods? It is known that they cannot be modified to be symplectic. However, as will be discussed, it is possible to use a predictor (that employs, as usual, previously stored trajectory data) along with an implicit symplectic Runge-Kutta method that serves the role of a corrector.

Finally, little seems to be known about possible symplectic modifications of extrapolation methods.

## 12.1 Splitting, $T + V$ Splitting, and Zassenhaus Formulas

This chapter is devoted to symplectic integration and the use of Zassenhaus formulas. In this section we begin with some background material, and then explore the case where the Hamiltonian is of the special but frequently encountered form  $H = T + V$ .

We have seen that if  $H(z)$  is a time-independent (autonomous) Hamiltonian, the transfer map  $\mathcal{M}$  associated with  $H$  can be written formally as

$$\mathcal{M} = \exp(t : -H :). \quad (12.1.1)$$

(Here, for convenience, we have taken the initial time to be  $t = 0$ , which can be done without loss of generality since  $H$  by assumption does not depend on  $t$ .) Suppose  $H$  is time dependent. Then we know that we may introduce a new independent variable  $\tau$ , extend the phase space to include  $t$  and  $p_t$  as dynamical variables, and introduce the new Hamiltonian  $K$  defined by the relation

$$K(t, q; p_t, p) = p_t + H(q, p, t). \quad (12.1.2)$$

(See Exercise 1.6.5.) Since  $K$  does not depend on  $\tau$ , we may write the transfer map associated with  $K$  formally as

$$\mathcal{M} = \exp(\tau : -K :). \quad (12.1.3)$$

At this point a remark is in order: Because (1.1) and (1.3) have identical structures, there seems to be no loss of generality in considering only the autonomous case. This assertion is correct if we have no particular concern about the form of the Hamiltonian. However, as seen in the last chapter, for the case (1.1) we are able to employ certain methods under the assumptions that  $H$  can be expanded about the origin  $z = 0$  and the term  $H_1$  is absent or

---

<sup>2</sup>Of course, in the limit  $h \rightarrow 0$  they are symplectic because they are then exact. They are also symplectic to machine precision, ignoring round-off error, when  $h$  is small enough for the integrator to be accurate to machine precision.

small, and these methods make it possible to find expansions for  $\mathcal{M}$  that can be evaluated explicitly. By contrast, these methods fail for the Hamiltonian (1.2) because the presence in  $K$  of the (linear) term  $p_t$  with a coefficient of *one* means that  $K_1$  cannot be regarded as being small. Nevertheless, if a method is capable of handling sufficiently general Hamiltonians, then there is no loss in generality in considering only autonomous Hamiltonians. Such will be the case for the methods in this section and therefore, without loss in generality, we assume that  $H$  is time independent.

As in Section 2.1, let us subdivide the  $t$  or  $\tau$  axis, whichever the case may be, into equal steps of duration  $h$ . Then, again employing the notation of Chapter 2, we have the exact marching rule

$$z^{n+1} = \exp(h : -H :) z^n. \quad (12.1.4)$$

Here  $H$  is to be viewed as a function of the variables  $z^n$ .

Equation (1.4) provides a stepping formula that can be used for numerical integration providing we have some method of computing or approximating  $\exp(h : -H :)$ . Suppose the Hamiltonian  $H$  can be *split* into (written as the sum of) two terms  $A$  and  $B$ ,

$$H(q, p) = A(q, p) + B(q, p), \quad (12.1.5)$$

in such a way that both the maps  $\exp(-h : A :)$  and  $\exp(-h : B :)$  can be evaluated explicitly or have some other desirable property. For example, suppose that  $H$  can be written as a sum of kinetic and potential energies,

$$H(q, p) = T(p) + V(q). \quad (12.1.6)$$

(See Exercise 1.1 for a review of when this is possible.) Then we have the exact results

$$\exp(-h : T :) q_i = q_i + h \partial T / \partial p_i, \quad (12.1.7)$$

$$\exp(-h : T :) p_i = p_i, \quad (12.1.8)$$

$$\exp(-h : V :) q_i = q_i, \quad (12.1.9)$$

$$\exp(-h : V :) p_i = p_i - h \partial V / \partial q_i. \quad (12.1.10)$$

Consider making the simple Zassenhaus approximation

$$\exp(h : -H :) = \exp(-h : A : -h : B :) \simeq \exp(-h : A :) \exp(-h : B :). \quad (12.1.11)$$

See Section 8.8. From the BCH formula (3.7.34) we have the result

$$\exp(-h : A :) \exp(-h : B :) = \exp[-h : A : -h : B : + (h^2/2) \{ : A :, : B : \} + O(h^3)]. \quad (12.1.12)$$

We see that, as a stepping formula, (1.4) with the approximation (1.11) makes local errors of order  $h^2$ . Thus, like the crude Euler method of Section 2.2, it can be used for numerical integration providing the step size is made sufficiently small. However, unlike crude Euler, the combination of (1.4) and (1.11) provides a method that is *exactly symplectic*. That is, the relation between  $z^f$  and  $z^i$  produced by this method is (apart from numerical roundoff errors) exactly a symplectic map for *any* value of  $h$ . See Exercise 1.2.

With this general background in mind, we now turn to the task of improving the approximation (1.11) to obtain integrators that are again exactly symplectic, but have higher order (in  $h$ ) accuracy. We will seek improved formulas of the Zassenhaus type, but will discover a method that is, in fact, more general. From the BCH formula we find the result

$$\begin{aligned} & \exp[-(h/2) : A :] \exp(-h : B :) \exp[-(h/2) : A :] \\ &= \exp[h : -(A + B) : + h^3(\{ : A :, \{ : A :, : B :\}) / 24 \\ &\quad - \{ : B :, \{ : B :, : A :\}) / 12) + O(h^4)]. \end{aligned} \quad (12.1.13)$$

Consequently, making the approximation

$$\exp(h : -H :) \simeq \exp[-(h/2) : A :] \exp(-h : B :) \exp[-(h/2) : A :], \quad (12.1.14)$$

sometimes called *Strang* splitting, produces a stepping formula that is again exactly symplectic (again assuming the individual factors can be evaluated exactly), but makes local errors of order  $h^3$ . This method, in somewhat different guise as the *leap-frog* algorithm, has been known for a long time, and long before the formal advent of symplectic integrators. See Exercise 1.3.

There is a reason why the coefficient of the  $h^2$  term on the right side of (1.13) vanishes. Introduce the notation

$$\mathcal{S}_2(h) = \exp[-(h/2) : A :] \exp(-h : B :) \exp[-(h/2) : A :]. \quad (12.1.15)$$

It is easily verified that the map  $\mathcal{S}_2$  satisfies the relation

$$\mathcal{S}_2^{-1}(h) = \mathcal{S}_2(-h). \quad (12.1.16)$$

A map that has the property

$$\mathcal{S}^{-1}(h) = \mathcal{S}(-h) \quad (12.1.17)$$

is called *symmetric*, and we have used the symbol  $\mathcal{S}$  for this reason.<sup>3</sup> We have also appended a subscript of 2 in (1.15) because, as we have seen, the use of  $\mathcal{S}_2$  furnishes us with an integrator

---

<sup>3</sup> Any integrator that is exact must satisfy (1.17) because exact integration backwards must send the final conditions back to the initial conditions. However, we are dealing with approximate integration that is only accurate through some power in  $h$ , and therefore (1.17) may or may not be true depending on what integration method is employed. It is not true for the explicit Runge-Kutta methods or Adams predictor-corrector methods or extrapolation methods described in Chapter 2. For any such method (1.17) holds only through terms of order  $h^m$  assuming any such method is locally accurate through terms of order  $h^m$ . We also remark that the property (1.17) is also sometimes referred to as *reversibility* or *time reversibility*. We avoid this usage, which can lead to confusion, because, properly speaking, reversibility and time reversibility are properties of particular *dynamical systems*, and not others. See Chapter 36. Being symmetric is a property of an integrator, and being reversible or time reversible is a property of a dynamical system, and therefore also of its associated (and exact) transfer map. Of course, if we are integrating a reversible or time reversible system by some approximate integrator, the resulting approximate transfer map may or may not be reversible or time reversible. In this context, it can be shown that there is an interplay between the symmetry of the integrator and the reversibility or time reversibility of the resulting approximate transfer map. Finally, we note that some authors call an integration method symmetric if its global error has an expansion in  $h$  that contains only even powers, as is the case for the extrapolation method of Section 2.6. See (2.6.10). This terminology is also potentially confusing because the extrapolation method is not symmetric in the sense (1.17) used here.

that is locally correct through terms of order  $h^2$ . We claim that *any* map that satisfies (1.17), when written in exponential form, must have an exponential expansion that involves only *odd* powers of  $h$ . That is, if we write

$$\mathcal{S}(h) = \exp[\mathcal{C}(h)], \quad (12.1.18)$$

then  $\mathcal{C}(h)$  must be odd in  $h$

$$\mathcal{C}(-h) = -\mathcal{C}(h). \quad (12.1.19)$$

To see the truth of this assertion, make the expansion

$$\mathcal{C}(h) = \sum_{m=1}^{\infty} \mathcal{C}_m h^m. \quad (12.1.20)$$

The series (1.20) has no constant term because, as can be seen from (1.4) and (1.15), we want to impose the condition

$$\mathcal{S}(0) = \mathcal{I}. \quad (12.1.21)$$

Now insert the representation (1.18) into (1.17). Doing so gives the result

$$\exp[-\mathcal{C}(h)] = \exp[\mathcal{C}(-h)]. \quad (12.1.22)$$

We conclude from (1.22) that (1.19) must hold as a result of the uniqueness of the exponent. See Appendix C. [At this point it may be remarked, in passing, that the  $O(h^4)$  term indicated in (1.13) must actually vanish so that the next possibly nonvanishing term must be of  $O(h^5)$ .]

We are now ready to describe a method, sometimes called the *triplet construction*, for parlaying a symmetric integrator of order  $2k$  into a symmetric integrator of order  $(2k+2)$ . It is also sometimes called the *Yoshida trick* or *Yoshida construction* in recognition of one of its discoverers. Suppose an order  $2k$  symmetric integrator  $\mathcal{S}_{2k}$  is known. Then the map  $\mathcal{S}_{2k+2}$  defined by writing

$$\mathcal{S}_{2k+2}(h) = \mathcal{S}_{2k}(\alpha h)\mathcal{S}_{2k}(\beta h)\mathcal{S}_{2k}(\alpha h) \quad (12.1.23)$$

is a symmetric integrator of order  $(2k+2)$  providing  $\alpha$  and  $\beta$  are the numbers

$$\alpha = 1/[2 - 2^{1/(2k+1)}], \quad (12.1.24)$$

$$\beta = -[2^{1/(2k+1)}]\alpha. \quad (12.1.25)$$

Why is this true? First we verify the relation

$$\begin{aligned} \mathcal{S}_{2k+2}^{-1}(h) &= \mathcal{S}_{2k}^{-1}(\alpha h)\mathcal{S}_{2k}^{-1}(\beta h)\mathcal{S}_{2k}^{-1}(\alpha h) \\ &= \mathcal{S}_{2k}(-\alpha h)\mathcal{S}_{2k}(-\beta h)\mathcal{S}_{2k}(-\alpha h) = \mathcal{S}_{2k+2}(-h), \end{aligned} \quad (12.1.26)$$

which follows from the assumed symmetry of  $\mathcal{S}_{2k}$ ,

$$\mathcal{S}_{2k}^{-1}(h) = \mathcal{S}_{2k}(-h). \quad (12.1.27)$$

We see that  $\mathcal{S}_{2k+2}$  is symmetric by construction. Next, because by hypothesis  $\mathcal{S}_{2k}$  is an integrator of order  $2k$ , we have the result

$$\mathcal{S}_{2k}(h) = \exp(h : -H : + \mathcal{C}_{2k+1} h^{2k+1} + \dots). \quad (12.1.28)$$

Consequently we conclude from (1.23), (1.28), the BCH series, and symmetry that  $\mathcal{S}_{2k+2}$  must have the form

$$\mathcal{S}_{2k+2} = \exp[(2\alpha + \beta)h : -H : + (2\alpha^{2k+1} + \beta^{2k+1})\mathcal{C}_{2k+1}h^{2k+1} + O(h^{2k+3})]. \quad (12.1.29)$$

Evidently  $\mathcal{S}_{2k+2}$  will be an integrator of order  $(2k+2)$  if  $\alpha$  and  $\beta$  satisfy the relations

$$2\alpha + \beta = 1, \quad (12.1.30)$$

$$2\alpha^{2k+1} + \beta^{2k+1} = 0. \quad (12.1.31)$$

Finally, the relations (1.30) and (1.31) have the solutions (1.24) and (1.25).

Note that nowhere does the demonstration just given depend on any property of  $\mathcal{S}_{2k}$  other than symmetry and that it is indeed an integrator of order  $2k$ . Therefore, if we can produce a symmetric integrator of order  $2k$  by any method whatsoever [not necessarily of Zassenhaus type and not necessarily making the splitting assumption (1.5)], then (1.23) through (1.25) produce from it a symmetric integrator of order  $(2k+2)$ . Finally, suppose  $\mathcal{S}_{2k}$  is a symplectic integrator. Then, since any product of symplectic maps is again a symplectic map, (1.23) shows that  $\mathcal{S}_{2k+2}$  will also be a symplectic integrator.<sup>4</sup>

As an example of the use of (1.23), let us construct a 4th-order Zassenhaus integrator  $\mathcal{S}_4$  from the known 2nd-order integrator  $\mathcal{S}_2$  given by (1.15). From (1.23) through (1.25) we get the result

$$\mathcal{S}_4(h) = \mathcal{S}_2(\alpha h)\mathcal{S}_2(\beta h)\mathcal{S}_2(\alpha h) \quad (12.1.32)$$

with

$$\alpha = 1/(2 - 2^{1/3}), \quad (12.1.33)$$

$$\beta = -(2^{1/3})/(2 - 2^{1/3}). \quad (12.1.34)$$

Now carry out the multiplications indicated in (1.32) to obtain the final Zassenhaus relation

$$\begin{aligned} \mathcal{S}_4(h) = & \exp(w_1 h : A :) \exp(w_2 h : B :) \exp(w_3 h : A :) \exp(w_4 h : B :) \times \\ & \exp(w_5 h : A :) \exp(w_6 h : B :) \exp(w_7 h : A :). \end{aligned} \quad (12.1.35)$$

Here the *weights*  $w_i$  have the values

$$\begin{aligned} w_1 = w_7 &= -1/[2(2 - 2^{1/3})], \quad w_3 = w_5 = (1 - 2^{1/3})w_1, \\ w_2 = w_6 &= 2w_1, \quad w_4 = -2^{1/3}w_2. \end{aligned} \quad (12.1.36)$$

By construction they must have the remarkable property that

$$\mathcal{S}_4(h) = \exp[h : -H : + O(h^5)]. \quad (12.1.37)$$

---

<sup>4</sup>This observation is due to Étienne Forest.

The procedure (1.23) gives a general way of constructing even-order symmetric integrators, but we might also desire odd-order integrators; and we might be willing to give up the prescription (1.23) or even the symmetry condition in favor of having fewer factors in the Zassenhaus product. How does one find general Zassenhaus formulas of the form (1.35)? That is, having decided in advance how many factors we wish to allow in a Zassenhaus product, how do we select weights  $w_i$  to achieve a formula of maximum order? For example, there is the 3rd-order formula

$$\begin{aligned} \exp(-h : H :) = & \exp(w_1 h : A :) \exp(w_2 h : B :) \exp(w_3 h : A :) \exp(w_4 h : B :) \times \\ & \exp(w_5 h : A :) \exp(w_6 h : B :) \times [1 + O(h^4)] \end{aligned} \quad (12.1.38)$$

with

$$\begin{aligned} w_1 &= -7/24, \quad w_3 = -3/4, \quad w_5 = 1/24, \\ w_2 &= -2/3, \quad w_4 = 2/3, \quad w_6 = -1. \end{aligned} \quad (12.1.39)$$

The general answer to this question is difficult, but results are known through modest order, and are discussed in some of the references cited at the end of this chapter.<sup>5</sup>

We also note that if some commutators of the form  $C = \{\cdot : A ; , \{\cdot : A ; , \cdot : B :\}\}$  or  $D = \{\cdot : B ; , \{\cdot : B ; , \cdot : A :\}\}$  are readily computable, and if the maps  $\exp(\tau C)$  or  $\exp(\tau D)$  can be computed exactly, then one can also construct higher-order symplectic integrators using these quantities. Such methods are sometimes called *force gradient* algorithms. They are known for some examples to have superior accuracy compared to triplet constructed algorithms of the same order. Moreover, there are symmetric force gradient algorithms which can then be employed in a triplet construction to go to still higher order.

## Exercises

**12.1.1.** Hamiltonians of the form  $H = T(p) + V(q)$  commonly occur in the case of nonrelativistic motion in a force field derivable from a potential. However, they can also occur in some cases of relativistic motion in an electromagnetic field. Verify from (1.5.30) in Exercise 1.5.3 that  $H$  is of the  $T + V$  form when there is only an electric field (no magnetic field so that  $\mathbf{A} = 0$ ) and the time  $t$  is taken to be the independent variable. Verify from (1.6.16) in Exercise 1.6.1 that  $H$  is of the  $T + V$  if  $\psi = 0$ , and the magnetic field  $\mathbf{B}$  is of the form that it can be derived from a vector potential such that only  $A_z \neq 0$ , and the coordinate  $z$  is taken to be the independent variable.

**12.1.2.** Show that crude Euler is not symplectic. Show that integration using (1.4) and (1.11) is symplectic.

**12.1.3.** Leapfrog exercise.

**12.1.4.** Verify that  $\mathcal{S}_2$ , as given by (1.46), satisfies (1.16).

---

<sup>5</sup>We remark that the methods we have been describing are sometimes referred to as *composition* methods because  $\exp(-h : H :)$  is written as the composition of several factors or as *fractional-step* methods because a step of duration  $h$  is accomplished by taking several steps whose durations are sometimes fractions of  $h$ .

**12.1.5.** Verify (1.29).

**12.1.6.** Verify that (1.30) and (1.31) have the solution (1.24) and (1.25).

**12.1.7.** Verify (1.35) and (1.36).

**12.1.8.** Exercise for case where  $V$  is time dependent.

**12.1.9.** For the  $\mathcal{S}_2$  given by (1.15), show that

$$\mathcal{S}_2 H = H + h^3 G + O(h^4) \quad (12.1.40)$$

where

$$G = -[A, [A, [A, B]]]/24 - [B, [B, [A, B]]]/12 - [A, [B, [A, B]]]/8. \quad (12.1.41)$$

Evaluate  $G$  for the case

$$A = p^2/2, \quad B = q^2/2. \quad (12.1.42)$$

You have verified a particular case of the general theorem that symplectic integration does not conserve the value of the Hamiltonian.

**12.1.10.** Verify that the weights of the “ $A$ ” and “ $B$ ” terms in (1.36) and (1.39) sum separately to -1. Why should this be?

**12.1.11.** Review the discussion of backward error analysis that appears near the end of Section 2.7. For a two-dimensional phase space, let  $H$  be the Hamiltonian

$$H = (1/2)ap^2 + (1/2)bq^2 = T(p) + V(q), \quad (12.1.43)$$

and let  $\mathcal{S}_2(H, h)$  be the symplectic integrator

$$\mathcal{S}_2(H, h) = \exp : -(h/2)T : \exp : -hV : \exp : -(h/2)T : . \quad (12.1.44)$$

Show that there are functions  $\alpha(h)$  and  $\beta(h)$  such that

$$\mathcal{S}_2(H, h) = \exp : -h\bar{H} : \quad (12.1.45)$$

where

$$\bar{H} = (1/2)\alpha p^2 + (1/2)\beta q^2. \quad (12.1.46)$$

Find  $\alpha(h)$  and  $\beta(h)$  explicitly. Thus,  $\mathcal{S}_2(H, h)$  produces exact trajectories for the modified Hamiltonian  $\bar{H}$ . Find a Hamiltonian  $\hat{H}$  of the form

$$\hat{H} = (1/2)A(h)p^2 + (1/2)B(h)q^2 \quad (12.1.47)$$

such that  $\mathcal{S}_2(\hat{H}, h)$  produces exact trajectories for the Hamiltonian  $H$ .

**12.1.12.** Exercise on using BCH to combine exponents and backward error analysis to find effective Hamiltonian.

## 12.2 Explicit Symplectic Integrator for Polynomial Hamiltonians

## 12.3 Symplectic Runge-Kutta Methods for $T + V$ Split Hamiltonians: Partitioned Runge Kutta and Nyström Runge Kutta

The integration methods of the Zassenhaus type, for the *special* Hamiltonians of the form  $T + V$ , are explicit and symplectic. They can also be viewed as being Runge-Kutta-like in that they involve multiple evaluations of the effects of  $T$  and  $V$  in the course of making a single integration step. For these special Hamiltonians there are also other methods, called *partitioned* Runge Kutta (PRK) and Nyström Runge Kutta (NRK), that are explicit and symplectic.

### 12.3.1 Partitioned Runge-Kutta

### 12.3.2 Nyström Runge-Kutta

## 12.4 Symplectic Runge-Kutta Methods for General Hamiltonians

So far, the construction of symplectic integrators has been based on the assumption that the Hamiltonian can be split in the  $T + V$  form (1.6). This assumption is not true for a broad class of problems of interest for Accelerator Physics, namely motion in a general electromagnetic field. It is not true for the Hamiltonian (1.5.30), assuming  $\mathbf{A} \neq 0$ , and it is not true for the Hamiltonian (1.6.16) except in the special case  $\psi = A_x = A_y = 0$ . In this subsection we will describe briefly symplectic Runge-Kutta methods that are applicable to general Hamiltonians.<sup>6</sup>

### 12.4.1 Background

Let us begin by setting up a notation that is convenient for working with differential equations in Hamiltonian form. Suppose the canonical variables are ordered as in (1.7.9) and Hamilton's equations of motion are written in the form (2.1.1). That is, we make the identification

$$\mathbf{y} = (q_1, q_2, \dots, q_\ell; p_1, p_2, \dots, p_\ell), \quad (12.4.1)$$

and therefore, from  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}, t)$ , it follows that

$$\mathbf{f} = J\partial H/\partial \mathbf{y}. \quad (12.4.2)$$

---

<sup>6</sup>We remark that the Hamiltonian for the restricted 3-body problem, when formulated in a rotating coordinate system in order to exploit the existence of the Jacobi integral that appears in this formulation, also cannot be split in the form  $T + V$ .

Write the Runge-Kutta procedure in terms of canonical variables by introducing the notation

$$\mathbf{y}^{n+1} = (Q_1, Q_2, \dots, Q_\ell; P_1, P_2, \dots, P_\ell) \quad (12.4.3)$$

and the slightly modified notation

$$\mathbf{y}^n = (q_1, q_2, \dots, q_\ell; p_1, p_2, \dots, p_\ell). \quad (12.4.4)$$

As in Sections 4.8 and 6.5.1, it is also convenient to employ the notation

$$z = (\mathbf{q}; \mathbf{p}), \quad (12.4.5)$$

$$Z = (\mathbf{Q}; \mathbf{P}). \quad (12.4.6)$$

In this terminology, performing a Runge-Kutta step from  $t = t^n$  to  $t = t^{n+1}$  sends the old pair  $\mathbf{q}, \mathbf{p}$  to the new pair  $\mathbf{Q}, \mathbf{P}$ . Or, equivalently, it sends  $z$  to  $Z$ . In map notation, this transformation can be expressed in the form

$$Z = \mathcal{M}_{\text{RK}} z. \quad (12.4.7)$$

We would like the map  $\mathcal{M}_{\text{RK}}$  to be exactly symplectic.

As further notation, decompose the vector  $\mathbf{f}$  into  $q$ -like and  $p$ -like components,

$$\mathbf{f} = (\mathbf{f}^q; \mathbf{f}^p). \quad (12.4.8)$$

Also, introduce the  $\ell$ -component vectors  $\mathbf{H}_q$  and  $\mathbf{H}_p$  by the rules

$$\mathbf{H}_q(\mathbf{q}, \mathbf{p}, t) = \partial H(\mathbf{q}, \mathbf{p}, t) / \partial \mathbf{q}, \quad (12.4.9)$$

$$\mathbf{H}_p(\mathbf{q}, \mathbf{p}, t) = \partial H(\mathbf{q}, \mathbf{p}, t) / \partial \mathbf{p}. \quad (12.4.10)$$

With these definitions, in view of (3.2), we have the relations

$$\mathbf{f}^q = \mathbf{H}_p, \quad (12.4.11)$$

$$\mathbf{f}^p = -\mathbf{H}_q. \quad (12.4.12)$$

Recall that a Runge-Kutta method is specified by a Butcher tableau. Review Section 2.3.4. From (2.3.6) we see that application of the Runge-Kutta stepping formula gives the relations

$$\mathbf{Q} = \mathbf{q} + h \sum_{i=1}^s b_i \mathbf{k}_i^q, \quad (12.4.13)$$

$$\mathbf{P} = \mathbf{p} + h \sum_{i=1}^s b_i \mathbf{k}_i^p. \quad (12.4.14)$$

Here we have introduced the notation

$$\mathbf{k}_i = (\mathbf{k}_i^q; \mathbf{k}_i^p) \quad (12.4.15)$$

to indicate that the  $\mathbf{k}_i$  also have  $q$ -like and  $p$ -like components. In terms of this notation, (2.3.7), (3.11), and (3.12) yield the definitions

$$\mathbf{k}_i^q = \mathbf{f}^q(\mathbf{q}_i, \mathbf{p}_i, t_i) = \mathbf{H}_p(\mathbf{q}_i, \mathbf{p}_i, t_i), \quad (12.4.16)$$

$$\mathbf{k}_i^p = \mathbf{f}^p(\mathbf{q}_i, \mathbf{p}_i, t_i) = -\mathbf{H}_q(\mathbf{q}_i, \mathbf{p}_i, t_i). \quad (12.4.17)$$

For each value of  $i$  the right sides of (3.16) and (3.17) are to be evaluated at the points  $(\mathbf{q}_i, \mathbf{p}_i, t_i)$  specified by the relations

$$\mathbf{q}_i = \mathbf{q} + h \sum_{j=1}^s a_{ij} \mathbf{k}_j^q, \quad (12.4.18)$$

$$\mathbf{p}_i = \mathbf{p} + h \sum_{j=1}^s a_{ij} \mathbf{k}_j^p, \quad (12.4.19)$$

$$t_i = t^n + c_i h. \quad (12.4.20)$$

We will see, in view of (3.16) and (3.17) and results to follow, that the relations (3.18) and (3.19) are implicit if the integrator is to be symplectic. They therefore must be solved by simple iteration or the more involved Newton's method.

### 12.4.2 Condition for Symplecticity

We will subsequently see that a necessary and sufficient condition for a Runge-Kutta method to be symplectic is that the entries in the matrix  $a$  and the vector  $b$  satisfy the relations

$$b_i a_{ij} + b_j a_{ji} - b_i b_j = 0 \text{ for } i, j = 1, \dots, s. \quad (12.4.21)$$

As usual, the entries  $c$  are given by (2.3.11). From the condition (3.21) it is easy to prove that there are no explicit symplectic Runge-Kutta methods. See Exercise 3.1.

Note that the condition (3.21) makes no mention of the number of equations being integrated. Of course, in the Hamiltonian case, one is always integrating an even number equations, say  $\ell$  for the  $q$ 's and  $\ell$  for the  $p$ 's. When the number of equations is odd, it makes no sense to talk about a symplectic condition. However, it can be shown that in general a Runge-Kutta integrator satisfying (3.21) has additional desirable stability properties compared to other Runge-Kutta integrators, and therefore the condition (3.21) is also of interest when any set of differential equations, including non-Hamiltonian or odd numbers of equations, is being integrated.

### 12.4.3 The Single-Stage Case

Before verifying the necessity and sufficiency of the condition (3.21), let us consider the simplest case, the one-stage case  $s = 1$ . In this case the Butcher tableau has the general form

$$\begin{array}{c|c} c_1 & a_{11} \\ \hline & b_1 \end{array}, \quad (12.4.22)$$

and use of (3.21) yields the relation

$$2a_{11}b_1 = (b_1)^2. \quad (12.4.23)$$

But, in order for the method to at least be of order 1, we must have  $b_1=1$ . See (2.3.31). It follows from (3.23) that  $a_{11} = 1/2$ . Thus, if (3.21) holds, the Butcher tableau for a one-stage symplectic Runge-Kutta method is

$$\begin{array}{c|cc} 1/2 & 1/2 \\ \hline & 1 \end{array} \quad \text{Gauss2.} \quad (12.4.24)$$

Here we have also used (2.3.11).

We have seen this Butcher tableau before. Look at (2.3.12) and (2.3.13) to see that this method is the implicit midpoint rule. And, as we learned from Exercise 2.3.7, this method is of order two. We will soon observe that this method is related to Gaussian quadrature, and for this reason we will refer to it as *Gauss2*. And, given equations of motion, we will call  $\mathcal{M}_{\text{G2}}$  the transfer map that arises from integrating these equations of motion using Gauss2.

Let us now verify, in the Hamiltonian context, that  $\mathcal{M}_{\text{G2}}$  is a symplectic map. For the coefficients (3.24), and in the Hamiltonian case, the Runge-Kutta relations become

$$\mathbf{Q} = \mathbf{q} + h\mathbf{k}_1^q, \quad (12.4.25)$$

$$\mathbf{P} = \mathbf{p} + h\mathbf{k}_1^p, \quad (12.4.26)$$

where

$$\mathbf{k}_1^q = \mathbf{H}_p(\mathbf{q}_1, \mathbf{p}_1, t_1), \quad (12.4.27)$$

$$\mathbf{k}_1^p = -\mathbf{H}_q(\mathbf{q}_1, \mathbf{p}_1, t_1), \quad (12.4.28)$$

with

$$\mathbf{q}_1 = \mathbf{q} + (h/2)\mathbf{k}_1^q, \quad (12.4.29)$$

$$\mathbf{p}_1 = \mathbf{p} + (h/2)\mathbf{k}_1^p, \quad (12.4.30)$$

$$t_1 = t^n + h/2. \quad (12.4.31)$$

Suppose small changes  $d\mathbf{q}$  and  $d\mathbf{p}$  are made in  $\mathbf{q}$  and  $\mathbf{p}$ . Then, there will be related small changes  $d\mathbf{Q}$  and  $d\mathbf{P}$  in  $\mathbf{Q}$  and  $\mathbf{P}$ . What we wish to verify is that the matrix  $M_{\text{G2}}$  in the relation

$$dZ = M_{\text{G2}}dz \quad (12.4.32)$$

is symplectic, and therefore the map  $\mathcal{M}_{\text{G2}}$  is symplectic.

According to (3.25) and (3.26), there will be the relations

$$d\mathbf{Q} = d\mathbf{q} + hd\mathbf{k}_1^q, \quad (12.4.33)$$

$$d\mathbf{P} = d\mathbf{p} + hd\mathbf{k}_1^p. \quad (12.4.34)$$

From (3.27) and (3.28) we find the relations

$$d\mathbf{k}_1^q = \mathcal{H}_{pq}(\mathbf{q}_1, \mathbf{p}_1, t_1)d\mathbf{q}_1 + \mathcal{H}_{pp}(\mathbf{q}_1, \mathbf{p}_1, t_1)d\mathbf{p}_1, \quad (12.4.35)$$

$$d\mathbf{k}_1^p = -\mathcal{H}_{qq}(\mathbf{q}_1, \mathbf{p}_1, t_1) d\mathbf{q}_1 - \mathcal{H}_{qp}(\mathbf{q}_1, \mathbf{p}_1, t_1) d\mathbf{p}_1. \quad (12.4.36)$$

Here  $\mathcal{H}_{qq}$ ,  $\mathcal{H}_{pq}$ , etc. are the  $\ell \times \ell$  Hessian block matrices,

$$\begin{aligned} \mathcal{H}_{qq}(\mathbf{q}, \mathbf{p}, t) &= \partial^2 H(\mathbf{q}, \mathbf{p}, t) / \partial \mathbf{q} \partial \mathbf{q}, \\ \mathcal{H}_{pq}(\mathbf{q}, \mathbf{p}, t) &= \mathcal{H}_{qp}(\mathbf{q}, \mathbf{p}, t) = \partial^2 H(\mathbf{q}, \mathbf{p}, t) / \partial \mathbf{p} \partial \mathbf{q}, \\ \mathcal{H}_{pp}(\mathbf{q}, \mathbf{p}, t) &= \partial^2 H(\mathbf{q}, \mathbf{p}, t) / \partial \mathbf{p} \partial \mathbf{p}. \end{aligned} \quad (12.4.37)$$

And, from (3.29) and (3.30), we find the relations

$$d\mathbf{q}_1 = d\mathbf{q} + (h/2)d\mathbf{k}_1^q, \quad (12.4.38)$$

$$d\mathbf{p}_1 = d\mathbf{p} + (h/2)d\mathbf{k}_1^p. \quad (12.4.39)$$

Now substitute (3.38) and (3.39) into (3.35) and (3.36) to yield the results

$$d\mathbf{k}_1^q = \mathcal{H}_{pq}[d\mathbf{q} + (h/2)d\mathbf{k}_1^q] + \mathcal{H}_{pp}[d\mathbf{p} + (h/2)d\mathbf{k}_1^p], \quad (12.4.40)$$

$$d\mathbf{k}_1^p = -\mathcal{H}_{qq}[d\mathbf{q} + (h/2)d\mathbf{k}_1^q] - \mathcal{H}_{qp}[d\mathbf{p} + (h/2)d\mathbf{k}_1^p], \quad (12.4.41)$$

which can be rewritten in the form

$$[1 - (h/2)\mathcal{H}_{pq}]d\mathbf{k}_1^q - (h/2)\mathcal{H}_{pp}d\mathbf{k}_1^p = \mathcal{H}_{pq}d\mathbf{q} + \mathcal{H}_{pp}d\mathbf{p}, \quad (12.4.42)$$

$$(h/2)\mathcal{H}_{qq}d\mathbf{k}_1^q + [1 + (h/2)\mathcal{H}_{qp}]d\mathbf{k}_1^p = -\mathcal{H}_{qq}d\mathbf{q} - \mathcal{H}_{qp}d\mathbf{p}. \quad (12.4.43)$$

Our goal is to solve the relations (3.42) and (3.43) for  $d\mathbf{k}_1^q$  and  $d\mathbf{k}_1^p$ , and then substitute the results into (3.33) and (3.34). To do so it is convenient to rewrite (3.42) and (3.43) in matrix/vector form. Let  $A$  be the matrix

$$A = \begin{pmatrix} \mathcal{H}_{pq} & \mathcal{H}_{pp} \\ -\mathcal{H}_{qq} & -\mathcal{H}_{qp} \end{pmatrix}. \quad (12.4.44)$$

Then (3.42) and (3.43) can be written in the form

$$[I - (h/2)A]d\mathbf{k}_1 = Adz, \quad (12.4.45)$$

and therefore

$$d\mathbf{k}_1 = [I - (h/2)A]^{-1}Adz. \quad (12.4.46)$$

Correspondingly, (3.33) and (3.34) become

$$dZ = M_{G2}dz \quad (12.4.47)$$

with

$$M_{G2} = I + h[I - (h/2)A]^{-1}A = [I - (h/2)A]^{-1}[I + (h/2)A]. \quad (12.4.48)$$

We claim that  $M_{G2}$  is a symplectic matrix. Correspondingly, in this Hamiltonian case,  $M_{G2}$  is a symplectic map. To verify this claim, observe that  $A$  can be written in the form

$$A = JS \quad (12.4.49)$$

where

$$S = \begin{pmatrix} \mathcal{H}_{qq} & \mathcal{H}_{qp} \\ \mathcal{H}_{pq} & \mathcal{H}_{pp} \end{pmatrix}. \quad (12.4.50)$$

As the notation is intended to indicate, and because of the equality of mixed partial derivatives,  $S$  is a symmetric matrix. Consequently  $M_{G2}$  can also be written in the form

$$M_{G2} = [I - (h/2)JS]^{-1}[I + (h/2)JS]. \quad (12.4.51)$$

Reference to Section 3.12 shows that (3.51) is a Cayley representation when we make the identification

$$W = (h/2)S, \quad (12.4.52)$$

and therefore  $M_{G2}$  is indeed a symplectic matrix.

#### 12.4.4 Two-, Three-, and More-Stage Methods

In Subsection 3.3 we studied the single-stage symplectic Runge-Kutta method, found its Butcher tableau (3.24), and observed that this method is of order two. Remarkably, it is also known that, for  $s$  stages, there are Runge-Kutta methods of order  $m = 2s$ , and these methods are symplectic when applied to Hamiltonian systems. See Exercise 3.3. Butcher tableaux for these methods, for the cases of two and three stages, are given below.

##### Gauss4

$$\begin{array}{c|cc} 1/2 - \sqrt{3}/6 & 1/4 & 1/4 - \sqrt{3}/6 \\ 1/2 + \sqrt{3}/6 & 1/4 + \sqrt{3}/6 & 1/4 \\ \hline & 1/2 & 1/2 \end{array}, \quad (12.4.53)$$

##### Gauss6

$$\begin{array}{c|ccc} 1/2 - \sqrt{15}/10 & 5/36 & 2/9 - \sqrt{15}/15 & 5/36 - \sqrt{15}/30 \\ 1/2 & 5/36 + \sqrt{15}/24 & 2/9 & 5/36 - \sqrt{15}/24 \\ 1/2 + \sqrt{15}/10 & 5/36 + \sqrt{15}/30 & 2/9 + \sqrt{15}/15 & 5/36 \\ \hline & 5/18 & 8/18 & 5/18 \end{array}. \quad (12.4.54)$$

They have orders 4 and 6, respectively. Observe that, in tableaux (3.24), (3.53), and (3.54), the  $b_i$  are weights and the  $c_i$  are evaluation points for Gaussian quadrature. This circumstance arises from the fact that the Runge-Kutta methods based on these tableaux are related to Gaussian quadrature. See Appendix T. For this reason these methods are sometimes referred to as Gauss2, Gauss4, and Gauss6. Butcher tableaux for Gauss8 and Gauss10 are also known. See the book of Sanz-Serna and Calvo listed in the Bibliography for this chapter.

We also remark that these methods are symmetric,

$$\mathcal{M}_{G2s}(-h) = [\mathcal{M}_{G2s}(h)]^{-1}. \quad (12.4.55)$$

See the book of Hairer, Nørsett, and Wanner listed in the Bibliography for this chapter.

## Exercises

**12.4.1.** The purpose of this exercise is to verify that a Runge-Kutta, in order to be symplectic, must be implicit. Set  $j = i$  in (3.21) to obtain the condition

$$b_i a_{ii} + b_i a_{ii} - b_i b_i = 0 \text{ for } i = 1, \dots, s, \quad (12.4.56)$$

from which it follows that

$$2b_i a_{ii} = (b_i)^2 \text{ for } i = 1, \dots, s. \quad (12.4.57)$$

Verify that, for a Runge-Kutta method to be explicit, the matrix  $a$  must be lower triangular. In particular, it must satisfy the condition

$$a_{ii} = 0 \text{ for } i = 1, \dots, s. \quad (12.4.58)$$

Combine (3.56) and (3.57) to conclude that

$$b_i = 0 \text{ for } i = 1, \dots, s, \quad (12.4.59)$$

in which case, according to (2.3.6), the associated Runge-Kutta method fails to advance the solution  $\mathbf{y}$  at all.

**12.4.2.** Repeat the calculations of Subsection 3.3 for the case of general  $a_{11}$  and general  $b_1$ . Show that (3.21) must be satisfied for  $M_{G2}$  to be a symplectic matrix.

**12.4.3.** Verify that the Butcher tableaux (3.24), (3.53), and (3.54) for the Runge-Kutta methods Gauss2, Gauss4, and Gauss6 satisfy the relations (3.21). These Runge-Kutta methods are therefore symplectic when applied to Hamiltonian-produced equations of motion.

## 12.5 Study of Single-Stage Method

We have already confessed that the integration formulas associated with each of the Butcher tableaux (3.24), (3.53), and (3.54) are implicit, and therefore must be made explicit (repeatedly solved) at each integration step in order to actually produce a trajectory. Let us see what is involved by first examining the simplest case, that of Gauss2 specified by the single-stage Butcher tableau (3.24). Application of (2.3.6) and (2.3.7) shows that Gauss2 employs the stepping formula

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h\mathbf{k}_1 \quad (12.5.1)$$

where, at each step,

$$\mathbf{k}_1 = \mathbf{f}[\mathbf{y}^n + h(1/2)\mathbf{k}_1, t^n + (1/2)h]. \quad (12.5.2)$$

Let us now examine how to deal with/solve the implicit relation (4.2). Suppose we have some initial guess, which we will call  $\mathbf{k}_1^0$ , for  $\mathbf{k}_1$ . It might be  $\mathbf{f}(\mathbf{y}^n, t^n)$ , but we could hope for something better. Let us convert (4.2) into a recursion relation by making the rule

$$\mathbf{k}_1^{j+1} = \mathbf{f}[\mathbf{y}^n + h(1/2)\mathbf{k}_1^j, t^n + (1/2)h]. \quad (12.5.3)$$

It can be verified that (4.3) is a contraction map for small enough  $h$ . Therefore, providing the initial guess  $\mathbf{k}_1^0$  is in the basin of attraction, we have the result

$$\mathbf{k}_1 = \lim_{j \rightarrow \infty} \mathbf{k}_1^j. \quad (12.5.4)$$

Finally, having found  $\mathbf{k}_1$ ,  $\mathbf{y}^{n+1}$  is given by (4.1).

To learn a bit more about the nature of the iteration process (4.3) and (4.4) for the solution of (4.2), it is instructive to study a simple example, that of the harmonic oscillator with Hamiltonian

$$H = (p^2 + q^2)/2. \quad (12.5.5)$$

For this case, with  $\mathbf{y} = (q, p)$ , we find the results

$$f_1 = \dot{y}_1 = \dot{q} = \partial H / \partial p = p = y_2, \quad (12.5.6)$$

$$f_2 = \dot{y}_2 = \dot{p} = -\partial H / \partial q = -q = -y_1, \quad (12.5.7)$$

which can be written more compactly in the matrix form

$$\dot{\mathbf{y}} = \mathbf{f} = J\mathbf{y}. \quad (12.5.8)$$

Here  $J$  is the matrix  $J_2$  given by (3.2.11). Correspondingly, application of (4.2) yields the relations

$$\{\mathbf{k}_1\}_1 = \{\mathbf{f}[\mathbf{y}^n + h(1/2)\mathbf{k}_1, t^n + (1/2)h]\}_1 = \{\mathbf{y}^n\}_2 + h(1/2)\{\mathbf{k}_1\}_2, \quad (12.5.9)$$

$$\{\mathbf{k}_1\}_2 = \{\mathbf{f}[\mathbf{y}^n + h(1/2)\mathbf{k}_1, t^n + (1/2)h]\}_2 = -\{\mathbf{y}^n\}_1 - h(1/2)\{\mathbf{k}_1\}_1, \quad (12.5.10)$$

which can be conveniently written together in the matrix form

$$\mathbf{k}_1 = J[\mathbf{y}^n + h(1/2)\mathbf{k}_1], \quad (12.5.11)$$

and solved for  $\mathbf{k}_1$  to yield the result

$$\mathbf{k}_1 = (I - hJ/2)^{-1}J\mathbf{y}^n. \quad (12.5.12)$$

Note that in general we are only able to explicitly solve for  $\mathbf{k}_1$  in the case that the right side of (4.2) is linear in  $\mathbf{k}_1$ , as is true for this special example.

At this point we cannot resist the urge to employ (4.12) in (4.1) to obtain the explicit stepping relation

$$\mathbf{y}^{n+1} = I\mathbf{y}^n + h(I - hJ/2)^{-1}J\mathbf{y}^n = (I + hJ/2)(I - hJ/2)^{-1}\mathbf{y}^n = M_{G2}\mathbf{y}^n \quad (12.5.13)$$

with

$$M_{G2} = (I + hJ/2)(I - hJ/2)^{-1}. \quad (12.5.14)$$

Observe that  $M_{G2}$  is in the Cayley form (3.12.5) with

$$W = (h/2)I. \quad (12.5.15)$$

Since  $W$  is symmetric, it follows that  $M_{G2}$  is symplectic, as expected. See (3.12.6).

We also know from Section 3.12 that  $M_{G2}$  can be written in the form

$$M_{G2} = \exp(JS) \quad (12.5.16)$$

with  $S$  given by

$$S = -2J \tanh^{-1}(JW). \quad (12.5.17)$$

See (3.12.4). Use of (4.15) in (4.17) yields the result

$$\begin{aligned} S &= -2J \tanh^{-1}[(h/2)J] = (-2J)[(hJ/2) + (hJ/2)^3/3 + (hJ/2)^5/5 + \dots] \\ &= 2[(h/2) - (h/2)^3/3 + (h/2)^5/5 + \dots]I \\ &= 2 \tan^{-1}(h/2)I = h(2/h) \tan^{-1}(h/2)I = h(1 - h^2/12 + h^4/80 - \dots)I. \end{aligned} \quad (12.5.18)$$

It follows that, for  $H$  given by (4.5),  $M_{G2}$  can be written in the form

$$M_{G2} = \exp(JS) = \exp[(2/h) \tan^{-1}(h/2)hJ]. \quad (12.5.19)$$

Consider the Hamiltonian  $H'$  defined by

$$H' = \omega(p^2 + q^2)/2. \quad (12.5.20)$$

It can be verified that the exact solution to the equation of motion generated by  $H'$  is given by the relation

$$\mathbf{y}_{\text{true}}(t) = \exp[\omega(t - t^0)J]\mathbf{y}_{\text{true}}^0, \quad (12.5.21)$$

and therefore

$$\mathbf{y}_{\text{true}}^{n+1} = \exp(\omega h J)\mathbf{y}_{\text{true}}^n. \quad (12.5.22)$$

Upon comparing (4.13) and (4.19) with (4.22), we conclude that use of Gauss2 to integrate the equations of motion generated by the Hamiltonian  $H$  gives the exact solution to the equations of motion generated by  $H'$  with

$$\omega = (2/h)\tan^{-1}(h/2) = 1 - h^2/12 + h^4/80 - \dots. \quad (12.5.23)$$

We observe that  $H'$  is conserved and therefore, since  $H$  and  $H'$  are proportional,  $H$  is also conserved by Gauss2. Finally, according to (4.23), the trajectory given by Gauss2, since it is the exact trajectory for  $H'$ , differs from the exact trajectory for  $H$  only by a reparameterization of the time.

Also, here we have an instance of backward error analysis. *Approximately* but symplectically integrating the equations of motion generated by  $H$  yields the *exact* trajectory for the equations of motion generated by  $H'$  with  $H'$  being a small (when  $h$  is small) modification of  $H$ . See the discussion of backward error analysis in Section 2.7. Conversely, given  $H$ , it should be possible to find a related Hamiltonian  $H''$  such that symplectically integrating the equations of motion generated by  $H''$  yields the *exact* trajectory for the equations of motion generated by  $H$ . See Exercise 4.1.

With this diversion behind us, let us return to an analysis of the iteration process (4.3). By reasoning analogous to that which produced (4.11), the iteration process (4.3) for the case of  $H$  given by (4.5) yields the matrix relation

$$\mathbf{k}_1^{j+1} = J[\mathbf{y}^n + h(1/2)\mathbf{k}_1^j]. \quad (12.5.24)$$

We may view it as a mapping with fixed point  $\mathbf{k}_1$ . How do points near this fixed point behave under the influence of this map? Introduce deviation variables  $\boldsymbol{\delta}^j$  by writing

$$\mathbf{k}_1^j = \mathbf{k}_1 + \boldsymbol{\delta}^j. \quad (12.5.25)$$

In terms of these variables, (4.24) takes the form

$$\boldsymbol{\delta}^{j+1} = (hJ/2)\boldsymbol{\delta}^j. \quad (12.5.26)$$

This recursion relation has the solution

$$\boldsymbol{\delta}^j = (hJ/2)^j \boldsymbol{\delta}^0. \quad (12.5.27)$$

The eigenvalues of  $J$  are  $\pm i$ , and therefore the eigenvalues of  $hJ/2$  are  $\pm ih/2$ . These eigenvalues lie within the unit circle as long as the step size satisfies  $|h/2| < 1$ , and therefore the fixed point  $\mathbf{k}_1$  is attracting under this condition. That is, if  $|h/2| < 1$ , then

$$\lim_{j \rightarrow \infty} \boldsymbol{\delta}^j = 0. \quad (12.5.28)$$

Moreover, examination of (4.27) shows that, for the  $H$  of this example, the basin of attraction is the entire  $\boldsymbol{\delta}^0$  plane, and therefore also the entire  $\mathbf{k}_1^0$  plane.

There remains the problem of constructing a good initial guess  $\mathbf{k}_1^0$ . Suppose we begin with the initial guess

$$\mathbf{k}_1^0 = J\mathbf{y}^n. \quad (12.5.29)$$

Doing so is equivalent to ignoring the order  $h$  terms in the argument of the right side of (4.2) thereby setting  $\mathbf{k}_1^0 = \mathbf{f}(\mathbf{y}^n, t^n)$ . We then find the results

$$\mathbf{k}_1^1 = J\mathbf{y}^n + (h/2)J\mathbf{k}_1^0 = J\mathbf{y}^n + Jh(1/2)J\mathbf{y}^n = (I + hJ/2)J\mathbf{y}^n, \quad (12.5.30)$$

$$\begin{aligned} \mathbf{k}_1^2 &= J\mathbf{y}^n + (h/2)J\mathbf{k}_1^1 = J\mathbf{y}^n + (hJ/2)(I + hJ/2)J\mathbf{y}^n \\ &= [I + (hJ/2) + (hJ/2)^2]J\mathbf{y}^n, \text{ etc.} \end{aligned} \quad (12.5.31)$$

Observe that (4.12) has the expansion

$$\mathbf{k}_1 = (I - hJ/2)^{-1}J\mathbf{y}^n = [I + (hJ/2) + (hJ/2)^2 + \dots]J\mathbf{y}^n. \quad (12.5.32)$$

Evidently, the iterative process, with the initial guess (4.29), reproduces this expansion in such a way that each iteration produces one more term in the expansion.

Suppose instead we begin with the guess

$$\mathbf{k}_1^0 = (I + hJ/2)J\mathbf{y}^n. \quad (12.5.33)$$

Then we find

$$\mathbf{k}_1^1 = J\mathbf{y}^n + (h/2)J\mathbf{k}_1^0 = J\mathbf{y}^n + (hJ/2)(I + hJ/2)J\mathbf{y}^n = [I + (hJ/2) + (hJ/2)^2]J\mathbf{y}^n. \quad (12.5.34)$$

Evidently, this is a better guess because it moves us one further step down the chain of iterations.

How could we have anticipated that this would be a better guess? We have remarked that the Butcher tableaux (3.24), (3.53), and (3.54) are related to Gaussian quadrature. Integrate both sides of (2.1.1) over the interval  $[t^n, t^{n+1}]$ . So doing yields the result

$$\mathbf{y}^{n+1} - \mathbf{y}^n = \int_{t^n}^{t^{n+1}} d\tau \dot{\mathbf{y}}(\tau) = \int_{t^n}^{t^{n+1}} d\tau \mathbf{f}[\mathbf{y}(\tau), \tau]. \quad (12.5.35)$$

Estimate the integral on the right side of (4.35) using lowest-order Gaussian quadrature, which amounts to the midpoint rule, to find the approximation

$$\mathbf{y}^{n+1} - \mathbf{y}^n = \int_{t^n}^{t^{n+1}} d\tau \mathbf{f}[\mathbf{y}(\tau), \tau] \simeq h \mathbf{f}[\mathbf{y}(t^n + h/2), t^n + h/2] = h \mathbf{f}[\mathbf{y}(t^n + c_1 h), t^n + c_1 h]. \quad (12.5.36)$$

Comparison of (4.36) with (4.1) suggests that a good first guess in the one-stage case would be

$$\mathbf{k}_1^0 = \mathbf{f}[\mathbf{y}(t^n + h/2), t^n + h/2]. \quad (12.5.37)$$

Here  $\mathbf{y}$  is the exact solution to (2.1.1). But, of course, we do not know the exact solution. However, we can imagine having computed and stored  $\mathbf{f}^n$  as given by (2.1.4). Then, in predictor-corrector terminology with  $N = 0$  (see Section 2.4), we can construct a predictor formula using jet formulation (see Section 2.5.3) that will produce  $\mathbf{y}_{\text{pred}}(t^n + h/2)$  with a local error of order  $h^2$  [see (2.4.38)].<sup>7</sup> With a knowledge of  $\mathbf{y}_{\text{pred}}(t^n + h/2)$  we can define  $\mathbf{k}_1^0$  by the rule

$$\mathbf{k}_1^0 = \mathbf{f}[\mathbf{y}_{\text{pred}}(t^n + h/2), t^n + h/2]. \quad (12.5.38)$$

Use of the  $N = 0$  predictor in jet formulation gives the result

$$\mathbf{y}_{\text{pred}}(t^n + h/2) = \mathbf{y}^n + (h/2)\mathbf{f}^n. \quad (12.5.39)$$

For the current example,

$$\mathbf{y}^n + (h/2)\mathbf{f}^n = (I + hJ/2)\mathbf{y}^n. \quad (12.5.40)$$

Combining (4.38) through (4.40) yields the result

$$\begin{aligned} \mathbf{k}_1^0 &= \mathbf{f}[\mathbf{y}_{\text{pred}}(t^n + h/2), t^n + h/2] \\ &= \mathbf{f}[\mathbf{y}^n + (h/2)\mathbf{f}^n, t^n + h/2] \\ &= \mathbf{f}[(I + hJ/2)\mathbf{y}^n] \\ &= (I + hJ/2)J\mathbf{y}^n, \end{aligned} \quad (12.5.41)$$

in agreement with (4.33). With this  $\mathbf{k}_1^0$  in hand, we can proceed to carry out the iterations (4.3) to yield, depending on the number of iterations made, some approximation to  $\mathbf{k}_1$ , and

---

<sup>7</sup>One might wonder about storing more previous  $\mathbf{f}$  values so that  $\mathbf{y}_{\text{pred}}(t^n + h/2)$  would be given with yet higher order accuracy. The use of an additional  $\mathbf{f}$  value would yield an  $h^2$  contribution to  $\mathbf{k}_1^0$  and an  $h^3$  contribution to  $\mathbf{k}_1^1$ . However, there would seem to be no point in doing so. We know that the predictor is attempting to integrate the trajectories associated with  $H$  and the symplectic Runge-Kutta procedure integrates the trajectories associated with  $H'$ , and these Hamiltonians differ by terms of order  $h^2$ . Therefore, the higher-order terms produced by a higher-order predictor would not be expected to improve the guess of  $\mathbf{k}_1^0$ .

then finally determine  $\mathbf{y}^{n+1}$  using (4.1). Note that, if we wish, we may view (4.3) as a kind of corrector formula.

Let us make two last comments about the solution of (4.2). First, as we have seen from the harmonic oscillator example, its solution by iteration, as in (4.3) and (4.4), requires an infinite number of iterations. Therefore if we, as we must, make only make a finite number iterations, the result of the integration method will not be exactly symplectic. It will only be symplectic to some high power of  $h$  depending on how many iterations are made at each step. Each iteration makes the method exactly symplectic through terms of yet one order higher in  $h$ . Of course, no matter how many iterations are made, the result is only locally accurate through terms of order  $h^2$ . That is, although the result may be highly symplectic, depending on the number of iterations, an error of order  $h^3$  is still made at each step. Second, in order to speed convergence, we might attempt to solve (5.2) by Newton's method. This is possible at the cost of extra programming.<sup>8</sup> For an introduction to Newton's method, see Section 29.4.3.

## Exercises

**12.5.1.** This is a study in backward error analysis for the harmonic oscillator when its equations of motion are integrated using Gauss2. We have learned, when integrating the equations of motion associated with the Hamiltonian  $H$  given by (4.5), that use of Gauss2 yields exact trajectories for the nearby Hamiltonian  $H'$  given by (4.20) and (4.23). Find a Hamiltonian  $H''$  whose equations of motion, when integrated using Gauss2, produces the exact trajectories for the Hamiltonian  $H$ .

**12.5.2.** This is a study in backward error analysis for the general static quadratic Hamiltonian when its associated equations of motion are integrated using Gauss2. Review Subsection 3.3. There we were able to find  $M_{G2}$ , the matrix for the linear part of  $\mathcal{M}_{G2}$ , for a general trajectory generated by a general Hamiltonian. Verify that if the equations of motion are linear, as will be the case if the Hamiltonian is quadratic, then the various implicit equations that have to be solved using Gauss2 are all linear. Moreover, again because the Hamiltonian is assumed to be quadratic, the matrix  $S$  given by (3.50) will have constant entries. Correspondingly, the map  $\mathcal{M}_{G2}$  will be linear, and will be completely represented by its linear part  $M_{G2}$  with  $M_{G2}$  given by (3.51).

Verify, at least when  $H$  is quadratic and static, that  $\mathcal{M}_{G2}$  is symmetric,

$$\mathcal{M}_{G2}(-h) = [\mathcal{M}_{G2}(h)]^{-1}. \quad (12.5.42)$$

As stated earlier, an analogous result is known to hold in general when any Gauss2s is used to integrate any set of differential equations.

How does  $\mathcal{M}_{G2}$  compare with the exact map  $\mathcal{M}$ ? In the case that  $H$  is static, we know that

$$\mathcal{M} = \exp(-h : H :). \quad (12.5.43)$$

---

<sup>8</sup>Of course, whatever method is used to solve (4.2), it can, if desired, be solved to machine precision in a finite number of steps.

Recall (1.4). Show that if  $H$  is quadratic, then  $\mathcal{M}$  will be linear and will be described by the matrix  $M$  given by

$$M = \exp(hJS). \quad (12.5.44)$$

We can now compare  $\mathcal{M}_{G2}$  and  $\mathcal{M}$ . From (3.12.1) through (3.12.5) show that

$$M_{G2} = \exp(hJS') \quad (12.5.45)$$

with

$$JS' = (2/h)\tanh^{-1}[(h/2)JS], \quad (12.5.46)$$

and therefore

$$S' = -J(2/h)\tanh^{-1}[(h/2)JS]. \quad (12.5.47)$$

Consequently, show that use of Gauss2 to integrate the equations of motion associated with the quadratic Hamiltonian  $H$  given by

$$H(z) = (1/2)(z, Sz) \quad (12.5.48)$$

produces the exact trajectory for the equations of motion associated with the Hamiltonian  $H'$  given by

$$H'(z) = (1/2)(z, S'z). \quad (12.5.49)$$

(Here, as before, we assume that  $S$  is time independent.)

Use the results of Exercise 3.12.1, the machinery of (5.5.1) through (5.5.13), and (4.45) to show that

$$[H, H'] = 0, \quad (12.5.50)$$

and therefore  $\mathcal{M}_{G2}$  conserves  $H$ .

Given a quadratic time independent  $H$ , find a Hamiltonian  $H''$  such that integrating its associated equations of motion using Gauss2 produces the exact trajectories for the Hamiltonian  $H$ .

## 12.6 Study of Two-Stage Method

Now that we have explored the behavior of the single-stage method Gauss2, let us make a similar exploration of the two-stage method Gauss4. Doing so will give us a general understanding of what to expect in the multi-stage case. For the two-stage case use of (2.3.6) and (2.3.7) and the Butcher tableau (3.53) provides the stepping formula

$$\mathbf{y}^{n+1} = \mathbf{y}^n + h(1/2)\mathbf{k}_1 + h(1/2)\mathbf{k}_2 \quad (12.6.1)$$

where, at each step,

$$\mathbf{k}_1 = \mathbf{f}[\mathbf{y}^n + h(1/4)\mathbf{k}_1 + h(1/4 - \sqrt{3}/6)\mathbf{k}_2, t^n + h(1/2 - \sqrt{3}/6)], \quad (12.6.2)$$

$$\mathbf{k}_2 = \mathbf{f}[\mathbf{y}^n + h(1/4 + \sqrt{3}/6)\mathbf{k}_1 + h(1/4)\mathbf{k}_2, t^n + h(1/2 + \sqrt{3}/6)]. \quad (12.6.3)$$

Like (4.2) in the single-stage case, the relations (5.2) and (5.3) are implicit. To solve them numerically, we may make initial guesses  $\mathbf{k}_1^0$  and  $\mathbf{k}_2^0$  for  $\mathbf{k}_1$  and  $\mathbf{k}_2$ , respectively, and set up the recursion relations

$$\mathbf{k}_1^{j+1} = \mathbf{f}[\mathbf{y}^n + h(1/4)\mathbf{k}_1^j + h(1/4 - \sqrt{3}/6)\mathbf{k}_2^j, t^n + h(1/2 - \sqrt{3}/6)], \quad (12.6.4)$$

$$\mathbf{k}_2^{j+1} = \mathbf{f}[\mathbf{y}^n + h(1/4 + \sqrt{3}/6)\mathbf{k}_1^j + h(1/4)\mathbf{k}_2^j, t^n + h(1/2 + \sqrt{3}/6)]. \quad (12.6.5)$$

It can be shown that the relations (5.4) and (5.5) constitute a contraction map for sufficiently small  $h$ . Consequently, assuming that the  $\mathbf{k}_i^0$  are in the basin of attraction, there will be the result

$$\mathbf{k}_i = \lim_{j \rightarrow \infty} \mathbf{k}_i^j. \quad (12.6.6)$$

Alternatively, to achieve more rapid convergence at the expense of a more involved procedure, we may solve (5.2) and (5.3) by Newton's method, which will also be convergent for sufficiently small  $h$  and a sufficiently good guess for the  $\mathbf{k}_i^0$ . Finally, having found the  $\mathbf{k}_i$  by whatever method,  $\mathbf{y}^{n+1}$  is given by (5.1).

It is again instructive to apply this method to the harmonic oscillator example with the Hamiltonian (4.5) and the equations of motion (4.8). In this case the relations (5.2) and (5.3) take the specific form

$$\mathbf{k}_1 = J[\mathbf{y}^n + h(1/4)\mathbf{k}_1 + h(1/4 - \sqrt{3}/6)\mathbf{k}_2], \quad (12.6.7)$$

$$\mathbf{k}_2 = J[\mathbf{y}^n + h(1/4 + \sqrt{3}/6)\mathbf{k}_1 + h(1/4)\mathbf{k}_2]. \quad (12.6.8)$$

As they stand these relations, because of their vector/matrix form, comprise four linear equations in four unknowns, namely the components of the two two-dimensional vectors  $\mathbf{k}_1$  and  $\mathbf{k}_2$ . With sufficient effort, linear equations can always be solved. In this case there is a trick that simplifies the problem. First rewrite (5.8) to bring all terms involving  $\mathbf{k}_2$  to the left and all other terms to the right. So doing yields the relation

$$(I - hJ/4)\mathbf{k}_2 = J[\mathbf{y}^n + h(1/4 + \sqrt{3}/6)\mathbf{k}_1]. \quad (12.6.9)$$

Next solve (5.9) for  $\mathbf{k}_2$  in terms of everything else to find the result

$$\mathbf{k}_2 = (I - hJ/4)^{-1} J[\mathbf{y}^n + h(1/4 + \sqrt{3}/6)\mathbf{k}_1]. \quad (12.6.10)$$

Now substitute (5.10) into (5.7) to obtain a relation involving only  $\mathbf{k}_1$ ,

$$\mathbf{k}_1 = J\{\mathbf{y}^n + h(1/4)\mathbf{k}_1 + h(1/4 - \sqrt{3}/6)(I - hJ/4)^{-1} J[\mathbf{y}^n + h(1/4 + \sqrt{3}/6)\mathbf{k}_1]\}. \quad (12.6.11)$$

Manipulate this relation to find, as intermediate steps, the results

$$\begin{aligned} & \mathbf{k}_1 - Jh(1/4)\mathbf{k}_1 + h(1/4 - \sqrt{3}/6)(I - hJ/4)^{-1}h(1/4 + \sqrt{3}/6)\mathbf{k}_1 \\ &= J[\mathbf{y}^n + h(1/4 - \sqrt{3}/6)(I - hJ/4)^{-1}J\mathbf{y}^n], \end{aligned} \quad (12.6.12)$$

or,

$$\begin{aligned} & [I - Jh(1/4) - h^2(1/48)(I - hJ/4)^{-1}]\mathbf{k}_1 \\ &= [J - h(1/4 - \sqrt{3}/6)(I - hJ/4)^{-1}]\mathbf{y}^n, \end{aligned} \quad (12.6.13)$$

and, as a last step, the explicit solution

$$\mathbf{k}_1 = [I - hJ/4 - (h^2/48)(I - hJ/4)^{-1}]^{-1} [J - h(1/4 - \sqrt{3}/6)(I - hJ/4)^{-1}] \mathbf{y}^n. \quad (12.6.14)$$

(Here we have used the relation  $J^2 = -I$ .) Finally, insert (5.14) into (5.10) to yield an explicit result for  $\mathbf{k}_2$ ,

$$\mathbf{k}_2 = . \quad (12.6.15)$$

## Exercises

**12.6.1.**

**12.6.2.**

## 12.7 Numerical Examples for One- and Two-Stage Methods

## 12.8 How Much Iteration Is Required for Implicit Methods?

Review Exercise 2.4.14, which treated the Predictor-Corrector method  $PECECEC\cdots$ , where  $P$  is an explicit calculation which evaluated the predictor formula, and  $ECECEC\cdots$  is a sequence of iterations intended to evaluate the implicit corrector formula. In that exercise we concluded that for the iteration part of the predictor-corrector method there is no point in carrying out the iteration procedure until convergence is achieved to machine precision. Employ the phrase *termination error* to denote the difference between the result obtained by terminating the iteration procedure and the result obtained by iterating to machine precision. We concluded that, once the termination error became smaller under iteration than the truncation error expected for the corrector formula, then there was no point in iterating further. And, if the step size  $h$  is sufficiently small, in practice this goal is achieved after just a few iterations.

What can be said about the number of iterations to be employed in the case of implicit Runge Kutta including implicit Gaussian? If the only consideration is truncation error, there is again no point in iterating further once the termination error becomes smaller than the truncation error. Typically the termination error decreases by a factor of  $h$  at each iteration. See, for example, (4.26). Therefore, since the truncation error is proportional to some fixed power of  $h$ , the termination error is expected to become smaller than the truncation error after only a few iterations provided the step size is sufficiently small. But suppose the integrator is expected, beyond being accurate to some order in  $h$ , to be *exactly symplectic* when employed to integrate Hamiltonian-derived equations of motion. For this to be achieved, the *implicit* integration formula must, in effect, be made *explicit*. See, for example, (4.12) where this goal is achieved analytically for a simple case. For this goal to

be achieved numerically when working to infinite precision, an infinite number of iterations is required. For this goal to be achieved to machine precision, sufficiently many iterations must be performed for the termination error to vanish to machine precision. This number of iterations may be substantially larger than that required to make the termination error smaller than than the truncation error. Consequently implicit symplectic methods, when employed to satisfy the symplectic condition to machine precision, are expected to be quite slow due to the large number of required iterations. Correspondingly, due to the need to iterate until convergence is achieved to machine precision, implicit symplectic integrators are expected to be considerably slower than explicit symplectic integrators (assuming an explicit symplectic integrator can be found for the Hamiltonian of interest).

## 12.9 Proof of Condition for Runge-Kutta to be Symplectic

There are at least three ways to verify that (3.21) is a necessary and sufficient condition for  $\mathcal{M}_{\text{RK}}$  to be symplectic. The first requires making a brute force calculation of the Jacobian matrix  $M_{\text{RK}}$  associated with  $\mathcal{M}_{\text{RK}}$  followed by a verification that  $M_{\text{RK}}$  is a symplectic matrix. That is what we have just done for the simplest case  $s = 1$ . See also Exercise 3.2. The second, far more elegant and compact, makes use of differential forms. The third, which we will employ here, uses the contents of the Butcher tableau to define a generating function  $F_2$ . It then demonstrates that the generating produced by  $F_2$  reproduces the Runge-Kutta step for the Butcher tableau provided the contents of the Butcher tableau satisfy (3.21).

Consider the generating function  $F_2(\mathbf{q}, \mathbf{P}, t^n; h)$  defined by writing

$$F_2(\mathbf{q}, \mathbf{P}, t^n; h) = \mathbf{q} \cdot \mathbf{P} + G_2(\mathbf{q}, \mathbf{P}, t^n; h), \quad (12.9.1)$$

where

$$G_2 = G_2^1 + G_2^2 \quad (12.9.2)$$

with

$$G_2^1 = h \sum_{i=1}^s b_i H(\mathbf{q}_i, \mathbf{p}_i, t_i), \quad (12.9.3)$$

$$G_2^2 = -h^2 \sum_{i,j=1}^s b_i a_{ij} [\mathbf{H}_q(\mathbf{q}_i, \mathbf{p}_i, t_i) \cdot \mathbf{H}_p(\mathbf{q}_j, \mathbf{p}_j, t_j)]. \quad (12.9.4)$$

The thoughtful reader may find the definition given by (7.1) through (7.4) puzzling because the terms on the right sides of (7.3) and (7.4) are functions of the old phase-space variables  $\mathbf{q}, \mathbf{p}$  while the phase-space arguments of  $G_2$  are specified as being the mixed pair  $\mathbf{q}, \mathbf{P}$ . Here is what is meant: The relation (3.14) specifies  $\mathbf{P}$  as a function of  $\mathbf{q}, \mathbf{p}, t^n$ ; and  $h$ ,

$$\mathbf{P} = \mathbf{P}(\mathbf{q}, \mathbf{p}, t^n; h). \quad (12.9.5)$$

This relation is to be partially inverted to yield  $\mathbf{p}$  as a function of  $\mathbf{q}, \mathbf{P}, t^n$  and  $h$ ,

$$\mathbf{p} = \mathbf{p}(\mathbf{q}, \mathbf{P}, t^n; h). \quad (12.9.6)$$

From the form of (3.14) we see, by the inverse function theorem, that such an inversion is possible for small enough  $h$  because then  $\mathbf{p} \simeq \mathbf{P}$ . Finally, the right side of (7.6) is to be substituted for  $\mathbf{p}$  in the right sides of (7.3) and (7.4) to yield  $G_2(\mathbf{q}, \mathbf{P}, t^n; h)$ .

Suppose we use  $F_2(\mathbf{q}, \mathbf{P}, t^n; h)$  to produce a transformation that sends  $\mathbf{q}, \mathbf{p}$  to  $\mathbf{Q}, \mathbf{P}$  by the standard rules (6.5.5),

$$p_k = \partial F_2 / \partial q_k, \quad Q_k = \partial F_2 / \partial P_k. \quad (12.9.7)$$

From the work of Section 6.5.1 we know that so doing will produce a symplectic map, which we will call  $\mathcal{M}$ . With a view to implementing (7.7), let us compute (save for holding  $t^n$  fixed) the total differential of  $G_2(\mathbf{q}, \mathbf{P}, t^n; h)$ .

## 12.10 Symplectic Integration of General Hamiltonians Using Generating Functions

Section 6.7.3 on the Hamilton-Jacobi equation studied the relation between Hamiltonians and generating functions. There it was shown, once a Darboux matrix  $\alpha$  has been selected, that there is a unique relation between the Hamiltonian  $H(Z, t)$  and the source function  $g(u, t)$ . Given  $H$ , some time  $t^n$ , some phase-space point  $z^n$ , and a time step  $h$ , we wish to find  $z^{n+1}$  at time  $t^{n+1} = t^n + h$  in such a way that the relation between  $z^{n+1}$  and  $z^n$  is symplectic and  $z^{n+1}$  is very nearly (with error of order  $h^{m+1}$ ) equal to the result found by exactly integrating the equations generated by  $H$  starting with initial conditions  $z^n$  at  $t = t^n$  and integrating to  $t = t^n + h$ . We assume that the trajectory generated by  $H$  is analytic in  $t$ , which will be the case if  $H(z, t)$  is analytic in the phase-space variables  $z$  and the time  $t$ . (See Poincaré's theorem in Section 1.3.) Then  $z(t^n + h)$  will have a Taylor expansion in  $h$ .

## 12.11 Explicit Symplectic Integrator for Motion in General Electromagnetic Fields

We have seen that there are *implicit* symplectic Runge-Kutta integrators for general Hamiltonians, and hence also for motion in general electromagnetic fields. Remarkably, there are *explicit* symplectic integrators for the Hamiltonian (1.6.192) and, by extension, for the Hamiltonian (1.6.77).

We begin with the simpler case, the Hamiltonian (1.6.192), which can be written in the form

$$H = H_x + H_y + H_z \quad (12.11.1)$$

where

$$H_x = (p_x - qA_x)^2 / (2m^*), \quad (12.11.2)$$

$$H_y = (p_y - qA_y)^2 / (2m^*), \quad (12.11.3)$$

$$H_z = (p_z - qA_z)^2/(2m^*). \quad (12.11.4)$$

Since  $H$  is time independent, the relation (1.1) still holds. Moreover, we may again subdivide the time axis into equal steps of duration  $h$  to obtain the exact marching rule (1.4). However, since in the present context the symbol  $z$  is being used to denote a coordinate, we rewrite (1.4) in the form

$$w^{n+1} = \exp(h : -H :) w^n \quad (12.11.5)$$

where  $w$  denotes the collection of phase-space variables

$$w = (x, y, z, ; p_x, p_y, p_z). \quad (12.11.6)$$

We next make the approximation

$$\exp(h : -H :) \cong \mathcal{S}_2(h) \quad (12.11.7)$$

where  $\mathcal{S}_2(k)$  is now defined by the rule

$$\begin{aligned} \mathcal{S}_2(h) &= \exp[-(h/2) : H_x :] \exp[-(h/2) : H_y :] \times \\ &\quad \exp[-h : H_z :] \exp[-(h/2) : H_y :] \exp[-(h/2) : H_x :]. \end{aligned} \quad (12.11.8)$$

Upon combining the exponents on the right side of (9.8) to first order in  $h$ , we see that the exponent on the left side of (9.7) is regained. Also, by construction,  $\mathcal{S}_2$  satisfies (1.16). Therefore, as the notation is intended to indicate,  $\mathcal{S}_2$  is a symmetric integrator that is locally correct through terms of order  $h^2$ . Note other permutations of  $H_x$ ,  $H_y$ , and  $H_z$  could have been used in the definition of  $\mathcal{S}_2$ . There are thus  $3!$  possible formulas of the kind (9.8).

We are still faced with the problem of evaluating the action of the individual factors on the right side of (9.8). We will see that this problem can be solved with the use of gauge transformations.<sup>9</sup> Define functions  $U_x(x, y, z)$ ,  $U_y(x, y, z)$ , and  $U_z(x, y, z)$  by the requirements

$$A_x = \partial U_x / \partial x, \quad A_y = \partial U_y / \partial y, \quad A_z = \partial U_z / \partial z. \quad (12.11.9)$$

There are many such functions, and we may choose among them at will at each integration step. For example, we may write

$$U_x = \int^x A_x(x', y, z) dx' \quad (12.11.10)$$

and add to it any function of  $y$  and  $z$ . Use the  $U$ 's to make symplectic maps  $\mathcal{A}_x$ ,  $\mathcal{A}_y$ , and  $\mathcal{A}_z$  defined by the relations

$$\mathcal{A}_x = \exp(-q : U_x :), \quad \mathcal{A}_y = \exp(-q : U_y :), \quad \mathcal{A}_z = \exp(-q : U_z :). \quad (12.11.11)$$

These maps produce gauge transformations. See Exercise 6.2.8. It is easily verified, for example, that  $\mathcal{A}_x$  and  $\mathcal{A}_x^{-1}$  have the phase-space actions

$$\mathcal{A}_x x = x, \quad \mathcal{A}_x y = y, \quad \mathcal{A}_x z = z, \quad (12.11.12)$$

---

<sup>9</sup>This ingenious idea is due to Wu, Forest, and Robin. See the references at the end of this chapter.

$$\mathcal{A}_x^{-1}x = x, \quad \mathcal{A}_x^{-1}y = y, \quad \mathcal{A}_x^{-1}z = z, \quad (12.11.13)$$

$$\mathcal{A}_x p_x = p_x - q : U_x : p_x = p_x - q[U_x, p_x] = p_x - q(\partial U_x / \partial x) = p_x - qA_x, \quad (12.11.14)$$

$$\mathcal{A}_x^{-1}p_x = p_x + q : U_x : p_x = p_x + qA_x, \quad (12.11.15)$$

$$\mathcal{A}_x p_y = p_y - q : U_x : p_y = p_y - q[U_x, p_y] = p_y - q(\partial U_x / \partial y), \quad (12.11.16)$$

$$\mathcal{A}_x^{-1}p_y = p_y + q : U_x : p_y = p_y + q(\partial U_x / \partial y), \quad (12.11.17)$$

$$\mathcal{A}_x p_z = p_z - q : U_x : p_z = p_z - q[U_x, p_z] = p_z - q(\partial U_x / \partial z), \quad (12.11.18)$$

$$\mathcal{A}_x^{-1}p_z = p_z + q : U_x : p_z = p_z + q(\partial U_x / \partial z). \quad (12.11.19)$$

In particular, it follows from (9.14) and (9.15) that  $\mathcal{A}_x$  has the property

$$\begin{aligned} \mathcal{A}_x \exp[-(h/2) : p_x^2 / (2m^*) :] \mathcal{A}_x^{-1} &= \exp[-(h/2) : (p_x - qA_x)^2 / (2m^*)] \\ &= \exp[-(h/2) : H_x :]. \end{aligned} \quad (12.11.20)$$

We note at this point that it does not matter what special choice is made for  $U_x$ , see (9.10), because from (9.20) it is evident that all allowed choices yield the same net result.

As a consequence of (9.20) and similar relations,  $\mathcal{S}_2$  can be rewritten in the factored product form

$$\begin{aligned} \mathcal{S}_2 = & \mathcal{A}_x \exp[-(h/2) : \bar{H}_x :] \mathcal{A}_x^{-1} \times \\ & \mathcal{A}_y \exp[-(h/2) : \bar{H}_y :] \mathcal{A}_y^{-1} \times \\ & \mathcal{A}_z \exp[-h : \bar{H}_z :] \mathcal{A}_z^{-1} \times \\ & \mathcal{A}_y \exp[-(h/2) : \bar{H}_y :] \mathcal{A}_y^{-1} \times \\ & \mathcal{A}_x \exp[-(h/2) : \bar{H}_x :], \end{aligned} \quad (12.11.21)$$

where we have used the notation

$$\bar{H}_x = p_x^2 / (2m^*), \quad (12.11.22)$$

$$\bar{H}_y = p_y^2 / (2m^*), \quad (12.11.23)$$

$$\bar{H}_z = p_z^2 / (2m^*). \quad (12.11.24)$$

We have already seen that the phase-space actions of the  $\mathcal{A}$ 's and  $\mathcal{A}^{-1}$ 's can be evaluated exactly using relations of the form (9.12) through (9.19). Evidently the actions of the maps  $\exp[-(h/2) : \bar{H}_x :]$ ,  $\exp[-(h/2) : \bar{H}_y :]$ , and  $\exp[-h : \bar{H}_z :]$  can also be evaluated exactly. See Exercises 5.4.1 and 5.4.2. Therefore, use of the approximation (9.7) with  $\mathcal{S}_2$  given by (9.21) produces a symmetric integrator that is locally correct through terms of order  $h^2$  and is exactly symplectic.

At this point, two remarks are in order. The first is that, with  $\mathcal{S}_2(k)$  in hand, the triplet construction can be used to produce higher-order symmetric and symplectic integrators. For example,  $\mathcal{S}_4$  is given by (1.32) through (1.34).

The second remark is less triumphant. For a symmetric integrator there is the general relation

$$\mathcal{S}_{2k}(h) = \exp[h : -H : +O(h^{2k+1})]. \quad (12.11.25)$$

Moreover, the error term does not commute with  $\dot{H}$  so that for each integration step there is the result

$$\mathcal{S}_{2k}(h)H = H + O(h^{2k+1}) \quad (12.11.26)$$

where the error term is *nonzero*. In fact *Ge* and *Marsden* have essentially shown that it is impossible to construct an integrator that is exactly symplectic and also exactly conserves  $H$ .<sup>10</sup> In some applications this may not much matter. Indeed, it is sometimes argued that a symplectic integrator can be used with a larger time step  $h$  than a nonsymplectic integrator of the same order because the symplectic integrator at least respects the underlying structure of any Hamiltonian system. And the fact (hope) that a larger time step can be used compensates for the relatively large amount of work associated with each time step. [Moreover, the variation in  $H$  during symplectic integration is often observed to be essentially periodic when the trajectory being integrated is essentially periodic, and this good behavior can be understood using the BCH series. By contrast, the value of  $H$  typically grows (or damps) linearly or quadratically, or eventually even exponentially, in time when nonsymplectic integration is employed.] However, for the Hamiltonian (1.6.192) we know that the only physically meaningful trajectories are those for which  $H$  has the value (1.6.193). Therefore in this case it is necessary to use a time step  $h$  that is sufficiently small to ensure that over the course of integration  $H$  obeys (1.6.193) to high accuracy. In particular, one should monitor the value of  $H$  during the course of integration to verify that (1.6.193) is met with sufficient accuracy.

While in the mode of exploring difficulties associated with this approach, we should also consider how much effort is required to carry out the integrations of the form (9.10) required to compute the functions  $U_x(x, y, z)$  through  $U_z(x, y, z)$ . If the vector potential  $\mathbf{A}(\mathbf{r})$  is known in analytic form and has a sufficiently simple structure, then these integrals can be evaluated analytically in terms of elementary functions prior to any use of the symplectic integrator. However for most if not all realistic applications, these integrations yield higher transcendental functions that are expensive to evaluate, or these integrations must be carried out numerically. And these evaluations/numerical integrations must be performed with high accuracy if the symplecticity of the overall integration process is to be assured. These evaluations/integrations further add to the already high computational overhead associated with symplectic integration. (See Exercise 9.3.) The considerations of this and the previous paragraph make one wonder if the computational burden for symplectic integration of the kind just described for trajectories in realistic electromagnetic fields is so high as to make nonsymplectic, but high-order, integration superior to symplectic integration. The answer to this question is presumably problem dependent. Its answer for any realistic problem would require the comparison of symplectic integration and high accuracy (probably not Runge-Kutta) nonsymplectic integration. When such explorations are made, one should also consider symplectic (but implicit) Runge-Kutta methods and symplectic generating

---

<sup>10</sup>Ge and Marsden have shown that if a symplectic integrator conserves  $H$ , it must be exact or, at worst, produce exact trajectories up to a reparameterization of the time. (See Section 4 for an example where this happens.) Now in general a symplectic integrator cannot produce exact trajectories because in general it makes errors of some order in  $h$ . Moreover, it is unlikely that these errors only amount to a reparameterization of the time. Therefore, generally  $H$  is not conserved. Of course, one can easily check during the course of an integration to see if the value  $H$  is changing, and in general, as expected, one finds that it is.

function methods of the kind described in the previous subsections.

We close this section with an analogous discussion of the Hamiltonian (1.6.77), which can be written in the form

$$H_R = H_x + H_y + H_z + H_t \quad (12.11.27)$$

where

$$H_x = (p_x - qA_x)^2/(2mc), \quad (12.11.28)$$

$$H_y = (p_y - qA_y)^2/(2mc), \quad (12.11.29)$$

$$H_z = (p_z - qA_z)^2/(2mc), \quad (12.11.30)$$

$$H_t = -(p_4 + qA^4)^2/(2mc), \quad (12.11.31)$$

and  $w$  becomes the collection of phase-space variables

$$w = (x, y, z, x^4; p_x, p_y, p_z, p_4). \quad (12.11.32)$$

Since  $H_R$  is  $\tau$  independent, the transfer map associated with  $H_R$  is given by the relation

$$\mathcal{M} = \exp(\tau : -H_R :). \quad (12.11.33)$$

As expected, we subdivide the  $\tau$  axis into steps of equal amount  $h$  to obtain the exact marching rule (9.5).

In analogy with (9.7) and (9.8), we next make the approximation

$$\exp(h : -H_R :) \cong \mathcal{S}_2(h) \quad (12.11.34)$$

where  $\mathcal{S}_2(h)$  is now defined by the rule

$$\begin{aligned} \mathcal{S}_2(h) = & \exp[-(h/2) : H_x :] \exp[-(h/2) : H_y :] \times \\ & \exp[-(h/2) : H_z :] \exp[-h : H_t :] \exp[-(h/2) : H_z :] \times \\ & \exp[-(h/2) : H_y :] \exp[-(h/2) : H_x :]. \end{aligned} \quad (12.11.35)$$

By construction this  $\mathcal{S}_2$  is a symmetric integrator that is locally correct through terms of order  $h^2$ . Other permutations of  $H_x$  through  $H_t$  could have been used in the definition of  $\mathcal{S}_2$ , and there are therefore  $4!$  possible integrators of this kind.

We again define functions  $U_x$  through  $U_z$  by the requirements (9.10), and we add to their collection the function  $U^4$  defined by the requirement

$$A^4 = -\partial U^4 / \partial x^4. \quad (12.11.36)$$

The symplectic maps  $\mathcal{A}_x$  through  $\mathcal{A}_z$  are also again defined by (9.11), and to their collection we add the symplectic map  $\mathcal{A}_4$  defined by

$$\mathcal{A}_4 = \exp : -qU^4 :. \quad (12.11.37)$$

This map has the property

$$\mathcal{A}_4 p_4 = p_4 - q : U^4 : p_4 = p_4 - q[U^4, p_4] = p_4 - q\partial U^4 / \partial x^4 = p_4 + qA^4, \quad (12.11.38)$$

from which it follows that

$$\exp(-h : H_t :) = \mathcal{A}_4 \exp[h : p_4^2/(2mc) :] \mathcal{A}_4^{-1}. \quad (12.11.39)$$

We are now able to proceed as before to express  $\mathcal{S}_2$  as a product of maps, all of which can be evaluated explicitly,

$$\begin{aligned} \mathcal{S}_2 = & \mathcal{A}_x \exp[-(h/2) : \bar{H}_x :] \mathcal{A}_x^{-1} \times \\ & \mathcal{A}_y \exp[-(h/2) : \bar{H}_y :] \mathcal{A}_y^{-1} \times \\ & \mathcal{A}_z \exp[-(h/2) : \bar{H}_z :] \mathcal{A}_z^{-1} \times \\ & \mathcal{A}_4 \exp[-h : \bar{H}_t :] \mathcal{A}_4^{-1} \times \\ & \mathcal{A}_z \exp[-(h/2) : \bar{H}_z :] \mathcal{A}_z^{-1} \times \\ & \mathcal{A}_y \exp[-(h/2) : \bar{H}_y :] \mathcal{A}_y^{-1} \times \\ & \mathcal{A}_x \exp[-(h/2) : \bar{H}_x :] \mathcal{A}_x^{-1}. \end{aligned} \quad (12.11.40)$$

Here we have used the notation

$$\begin{aligned} \bar{H}_x &= p_x^2/(2mc), \\ \bar{H}_y &= p_y^2/(2mc), \\ \bar{H}_z &= p_z^2/(2mc), \\ \bar{H}_t &= -p_4^2/(2mc). \end{aligned} \quad (12.11.41)$$

We must again be aware that trajectories generated by  $H_R$  are only physically meaningful when  $H_R$  has a special value, namely that given by the (mass shell) condition (1.6.92), and that this value cannot be maintained exactly by any symplectic integrator. Therefore it is again necessary to choose  $h$  sufficiently small to ensure that over the course of integration  $H_R$  obeys (1.6.92) to high accuracy. Moreover there is again added overhead. Now we must compute the functions  $U_x$  through  $U^4$ . See Exercise 9.3.

## Exercises

**12.11.1.** Exercise on what happens when, in  $H_R$ ,  $A^4 = 0$  and  $\mathbf{A}$  is static.

**12.11.2.** Show that in the nonrelativistic approximation the Lagrangian (1.5.1) may be replaced by the Langrangian

$$L_{NR} = (1/2)mv^2 - q\psi(\mathbf{r}, t) + q\mathbf{v} \cdot \mathbf{A}(\mathbf{r}, t). \quad (12.11.42)$$

Find the associated Hamiltonian  $H_{NR}$ . Show that  $H_{NR}$  is conserved if the electromagnetic fields are static. Construct a symplectic integrator for  $H_{NR}$ . Show that this integrator does not conserve  $H_{NR}$ . Perhaps this nonconservation is not so important in the nonrelativistic case because it might be argued that  $H_{NR}$  has no *fundamental* significance. Moreover,  $H_{NR}$  is not conserved anyway for the exact motion if the electromagnetic fields are time dependent.

**12.11.3.** Exercise on what exactly is involved in computing the integrals  $U_x$  through  $U_z$  or  $U_x$  through  $U^4$ .

## 12.12 Zassenhaus Formulas and Map Computation

The discussion so far has dealt mostly with the use of Zassenhaus formulas of the kinds (1.14), (1.35), (1.38), (9.21), and (9.40) as *symplectic integrators*. However, Zassenhaus formulas can also be used to *compute maps* both in Taylor and factored product form.

### 12.12.1 Case of $T + V$ or General Electromagnetic Field Hamiltonians

As a simple example, suppose that  $H$  has the  $T + V$  decomposition (1.6). Then, the Taylor maps for  $\exp(\sigma : T :)$  and  $\exp(\sigma : V :)$ , where  $\sigma$  is some parameter, can be computed exactly by formulas analogous to (1.7) through (1.10); and therefore we can find the net Taylor map for any factored map of the kinds (1.14), (1.35), or (1.38). Also, we can expand  $T$  and  $V$  in homogeneous polynomials. For example, we can write

$$T = \sum_{m=0}^{\infty} T_m \quad (12.12.1)$$

with a similar expansion for  $V$ . Then we have the factored product representation

$$\exp(\sigma : T :) = \exp(\sigma : T_1 :) \exp(\sigma : T_2 :) \exp(\sigma : T_3 :) \cdots, \quad (12.12.2)$$

with a similar representation for  $\exp(\sigma : V :)$ . It follows that each of the factors appearing in any approximation of the kinds (1.14), (1.35), or (1.38) for the map  $\exp(h : -H :)$  can be written in factored product form. The resulting factors can then be concatenated together to yield for the map  $\exp(h : -H :)$  a final approximation that is also in factored product form.

After some reflection, we see that the same procedure can be applied to the symplectic integrators (9.21) and (9.40). Each of the factors can be expanded in Taylor form or written in factored product form, and these maps can then be concatenated to yield a net map either in Taylor or factored product form.

There is one last remark to be made: As explained in Section 1.6, it is often convenient to have maps for which some coordinate is the independent variable, and in this subsection we have been using the time  $t$  or some world-line parameter  $\tau$  as the independent variable. If we wish to compute maps rather than trajectories (apart from the reference trajectory) this problem can be overcome with the use of *matching* maps. When the map for some element has been computed using  $t$  or  $\tau$  as an independent variable, the necessary conversion can be made by preceding the map with a matching map that transforms from phase-space variables for which some coordinate is the independent variable to phase-space variables for which  $t$  or  $\tau$  is the independent variable, and following the map by a second matching map that transforms back to the phase-space variables for which some coordinate is the independent variable. See Section \*.\*.

### 12.12.2 Case of Hamiltonians Expanded in Homogeneous Polynomials

Zassenhaus formulas can also be used to provide factored product approximations for  $\mathcal{M} = \exp(h : -H :)$  when  $H$  is decomposed into homogeneous polynomials as in (10.9.1). Here we will consider the autonomous case. The nonautonomous case is best treated using the methods of Sections 10.5.2 and 10.6.2.

#### Derivation

Define  $A$  and  $B$  by writing the equations

$$A = H_1 + H_2, \quad (12.12.3)$$

$$B = H_r = H_3 + H_4 + \dots . \quad (12.12.4)$$

Evidently, any map of the kind  $\exp(-h : A :)$  with  $A$  given by (10.3) can be written in factored product form using the methods of Section 9.2. See (9.2.4), (9.2.7), and (9.2.9). What about maps of the kind  $\exp(-h : B :)$  with  $B$  given by (10.4)? How do we find generators  $f_m$  such that

$$\exp(-h : B :) = \exp(-h : H_3 + H_4 + \dots :) = \exp(: f_3 :) \exp(: f_4 :) \dots ? \quad (12.12.5)$$

We note that, since there is no  $H_2$  term in (10.4), we may use the methods of Section 10.6.2 with the understanding that

$$\mathcal{M}_2 = \mathcal{I} \quad (12.12.6)$$

and therefore

$$H_m^{\text{int}} = H_m. \quad (12.12.7)$$

It follows from (10.6.14) through (10.6.20) that

$$f_3 = -hH_3, \quad (12.12.8)$$

$$f_4 = -hH_4, \quad (12.12.9)$$

$$f_5 = -hH_5 - (h^2/2)[H_3, H_4], \quad (12.12.10)$$

$$f_6 = -hH_6 - (h^2/2)[H_3, H_5] - (h^3/6)[H_3, [H_3, H_4]], \quad (12.12.11)$$

$$\begin{aligned} f_7 &= -hH_7 - (h^2/2)([H_3, H_6] + [H_4, H_5]) - h^3(H_3, [H_3, H_5])/6 + [H_4, [H_3, H_4]]/3 \\ &\quad - (h^4/24)[H_3, [H_3, [H_3, H_4]]], \end{aligned} \quad (12.12.12)$$

$$\begin{aligned} f_8 &= -hH_8 - (h^2/2)([H_3, H_7] + [H_4, H_6]) \\ &\quad - h^3([H_3, [H_3, H_6]]/6 + [H_4, [H_3, H_5]]/3 + [H_5, [H_3, H_4]]/12) \\ &\quad - h^4([H_3, [H_3, [H_3, H_5]]]/24 + [H_4, [H_3, [H_3, H_4]]]/8) \\ &\quad - (h^5/120)[H_3, [H_3, [H_3, [H_3, H_4]]]] \end{aligned} \quad (12.12.13)$$

$$f_m = \text{expression involving } H_m \text{ and the } H_\ell \text{ with } \ell < m. \quad (12.12.14)$$

See Exercise 10.1.

We conclude that all the factors in a Zassenhaus representation can themselves be written in factored product form. These maps can now be concatenated together to yield a final approximation for  $\exp(h : -H :)$  that is also in factored product form and is accurate through some order in  $h$ . (Note that if  $H_1$  terms are present in  $H$ , then we must at this point assume they are small in order to use the concatenation formulas of Section 9.3.) The net result of our discussion is that the use of Zassenhaus symplectic integrator formulas makes it possible to find “linear” maps  $\mathcal{R}(h)$  and generators  $f_m(h)$  such that  $\exp(-h : H :)$  has the factorization

$$\exp(-h : H :) = \exp[: f_1(h) :] \mathcal{R}(h) \exp[: f_3(h) :] \exp[: f_4(h) :] \cdots \times [1 + O(h^{N+1})] \quad (12.12.15)$$

where  $(N + 1)$  is the order of the error in the Zassenhaus formula.

### Application to Scaling, Splitting, and Squaring

For the autonomous case that we have just been considering, we are now able to use scaling, splitting, and squaring (as in Section 10.7) with (10.17) now playing the role of a splitting formula. As before, we define  $\tau$  by writing

$$\tau = t/2^n, \quad (12.12.16)$$

and find the approximation

$$\begin{aligned} \mathcal{M} &= \exp(t : -H :) = [\exp(\tau : -H :)]^{2^n} \\ &= \{\cdots \{ \{ \exp[: f_1(\tau) :] \mathcal{R}(\tau) \exp[: f_3(\tau) :] \exp[: f_4(\tau) :] \cdots \}^2 \cdots \}^2 \cdots \}^2 \\ &\quad (n \text{ squarings}). \end{aligned} \quad (12.12.17)$$

[Note: The quantity  $\tau$  as given by (10.16) should not be confused with that used in (1.3) or (9.33).]

What will be the relative error for this approximation? We expect that it will scale as  $(1/2^n)^N$ . For a more precise result we need an estimate for the error term in the underlying Zassenhaus formula. Suppose, for example, we use the Zassenhaus formula  $\mathcal{S}_4$  as given by (1.35). Although we have discussed Zassenhaus formulas in the context of symplectic integrators, they are really *operator identities* that hold for any set of linear operators. (Consequently, they may be used in other contexts including the construction of integrators designed to preserve group properties for any Lie group. For example, they may be used in the context of rigid-body motion to preserve the orthogonality condition and in the context of quantum dynamics to preserve the unitarity condition.) To emphasize this fact, let us introduce the notation

$$\begin{aligned} \mathcal{S}_4(h\mathcal{A}, h\mathcal{B}) &= \exp(w_1 h\mathcal{A}) \exp(w_2 h\mathcal{B}) \exp(w_3 h\mathcal{A}) \exp(w_4 h\mathcal{B}) \times \\ &\quad \exp(w_5 h\mathcal{A}) \exp(w_6 h\mathcal{B}) \exp(w_7 h\mathcal{A}), \end{aligned} \quad (12.12.18)$$

where  $\mathcal{A}$  and  $\mathcal{B}$  are any pair of linear operators. Then we have a result of the form

$$\mathcal{S}_4(h\mathcal{A}, h\mathcal{B}) = \exp[-h(\mathcal{A} + \mathcal{B}) + \mathcal{C}_5 h^5 + O(h^7)]. \quad (12.12.19)$$

What we need for error analysis is an estimate for the term  $\mathcal{C}_5$ . In analogy to the error term in (9.13) and based on the general properties of the BCH series, we expect that  $\mathcal{C}_5$  will be made of multiple commutators of  $\mathcal{A}$  and  $\mathcal{B}$ . First there will be a term with four  $\mathcal{A}$ 's and one  $\mathcal{B}$ . Next there will be terms with three  $\mathcal{A}$ 's and two  $\mathcal{B}$ 's, etc. Finally, there will be a term with one  $\mathcal{A}$  and four  $\mathcal{B}$ 's. Consequently,  $\mathcal{C}_5$  can be written in the form

$$\mathcal{C}_5(\mathcal{A}, \mathcal{B}) = d_1 \# \mathcal{A} \#^4 \mathcal{B} + \cdots + d_{\text{last}} \# \mathcal{B} \#^4 \mathcal{A}. \quad (12.12.20)$$

Here, in accord with the notation of Chapter 8,  $\#\mathcal{A}\#$  denotes the adjoint of  $\mathcal{A}$  as defined in terms of the commutator,

$$\#\mathcal{A}\#\mathcal{B} = \{\mathcal{A}, \mathcal{B}\}. \quad (12.12.21)$$

Thus, to specify  $\mathcal{C}_5$ , we need to determine the coefficients  $d_1 \cdots d_{\text{last}}$ .

The determination of all the coefficients  $d_1 \cdots d_{\text{last}}$  is a sizable algebraic task. However, we can find  $d_1$  and  $d_{\text{last}}$  fairly easily. Since (10.16) is an operator identity, as we have just discussed, it must hold for any linear operators  $\mathcal{A}$  and  $\mathcal{B}$ . In particular, it must hold for  $2 \times 2$  matrices. Let  $F$  and  $G$  be the matrices

$$F = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (12.12.22)$$

$$G = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \quad (12.12.23)$$

They satisfy the commutation rules

$$\#F\#G = \{F, G\} = 2G, \quad (12.12.24)$$

$$\#G\#F = \{G, F\} = -\{F, G\} = -2G. \quad (12.12.25)$$

It follows from these rules that

$$\mathcal{C}_5(F, G) = d_1 \# F \#^4 G + \cdots + d_{\text{last}} \# G \#^4 F = 2^4 d_1 G, \quad (12.12.26)$$

$$\mathcal{C}_5(G, F) = d_1 \# G \#^4 F + \cdots + d_{\text{last}} \# F \#^4 G = 2^4 d_{\text{last}} G. \quad (12.12.27)$$

That is, only the first term in the expansion (10.17) contributes to  $\mathcal{C}_5(F, G)$ , and only the last term contributes to  $\mathcal{C}_5(G, F)$ . Consequently, we have the matrix identity

$$\begin{aligned} \mathcal{S}_4(hF, hG) &= \exp[-h(F + G) + 16d_1 Gh^5 + O(h^7)] \\ &= \exp[-h(F + G)] \exp[16d_1 Gh^5 + O(h^6)] \\ &= \exp[-h(F + G)][1 + 16d_1 Gh^5 + O(h^6)]. \end{aligned} \quad (12.12.28)$$

From (10.25) it follows that

$$16d_1 Gh^5 = \mathcal{S}_4(hF, hG) - \exp[-h(F + G)] + O(h^6). \quad (12.12.29)$$

Similarly, we have the result

$$16d_{\text{last}} Gh^5 = \mathcal{S}_4(hG, hF) - \exp[-h(F + G)] + O(h^6). \quad (12.12.30)$$

At this point we observe that the right sides of (10.26) and (10.27) can be evaluated exactly. First, it is easily verified that the matrix  $(F + G)$  has the property

$$(F + G)^2 = I. \quad (12.12.31)$$

Consequently, we have the relation

$$\begin{aligned} \exp[-h(F + G)] &= \cosh[h(F + G)] - \sinh[h(F + G)] = I \cosh(h) - (F + G) \sinh(h) \\ &= \begin{pmatrix} \exp(-h) & -\sinh(h) \\ 0 & \exp(h) \end{pmatrix}. \end{aligned} \quad (12.12.32)$$

Now take 1,2 matrix elements of both sides of (10.26) and (10.27) to get the results

$$16d_1 h^5 = [\mathcal{S}_4(hF, hG)]_{12} + \sinh(h), \quad (12.12.33)$$

$$16d_{\text{last}} h^5 = [\mathcal{S}_4(hG, hF)]_{12} + \sinh(h). \quad (12.12.34)$$

Next, we have the relations

$$\exp(\sigma F) = \begin{pmatrix} \exp(\sigma) & 0 \\ 0 & \exp(-\sigma) \end{pmatrix}, \quad (12.12.35)$$

$$\exp(\sigma G) = \begin{pmatrix} 1 & \sigma \\ 0 & 1 \end{pmatrix}. \quad (12.12.36)$$

Consequently, all the factors in  $\mathcal{S}_4(hF, hG)$  and  $\mathcal{S}_4(hG, hF)$  are known, and the multiplications indicated in (10.15) can be carried out exactly. Finally, we can expand the right sides of (10.30) and (10.31) as power series in  $h$  and extract the terms of degree 5. Doing so gives the results

$$d_1 = (1/16)[18(2)^{2/3} - 40(2^{1/3}) + 22]/[9360(2)^{2/3} - 7200(2)^{1/3} - 5760] \simeq 4.14 \times 10^{-4}, \quad (12.12.37)$$

$$d_{\text{last}} = (1/16)(1/12)\{(1/10) + [14(2^{1/3}) - 11(2^{2/3})]/[2 - (2^{1/3})]^5 \simeq 4.68 \times 10^{-3}. \quad (12.12.38)$$

We are now prepared to make an error analysis similar to that of Section 10.7. For  $A$  and  $B$  given by (10.3) and (10.4), and assuming  $H_1 = 0$  for simplicity, we find for  $\mathcal{C}_5$  the result

$$\mathcal{C}_5 = d_1 \# H_2 \#^4 : H_r : + \cdots + d_{\text{last}} \# H_r \#^4 : H_2 : . \quad (12.12.39)$$

We know, in this context, that  $\mathcal{C}_5$  must be a Lie operator. Let  $C$  be the function associated with  $\mathcal{C}_5$ ,

$$\mathcal{C}_5 =: C : . \quad (12.12.40)$$

With this notation, (10.39) is equivalent to the relation

$$C = d_1 : H_2 :^4 H_r + \cdots + d_{\text{last}} : H_r :^4 H_2. \quad (12.12.41)$$

We will also focus our attention on the first term in (10.41), which is equivalent to the assumption that  $H_2$  has a larger effect than  $H_r$ . Then, in view of (9.36) and in line with our assumption, we want to make the comparison

$$hH_r \xrightarrow{?} h^5 d_1 : H_2 :^4 H_r \quad (12.12.42)$$

where, according to (10.13) and (10.14),

$$h = \tau = t/2^n. \quad (12.12.43)$$

By making use of (7.52) and (7.55), and looking at (10.39), we see that the relative error can be written in the form

$$\text{relative error} \sim d_1(m\lambda/2^n)^4. \quad (12.12.44)$$

Suppose, as in Section 10.7, we limit our attention to the case  $m \leq 8$  and select  $n$  so that (7.57) is satisfied. Then we find the result

$$\text{relative error} \sim d_1(1/20)^4 \simeq 3 \times 10^{-9}. \quad (12.12.45)$$

This error is somewhat larger than that given by (7.59). However, the relations (10.6) through (10.11) are simpler than the relations (7.28) through (7.32). Consequently, the use of Zassenhaus formulas for splitting is easier to program. Of course both errors can be decreased substantially, if desired, by a modest increase in the number of squarings  $n$ .

## Exercises

**12.12.1.** The aim of this exercise is to show that the results (10.8) through (10.14) can be obtained from (10.6.14) through (10.6.20). What you are to do is to integrate the equations (10.6.14) through (10.6.20) with respect to  $t$  over the interval  $t = 0$  to  $t = h$ . Let us begin. Verify that integrating (10.6.14), remembering (10.7) and that  $H_3$  is assumed to be time independent, gives the result

$$f_3 = \int_0^h dt \dot{f}_3 = - \int_0^h dt H_3 = -hH_3. \quad (12.12.46)$$

Therefore (10.8) follows from (10.6.14). Next consider  $f_4$ . From (10.46) we know that

$$f_3(t) = -tH_3. \quad (12.12.47)$$

With regard to the ingredients of (10.5.15), show, in view of (10.7) and (10.47), that in our case

$$(: f_3 : /2)(-H_3^{\text{int}}) = 0. \quad (12.12.48)$$

Verify, therefore, that integration of (10.6.15) in our case yields the result

$$f_4 = \int_0^h dt \dot{f}_4 = - \int_0^h dt H_4 = -hH_4, \quad (12.12.49)$$

in agreement with (10.9). Move on to the case of  $f_5$ . Verify that in our case some of the ingredients in (10.6.16) also vanish so that there is the result

$$f_5 = \int_0^h dt \dot{f}_5 = - \int_0^h dt H_5 + \int_0^h dt t[H_4, H_3] = -hH_5 - (h^2/2)[H_3, H_4], \quad (12.12.50)$$

in agreement with (10.10). What about  $f_6$ ? Verify that in our case several of the ingredients in (10.6.17) vanish so that there is the result

$$\begin{aligned} f_6 &= \int_0^h dt \dot{f}_6 = - \int_0^h dt H_6 + \int_0^h dt :f_5:(-H_3) \\ &= -hH_6 - \int_0^h dt :tH_5 + (t^2/2)[H_3, H_4]:(-H_3) \\ &= -hH_6 - (h^2/2)[H_3, H_5] - (h^3/6)[H_3, [H_3, H_4]], \end{aligned} \quad (12.12.51)$$

in agreement with (10.11). Show that (10.12) and (10.13) can be obtained analogously.

### 12.12.2.

## 12.13 Other Zassenhaus Formulas and Their Use

There is a somewhat different class of Zassenhaus formulas that also merits discussion. Suppose we rewrite (9.13) in the more general form

$$\exp[(s/2)\mathcal{A}] \exp(t\mathcal{B}) \exp[(s/2)\mathcal{A}] = \exp[(s\mathcal{A} + t\mathcal{B}) + \mathcal{C}(s\mathcal{A}, t\mathcal{B})] \quad (12.13.1)$$

where

$$\mathcal{C} = (s^2 t / 24) \{\mathcal{A}, \{\mathcal{A}, \mathcal{B}\}\} - (st^2 / 12) \{\mathcal{B}, \{\mathcal{B}, \mathcal{A}\}\} + \dots \quad (12.13.2)$$

Here, as before,  $\mathcal{A}$  and  $\mathcal{B}$  are any pair of linear operators, and  $s$  and  $t$  are expansion parameters. We see that the left side of (11.1) produces the desired result  $\exp(s\mathcal{A} + t\mathcal{B})$  save for an error  $\mathcal{C}$  that contains, among other things, terms linear in  $s$  (but higher order in  $t$ ) and terms linear in  $t$  (but higher order is  $s$ ). Examination of (10.15) through (10.17) shows that the higher order Zassenhaus integrator formulas have similar properties. For example, reference to (10.16) and (10.17) shows that the  $\mathcal{S}_4$  given by (10.15) has errors linear in  $t$  that are proportional to  $s^4$ , errors quadratic in  $t$  that are proportional to  $s^3$ , etc.

Suppose we set for ourselves what will turn out to be an easier goal: find Zassenhaus approximations that are only correct through terms linear in  $t$ , but the term that is independent of  $t$  and the term that is linear in  $t$  should be correct to *high* order in  $s$ . Our starting point is the relation (8.8.13). See Section 8.8. This relation has the generalization

$$\exp(s\mathcal{A} + t\mathcal{B}) = \exp[O(t^2)] \exp[\text{dex}(s\#\mathcal{A}\#)(t\mathcal{B})] \exp(s\mathcal{A}) \quad (12.13.3)$$

where

$$\text{dex}(s\#\mathcal{A}\#)(t\mathcal{B}) = \int_0^1 d\tau \exp(\tau s\#\mathcal{A}\#)(t\mathcal{B}). \quad (12.13.4)$$

By construction we know that the term in (11.3) that is independent of  $t$  and the term that is linear in  $t$  are both *exact* in  $s$ .

The next step is to convert the integral (11.4) into a finite sum with the aid of a *quadrature formula*. Suppose we wish to integrate some function (operator)  $\mathcal{G}(\tau)$ . A quadrature formula is a set of  $k$  successive *sampling points*  $\tau_i$  in the interval  $[0, 1]$  and *weights*  $w_i$  such that

$$\int_0^1 d\tau \mathcal{G}(\tau) \simeq \sum_{i=1}^k w_i \mathcal{G}(\tau_i). \quad (12.13.5)$$

In our case

$$\mathcal{G}(\tau) = \exp(\tau s \# \mathcal{A} \#)(t\mathcal{B}). \quad (12.13.6)$$

Shortly we will consider how the  $\tau_i$  and  $w_i$  might be chosen. First, let us see how a quadrature formula can be used. With the aid of (11.5) and (11.6) we find the result

$$\begin{aligned} \exp[iex(s \# \mathcal{A} \#)(t\mathcal{B})] &= \exp\left[\int_0^1 d\tau \mathcal{G}(\tau)\right] \simeq \exp\left[\sum_{i=1}^k w_i \mathcal{G}(\tau_i)\right] \\ &= \exp[O(t^2)] \exp[w_1 \mathcal{G}(\tau_1)] \exp[w_2 \mathcal{G}(\tau_2)] \cdots \exp[w_k \mathcal{G}(\tau_k)]. \end{aligned} \quad (12.13.7)$$

Note that the operators  $\mathcal{G}(\tau_i)$  generally do not commute. Therefore the conversion of the exponential of a sum into a product of exponentials, which occurs as the last step in (11.7), produces correction terms that involve commutators. However, since the  $\mathcal{G}(\tau_i)$  are linear in  $t$ , these commutators are  $O(t^2)$  as indicated in (11.7).

For each factor in (11.7) we have the result

$$\begin{aligned} \exp[w_i \mathcal{G}(\tau_i)] &= \exp[\exp(\tau_i s \# \mathcal{A} \#)(w_i t \mathcal{B})] \\ &= \exp[\exp(\tau_i s \mathcal{A})(w_i t \mathcal{B}) \exp(-\tau_i s \mathcal{A})] \\ &= \exp(\tau_i s \mathcal{A}) \exp(w_i t \mathcal{B}) \exp(-\tau_i s \mathcal{A}). \end{aligned} \quad (12.13.8)$$

Putting all these results together gives the relation

$$\begin{aligned} \exp(s\mathcal{A} + t\mathcal{B}) &\simeq \exp[O(t^2)] \exp(\tau_1 s \mathcal{A}) \exp(w_1 t \mathcal{B}) \exp(-\tau_1 s \mathcal{A}) \times \\ &\quad \exp(\tau_2 s \mathcal{A}) \exp(w_2 t \mathcal{B}) \exp(-\tau_2 s \mathcal{A}) \cdots \times \\ &\quad \exp(\tau_k s \mathcal{A}) \exp(w_k t \mathcal{B}) \exp(-\tau_k s \mathcal{A}) \exp(s\mathcal{A}). \end{aligned} \quad (12.13.9)$$

Finally, carrying out the indicated multiplications gives the result

$$\begin{aligned} \exp(s\mathcal{A} + t\mathcal{B}) &\simeq \exp[O(t^2)] \exp(\tau_1 s \mathcal{A}) \exp(w_1 t \mathcal{B}) \\ &\quad \exp[(\tau_2 - \tau_1)s\mathcal{A}] \exp(w_2 t \mathcal{B}) \exp[(\tau_3 - \tau_2)s\mathcal{A}] \cdots \\ &\quad \exp(w_k t \mathcal{B}) \exp[(1 - \tau_k)s\mathcal{A}]. \end{aligned} \quad (12.13.10)$$

This is the desired Zassenhaus approximation.

We must still consider how to select the  $\tau_i$  and  $w_i$ . One possibility is to space the  $\tau_i$  evenly with  $\tau_1 = 0$  and  $\tau_k = 1$ ,

$$\tau_i = (i - 1)/(k - 1). \quad (12.13.11)$$

In this case we should use *Newton-Cotes* weights. See Appendix T. For example, for the case  $k = 3$  we have the celebrated *Simpson's rule* 1-4-1 formula

$$\int_0^1 d\tau \mathcal{G}(\tau) \simeq (1/6)\mathcal{G}(0) + (4/6)\mathcal{G}(1/2) + (1/6)\mathcal{G}(1). \quad (12.13.12)$$

Another appealing possibility is not to space the  $\tau_i$  evenly, but rather to select them (as well as the weights  $w_i$ ) in such a way that (for a fixed  $k$ ) the order is maximized. This choice

produces the family of *Legendre-Gauss* quadrature formulas. Again see Appendix T. For example, for  $k = 3$  there is the formula

$$\int_0^1 d\tau \mathcal{G}(\tau) \simeq (5/18)\mathcal{G}(1/2 - \sqrt{15}/10) + (8/18)\mathcal{G}(1/2) + (5/18)\mathcal{G}(1/2 + \sqrt{15}/10). \quad (12.13.13)$$

As another example, consider the case  $k = 2$ . Then there is the formula

$$\int_0^1 d\tau \mathcal{G}(\tau) \simeq (1/2)\mathcal{G}(1/2 - \sqrt{3}/6) + (1/2)\mathcal{G}(1/2 + \sqrt{3}/6). \quad (12.13.14)$$

Because (11.5) replaces an integral by a sum, the term in the Zassenhaus approximation (11.10) that is linear in  $t$  is no longer exact in  $s$ . (However, the term independent of  $t$  still is exact.) We can estimate the error made in  $s$  from formulas of the kind (11.14), (11.16), and (11.18). Taylor series expansion of  $G(\tau)$  as given by (11.6) provides the result

$$\sum_{i=1}^k w_i G(\tau_i) = \sum_{n=0}^{\infty} (s^n/n!) \# \mathcal{A} \#^n t \mathcal{B} \sum_{i=1}^k w_i \tau_i^n. \quad (12.13.15)$$

Let  $c_k$  be the error term in the relation

$$\sum_{i=1}^k w_i (\tau_i)^{\ell_{\max}+1} = 1/(\ell_{\max} + 2) + c_k. \quad (12.13.16)$$

Again see Appendix T, and the example relations (T.1.6), (T.1.12), and (T.1.15). From (12.15) and (12.16) we find the error estimate

$$\sum_{i=1}^k w_i G(\tau_i) = \text{iex}(s \# \mathcal{A} \#)(t \mathcal{B}) + [c_k/(\ell_{\max} + 1)!] s^{(\ell_{\max}+1)} (\# \mathcal{A} \#)^{(\ell_{\max}+1)} (t \mathcal{B}) + O[t(s)^{(\ell_{\max}+2)}]. \quad (12.13.17)$$

Consequently, to examine relative error, we must make the comparison

$$t \mathcal{B} \xrightarrow{?} [c_k/(\ell_{\max}+1)!] s^{(\ell_{\max}+1)} (\# \mathcal{A} \#)^{\ell_{\max}+1} (t \mathcal{B}). \quad (12.13.18)$$

For example, suppose that  $\mathcal{A}$  is the Lie operator :  $H_2$  : for a quadratic Hamiltonian, and  $(t \mathcal{B})$  is the Lie operator :  $H_r$  : for the remaining piece as in (10.4). Let  $JS$  be the matrix associated with  $H_2$ . In analogy to (7.56), define  $\lambda$  by the relation

$$\lambda = \|sJS\|. \quad (12.13.19)$$

Also, suppose that  $H_r$  does not contain terms beyond degree  $m$ . Then, in analogy to (6.5.2), we have the estimate

$$s^{\ell_{\max}+1} (\# \mathcal{A} \#)^{\ell_{\max}+1} (t \mathcal{B}) \sim (m\lambda)^{\ell_{\max}+1} (t \mathcal{B}). \quad (12.13.20)$$

Consequently, we conclude that the relative error in the Zassenhaus approximation (11.10) has the estimate

$$\text{relative error} \sim [c_k/(\ell_{\max} + 1)!] (m\lambda)^{(\ell_{\max}+1)}. \quad (12.13.21)$$

What uses can be made of Zassenhaus approximations of the form (11.10)? In the context of Accelerator Physics, we will see in Chapters 12 and 12 that  $H_r$  becomes small in the limit of high energies. Therefore, at least three possible uses come to mind.

First, in the autonomous case, these Zassenhaus approximations can be used as splitting formulas for map computation by scaling, splitting, and squaring. Second, they can be used as symplectic integrators. The autonomous case can be treated using the form (11.10), and the nonautonomous case can be treated using related formulas. See Exercise 11.\*. In the context of symplectic integrators, employing the Gaussian sampling points and weights seems particularly attractive because doing so minimizes (for a given  $k$ ) the number of operators  $\exp(w_i t \mathcal{B})$ , whose evaluation is relatively expensive.

Finally, these Zassenhaus approximations can be used as the basis of an accelerator lattice *correction* scheme. Suppose we find  $H_r$ , or some terms in  $H_r$ , to be offensive. Then we can counter the effect of these terms on the performance of an accelerator lattice by placing, at the sampling points, local correctors having these same offensive properties. These correctors should be powered with strengths proportional to  $(-w_i)$  in such a way that the *net* effect of the correctors and the offensive terms in  $H_r$  *cancel* to first order in  $t$  and high order in  $s$ . See Exercise \*.\*. In this case perhaps, although not necessarily, use of the Newton-Cotes sampling points might be more convenient from an engineering perspective.

## Exercises

### 12.13.1.

# Bibliography

## General References

- [1] E. Forest, *Beam Dynamics: A New Attitude and Framework*, Harwood Academic Press (1998).
- [2] E. Hairer, S. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations I. Non-stiff Problems*, Springer (1993).
- [3] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II. Stiff and Differential Algebraic Problems*, Springer (2002).

## Geometric/Symplectic Integration and Splitting Methods

- [4] R.D. Ruth, *IEEE Trans. Nucl. Sci.* **30**, 2669 (1983).
- [5] F. Neri, “Lie algebras and canonical integration”, Department of Physics Technical Report, University of Maryland, 1988 (unpublished). This paper initiated the use of Lie methods to find symplectic integrators. It was the first to derive the results (1.32) through (1.37) using the Lie methods presented in Section 1.
- [6] J.M. Sanz-Serna and M.P. Calvo, *Numerical Hamiltonian Problems*, Chapman and Hall (1994) or Dover (1994).
- [7] J.E. Marsden, G.W. Patrick, W.F. Shadwick, eds., *Integration Algorithms and Classical Mechanics*, Fields Institute Communications, American Mathematical Society (1996).
- [8] A.M. Stuart and A.R. Humphries, *Dynamical Systems and Numerical Analysis*, Cambridge University Press (1996).
- [9] H. Yoshida, “Recent Progress in the Theory and Application of Symplectic Integrators”, *Celestial Mechanics and Dynamical Astronomy* **56**, 27 (1993).
- [10] E. Forest and R.D. Ruth, “Fourth-Order Symplectic Integration”, *Physica D* **43**, 105 (1990).
- [11] E. Forest, “Sixth-Order Lie Group Integrators”, *J. Comp. Phys.* **99**, 209 (1992).
- [12] E. Forest, J. Bengtsson, and M.F. Reusch, “Application of the Yoshida-Ruth techniques to implicit integration and multi-map explicit integration”, *Phys. Lett. A* **158**, 99 (1991).

- [13] E. Forest, “Geometric Integration for Particle Accelerators”, *Journal of Physics A* **39**, 5321-5377 (2006).
- [14] P.-V. Koseleff, “Recherche d’approximants d’exponentielles”, preprint, École Polytechnique (1993).
- [15] P.J. Channell and C. Scovel, “Symplectic integration of Hamiltonian systems”, *Nonlinearity* **3**, 231 (1990).
- [16] M. Campostrini and P. Rossi, “A comparison of numerical algorithms for dynamical fermions”, *Nuclear Physics B* **329**, 753 (1990).
- [17] M. Suzuki, “On the convergence of exponential operators – the Zassenhaus formula, BCH formula and systematic approximants”, *Commun. Math. Phys.* **57**, 193-200 (1977).
- [18] M. Suzuki, “Symmetry and Coherent Approximations in Non-Equilibrium Systems”, preprint, University of Tokyo.
- [19] M. Suzuki, “General Theory of higher-order decomposition of exponential operators and symplectic integrators”, *Phys. Lett. A* **165**, 387 (1992).
- [20] M. Suzuki, “General Nonsymmetric Higher-Order Decomposition of Exponential Operators and Symplectic Integrators”, *J. Phys. Soc. of Japan* **61**, 3015 (1992).
- [21] M. Suzuki, “Fractal Decomposition of Exponential Operators with Applications to Many-body Theories and Monte Carlo Simulation”, *Phys. Lett. A* **146**, 319 (1990).
- [22] M. Suzuki, “General theory of fractal path integrals with applications to many-body theories and statistical physics”, *J. Math. Phys.* **32**, 400 (1991). <http://chaosbook.org/library/SuzukiJMP91.pdf>
- [23] M. Suzuki, “Symmetry and Systematics”, preprint, University of Japan.
- [24] M. Suzuki, “Improved Trotter-like Formula”, *Phys. Lett. A* **180**, 232 (1993).
- [25] M. Suzuki and K. Umeno, “Higher-Order Decomposition Theory of Exponential Operators and Its Applications to QMC and Nonlinear Dynamics”, *Computer Simulation Studies in Condensed Matter Physics VI*, D.P. Landau, K.K. Mon, and H.-B. Schüttler, eds., pp. 74-86, Springer-Verlag (1993).
- [26] M. Glasner, D. Yevick, and B. Hermansson, “Sixth Order generalized propagation techniques”, *Electron. Lett.*, vol. 27, pp. 475-478, 1991.
- [27] M. Glasner, D. Yevick, and B. Hermansson, “High Order generalized propagation techniques”, *J. Opt. Soc. B*, vol. 8, pp. 413-415, 1991.
- [28] M. Glasner, D. Yevick, and B. Hermansson, “Generalized propagation formulas of arbitrarily high order”, *J. Chem. Phys.*, vol. 95, pp. 8266-8272, 1991.

- [29] M. Glasner, D. Yevick, and B. Hermansson, “Computer generated generalized propagation techniques”, *Appl. Math. Lett.*, vol. 4, pp. 85-90, 1991.
- [30] M. Glasner, D. Yevick, and B. Hermansson, “Generalized propagation techniques for longitudinally varying refractive index distributions”, *Math. and Comp. Modelling*, vol. 16, pp. 179-184, 1992.
- [31] P. Saha and S. Tremaine, “Symplectic integrators for solar system dynamics”, *Astron. J.* **104**, 1633 (1992).
- [32] M. Austin, P.S. Krishnaprasad, and L.-S. Wang, “On Symplectic and Almost Poisson Integration of Rigid Body Systems”, University of Maryland Systems Research Center, technical report TR91-45 (1991).
- [33] Qin Meng-Zhao and Zhu Wen-Jie, “Construction of Higher Order Schemes for Ordinary Differential Equations by Composing Self-adjoint Lower Order Ones”, preprint, Computing Center, Academia Sinica, Beijing.
- [34] Qin Meng-Zhao, “A Difference Scheme for the Hamiltonian Equation”, *J. Comp. Math.* **5**, 203 (1987).
- [35] Feng Kang and Qin Meng-Zhao, “The Symplectic Methods for the Computation of Hamiltonian Equations”, in *Numerical Methods for partial differential equations*, Lect. Notes Math **1297**, 1, Springer (1987).
- [36] Feng Kang and Qin Meng-Zhao, “Hamiltonian algorithms for Hamiltonian systems and a comparative numerical study”, *Computer Physics Communications* **65**, 173-187 (1991).
- [37] H. De Raedt, *Product Formula Algorithms for Solving the Time Dependent Schrödinger Equation*, Computer Physics Reports 7, (North Holland, Amsterdam 1987).
- [38] H. De Raedt and B. De Raedt, *Phys. Rev. A* **28**, 3575 (1983).
- [39] H. De Raedt, “Quantum Dynamics in Nanoscale Devices”, in *Computational Physics*, K.H. Hoffmann and M. Schreiber, Eds., Springer (1996).
- [40] J. Candy and W. Rozmus, “A Symplectic Integration Algorithm for Separable Hamiltonian Functions”. *J. Comput. Phys.* **92**, 230 (1991).
- [41] S. Mikkola and P. Wiegert, “Regularizing time transformations in symplectic and composite integration”, *Celest. Mech. Astron.* **82**, 375 (2002).
- [42] J. Wisdom and M. Holman, “Symplectic maps for the n-body problem: Stability analysis”, *Astron. J.* **104**, 2022 (1992).
- [43] R.I. McLachlan, “On the numerical integration of ordinary differential equations by symmetric composition methods”, *SIAM J. Sci. Comp.* **16**, 151-168, (1995). Also available on the Web at <http://www.massey.ac.nz/~rmclachl/sisc95.pdf>.

- [44] R.I. McLachlan, “Composition methods in the presence of small parameters”, *BIT* **35**, 258-268, (1995).
- [45] R. I. McLachlan and P. Atela, “The accuracy of symplectic integrators”, *Nonlinearity* **5**, 541-562 (1992).
- [46] R.I. McLachlan and G.R.W. Quispel, “Splitting methods”, *Acta Numerica* **11**, 341-434 (2002).
- [47] R.I. McLachlan and G.R.W. Quispel, “Explicit geometric integration of polynomial vector fields”, *BIT* **44**, 515-538, (2004).
- [48] R.I. McLachlan and G.R.W. Quispel, “Geometric integrators for ODEs”, *Journal of Physics A* **39**, 5251-5285, (2006).
- [49] G.R.W. Quispel and R.I. McLachlan, Eds., Special Issue on Geometric Numerical Integration of Differential Equations, *Journal of Physics A* **39** (2006).
- [50] S. Blanes, F. Casas, and J. Ros, “Symplectic Integration with Processing: A General Study”, *SIAM J. Sci. Comput.* (1998).
- [51] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure Preserving Algorithms for Ordinary Differential Equations*, Springer (2002), corrected second printing 2004.
- [52] B. Leimkuhler and S. Reich, *Simulating Hamiltonian Dynamics*, Cambridge University Press (2004).
- [53] S. Reich, “Backward error analysis for numerical integrators”, *SIAM J. Numer. Anal.* **36**, 1549-1570 (1999).
- [54] A. Iserles, H. Munthe-Kaas, S. Nørsett, and A. Zanna, “Lie-group methods”, *Acta Numerica* **14**, 1–148, (2005). Also available on the Web at [http://www.damtp.cam.ac.uk/user/na/NA\\_papers/NA2000\\_03.pdf](http://www.damtp.cam.ac.uk/user/na/NA_papers/NA2000_03.pdf).
- [55] I.P. Omelyan, I.M. Mryglod, and R. Folk, “Symplectic Analytically Integrable Decomposition Algorithms: Classification, Derivation, and Application to Molecular Dynamics, Quantum and Celestial Mechanics Simulations”, *Computer Physics Communications*, vol. 151, 273-314 (2003).
- [56] S.-H. Tsai, H.K. Lee, and D.P. Landau, “Molecular and Spin Dynamics Simulations using Modern Integration Methods”, *Am. J. Physics*, vol. 73, 615-624 (2005).
- [57] L. Gauckler, E. Hairer, and C. Lubich, “Dynamics, Numerical Analysis, and some Geometry”, preprint (10 October 2017) <https://arxiv.org/abs/1710.03946>
- [58] S. Blanes and F. Casas, *A Concise Introduction to Geometric Numerical Integration*, CRC Press (2016).

- [59] M. Creutz and A. Gocksch, “Higher-order Hybrid Monte Carlo Algorithms”, *Phys. Rev. Lett.* **63**, 9 (1989).
- [60] H. Yoshida, “Construction of higher order symplectic integrators”, *Phys. Lett. A* **150**, 262 (1990).
- [61] E. Forest, J. Bengtsson, and M. Reusch, “Application of the Yoshida-Ruth techniques to implicit integration and multi-map explicit integration”, *Phys. Lett. A* **158**, 99 (1991).
- Explicit Symplectic Integrator for Polynomial Hamiltonians
- [62] P. Channell and F. Neri, “An Introduction to Symplectic Integrators”, *Integration Algorithms and Classical Mechanics*, J.E. Marsden, G.W. Patrick, W.F. Shadwick, eds., Fields Institute Communications, American Mathematical Society (1996).
- Force Gradient Algorithms
- [63] S.A. Chin and D.W. Kidwell, “Higher-order Force Gradient Symplectic Algorithms”, *Phys. Rev. E* **62**, 8746 (2000).
- [64] S.A. Chin and C.R. Chen, “Forward Symplectic Integrators for Solving Gravitational Few-Body Problems”, *Celestial Mechanics and Dynamical Astronomy* **91**, 301-322 (2005).
- Symplectic Runge-Kutta, Partitioned Runge-Kutta, and Nyström Runge-Kutta
- [65] B. Blanes and P.C. Moan, “Practical symplectic partitioned Runge-Kutta and Runge-Kutta-Nyström methods”, *Journal of Computational and Applied Mathematics* **142**, 313-330 (2002).
- [66] A. Iserles, “Efficient Runge-Kutta Methods for Hamiltonian Equations”, in *Advances on Computer Mathematics and its Applications*, E. Lipitakis, Ed., World Scientific (1993).
- [67] M. Sofroniou and W. Oevel, “Symplectic Runge-Kutta Schemes I: Order Conditions”, *SIAM Journal of Numerical Analysis* **34(5)** (1997).
- [68] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure Preserving Algorithms for Ordinary Differential Equations*, Springer (2002), corrected second printing 2004.
- [69] X. Tan, “Almost symplectic Runge-Kutta schemes for Hamiltonian systems”, *Journal of Computational Physics* **203**, 250-273 (2005).
- [70] R. McLachlan, “A New Implementation of Symplectic Runge-Kutta Methods”, *SIAM Journal on Scientific Computing*, Volume 29, Issue 4, 1637-1649 (2007).
- [71] J.C. Butcher, *Numerical Methods for Ordinary Differential Equations*, Second Edition, John Wiley (2008). <http://www.math.auckland.ac.nz/~butcher/ODE-book-2008/>.

[72] D. Okunbor and R.D. Skeel, “Explicit Canonical Methods for Hamiltonian Systems”, *Mathematics of Computation* **59**, 439-455 (1992).

[73] Lin-Yi Chou and P. W. Sharp, “On order 5 symplectic explicit Runge-Kutta Nyström methods”, *Journal of Applied Mathematics & Decision Sciences* **4**, 143-150 (2000).

#### Symplectic Integration Using Generating Functions

[74] Feng Kang and Qin Meng-Zhao, “The Symplectic Methods for the Computation of Hamiltonian Equations”, in *Numerical Methods for partial differential equations*, Lect. Notes Math **1297**, 1, Springer (1987).

[75] 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).

[76] Feng Kang, “The Calculus of Generating Functions and the Formal Energy for Hamiltonian Algorithms”, *Journal of Computational Mathematics* **16**, p. 481 (1998).

[77] Feng Kang and Mengzhao Qin, *Symplectic Geometric Algorithms for Hamiltonian Systems*, Zhejiang Publishing and Springer-Verlag (2010).

[78] P. J. Channell and C. Scovel, “Symplectic integration of Hamiltonian systems”, *Nonlinearity* **3**, 231-259 (1990).

#### Symplectic Integrator for Motion in a Magnetic Field

[79] Y.K. Wu, E. Forest, and D.S. Robin, “Explicit symplectic integrator for s-dependent static magnetic field”, *Phys. Rev. E* **68**, 046502, (2003).

[80] M. Aichinger, S.A. Chin, and E. Krotscheck, “Fourth-order Algorithms for Solving Local Schroedinger Equations in a Strong Magnetic Field”, *Comp. Phys. Comm.* **171**, 197 (2005).

[81] E. Chacon-Golcher and F. Neri, “A symplectic integrator with arbitrary vector and scalar potentials”, *Physics Letters A* **372**, 4661-4666, (2008).

[82] M. Tao, “Explicit high-order symplectic integrators for charged particles in general electromagnetic fields”, *Journal of Computational Physics* **327**, 245 (2016).

[83] M. Tao, “Explicit symplectic approximation of nonseparable Hamiltonians: Algorithm and long time performance”, *Physical Review E* **94**, 043303 (2016).

#### Conservation of $H$

[84] Z. Ge and J. Marsden, “Lie-Poisson Hamilton-Jacobi theory and Lie-Poisson integrators”, *Phys. Lett. A*, **133**, 134-139, (1988).

[85] G. Benettin and A. Giorgilli, “On the Hamiltonian interpolation of near to the identity symplectic mappings with application to symplectic integration algorithms”, *J. Statist. Phys.* **74**, 1117 (1994).

- [86] S. Reich, “Backward error analysis for numerical integrators”, *SIAM J. Numer. Anal.* **36**, 1549-1570 (1999).

Local Correction

- [87] D. Neuffer and E. Forest, *Phys. Lett. A* **135**, 197, (1989).

- [88] D. Neuffer, *Nucl. Instr. and Meth. A* **274**, 400, (1989).

- [89] D. Neuffer, Proceedings of the Workshop on Effects of Errors in Accelerators, Corpus Christi, TX, 1991, A. Chao, Ed., AIP Conf. Proc. **255**, 215, (1992).

- [90] E. Forest, *Beam Dynamics*, Section 11.4, Harwood Academic (1998).



# Chapter 13

## Transfer Maps for Idealized Straight Beam-Line Elements

### 13.1 Background

In this chapter we will describe the computation of transfer maps for idealized straight beam-line elements. Here we make two assumptions: First, the element geometry is such that Cartesian coordinates can conveniently be employed. Second, the design orbit is a straight line, which we take to be the  $z$  axis. For simplicity we will treat only magnetic elements, but the case of straight electric elements or straight electromagnetic elements can be handled similarly. Since the design orbit is assumed to be a straight line, the case of magnetic dipole fields is excluded.

We further assume that the design orbit is traversed in time with *constant* velocity  $v_z^0$ . In particular, to ensure that the velocity  $v_z^0$  is indeed constant we also assume that the electric scalar potential vanishes,  $\psi = 0$ . For the same reason, the vector potential  $\mathbf{A}$  is taken to be *time independent*. See (1.5.2). More complicated situations where  $v_z^0$  is not constant, such as occurs in RF accelerating cavities, can be treated in an analogous way.

Since the design orbit is taken to lie along the  $z$  axis, it is convenient to take  $z$  as the independent variable. In this case, according to (1.6.16), the Hamiltonian  $K$  is given by the relation

$$K = -[(p_t^{\text{can}})^2/c^2 - m^2c^2 - (p_x^{\text{can}} - qA_x)^2 - (p_y^{\text{can}} - qA_y)^2]^{1/2} - qA_z. \quad (13.1.1)$$

Here we have taken care to indicate that the momenta employed are *canonical*. The idealizations we will make in this chapter are that the vector potential  $\mathbf{A}$  is  $z$  *independent* and fringe-field effects can be neglected. These restrictions will be removed in the subsequent Chapters 15 through 21.

#### 13.1.1 Specification of Design Orbit

Under the assumptions made about the *design* orbit, we may write the relations

$$x^d = y^d = 0. \quad (13.1.2)$$

Also, the transverse velocities and therefore the transverse mechanical momenta will vanish on the design orbit. We further assume that the fields and selected gauge for the vector potential are such that the transverse components of the vector potential vanish on the design orbit. It follows that in this case the transverse canonical momenta will also vanish on the design orbit,

$$p_x^{\text{can d}} = p_y^{\text{can d}} = 0. \quad (13.1.3)$$

With regard to the momentum  $p_t^{\text{can}}$ , since we have assumed that  $\psi = 0$  and  $v_z^0$  is constant, we may write

$$p_t^{\text{can d}} = p_t^0 \quad (13.1.4)$$

where  $p_t^0$  is the value of  $p_t^{\text{can}}$  on the design orbit.

To complete our description of the design orbit, we need to find the time  $t$  as a function of  $z$  on this orbit. We may write

$$(dt/dz)|_{\text{design orbit}} = 1/[(dz/dt)|_{\text{design orbit}}] = 1/v_z^0. \quad (13.1.5)$$

Here, as described earlier,  $v_z^0$  is the design velocity on the design orbit, from which it follows that there is the relation

$$v_z^0 = (c^2 \gamma m v_z^0) / (\gamma m c^2) = -c^2 p_t^0 / p_t^0 \quad (13.1.6)$$

where  $\gamma$  denotes the usual relativistic factor and  $p_t^0$  is the value of  $p_z^{\text{mech}}$  on the design orbit. [This same result can be obtained from (1.1) by observing that  $p_x^{\text{can}}$ ,  $p_y^{\text{can}}$ ,  $A_x$ , and  $A_y$  all vanish on the design orbit. See Exercise \*.] Note also that there are the relations

$$p_t^0 = -[m^2 c^4 + (p_t^0)^2 c^2]^{1/2}, \quad (13.1.7)$$

$$p_t^0 = [(p_t^0/c)^2 - m^2 c^2]^{1/2}. \quad (13.1.8)$$

Since  $v_z^0$  is constant, integration of (1.5) yields the relation

$$t^d(z) = z/v_z^0 \quad (13.1.9)$$

where we have taken the origin in time to be such that the design orbit passes through  $z = 0$  at the time  $t = 0$ .

### 13.1.2 Deviation Variables

Under the assumptions just made about the design orbit, we may introduce transverse coordinate deviation variables  $\xi, \eta$  by the definitions

$$\xi = x - x^d = x, \quad (13.1.10)$$

$$\eta = y - y^d = y. \quad (13.1.11)$$

We also introduce transverse momentum deviation variables  $p_\xi, p_\eta$  by the definitions

$$p_\xi = p_x^{\text{can}} - p_x^{\text{can d}} = p_x^{\text{can}}, \quad (13.1.12)$$

$$p_\eta = p_y^{\text{can}} - p_y^{\text{can d}} = p_y^{\text{can}}. \quad (13.1.13)$$

With the results and definitions (1.2) through (1.4) and (1.9) through (1.13) in hand, we are now able to introduce a full set of deviation variables  $(\xi, \eta, T; p_\xi, p_\eta, p_T)$  by adding to the definitions (1.10) through (1.13) the definitions

$$T = t - t^d = t - z/v_z^0, \quad (13.1.14)$$

$$p_T = p_t^{\text{can}} - p_t^{\text{can d}} = p_t^{\text{can}} - p_t^0. \quad (13.1.15)$$

Note that by construction the deviation variables all vanish on the design orbit as desired. Since the relations (1.9) through (1.14) simply amount to a phase-space translation, it follows that the relation between the original variables and the deviation variables is a canonical transformation.

### 13.1.3 Deviation Variable Hamiltonian

Since the transformation given by (1.9) through (1.14) is canonical, the equations of motion for the deviation variables must also arise from a Hamiltonian, which we will call  $K^{\text{new}}(\xi, \eta, T, p_\xi, p_\eta, p_T; z)$ . Our task is to find  $K^{\text{new}}$  in terms of  $K$ . To do so, we will use the machinery of Subsection 10.4.1.

Following (10.4.2), define a function  $\bar{K}(\xi, \eta, T, p_\xi, p_\eta, p_T; z)$  by the rule

$$\bar{K}(\xi, \eta, T, p_\xi, p_\eta, p_T; z) = K(\xi, \eta, z/v_z^0 + T, p_\xi, p_\eta, p_t^0 + p_T; z). \quad (13.1.16)$$

Then, according to (10.4.20), the new Hamiltonian is given by the rule

$$\begin{aligned} K^{\text{new}}(\xi, \eta, T, p_\xi, p_\eta, p_T; z) &= \\ \bar{K}(\xi, \eta, T, p_\xi, p_\eta, p_T; z) - \bar{K}_1(\xi, \eta, T, p_\xi, p_\eta, p_T; z). \end{aligned} \quad (13.1.17)$$

Let us work out the implications of this rule for the case where  $K$  is given by (1.1). For  $\bar{K}$  we find the result

$$\begin{aligned} \bar{K}(\xi, \eta, T, p_\xi, p_\eta, p_T; z) &= \\ -[(p_t^0 + p_T)^2/c^2 - m^2 c^2 - (p_\xi - qA_x)^2 - (p_\eta - qA_y)^2]^{1/2} - qA_z. \end{aligned} \quad (13.1.18)$$

We next assume that  $\mathbf{A}$  is time independent, the expansions of  $A_x$  and  $A_y$  begin with linear terms, and the expansion of  $A_z$  begins with quadratic terms. In that case  $\mathbf{A}$  does not contribute to  $\bar{K}_1(\xi, \eta, T, p_\xi, p_\eta, p_T; z)$ , and we find from (1.17) the result

$$\begin{aligned} \bar{K}_1(\xi, \eta, T, p_\xi, p_\eta, p_T; z) &= -(p_t^0/c^2)[(p_t^0/c)^2 - m^2 c^2]^{-1/2} p_T \\ &= -(p_t^0/c^2)(1/p^0)p_T \\ &= p_T/v_z^0. \end{aligned} \quad (13.1.19)$$

Here we have used (1.6) and (1.8). Finally, combining (1.17) through (1.19) gives the result

$$\begin{aligned} K^{\text{new}}(\xi, \eta, T, p_\xi, p_\eta, p_T; z) &= \\ -[(p_t^0 + p_T)^2/c^2 - m^2 c^2 - (p_\xi - qA_x)^2 - (p_\eta - qA_y)^2]^{1/2} - qA_z - p_T/v_z^0. \end{aligned} \quad (13.1.20)$$

### 13.1.4 Dimensionless Scaled Deviation Variables

For many purpose it is useful to describe trajectories and maps in terms of dimensionless variables. Let  $\ell$  be some convenient scale length.<sup>1</sup> Introduce dimensionless variables  $(X, Y, \tau; P_x, P_y, P_\tau)$ , defined in terms of the deviation variables and the scale length, by the rules

$$X = \xi/\ell, \quad (13.1.21)$$

$$Y = \eta/\ell, \quad (13.1.22)$$

$$\tau = cT/\ell; \quad (13.1.23)$$

$$P_x = p_\xi/p^0, \quad (13.1.24)$$

$$P_y = p_\eta/p^0, \quad (13.1.25)$$

$$P_\tau = p_T/(p^0 c). \quad (13.1.26)$$

At this point it is useful to relate  $P_\tau$ , which may be viewed as a scaled energy deviation (with a minus sign), to the momentum deviation parameter  $\delta$  of Exercise 1.7.6. They are connected by the relations

$$\begin{aligned} P_\tau &= -(1/\beta_0)\{[1 + (2\delta + \delta^2)\beta_0^2]^{1/2} - 1\} \\ &= -\beta_0\delta + (\delta^2/2)(\beta_0^3 - \beta_0) - (\delta^3/2)(\beta_0^5 - \beta_0^3) + \dots, \end{aligned} \quad (13.1.27)$$

$$\begin{aligned} \delta &= (1 - 2P_\tau/\beta_0 + P_\tau^2)^{1/2} - 1 \\ &= -P_\tau/\beta_0 + (P_\tau^2/2)(1 - \beta_0^{-2}) + \dots \end{aligned} \quad (13.1.28)$$

Here  $\beta_0$  is the usual relativistic factor evaluated on the design orbit,

$$\beta_0 = v_z^0/c = -cp^0/p_t^0. \quad (13.1.29)$$

Note that in the ultra relativistic limit  $\beta_0 \rightarrow 1$  there are the relations

$$P_\tau = -\delta, \quad (13.1.30)$$

$$\delta = -P_\tau. \quad (13.1.31)$$

See Exercise \*.

### 13.1.5 Scaled Deviation-Variable Hamiltonian

We will now seek equations of motion for the scaled deviation variables. We will learn that they also can be derived from what we will call *scaled* deviation-variable Hamiltonian and will denote by the symbol  $H^s$ . To do so requires some care.

Although the Poisson brackets of the new coordinates with each other and the Poisson brackets of the new momenta with each other all vanish, the transformation given by (2.54)

---

<sup>1</sup>Here the coordinate scale factor  $\ell$  is not to be confused with the “path length”  $\ell$  of Exercise 1.7.6. Note also that in this subsection we use the notation  $p^0$  to denote the quantity  $p_0^{\text{mech}}$  of Exercise 1.7.6.

through (2.59) is not canonical because the Poisson brackets of the new coordinates with their corresponding new momenta do not have the value 1. Instead they have the common value

$$[X, P_x] = [Y, P_y] = [\tau, P_\tau] = (p^0 \ell)^{-1}. \quad (13.1.32)$$

Nevertheless, it is still possible to obtain the equations of motion for the deviation variables from a Hamiltonian providing the Hamiltonian, which we now denote by  $H$ , is taken to be the function

$$H^s = K^{\text{new}} / (p^0 \ell), \quad (13.1.33)$$

and we treat the deviation variables as being canonically conjugate. That is, the Poisson brackets of the new coordinates with each other and the Poisson brackets of the new momenta with each other all vanish, and the Poisson brackets of the new coordinates with their corresponding new momenta are taken to have the value 1,

$$[X, P_x] = [Y, P_y] = [\tau, P_\tau] = 1. \quad (13.1.34)$$

See Exercise \*.

Let us apply the Ansatz (2.66) to the Hamiltonian  $\bar{H}$  given by (2.53) and (1.1). So doing gives the preliminary result

$$\begin{aligned} H(X, Y, \tau, P_x, P_y, P_\tau; z) &= [1/(p^0 \ell)][K - (p_T + p_t^0)/v_z^0] \\ &= [1/(p^0 \ell)]K - [1/(p^0 \ell)](p_T + p_t^0)/v_z^0. \end{aligned} \quad (13.1.35)$$

We will work separately on each of the two terms appearing on the far right side of (2.68).

Save for the  $1/\ell$  factor the first term takes the form

$$(1/p^0)K = -\{[(p_t^{\text{can}}/p^0 c)^2 - (mc/p^0)^2 - (p_x^{\text{can}}/p^0 - A_x^s)^2 - (p_y^{\text{can}}/p^0 - A_y^s)^2]^{1/2} + A_z^s\} \quad (13.1.36)$$

where  $\mathbf{A}^s$  is a *scaled* vector potential given by

$$\mathbf{A}^s(X, Y, z) = (q/p^0)\mathbf{A}(\ell X, \ell Y, z). \quad (13.1.37)$$

Further manipulation employing (2.45) and (2.59) produces the relation

$$\begin{aligned} [p_t^{\text{can}}/(p^0 c)]^2 - (mc/p^0)^2 &= [(p_T + p_t^0)/(p^0 c)]^2 - (mc/p^0)^2 \\ &= [(p^0 c P_\tau + p_t^0)/(p^0 c)]^2 - (mc/p^0)^2 \\ &= [P_\tau + p_t^0/(p^0 c)]^2 - (mc/p^0)^2. \end{aligned} \quad (13.1.38)$$

From (2.62) there is the relation

$$p_t^0/(p^0 c) = -1/\beta_0. \quad (13.1.39)$$

There is also the relation

$$mc/p^0 = mc/(m\gamma_0\beta_0 c) = 1/(\gamma_0\beta_0) \quad (13.1.40)$$

where  $\gamma_0$  is the usual relativistic factor evaluated on the design orbit. It follows that there is the relation

$$\begin{aligned} [p_t^{\text{can}}/(p^0 c)]^2 - (mc/p^0)^2 &= (P_\tau - 1/\beta_0)^2 - 1/(\gamma_0 \beta_0)^2 \\ &= P_\tau^2 - (2P_\tau/\beta_0) + 1/\beta_0^2 - 1/(\gamma_0 \beta_0)^2. \end{aligned} \quad (13.1.41)$$

But, there is the identity

$$1/\beta_0^2 - 1/(\gamma_0 \beta_0)^2 = (1/\beta_0)^2[1 - 1/\gamma_0^2] = (1/\beta_0)^2[1 - (1 - \beta_0^2)] = 1. \quad (13.1.42)$$

Consequently there is the net result

$$[p_t^{\text{can}}/(p^0 c)]^2 - (mc/p^0)^2 = P_\tau^2 - (2P_\tau/\beta_0) + 1. \quad (13.1.43)$$

We conclude that

$$[1/(p_0 \ell)]K = -(1/\ell)\{[1 - (2P_\tau/\beta_0) + P_\tau^2 - (P_x - A_x^s)^2 - (P_y - A_y^s)^2]^{1/2} + A_z^s\} \quad (13.1.44)$$

where we have also used (2.38), (2.39), (2.57), and (2.58).

What remains is to work on the second term on the right side of (2.68). Save for a  $(-1/\ell)$  factor it takes the form

$$\begin{aligned} (1/p^0)(p_T + p_t^0)/v_z^0 &= (1/p^0)(p_0 c P_\tau + p_t^0)[-p_t^0/(p_0 c^2)] \\ &= -[p_t^0/(p^0 c)]P_\tau - (p_t^0)^2/(p^0 c)^2 \\ &= (P_\tau/\beta_0) - (1/\beta_0^2). \end{aligned} \quad (13.1.45)$$

Here we have used (2.59) and (2.72). We conclude that

$$-[1/(p_0 \ell)](p_T + p_t^0)/v_z^0 = (1/\ell)[(P_\tau/\beta_0) - (1/\beta_0^2)]. \quad (13.1.46)$$

We are, at last, ready to compute  $H$ . Upon combining (2.68), (2.77), and (2.79) we find the final result

$$H = -(1/\ell)\{[1 - (2P_\tau/\beta_0) + P_\tau^2 - (P_x - A_x^s)^2 - (P_y - A_y^s)^2]^{1/2} + A_z^s + (P_\tau/\beta_0)\}. \quad (13.1.47)$$

In the case of no magnetic field (1.45) takes the form

$$H = -(1/\ell)\{[1 - (2P_\tau/\beta_0) + P_\tau^2 - P_x^2 - P_y^2]^{1/2} + (P_\tau/\beta_0)\}. \quad (13.1.48)$$

Let us use this Hamiltonian to compute  $x'$ . From Hamilton's equations of motion we find the result

$$\begin{aligned} x' &= dx/dz = \ell dX/dz = \ell \partial H / \partial P_x \\ &= P_x[1 - (2P_\tau/\beta_0) + P_\tau^2 - P_x^2 - P_y^2]^{-1/2} \\ &= P_x + P_x P_\tau / \beta_0 + (1/2)P_x[P_\tau^2(3\beta_0^{-2} - 1) + P_x^2 + P_y^2] + \dots \end{aligned} \quad (13.1.49)$$

We see that  $x'$  agrees with  $P_x$  in lowest order; but there are second-order chromatic differences, and third- and higher-order geometric and chromatic differences. Also,  $X$  and  $x'$  are *not* canonically conjugate,  $[X, x'] \neq 1$ .

## Exercises

### 13.2 Axial Rotation

### 13.3 Drift

In this subsection we will compute the transfer map for a drift. To do so we begin with the Hamiltonian (1.1) and employ the vector potential given by (2.7) through (2.10). We then introduce deviation variables followed by scaled deviation variables. Next we find the scaled deviation-variable Hamiltonian. Finally, we expand the scaled deviation-variable Hamiltonian in a Taylor series, and employ this Taylor series to compute the transfer map.

### 13.4 Solenoid

$$R = *. \quad (13.4.1)$$

### 13.5 Wiggler/Undulator

### 13.6 Quadrupole

### 13.7 Sextupole

### 13.8 Octupole

### 13.9 Higher-Order Multipoles

### 13.10 Thin Lens Multipoles

### 13.11 Combined Function Quadrupole

### 13.12 Radio Frequency Cavity



# Bibliography



# Chapter 14

## Transfer Maps for Idealized Curved Beam-Line Elements

### 14.1 Background

### 14.2 Sector Bend

### 14.3 Parallel Faced (Rectangular) Bend

#### 14.3.1 Preliminaries

The geometry of a parallel faced (rectangular) bending/dipole magnet is illustrated in Figures 1.6.1 and 1.6.2. Using global Cartesian coordinates, the entry and exit faces of the magnet are described by equations of the form  $z = \text{constant}$ . Since we want to know when a particle is inside or outside such a magnet, it is convenient to take  $z$  to be the independent variable. In this case the Hamiltonian for charged particle motion takes the form

$$K = -[(p_t + q\psi)^2/c^2 - m^2c^2 - (p_x - qA_x)^2 - (p_y - qA_y)^2]^{1/2} - qA_z. \quad (14.3.1)$$

See Exercise 1.6.1.

We must next specify the scalar and vector potentials. Since we assume there is no electric field, we require that

$$\psi = 0 \text{ and } \partial \mathbf{A} / \partial t = 0. \quad (14.3.2)$$

See (1.5.2). What remains is to specify  $\mathbf{A}(\mathbf{r})$ . (Note that  $\mathbf{A}$  must have the property that the  $\mathbf{B}$  it produces must be divergence and curl free both within and outside the magnet.) To continue we assume, as an idealization, that the magnet is *infinitely* wide in the  $x$  direction. Again see Figure 1.6.2. This idealization is explored in Exercise 15.4.4. There it is shown that, without loss of generality and by suitable gauge transformations,  $\mathbf{A}$  may be taken to be of the form

$$A_x = \sum_{n=1}^{\infty} (-1)^n [1/(2n)!] y^{2n} O^{[2n-1]}(z) = -(1/2)y^2 O^{[1]}(z) + (1/24)y^4 O^{[3]}(z) + \dots, \quad (14.3.3)$$

$$A_y = 0, \quad (14.3.4)$$

$$A_z = -xO^{[0]}(z). \quad (14.3.5)$$

Correspondingly, the magnetic field is given by the relations

$$B_x = 0, \quad (14.3.6)$$

$$\begin{aligned} B_y &= \sum_{n=0}^{\infty} (-1)^n [1/(2n)!] y^{2n} O^{[2n]}(z) \\ &= O^{[0]}(z) - (1/2)y^2 O^{[2]}(z) + (1/24)y^4 O^{[4]}(z) + \dots, \end{aligned} \quad (14.3.7)$$

$$\begin{aligned} B_z &= \sum_{n=0}^{\infty} (-1)^n [1/(2n+1)!] y^{2n+1} O^{[2n+1]}(z) \\ &= yO^{[1]}(z) - (1/6)y^3 O^{[3]}(z) + (1/120)y^5 O^{[5]}(z) + \dots. \end{aligned} \quad (14.3.8)$$

Here  $O^{[0]}(z)$  is in principle an arbitrary profile function and we have employed the notation

$$O^{[n+1]}(z) = (d/dz)O^{[n]}(z), \text{ etc.} \quad (14.3.9)$$

For our application  $O^{[0]}(z)$  will have the general “bump function” shape depicted in Figure 16.2.5. From (3.6) through (3.8) we see that  $O^{[0]}(z)$  is the  $B_y$  component of the magnetic field in the mid plane  $y = 0$ . If we assume this field component is nearly constant within the body of the magnet, then there are the relations

$$O^{[0]}(z) \simeq \text{constant} \text{ and } O^{[>0]}(z) \simeq 0 \text{ within the magnet.} \quad (14.3.10)$$

Thus,  $\mathbf{B}$  is primarily in the  $y$  direction, has no  $x$  component, but does have a  $z$  component in the fringe-field regions where  $O^{[n]}(z) \neq 0$  for  $n > 0$ .

Let us now assume that we are only interested in finding the part of the trajectory that lies *within* the magnet, and we further assume the magnetic field is constant within the body of the magnet. That is, we make the assumption

$$O^{[0]}(z) = \text{constant} = O \text{ within the magnet.} \quad (14.3.11)$$

Then the vector potential and associated magnetic field are given (within the magnet) by the relations

$$A_x = 0, \quad (14.3.12)$$

$$A_y = 0, \quad (14.3.13)$$

$$A_z = -xO. \quad (14.3.14)$$

Correspondingly, the magnetic field (within the magnet) is given by the relations

$$B_x = 0, \quad (14.3.15)$$

$$B_y = O, \quad (14.3.16)$$

$$B_z = 0. \quad (14.3.17)$$

Note if we assume the particle charge is positive ( $q > 0$ ), that  $\dot{z} > 0$ , and that the trajectory curves as shown in Figure 1.6.2, then according to the Lorentz force law  $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$  the magnetic field must point in the  $y$  direction. That is, we must have  $O > 0$  in (3.16). Therefore, according to (3.15) through (3.17), we may write

$$O = B \quad (14.3.18)$$

where  $B$  is the (positive and constant) *magnitude* of the magnetic field within the magnet. Upon combining (3.1), (3.2), (3.12) through (3.14), and (3.18), we see that, under the assumptions made, the Hamiltonian  $K$  becomes

$$K = -[(p_t)^2/c^2 - m^2c^2 - (p_x)^2 - (p_y)^2]^{1/2} + qBx. \quad (14.3.19)$$

### 14.3.2 Determination of Trajectories

Our task is to find the trajectories generated by the Hamiltonian  $K$  as given by (3.19). Let  $z^{\text{in}}$  be the (*initial*) value of  $z$  on the entry face of the magnet. Given the initial conditions

$$\{x^{\text{in}}, p_x^{\text{in}}, y^{\text{in}}, p_y^{\text{in}}, t^{\text{in}}, p_t^{\text{in}}\}, \quad (14.3.20)$$

we seek to find (within the magnet) the functions

$$\{x(z), p_x(z), y(z), p_y(z), t(z), p_t(z)\}. \quad (14.3.21)$$

As a result of employing this notation we have the relations

$$x(z^{\text{in}}) = x^{\text{in}}, \text{ etc.} \quad (14.3.22)$$

Since the trajectory of a charged particle moving in a uniform magnetic field is a helix, we expect our task to be a soluble, although perhaps algebraically messy, problem.

Evidently the coordinates  $y$  and  $t$  are *ignorable*,

$$\partial K / \partial y = 0 \text{ and } \partial K / \partial t = 0. \quad (14.3.23)$$

It follows that  $p_y$  and  $p_t$  are integrals of motion,

$$p_y(z) = \text{constant} = p_y^{\text{in}} \text{ and } p_t(z) = \text{constant} = p_t^{\text{in}}. \quad (14.3.24)$$

Also, in our idealization,  $K$  does not depend on  $z$  within the magnet,

$$\partial K / \partial z = 0. \quad (14.3.25)$$

Therefore  $K$  is also an integral of motion,

$$K = \text{constant}. \quad (14.3.26)$$

Next employ the Hamiltonian equation of motion for  $p_x(z)$  to find the result

$$p'_x = -\partial K / \partial x = -qB \quad (14.3.27)$$

where a *prime* denotes  $(d/dz)$ . Since the entry face of the magnet is located at  $z = z^{\text{in}}$ , (3.27) has the solution

$$p_x(z) = p_x^{\text{in}} - qB(z - z^{\text{in}}). \quad (14.3.28)$$

We are now prepared to find  $y(z)$  and  $t(z)$ . For Hamilton's equation of motion for  $y(z)$  we find the result

$$y' = \partial K / \partial p_y = p_y [(p_t)^2 / c^2 - m^2 c^2 - (p_x)^2 - (p_y)^2]^{-1/2}. \quad (14.3.29)$$

Insertion of (3.24) and (3.28) into (3.29) gives the further result

$$y' = p_y^{\text{in}} \{ (p_t^{\text{in}})^2 / c^2 - m^2 c^2 - [p_x^{\text{in}} - qB(z - z^{\text{in}})]^2 - (p_y^{\text{in}})^2 \}^{-1/2}. \quad (14.3.30)$$

This equation can be integrated to yield the result

$$y(z) = y^{\text{in}} + \int_{z^{\text{in}}}^z d\hat{z} p_y^{\text{in}} \{ (p_t^{\text{in}})^2 / c^2 - m^2 c^2 - [p_x^{\text{in}} - qB(\hat{z} - z^{\text{in}})]^2 - (p_y^{\text{in}})^2 \}^{-1/2}. \quad (14.3.31)$$

Similarly, *mutatis mutandis*,  $t'$  is given by the relation

$$t' = -p_t^{\text{in}} \{ (p_t^{\text{in}})^2 / c^2 - m^2 c^2 - [p_x^{\text{in}} - qB(z - z^{\text{in}})]^2 - (p_y^{\text{in}})^2 \}^{-1/2} \quad (14.3.32)$$

and  $t(z)$  is given by the integral

$$t(z) = t^{\text{in}} - \int_{z^{\text{in}}}^z d\hat{z} p_t^{\text{in}} \{ (p_t^{\text{in}})^2 / c^2 - m^2 c^2 - [p_x^{\text{in}} - qB(\hat{z} - z^{\text{in}})]^2 - (p_y^{\text{in}})^2 \}^{-1/2}. \quad (14.3.33)$$

If needed, the integrals (3.31) and (3.33) can be evaluated in terms of  $\sin^{-1}$  functions.

Note also that there are the relations

$$y'/t' = (dy/dz)/(dt/dz) = dy/dt \quad (14.3.34)$$

and, from (3.30) and (3.32),

$$y'/t' = -p_y^{\text{in}}/p_t^{\text{in}}. \quad (14.3.35)$$

Upon combining these relations we find

$$dy/dt = -p_y^{\text{in}}/p_t^{\text{in}} \quad (14.3.36)$$

and therefore

$$y(t) = y^{\text{in}} - (p_y^{\text{in}}/p_t^{\text{in}})(t - t^{\text{in}}), \quad (14.3.37)$$

a result consistent with the expected helical nature of trajectories in a uniform magnetic field.<sup>1</sup> Note that if a particle initially moves in the plane  $y = y^{\text{in}}$ , so that initially  $p_y^{\text{in}} = 0$ , then it must remain in this plane. According to (3.37), in the body of the magnet, there is *no* focussing or defocussing in the  $y$ /vertical direction. (As we will see in the next section, there is vertical focusing in the fringe-field regions.)

---

<sup>1</sup>Recall that  $p_t^{\text{in}} < 0$ .

What remains is to find  $x(z)$ . For Hamilton's equation of motion for  $x(z)$  we find the result

$$x' = \partial K / \partial p_x = p_x [(p_t)^2 / c^2 - m^2 c^2 - (p_x)^2 - (p_y)^2]^{-1/2}. \quad (14.3.38)$$

Insertion of (3.24) and (3.28) into (3.38) gives the further result

$$x' = [p_x^{\text{in}} - qB(z - z^{\text{in}})] \{ (p_t^{\text{in}})^2 / c^2 - m^2 c^2 - [p_x^{\text{in}} - qB(z - z^{\text{in}})]^2 - (p_y^{\text{in}})^2 \}^{-1/2}. \quad (14.3.39)$$

This equation can be integrated to yield the result

$$x(z) = x^{\text{in}} + \int_{z^{\text{in}}}^z d\hat{z} [p_x^{\text{in}} - qB(\bar{z} - z^{\text{in}})] \{ (p_t^{\text{in}})^2 / c^2 - m^2 c^2 - [p_x^{\text{in}} - qB(\hat{z} - z^{\text{in}})]^2 - (p_y^{\text{in}})^2 \}^{-1/2}. \quad (14.3.40)$$

When needed, this integral can be evaluated in terms of square root and  $\sin^{-1}$  functions.

Let  $z^{\text{fin}}$  be the *final* value of  $z$  on the exit face of the magnet. Before leaving the subject of the integrals (3.31), (3.33), and (3.40) for  $x(z)$ , etc., we note that they can be written in a more symmetric form if they are to be evaluated at the upper limit of integration  $z = z^{\text{fin}}$  to compute  $x^{\text{fin}} = x(z^{\text{fin}})$ , etc. To begin, observe that the *distance*  $d$  between the magnet faces (the length of the magnet) is given by the relation

$$d = z^{\text{fin}} - z^{\text{in}}. \quad (14.3.41)$$

Introduce a new integration variable, to be called  $w$ , by the rule

$$w = \hat{z} - (1/2)(z^{\text{in}} + z^{\text{fin}}) \Leftrightarrow \hat{z} = w + (1/2)(z^{\text{in}} + z^{\text{fin}}). \quad (14.3.42)$$

With this change the integration end points become

$$w^{\text{in}} = z^{\text{in}} - (1/2)(z^{\text{in}} + z^{\text{fin}}) = -(1/2)(z^{\text{fin}} - z^{\text{in}}) = -d/2 \quad (14.3.43)$$

and

$$w^{\text{fin}} = z^{\text{fin}} - (1/2)(z^{\text{in}} + z^{\text{fin}}) = (1/2)(z^{\text{fin}} - z^{\text{in}}) = d/2. \quad (14.3.44)$$

Moreover, we see that

$$\hat{z} - z^{\text{in}} = w + (1/2)(z^{\text{in}} + z^{\text{fin}}) - z^{\text{in}} = w + (1/2)(z^{\text{fin}} - z^{\text{in}}) = w + d/2 \quad (14.3.45)$$

so that

$$[d/2 - (\hat{z} - z^{\text{in}})] = -w. \quad (14.3.46)$$

Jumping ahead a bit, in Subsections 3.3 and 3.4 we will define *initial design* and *initial deviation* quantities  $p_x^{\text{ind}}$  and  $\bar{p}_x^{\text{in}}$  by the relations

$$p_x^{\text{ind}} = (1/2)qBd, \quad (14.3.47)$$

$$p_x^{\text{in}} = p_x^{\text{ind}} + \bar{p}_x^{\text{in}} = (1/2)qBd + \bar{p}_x^{\text{in}}. \quad (14.3.48)$$

Consequently, there is the relation

$$\begin{aligned} [p_x^{\text{in}} - qB(\hat{z} - z^{\text{in}})] &= \bar{p}_x^{\text{in}} + [(1/2)qBd - qB(\hat{z} - z^{\text{in}})] = \\ \bar{p}_x^{\text{in}} + qB[d/2 - (\hat{z} - z^{\text{in}})] &= \bar{p}_x^{\text{in}} - qBw \end{aligned} \quad (14.3.49)$$

It follows that the integrals (3.31), (3.33), and (3.40), when evaluated at the upper limit of integration  $z = z^{\text{fin}}$ , take the somewhat simpler forms

$$y^{\text{fin}} = y^{\text{in}} + \int_{-d/2}^{d/2} dw p_y^{\text{in}} \{(p_t^{\text{in}})^2/c^2 - m^2 c^2 - [\bar{p}_x^{\text{in}} - qBw]^2 - (p_y^{\text{in}})^2\}^{-1/2}, \quad (14.3.50)$$

$$t^{\text{fin}} = t^{\text{in}} - \int_{-d/2}^{d/2} dw p_t^{\text{in}} \{(p_t^{\text{in}})^2/c^2 - m^2 c^2 - [\bar{p}_x^{\text{in}} - qBw]^2 - (p_y^{\text{in}})^2\}^{-1/2}, \quad (14.3.51)$$

$$x^{\text{fin}} = x^{\text{in}} + \int_{-d/2}^{d/2} dw [\bar{p}_x^{\text{in}} - qBw] \{(p_t^{\text{in}})^2/c^2 - m^2 c^2 - [\bar{p}_x^{\text{in}} - qBw]^2 - (p_y^{\text{in}})^2\}^{-1/2}. \quad (14.3.52)$$

But we note, *en passant*, that we shall try to avoid evaluating any of these integrals as much as possible.

As an alternative to evaluating the integrals (3.40) or (3.52), we can gain insight into the nature of  $x(z)$  using the fact that the Hamiltonian  $K$  is conserved. Recall (3.26). On the entry face, using (3.19), (3.24), and (3.28), we find the result

$$K|_{z=z^{\text{in}}} = -[(p_t^{\text{in}})^2/c^2 - m^2 c^2 - (p_x^{\text{in}})^2 - (p_y^{\text{in}})^2]^{1/2} + qBx^{\text{in}}. \quad (14.3.53)$$

And at a general interior point  $z$  we find from (3.19), (3.24), and (3.28) the result

$$K|_z = -\{(p_t^{\text{in}})^2/c^2 - m^2 c^2 - [\bar{p}_x^{\text{in}} - qB(z - z^{\text{in}})]^2 - (p_y^{\text{in}})^2\}^{1/2} + qBx(z). \quad (14.3.54)$$

Upon equating the right sides of (3.54) and (3.53) we find the result

$$\begin{aligned} & -\{(p_t^{\text{in}})^2/c^2 - m^2 c^2 - [\bar{p}_x^{\text{in}} - qB(z - z^{\text{in}})]^2 - (p_y^{\text{in}})^2\}^{1/2} + qBx(z) = \\ & -[(p_t^{\text{in}})^2/c^2 - m^2 c^2 - (p_x^{\text{in}})^2 - (p_y^{\text{in}})^2]^{1/2} + qBx^{\text{in}}. \end{aligned} \quad (14.3.55)$$

What does this result say about the trajectory? It can be rewritten in the form

$$\begin{aligned} & -\{(p_t^{\text{in}})^2/c^2 - m^2 c^2 - [\bar{p}_x^{\text{in}} - qB(z - z^{\text{in}})]^2 - (p_y^{\text{in}})^2\}^{1/2} = \\ & -[(p_t^{\text{in}})^2/c^2 - m^2 c^2 - (p_x^{\text{in}})^2 - (p_y^{\text{in}})^2]^{1/2} - qB[x(z) - x^{\text{in}}]. \end{aligned} \quad (14.3.56)$$

Now square and equate both sides of (3.56) and then move some terms across the equal sign to find the result

$$\begin{aligned} & \{(p_t^{\text{in}})^2/c^2 - m^2 c^2 - [\bar{p}_x^{\text{in}} - qB(z - z^{\text{in}})]^2 - (p_y^{\text{in}})^2\} - [(p_t^{\text{in}})^2/c^2 - m^2 c^2 - (p_x^{\text{in}})^2 - (p_y^{\text{in}})^2] \\ & = 2[(p_t^{\text{in}})^2/c^2 - m^2 c^2 - (p_x^{\text{in}})^2 - (p_y^{\text{in}})^2]^{1/2} qB[x(z) - x^{\text{in}}] + \{qB[x(z) - x^{\text{in}}]\}^2. \end{aligned} \quad (14.3.57)$$

The left side of (3.57) can be simplified to yield the final result

$$\begin{aligned} & 2qB\bar{p}_x^{\text{in}}(z - z^{\text{in}}) - [qB(z - z^{\text{in}})]^2 = \\ & 2[(p_t^{\text{in}})^2/c^2 - m^2 c^2 - (p_x^{\text{in}})^2 - (p_y^{\text{in}})^2]^{1/2} qB[x(z) - x^{\text{in}}] + \{qB[x(z) - x^{\text{in}}]\}^2. \end{aligned} \quad (14.3.58)$$

This result is an implicit equation of the form

$$Ax^2 + Bxz + Cz^2 + Dx + Ez + F = 0. \quad (14.3.59)$$

Inspection and some calculation reveals that  $A = C \neq 0$ ,  $B = 0$ , and  $F/A < 0$ ; therefore (3.59) specifies a *circle* in the  $x, z$  variables. Consequently the *projection* of the trajectory within the magnet and onto the  $x, z$  plane is a circular arc, again as expected for a helical trajectory.

### 14.3.3 Specification of Design Orbit

Having found the nature of all possible orbits, we now turn to specifying a particular design orbit. Figure 3.1 shows what we will call the standard *symmetric* design orbit for a parallel faced bend. We call this orbit symmetric because, as illustrated, the entry and exit angles both have the same value, namely  $\theta/2$ . Let us next relate the geometric features of the design orbit as exhibited in Figure 3.1 to phase-space quantities.

Let  $p_y^d$  and  $y^d$  be the values of  $p_y$  and  $y$  on the *design* orbit. We know that  $p_y$  is an integral of motion. Recall (3.24). Consistent with (3.37), we specify that the design orbit has the properties

$$p_y^d = 0 \text{ and } y^d = 0. \quad (14.3.60)$$

That is, the design orbit lies in the mid plane. Let  $p_y^{\text{ind}}, y^{\text{ind}}$  and  $p_y^{\text{find}}, y^{\text{find}}$  be the values of  $p_y^d, y^d$  on entry and exit, respectively. Then, in view of (3.60), there are the relations

$$y^{\text{ind}} = 0 \text{ and } y^{\text{find}} = 0, \quad (14.3.61)$$

$$p_y^{\text{ind}} = 0 \text{ and } p_y^{\text{find}} = 0. \quad (14.3.62)$$

We also know that  $p_t$  is an integral of motion. Again recall (3.24). Let  $p_t^d$  be its value on the design orbit. From (1.6.17), since we have assumed there is no electric field [recall (3.2)], there is the relation

$$p_t^d = -\gamma^d mc^2. \quad (14.3.63)$$

Here  $\gamma^d$  is the (*constant*) value of the relativistic  $\gamma$  factor given by (1.5.29) evaluated with  $v = v^d$  where  $v^d$  is the (*constant*) value of the particle velocity (actually *speed*) on the design orbit. Let  $p_t^{\text{ind}}$  and  $p_t^{\text{find}}$  be the values of  $p_t^d$  on entry and exit, respectively. Then, since  $p_t$  is an integral of motion, there are the relations

$$p_t^{\text{find}} = p_t^{\text{ind}} = -\gamma^d mc^2. \quad (14.3.64)$$

Let  $t^{\text{ind}}$  and  $t^{\text{find}}$  be the entry and exit values, respectively, of  $t$  on the design orbit. We will specify their values at the end of this subsection.

Let  $p_x^{\text{ind}}, x^{\text{ind}}$  and  $p_x^{\text{find}}, x^{\text{find}}$  be the values of  $p_x^d, x^d$  on entry and exit, respectively. Then, in view of Figure 3.1, there are the relations

$$x^{\text{ind}} = 0 \text{ and } x^{\text{find}} = 0, \quad (14.3.65)$$

$$p_x^{\text{ind}} = \gamma^d mv^d \sin(\theta/2) \text{ and } p_x^{\text{find}} = -p_x^{\text{ind}} = -\gamma^d mv^d \sin(\theta/2). \quad (14.3.66)$$

Also, from (3.28) and (3.41), there is the relation

$$p_x^{\text{find}} = p_x^d(z^{\text{fin}}) = p_x^{\text{ind}} - qB(z^{\text{fin}} - z^{\text{in}}) = p_x^{\text{ind}} - qBd. \quad (14.3.67)$$

Combining (3.47) and (3.67) gives the further results

$$p_x^{\text{ind}} = (1/2)qBd \text{ and } p_x^{\text{find}} = -(1/2)qBd, \quad (14.3.68)$$

from which it follows using (3.66) that there is the result

$$\gamma^d mv^d \sin(\theta/2) = (1/2)qBd. \quad (14.3.69)$$

Since the magnitude of the design momentum,  $p^d$ , satisfies the relation

$$p^d = \gamma^d m v^d, \quad (14.3.70)$$

see (1.5.28) and Exercise 1.5.9, the relation (3.69) can also be written as

$$p^d \sin(\theta/2) = (1/2)qBd. \quad (14.3.71)$$

This equation interrelates the design energy ( $-p_t^d$ ), the magnetic field, and geometric parameters. To further elaborate on geometric relations, let  $\rho$  be the radius of the design orbit. From Figure 3.1 we see that

$$\rho \sin(\theta/2) = d/2. \quad (14.3.72)$$

Correspondingly, the length  $s$  of the design orbit is given by

$$s = \rho\theta. \quad (14.3.73)$$

Note that (3.71) and (3.72) can be combined to give the relation

$$p^d/q = B\rho, \quad (14.3.74)$$

in agreement with (1.5.81). Remember too that the quantity  $B\rho$  is called the magnetic rigidity. Evidently we may also express  $\rho$  in terms of the magnetic rigidity and the magnetic field  $B$  by writing

$$\rho = B\rho/B = p^d/(qB) = (\gamma^d m v^d)/(qB). \quad (14.3.75)$$

Finally let  $t^{\text{ind}}$  and  $t^{\text{find}}$  be the initial and final times on the design orbit. Then there is the relation

$$t^{\text{find}} - t^{\text{ind}} = s/v^d. \quad (14.3.76)$$

As stated earlier, and will be verified in the next subsection, the quantity  $\bar{p}_x^{\text{in}}$  is zero for the design orbit. And, according to (3.60) and (3.61),  $p_y^{\text{in}}$  is also zero for the design orbit. Look at the the integral (3.51). Inspection reveals that, when evaluated for the design orbit, the time interval  $t^{\text{find}} - t^{\text{ind}}$  can only depend on the remaining quantities on the right side of (3.51), namely  $d$ ,  $p_t^{\text{in}}$ ,  $m$ ,  $c$ ,  $q$ , and  $B$ . Let us try to rewrite (3.76) in terms of these quantities. Evidently the content of (3.71) can be re-expressed in the form

$$\theta = 2 \sin^{-1} \left\{ (1/2) [(qBd)/p^d] \right\} = 2 \sin^{-1} \left\{ (1/2) [(qBd)/(\gamma^d m v^d)] \right\}. \quad (14.3.77)$$

Upon combining (3.75) and (3.77) in (3.73) we see that

$$s = 2[(\gamma^d m v^d)/(qB)] \sin^{-1} \left\{ (1/2) [(qBd)/(\gamma^d m v^d)] \right\} \quad (14.3.78)$$

and therefore, by (3.76),

$$t^{\text{find}} - t^{\text{ind}} = (2/v^d) [(\gamma^d m v^d)/(qB)] \sin^{-1} \left\{ (1/2) [(qBd)/(\gamma^d m v^d)] \right\}. \quad (14.3.79)$$

What remains is to express  $v^d$  and  $\gamma^d$  in terms of  $p_t^{\text{in}}$ . From (3.63) and (3.64) there is the relation

$$\gamma^d = -p_t^{\text{in}}/(mc^2). \quad (14.3.80)$$

Finally, there are the relations

$$v^d = \beta^d c \quad (14.3.81)$$

and

$$\gamma^d = [1 - (\beta^d)^2]^{-1/2} \Leftrightarrow \beta^d = [1 - (1/\gamma^d)^2]^{1/2} = (1/\gamma^d)(\gamma^d + 1)^{1/2}(\gamma^d - 1)^{1/2}. \quad (14.3.82)$$

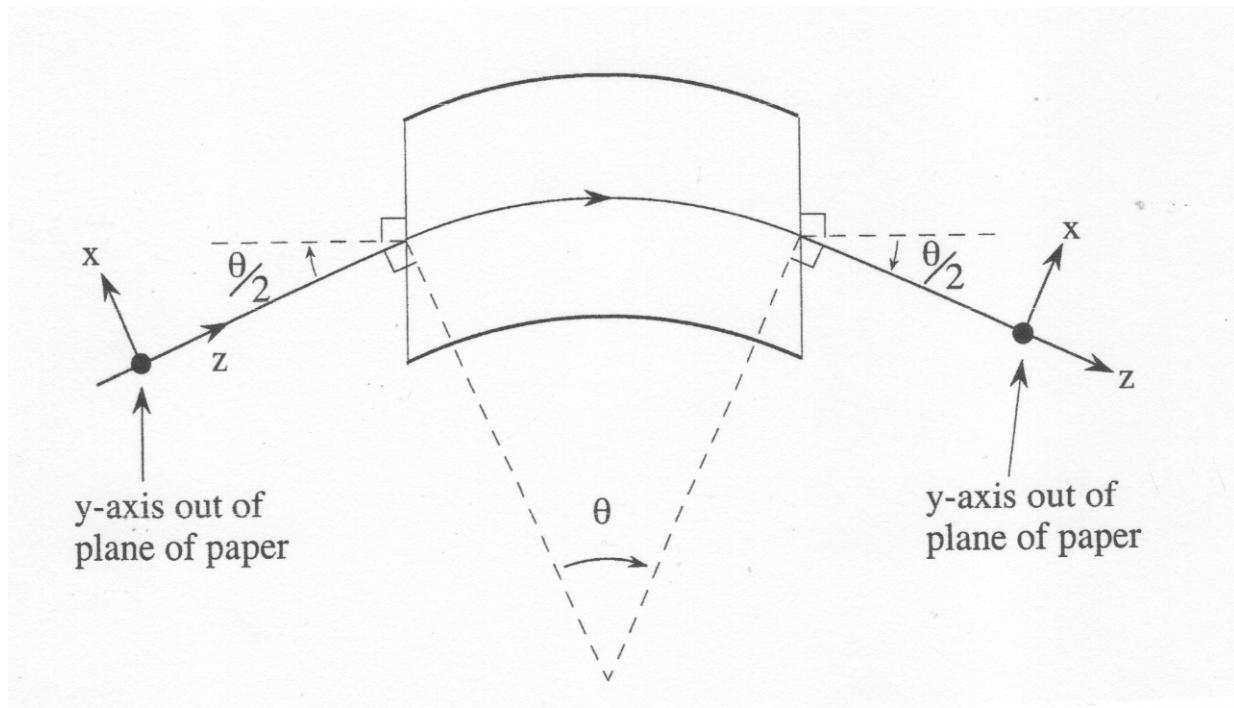


Figure 14.3.1: Top view of symmetric design orbit in a rectangular bend magnet showing *local* Cartesian coordinates attached to the design orbit. See also Figure 1.6.2 where *global* Cartesian coordinates are displayed.

### 14.3.4 Expansion About the Design Orbit

#### Our Goal

Introduce the quantities

$$\{x^{\text{fin}}, p_x^{\text{fin}}, y^{\text{fin}}, p_y^{\text{fin}}, t^{\text{fin}}, p_t^{\text{fin}}\} \quad (14.3.83)$$

where

$$x^{\text{fin}} = x(z^{\text{fin}}), \text{ etc.} \quad (14.3.84)$$

Our goal is to express these *final* quantities in terms of the *initial* quantities (3.20) by way of the *design* orbit quantities

$$\{x^{\text{ind}}, p_x^{\text{ind}}, y^{\text{ind}}, p_y^{\text{ind}}, t^{\text{ind}}, p_t^{\text{ind}}\} \quad (14.3.85)$$

and

$$\{x^{\text{find}}, p_x^{\text{find}}, y^{\text{find}}, p_y^{\text{find}}, t^{\text{find}}, p_t^{\text{find}}\}, \quad (14.3.86)$$

which have just been specified in the previous subsection. To do so, introduce (with respect to the design orbit) *deviation* variables denoted by an overbar and defined by

$$x^{\text{in}} = x^{\text{ind}} + \bar{x}^{\text{in}} \text{ etc. and } x^{\text{fin}} = x^{\text{find}} + \bar{x}^{\text{fin}} \text{ etc.} \quad (14.3.87)$$

More precisely, our goal is to express the final deviation variables in terms of the initial deviation variables. Note that since the design quantities appearing in (3.87) are fixed numbers describing the beam element (dipole) in question, the relations (3.87) between the deviation variables and the original variables comprise canonical transformations. We also know that the relation between the initial variables (3.20) and the final variables (3.83) is a canonical transformation. It follows that the relation between initial and final deviation variables is a canonical transformation. Why introduce deviation variables? By construction, the final deviation variables will vanish if the initial deviation variables vanish. Therefore, when deviation variables are employed, the transfer map  $\mathcal{M}$  for any element sends the origin into itself,

$$\mathcal{M}\mathbf{0} = \mathbf{0}. \quad (14.3.88)$$

#### Effect of $\mathcal{M}$ on Momenta

For a parallel-faced dipole it is easiest to begin with relations for momentum variables. For  $p_x$  we have, from (3.28) and (3.41), the relation

$$p_x^{\text{fin}} = p_x(z^{\text{fin}}) = p_x^{\text{in}} - qB(z^{\text{fin}} - z^{\text{in}}) = p_x^{\text{in}} - qBd, \quad (14.3.89)$$

which can be rewritten in the form

$$p_x^{\text{fin}} + (1/2)qBd = p_x^{\text{in}} - (1/2)qBd. \quad (14.3.90)$$

Next recall (3.67) to rewrite (3.90) in the form

$$p_x^{\text{fin}} - p_x^{\text{find}} = p_x^{\text{in}} - p_x^{\text{ind}}, \quad (14.3.91)$$

which, in view of (3.87), takes the amazingly simple final form

$$\bar{p}_x^{\text{fin}} = \bar{p}_x^{\text{in}}. \quad (14.3.92)$$

For  $p_y$  there are the relations

$$p_y^{\text{fin}} = p_y^{\text{in}} \text{ and } p_y^{\text{find}} = p_y^{\text{ind}} = 0. \quad (14.3.93)$$

See (3.24) and (3.62). It follows from these relations and (3.87) that there is the result

$$\bar{p}_y^{\text{fin}} = \bar{p}_y^{\text{in}}. \quad (14.3.94)$$

For  $p_t$  there are the relations

$$p_t^{\text{fin}} = p_t^{\text{in}} \text{ and } p_t^{\text{find}} = p_t^{\text{ind}} = p_t^{\text{d}} = -\gamma^{\text{d}} mc^2. \quad (14.3.95)$$

See (3.24) and (3.70). It follows from these relations and (3.87) that there is the result

$$\bar{p}_t^{\text{fin}} = \bar{p}_t^{\text{in}}. \quad (14.3.96)$$

From (3.92), (3.94), and (3.96) we see that, for the *body* of a parallel-faced dipole, *all* the *momentum deviation* variables are *unchanged* under the action of its transfer map  $\mathcal{M}$ ,

$$\mathcal{M}\bar{p}_x^{\text{in}} = \bar{p}_x^{\text{in}}, \quad \mathcal{M}\bar{p}_y^{\text{in}} = \bar{p}_y^{\text{in}}, \quad \text{and} \quad \mathcal{M}\bar{p}_t^{\text{in}} = \bar{p}_t^{\text{in}}. \quad (14.3.97)$$

### Effect of $\mathcal{M}$ on Position Variables

#### Results for $x$

Finding results for position variables is more complicated. Let us begin with  $x$ . Capitalize on the relation (3.58). Evaluate both sides for the  $z$  value  $z = z^{\text{fin}}$  to find the result

$$\begin{aligned} & 2qBp_x^{\text{in}}d - (qBd)^2 = \\ & 2[(p_t^{\text{in}})^2/c^2 - m^2c^2 - (p_x^{\text{in}})^2 - (p_y^{\text{in}})^2]^{1/2}qB(x^{\text{fin}} - x^{\text{in}}) + [qB(x^{\text{fin}} - x^{\text{in}})]^2. \end{aligned} \quad (14.3.98)$$

Here we have used (3.41) and (3.84). Evidently we can solve (3.98) for  $x^{\text{fin}}$ . Add the appropriate term to both sides of (3.98) to complete the square of the term on the right. So doing produces the relation

$$\begin{aligned} & 2qBp_x^{\text{in}}d - (qBd)^2 + [(p_t^{\text{in}})^2/c^2 - m^2c^2 - (p_x^{\text{in}})^2 - (p_y^{\text{in}})^2] = \\ & \{[(p_t^{\text{in}})^2/c^2 - m^2c^2 - (p_x^{\text{in}})^2 - (p_y^{\text{in}})^2]^{1/2} + qB(x^{\text{fin}} - x^{\text{in}})\}^2. \end{aligned} \quad (14.3.99)$$

Observe that

$$-(p_x^{\text{in}})^2 + 2qBp_x^{\text{in}}d - (qBd)^2 = -(p_x^{\text{in}} - qBd)^2. \quad (14.3.100)$$

Therefore (3.99) can be rewritten in the simpler form

$$\begin{aligned} -(p_x^{\text{in}} - qBd)^2 + [(p_t^{\text{in}})^2/c^2 - m^2c^2 - (p_y^{\text{in}})^2] = \\ \{[(p_t^{\text{in}})^2/c^2 - m^2c^2 - (p_x^{\text{in}})^2 - (p_y^{\text{in}})^2]^{1/2} + qB(x^{\text{fin}} - x^{\text{in}})\}^2. \end{aligned} \quad (14.3.101)$$

Further manipulate (3.101) by taking square roots of both sides and moving some terms across the equal sign to find the result

$$\begin{aligned} qB(x^{\text{fin}} - x^{\text{in}}) = \\ -[(p_t^{\text{in}})^2/c^2 - m^2c^2 - (p_x^{\text{in}})^2 - (p_y^{\text{in}})^2]^{1/2} \\ + [(p_t^{\text{in}})^2/c^2 - m^2c^2 - (p_x^{\text{in}} - qBd)^2 - (p_y^{\text{in}})^2]^{1/2}. \end{aligned} \quad (14.3.102)$$

Next specify quantities on both sides of (3.102) by their design values plus deviation values. That is, bring to mind the relations,

$$x^{\text{fin}} = x^{\text{ind}} + \bar{x}^{\text{fin}} = \bar{x}^{\text{fin}}, \quad x^{\text{in}} = x^{\text{ind}} + \bar{x}^{\text{in}} = \bar{x}^{\text{in}}, \quad (14.3.103)$$

$$p_x^{\text{in}} = p_x^{\text{ind}} + \bar{p}_x^{\text{in}} = (1/2)qBd + \bar{p}_x^{\text{in}}, \quad p_t^{\text{in}} = p_t^{\text{ind}} + \bar{p}_t^{\text{in}} = -\gamma^{\text{d}}mc^2 + \bar{p}_t^{\text{in}}, \quad (14.3.104)$$

$$p_y^{\text{in}} = p_y^{\text{ind}} + \bar{p}_y^{\text{in}} = \bar{p}_y^{\text{in}}. \quad (14.3.105)$$

See (3.61), (3.62), (3.64), (3.65), (3.68), and (3.87). Now employ these relations in (3.102). to find the the result

$$\begin{aligned} qB(\bar{x}^{\text{fin}} - \bar{x}^{\text{in}}) = \\ -\{(-\gamma^{\text{d}}mc^2 + \bar{p}_t^{\text{in}})^2/c^2 - m^2c^2 - [\bar{p}_x^{\text{in}} + (1/2)qBd]^2 - (\bar{p}_y^{\text{in}})^2\}^{1/2} \\ +\{(-\gamma^{\text{d}}mc^2 + \bar{p}_t^{\text{in}})^2/c^2 - m^2c^2 - [\bar{p}_x^{\text{in}} - (1/2)qBd]^2 - (\bar{p}_y^{\text{in}})^2\}^{1/2}. \end{aligned} \quad (14.3.106)$$

Note that in passing from (3.101) to (3.102) it was necessary to chose a particular sign for the square root. Looking at (3.106), we see that if both sides are evaluated with *vanishing* deviation variables, we find the desirable result  $0 = 0$ . This would have not been the result had we chosen the alternate sign for the square root. Therefore we made the proper choice.

We are almost ready to make Taylor expansions, in terms of deviation variables, of the right side of (3.106). The first curly bracket term on the right side of (3.106) can be rewritten as

$$\begin{aligned} \{(-\gamma^{\text{d}}mc^2 + \bar{p}_t^{\text{in}})^2/c^2 - m^2c^2 - [\bar{p}_x^{\text{in}} + (1/2)qBd]^2 - (\bar{p}_y^{\text{in}})^2\} = \\ \{(-\gamma^{\text{d}}mc^2)^2/c^2 - m^2c^2 - [(1/2)qBd]^2\} \\ + \{-2\gamma^{\text{d}}m\bar{p}_t^{\text{in}} - \bar{p}_x^{\text{in}}qBd\} \\ + \{(\bar{p}_t^{\text{in}})^2/c^2 - (\bar{p}_x^{\text{in}})^2 - (\bar{p}_y^{\text{in}})^2\}. \end{aligned} \quad (14.3.107)$$

Rewrite this result in the more compact form

$$\begin{aligned} \{(-\gamma^{\text{d}}mc^2 + \bar{p}_t^{\text{in}})^2/c^2 - m^2c^2 - [\bar{p}_x^{\text{in}} + (1/2)qBd]^2 - (\bar{p}_y^{\text{in}})^2\} = \\ A_0 + A_{1+} + A_2 = \\ A_0[1 + A_{1+}/A_0 + A_2/A_0] = \\ A_0(1 + B_{1+} + B_2) \end{aligned} \quad (14.3.108)$$

where

$$A_0 = \{(-\gamma^d mc^2)^2/c^2 - m^2 c^2 - [(1/2)qBd]^2\}, \quad (14.3.109)$$

$$A_{1+} = \{-2\gamma^d m \bar{p}_t^{\text{in}} - \bar{p}_x^{\text{in}} qBd\}, \quad (14.3.110)$$

$$A_2 = \{(\bar{p}_t^{\text{in}})^2/c^2 - (\bar{p}_x^{\text{in}})^2 - (\bar{p}_y^{\text{in}})^2\}, \quad (14.3.111)$$

and

$$B_{1+} = A_{1+}/A_0, \quad (14.3.112)$$

$$B_2 = A_2/A_0. \quad (14.3.113)$$

From (3.108) we conclude that there is the first curly bracket relation

$$\begin{aligned} & \{(-\gamma^d mc^2 + \bar{p}_t^{\text{in}})^2/c^2 - m^2 c^2 - [\bar{p}_x^{\text{in}} + (1/2)qBd]^2 - (\bar{p}_y^{\text{in}})^2\}^{1/2} = \\ & (A_0)^{1/2}(1 + B_{1+} + B_2)^{1/2}. \end{aligned} \quad (14.3.114)$$

Similarly, the second curly bracket on the right side of (3.106) can be rewritten in the more compact form

$$\begin{aligned} & \{(-\gamma^d mc^2 + \bar{p}_t^{\text{in}})^2/c^2 - m^2 c^2 - [\bar{p}_x^{\text{in}} - (1/2)qBd]^2 - (\bar{p}_y^{\text{in}})^2\} = \\ & A_0 + A_{1-} + A_2 = \\ & A_0[1 + A_{1-}/A_0 + A_2/A_0] = \\ & A_0(1 + B_{1-} + B_2) \end{aligned} \quad (14.3.115)$$

where

$$A_{1-} = \{-2\gamma^d m \bar{p}_t^{\text{in}} + \bar{p}_x^{\text{in}} qBd\}, \quad (14.3.116)$$

$$B_{1-} = A_{1-}/A_0. \quad (14.3.117)$$

Consequently, for the second curly bracket, there is the relation

$$\begin{aligned} & \{(-\gamma^d mc^2 + \bar{p}_t^{\text{in}})^2/c^2 - m^2 c^2 - [\bar{p}_x^{\text{in}} - (1/2)qBd]^2 - (\bar{p}_y^{\text{in}})^2\}^{1/2} = \\ & (A_0)^{1/2}(1 + B_{1-} + B_2)^{1/2}. \end{aligned} \quad (14.3.118)$$

What essentially remains is to expand the quantities  $(1 + B_{1\pm} + B_2)^{1/2}$  as power series in the deviation variables. One finds through third order, from the binomial theorem for fractional exponents or *Mathematica* for example, the result

$$(1 + B_{1\pm} + B_2)^{1/2} = 1 + C_{1\pm} + C_{2\pm} + C_{3\pm} + \dots \quad (14.3.119)$$

where

$$C_{1\pm} = (1/2)B_{1\pm}, \quad (14.3.120)$$

$$C_{2\pm} = (1/8)[-(B_{1\pm})^2 + 4B_2], \quad (14.3.121)$$

$$C_{3\pm} = (1/16)[(B_{1\pm})^3 - 4B_{1\pm}B_2]. \quad (14.3.122)$$

We are now prepared to employ the results (3.107) through (3.122) in (3.106). Using (3.114), (3.118), and (3.119) through (3.122), the relation (3.106) can be brought to the form

$$\begin{aligned}
qB(\bar{x}^{\text{fin}} - \bar{x}^{\text{in}}) = & \\
& -\{(-\gamma^d mc^2 + \bar{p}_t^{\text{in}})^2/c^2 - m^2 c^2 - [\bar{p}_x^{\text{in}} + (1/2)qBd]^2 - (\bar{p}_y^{\text{in}})^2\}^{1/2} \\
& +\{(-\gamma^d mc^2 + \bar{p}_t^{\text{in}})^2/c^2 - m^2 c^2 - [\bar{p}_x^{\text{in}} - (1/2)qBd]^2 - (\bar{p}_y^{\text{in}})^2\}^{1/2} = \\
& -(A_0)^{1/2}[(1 + B_{1+} + B_2)^{1/2} - (1 + B_{1-} + B_2)^{1/2}] = \\
& -(A_0)^{1/2}\{[1 + C_{1+} + C_{2+} + C_{3+}] + \dots - [1 + C_{1-} + C_{2-} + C_{3-}] + \dots\} = \\
& -(A_0)^{1/2}\{[C_{1+} - C_{1-}] + [C_{2+} - C_{2-}] + [C_{3+} - C_{3-}] + \dots\} = \\
& -(A_0)^{1/2}\{(1/2)(B_{1+} - B_{1-}) - (1/8)[(B_{1+})^2 - (B_{1-})^2] + (1/16)\} + \dots
\end{aligned} \tag{14.3.123}$$

with

$$(1/16)\} = (1/16)\{[(B_{1+})^3 - 4B_{1+}B_2] - [(B_{1-})^3 - 4B_{1-}B_2]\}. \tag{14.3.124}$$

For the moment let us concentrate our attention only on the terms on the right side of (3.123) that are *linear* in the deviation variables since these are the only terms that contribute to the matrix part of the transfer map. For the linear terms there is the result

$$\begin{aligned}
\text{linear terms} = & -(A_0)^{1/2}(1/2)(B_{1+} - B_{1-}) = -(A_0)^{-1/2}(1/2)(A_{1+} - A_{1-}) = \\
& -(A_0)^{-1/2}(1/2)[(-2\gamma^d m \bar{p}_t^{\text{in}} - \bar{p}_x^{\text{in}} q Bd) - (-2\gamma^d m \bar{p}_t^{\text{in}} + \bar{p}_x^{\text{in}} q Bd)] = \\
& (A_0)^{-1/2} q Bd \bar{p}_x^{\text{in}}.
\end{aligned} \tag{14.3.125}$$

Consequently, putting everything together, (3.106) becomes

$$\bar{x}^{\text{fin}} = \bar{x}^{\text{in}} + (A_0)^{-1/2} d \bar{p}_x^{\text{in}} + \text{higher order terms.} \tag{14.3.126}$$

### Linear Results for $y$

We have found results for  $x$ . We next search for results for  $y$ . For  $y$  we could evaluate the integral (3.50) and then expand the result as a Taylor series in the deviation variables. However, if for the moment we are only interested in the linear term, it is simpler to use (3.37). Setting  $t = t^{\text{fin}}$  in (3.37) gives the result

$$y^{\text{fin}} = y^{\text{in}} - (\bar{p}_y^{\text{in}} / p_t^{\text{in}})(t^{\text{fin}} - t^{\text{in}}). \tag{14.3.127}$$

In view of (3.61) and (3.62) this result can also be written in the form

$$\bar{y}^{\text{fin}} = \bar{y}^{\text{in}} - (\bar{p}_y^{\text{in}} / p_t^{\text{in}})(t^{\text{fin}} - t^{\text{in}}). \tag{14.3.128}$$

Moreover,  $t^{\text{fin}}$ ,  $t^{\text{in}}$ , and  $p_t^{\text{in}}$  differ from their design values only by terms that are linear and higher in the deviation variables. Therefore, we may write

$$\begin{aligned}
\bar{y}^{\text{fin}} &= \bar{y}^{\text{in}} - (\bar{p}_y^{\text{in}} / p_t^{\text{ind}})(t^{\text{find}} - t^{\text{ind}}) + \text{higher order terms} \\
&= \bar{y}^{\text{in}} + \bar{p}_y^{\text{in}}[(t^{\text{find}} - t^{\text{ind}}) / (-p_t^{\text{ind}})] + \text{higher order terms.}
\end{aligned} \tag{14.3.129}$$

Note that the value of the bracketed term in (3.129) is known from (3.64) and (3.79).

### Results for $t$ and $y$

The last task is to find results for  $t$ . This can be done using (3.51). Note that once the integral has been manipulated/evaluated to yield  $t^{\text{fin}} - t^{\text{in}}$  in terms of design values and deviation variables, the results can also be used in (3.128) to provide a complete expansion for  $\bar{y}^{\text{fin}}$ . Examine the integral (3.51). Evidently, since  $p_t^{\text{in}}$  does not depend on  $w$ , it may be factored out of the integral so that we are left to evaluate the integral given by

$$\text{integral} = \int_{-d/2}^{d/2} dw \left\{ (p_t^{\text{in}})^2/c^2 - m^2 c^2 - [\bar{p}_x^{\text{in}} - qBw]^2 - (\bar{p}_y^{\text{in}})^2 \right\}^{-1/2}. \quad (14.3.130)$$

[Note that here we have used the result (3.93) so that  $p_y^{\text{in}} = \bar{p}_y^{\text{in}}$ .]

Rather than evaluate the integral (3.130) and then expand the results in terms of design values and deviation variables, it is easier to take the opposite approach: first express/expand the integrand in terms of design values and deviation variables and then integrate the results. The integrand is given by the expression

$$\text{integrand} = \left\{ (p_t^{\text{in}})^2/c^2 - m^2 c^2 - [\bar{p}_x^{\text{in}} - qBw]^2 - (\bar{p}_y^{\text{in}})^2 \right\}^{-1/2} \quad (14.3.131)$$

Recall that there is the relation

$$p_t^{\text{in}} = p_t^{\text{ind}} + \bar{p}_t^{\text{in}} = -\gamma^d mc^2 + \bar{p}_t^{\text{in}} \quad (14.3.132)$$

so that

$$(p_t^{\text{in}})^2 = (p_t^{\text{ind}})^2 + 2p_t^{\text{ind}}\bar{p}_t^{\text{in}} + (\bar{p}_t^{\text{in}})^2 = (\gamma^d mc^2)^2 - 2\gamma^d mc^2 \bar{p}_t^{\text{in}} + (\bar{p}_t^{\text{in}})^2. \quad (14.3.133)$$

Consequently the integrand (3.131) can also be written in the form

$$\text{integrand} = \{a_0 + a_1 + a_2\}^{-1/2} \quad (14.3.134)$$

where

$$a_0 = [(\gamma^d)^2 - 1]m^2 c^2 - (qBw)^2, \quad (14.3.135)$$

$$a_1 = -2\gamma^d mc^2 \bar{p}_t^{\text{in}} + 2qBw \bar{p}_x^{\text{in}}, \quad (14.3.136)$$

and

$$a_2 = (\bar{p}_t^{\text{in}}/c)^2 - (\bar{p}_y^{\text{in}})^2 - (\bar{p}_y^{\text{in}})^2. \quad (14.3.137)$$

Begin to expand the integrand (3.134) by writing

$$a_0 + a_1 + a_2 = a_0[1 + a_1/a_0 + a_2/a_0] = a_0(1 + b_1 + b_2) \quad (14.3.138)$$

where

$$b_1 = a_1/a_0 \quad (14.3.139)$$

and

$$b_2 = a_2/a_0. \quad (14.3.140)$$

Correspondingly, the integrand takes the form

$$\text{integrand} = (a_0)^{-1/2}(1 + b_1 + b_2)^{-1/2}. \quad (14.3.141)$$

Continue the expansion process by using the binomial theorem to write

$$(1 + b_1 + b_2)^{-1/2} = 1 + c_1 + c_2 + c_3 + \dots \quad (14.3.142)$$

where

$$c_1 = -(1/2)b_1, \quad (14.3.143)$$

$$c_2 = (1/8)[3(b_1)^2 - 4b_2], \quad (14.3.144)$$

and

$$c_3 = (1/16)[(12)b_1b_2 - 5(b_1)^3], \text{ etc.} \quad (14.3.145)$$

Also observe that, using design values and deviation variables, we may write

$$t^{\text{fin}} - t^{\text{in}} = (t^{\text{find}} + \bar{t}^{\text{fin}}) - (t^{\text{ind}} + \bar{t}^{\text{in}}) = (t^{\text{find}} - t^{\text{ind}}) + (\bar{t}^{\text{fin}} - \bar{t}^{\text{in}}) \quad (14.3.146)$$

and

$$p_t^{\text{in}} = p_t^{\text{ind}} + \bar{p}_t^{\text{in}}. \quad (14.3.147)$$

We are now ready to combine the fruits of our labor. Putting everything together, we see that the relation (3.51) may be rewritten in the form

$$\begin{aligned} (t^{\text{find}} - t^{\text{ind}}) + (\bar{t}^{\text{fin}} - \bar{t}^{\text{in}}) &= \\ (p_t^{\text{ind}} + \bar{p}_t^{\text{in}}) \int_{-d/2}^{d/2} dw (a_0)^{-1/2} (1 + c_1 + c_2 + c_3 + \dots). \end{aligned} \quad (14.3.148)$$

Upon equating like powers we conclude that

$$t^{\text{find}} - t^{\text{ind}} = p_t^{\text{ind}} \int_{-d/2}^{d/2} dw (a_0)^{-1/2} \quad (14.3.149)$$

and

$$\begin{aligned} \bar{t}^{\text{fin}} - \bar{t}^{\text{in}} &= \\ p_t^{\text{ind}} \int_{-d/2}^{d/2} dw (a_0)^{-1/2} (c_1 + c_2 + c_3 + \dots) + \bar{p}_t^{\text{in}} \int_{-d/2}^{d/2} dw (a_0)^{-1/2} (1 + c_1 + c_2 + c_3 + \dots). \end{aligned} \quad (14.3.150)$$

Let us evaluate the zeroth and first order terms of our results. For the zeroth order term given by (3.149) the required integral is

$$\int_{-d/2}^{d/2} dw (a_0)^{-1/2} = \int_{-d/2}^{d/2} dw \{[(\gamma^d)^2 - 1]m^2c^2 - (qBw)^2\}^{-1/2}, \quad (14.3.151)$$

In Exercise 3.1 you will have the pleasure of verifying that, taken together, (3.149) and (3.151) reproduce (3.79).

Next look at first-order terms in the deviation variables. The first-order term arising from the first term on the right side of (3.150) involves the integral

$$\begin{aligned} \int_{-d/2}^{d/2} dw (a_0)^{-1/2} (c_1) &= \int_{-d/2}^{d/2} dw (a_0)^{-3/2} [(-1/2)a_1] = \\ \int_{-d/2}^{d/2} dw (a_0)^{-3/2} (-1/2) \{ -2\gamma^d mc^2 \bar{p}_t^{\text{in}} + 2qBw \bar{p}_x^{\text{in}} \}. \end{aligned} \quad (14.3.152)$$

Note that the function  $a_0$  is an *even* function of  $w$ . See (3.135). Therefore, in the integration from  $-d/2$  to  $+d/2$  in (3.152), any term in the curly bracket on the right side of (3.152) that is *odd* in  $w$  integrates to zero. It follows that

$$\int_{-d/2}^{d/2} dw (a_0)^{-1/2} (c_1) = (\gamma^d mc^2 \bar{p}_t^{\text{in}}) \int_{-d/2}^{d/2} dw (a_0)^{-3/2}. \quad (14.3.153)$$

We conclude that the first-order term arising from the first term on the right side of (3.150) yields the result

$$\text{result of first term} = (\bar{p}_t^{\text{ind}}) (\gamma^d mc^2 \bar{p}_t^{\text{in}}) \int_{-d/2}^{d/2} dw (a_0)^{-3/2}. \quad (14.3.154)$$

Finally, determination of the first-order term arising from the second term on the right side of (3.150) is easy. Inspection shows that it has the value

$$\text{result of second term} = (\bar{p}_t^{\text{in}}) \int_{-d/2}^{d/2} dw (a_0)^{-1/2}. \quad (14.3.155)$$

[Note that, in view of (3.149), this result is proportional to  $\{(t^{\text{find}} - t^{\text{ind}})/(\bar{p}_t^{\text{ind}})\}$ .] We see from (3.154) and (3.155) that the dependence of  $(\bar{t}^{\text{fin}} - \bar{t}^{\text{in}})$  that is first order in the deviation variables involves *only* the deviation variable  $\bar{p}_t^{\text{in}}$ .

### Summary of Linear Part of $\mathcal{M}$

We have found the linear part  $\mathcal{R}$  of the transfer map  $\mathcal{M}$  for the body of a parallel-faced bend. The relations (3.97) specify the effect of  $\mathcal{M}$  on momenta. The relations (3.126), (3.129), (3.154), and (3.155) specify the effect of  $\mathcal{R}$  on position variables. We see that the matrix  $R$  associated with  $\mathcal{R}$  has the general form

$$R = \begin{pmatrix} 1 & R_{12} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & R_{34} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & R_{56} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (14.3.156)$$

It consists of three  $2 \times 2$  blocks on the diagonal and each block is “*drift like*”. The values of  $R_{12}$  and  $R_{34}$  are given by (3.126) and (3.129), respectively. The value of  $R_{56}$  is given by (3.154) and (3.155).

### Nonlinear Part of $\mathcal{M}$

Make the factorization

$$\mathcal{M} = \mathcal{R}\mathcal{N} \quad (14.3.157)$$

so that, by definition,  $\mathcal{N}$  is the *nonlinear* part of  $\mathcal{M}$ . According to the factorization theorem, recall Section 7.6,  $\mathcal{N}$  can be written in the product form

$$\mathcal{N} = \exp(: f_3 :) \exp(: f_4 :) \cdots . \quad (14.3.158)$$

Moreover, the  $f_{>2}$  can be found from the Taylor series (symplectic jet) for  $\mathcal{M}$ . The preceding work of this subsection is sufficient to determine, in the case of the body of a parallel-faced bend, the Taylor series for  $\mathcal{M}$  through terms of degree three (and hence sufficient to specify  $f_3$  and  $f_4$ ). See (3.120) through (3.122) and (3.143) through (3.145). Moreover, complete machinery involving the binomial series [see (3.119) and (3.142)] and standard known integrals of known functions [see (3.150)] has been developed to compute for this case the Taylor series for  $\mathcal{M}$  to any desired order (and hence the associated  $f_{n>2}$  for any desired  $n > 2$ ).

#### 14.3.5 Scaled and Dimensionless Deviation Variables

### Exercises

**14.3.1.** Here is a consistency check: Verify, by performing the requisite integration, that use of (3.149) and (3.151) reproduces (3.79).

**14.4 Hard-Edge Fringe Fields****14.5 Pole Face Rotations****14.6 General Bend****14.7 Combined Function Bend**



# Chapter 15

## Taylor and Spherical and Cylindrical Harmonic Expansions

### 15.1 Introduction

Chapters 13 and 14 treated idealized beam-line elements for which variations in the field with position along the beam-line element, and fringe-field effects, were neglected. There are situations for which these neglected effects can be important when accurate modeling is desired. By developing various mathematical tools, this chapter prepares the way for Chapters 16 through 21 that describe the calculation of realistic transfer maps for straight beam-line elements, and Chapter 22 that describes the calculation of realistic transfer maps for general curved beam-line elements.

#### Restrictions Discovered by Hamilton (Symplecticity)

In previous chapters we learned that the motion of charged particles through any beam-line element can be described by the transfer map  $\mathcal{M}$  for that element. We also learned that the equations of motion for charged particle motion can be derived from a Hamiltonian, and therefore  $\mathcal{M}$  cannot be an arbitrary map, but must be a symplectic map. Consequently, through aberrations of order  $(n - 1)$ , such a map has the Lie representation

$$\mathcal{M} = \mathcal{R}_2 \exp(: f_3 :) \exp(: f_4 :) \cdots \exp(: f_n :) \quad (15.1.1)$$

where  $\mathcal{R}_2$  describes the linear part of the map. The linear map  $\mathcal{R}_2$  and the Lie generators  $f_\ell$  are determined by solving the equation of motion

$$\dot{\mathcal{M}} = \mathcal{M} : -H : \quad (15.1.2)$$

where

$$H = H_2 + H_3 + H_4 + \cdots \quad (15.1.3)$$

is the Hamiltonian expressed in terms of deviation variables and expanded in a homogeneous polynomial series. See Sections 10.1 and 10.4. The deviation variable Hamiltonian

$H$  is determined in turn by the Hamiltonian  $K$  for which some coordinate is the independent variable. For example, in Cartesian coordinates and with  $z$  taken as the independent variable,  $K$  is given by the relation

$$K = -[(p_t + q\psi)^2/c^2 - m^2c^2 - (p_x - qA_x)^2 - (p_y - qA_y)^2]^{1/2} - qA_z. \quad (15.1.4)$$

Here  $\psi$  and  $\mathbf{A}$  are the electric scalar and magnetic vector potentials, respectively. See (1.6.16). We conclude that (in the case of no electric fields,  $\psi = 0$ ) what we need are Taylor expansions for the vector potential components  $A_x$ ,  $A_y$ ,  $A_z$  in the deviation variables  $x$  and  $y$ .

### Restrictions Associated with Maxwell's Equations

For common beam-line elements the charged particles move in an evacuated beam pipe, and therefore the electric and magnetic fields controlling particle motion are *source free* in the vicinity of the beam. Correspondingly, the source-free Maxwell equations impose restrictions on what fields can be employed. They also provide, through the use of the Laplace equation and its associated mathematical machinery, powerful tools for the description of electric and magnetic fields.

This chapter begins with a discussion of harmonic functions and the use of spherical coordinates to obtain *spherical harmonic* expansions thereby leading to suitable Taylor expansions for source-free magnetic fields and their scalar and vector potentials. For our purposes this material is of particular use in terminating end fields. See Sections 16.7 and 22.8. It has other general uses as well including the treatment of ambient fields.

Next we will find, for the case of straight beam-line elements, expressions for the required Taylor expansions in terms of *on-axis gradients* with the aid of *cylindrical harmonic* expansions.<sup>1</sup> The on-axis gradients themselves are generally unspecified functions of  $z$ . In some simple cases they can be found analytically, as illustrated in Chapter 16. However, in general they must be determined numerically. Chapters 17 through 21 describe how this can be done in terms of magnetic field or magnetic potential values determined numerically at points on some regular 3-dimensional grid with the aid of some electromagnetic code.

In this part of the present chapter we will first learn how to characterize the magnetic scalar potential in terms of cylindrical harmonics described by on-axis gradients. Next we will find field expansions in terms of cylindrical harmonics. Then we will relate vector potentials to on-axis gradients. The work of this part of the chapter continues with the treatment of an analytically soluble model problem, that of the magnetic monopole doublet, which will be used in Chapter 19 to benchmark the methods to be developed in Chapters 17 through 21. Subsequently, further mathematical machinery is developed for the general representation of on-axis gradients in terms of cylindrical sources.

The chapter ends with some closing remarks meant to provide added perspective.

---

<sup>1</sup>In this chapter we will use the word *cylindrical* to mean *circular* cylindrical. In subsequent chapters we will distinguish between circular, elliptic, and rectangular cylinders.

## 15.2 Spherical Expansion

### 15.2.1 Harmonic Functions and Absolute and Expansion Coordinates

In a current-free region the magnetic field  $\mathbf{B}$  is curl free, and can therefore be described most simply in terms of a *magnetic scalar potential*  $\Psi$ . In analogy with the case of electrostatics, we write

$$\mathbf{B} = -\nabla\Psi. \quad (15.2.1)$$

Because  $\mathbf{B}$  is also divergence free,  $\Psi$  must obey the Laplace equation,

$$\nabla^2\Psi = -\nabla \cdot \mathbf{B} = 0. \quad (15.2.2)$$

Functions  $\Psi$  that obey the Laplace equation are said to be *harmonic*.<sup>2</sup>

At this point it is convenient to introduce an “absolute” coordinate  $\mathbf{R}$  and an “expansion” coordinate  $\mathbf{r}$  about some “reference/expansion” point  $\mathbf{R}_0$  by writing

$$\mathbf{R} = \mathbf{R}_0 + \mathbf{r}. \quad (15.2.3)$$

In terms of these variables we may define a related scalar potential  $\psi$  by writing

$$\psi(x, y, z; \mathbf{R}_0) = \Psi(\mathbf{R}_0 + \mathbf{r}) \quad (15.2.4)$$

where

$$\mathbf{R} = X\mathbf{e}_x + Y\mathbf{e}_y + Z\mathbf{e}_z \text{ and } \mathbf{r} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z. \quad (15.2.5)$$

Like  $\Psi$ , the related scalar potential  $\psi$  satisfies the relations

$$\mathbf{B} = -\nabla\psi, \quad (15.2.6)$$

and

$$\nabla^2\psi = 0. \quad (15.2.7)$$

Here,  $\psi$  is not to be confused with the  $\psi$  that was used in other sections to describe an electric field. Also, strictly speaking, the derivatives in (2.1) and (2.2) are to be taken with respect to the components of  $\mathbf{R}$ , and the derivatives in (2.6) and (2.7) are to be taken with respect to the components of  $\mathbf{r}$ . Moreover, in subsequent work, we will sometimes suppress the dependence of  $\psi$  on  $\mathbf{R}_0$  and simply write  $\psi(x, y, z)$ .

In summary quantities, such as  $\mathbf{B}$ ,  $\psi$ , and any vector potential  $\mathbf{A}$ , will be viewed as functions of both the expansion point  $\mathbf{R}_0$  and the expansion variable  $\mathbf{r}$ . Also we note that, since we will assume that  $\mathbf{R}_0$  is in a current free region,  $\mathbf{B}$  will be analytic in this region. Consequently,  $\Psi$  will be analytic in this region. Correspondingly,  $\mathbf{B}(\mathbf{r})$ ,  $\psi(\mathbf{r})$ , and  $\mathbf{A}(\mathbf{r})$  will be analytic about  $\mathbf{r} = 0$ . See Chapter 35 and Appendix F.

We should also say a word about units. Let  $B$  denote a quantity that has units of magnetic field (e.g. Tesla) and let  $L$  denote a quantity that has units of length. Then  $\Psi$ ,  $\psi$ , and  $\mathbf{A}$  have units  $BL$ .

---

<sup>2</sup>Note the *minus* sign in (2.1), and (2.6) which follows. Many authors use a *plus* sign instead. Either choice is arbitrary. A virtue of the plus sign choice is that many subsequent formulas are neater in that they are also minus sign free. We have made the minus sign choice because we wish to make magnetostatics resemble electrostatics as much as possible in order to exploit the intuitions we frequently already have about electrostatic concepts and relations.

### 15.2.2 Spherical and Cylindrical Coordinates

Introduce spherical coordinates  $r, \theta, \phi$  by the usual rules

$$r^2 = x^2 + y^2 + z^2, \quad (15.2.8)$$

$$x = r \sin(\theta) \cos(\phi), \quad (15.2.9)$$

$$y = r \sin(\theta) \sin(\phi), \quad (15.2.10)$$

$$z = r \cos \theta. \quad (15.2.11)$$

Also, for future use, introduce cylindrical coordinates  $\rho, \phi$ , and  $z$  by the usual rules

$$\rho^2 = x^2 + y^2, \quad (15.2.12)$$

$$x = \rho \cos \phi, \quad (15.2.13)$$

$$y = \rho \sin \phi. \quad (15.2.14)$$

Note these two coordinate systems have the coordinate  $\phi$  in common. In both cases there is the  $\phi$ -defining pair of relations

$$\sin \phi = y / \sqrt{x^2 + y^2}, \quad (15.2.15)$$

$$\cos \phi = x / \sqrt{x^2 + y^2}. \quad (15.2.16)$$

The other coordinates are related by (2.11) and the equations

$$r^2 = \rho^2 + z^2, \quad (15.2.17)$$

$$\rho = r \sin \theta. \quad (15.2.18)$$

We also record results for the orthonormal triads  $\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi$  and  $\mathbf{e}_\rho, \mathbf{e}_\phi, \mathbf{e}_z$ , and their relation to  $\mathbf{r}$ . For the spherical orthonormal triad there are the results

$$\begin{aligned} \mathbf{e}_r &= \sin(\theta) \cos(\phi) \mathbf{e}_x + \sin(\theta) \sin(\phi) \mathbf{e}_y + \cos(\theta) \mathbf{e}_z \\ &= (1/r)(x \mathbf{e}_x + y \mathbf{e}_y + z \mathbf{e}_z) = \mathbf{r}/r, \\ \mathbf{e}_\theta &= \cos(\theta) \cos(\phi) \mathbf{e}_x + \cos(\theta) \sin(\phi) \mathbf{e}_y - \sin(\theta) \mathbf{e}_z, \\ \mathbf{e}_\phi &= -\sin(\phi) \mathbf{e}_x + \cos(\phi) \mathbf{e}_y, \\ \mathbf{r} &= r \mathbf{e}_r. \end{aligned} \quad (15.2.19)$$

For the cylindrical orthonormal triad there are the results

$$\begin{aligned} \mathbf{e}_\rho &= \cos \phi \mathbf{e}_x + \sin \phi \mathbf{e}_y = (1/\rho)(x \mathbf{e}_x + y \mathbf{e}_y), \\ \mathbf{e}_\phi &= -\sin \phi \mathbf{e}_x + \cos \phi \mathbf{e}_y = (1/\rho)(-y \mathbf{e}_x + x \mathbf{e}_y), \\ \mathbf{r} &= \rho \mathbf{e}_\rho + z \mathbf{e}_z. \end{aligned} \quad (15.2.20)$$

See Exercises 2.1 and 2.2.

Finally, with regard to the relation between Cartesian and cylindrical coordinates, if one defines Cartesian and cylindrical components for any vector  $\mathbf{A}$  by writing

$$\mathbf{A} = A_x \mathbf{e}_x + A_y \mathbf{e}_y + A_z \mathbf{e}_z = A_\rho \mathbf{e}_\rho + A_\phi \mathbf{e}_\phi + A_z \mathbf{e}_z, \quad (15.2.21)$$

then there are the component relations

$$A_\rho = \mathbf{e}_\rho \cdot \mathbf{A} = \cos \phi A_x + \sin \phi A_y, \quad (15.2.22)$$

$$A_\phi = \mathbf{e}_\phi \cdot \mathbf{A} = -\sin \phi A_x + \cos \phi A_y, \quad (15.2.23)$$

and their inverses

$$A_x = \mathbf{e}_x \cdot \mathbf{A} = \cos \phi A_\rho - \sin \phi A_\phi, \quad (15.2.24)$$

$$A_y = \mathbf{e}_y \cdot \mathbf{A} = \sin \phi A_\rho + \cos \phi A_\phi. \quad (15.2.25)$$

### 15.2.3 Harmonic Polynomials, Harmonic Polynomial Expansions, and General Spherical Polynomials

Polynomials in  $x$ ,  $y$ , and  $z$  that are harmonic are called *harmonic polynomials* or *solid harmonics*. In *complex* form these polynomials, call them  $H_\ell^m$ , can be defined in terms of the spherical harmonics  $Y_\ell^m(\theta, \phi)$  and the associated Legendre functions  $P_\ell^m$  by the rule

$$\begin{aligned} H_\ell^m(\mathbf{r}) &= r^\ell Y_\ell^m(\theta, \phi) \\ &= r^\ell \{[(2\ell+1)(\ell-m)!]/[4\pi(\ell+m)!]\}^{1/2} P_\ell^m(\cos \theta) \exp(im\phi) \\ &\quad \text{with } -\ell \leq m \leq \ell. \end{aligned} \quad (15.2.26)$$

The  $H_\ell^m$  are homogeneous polynomials of degree  $\ell$  in the variables  $x$ ,  $y$ , and  $z$ .<sup>3</sup> For example, there are the definitions

$$H_0^0(\mathbf{r}) = 1/\sqrt{4\pi}; \quad (15.2.27)$$

$$\begin{aligned} H_1^1(\mathbf{r}) &= \sqrt{3/(4\pi)}(-1/\sqrt{2})(x+iy) = -\sqrt{3/(8\pi)}(x+iy), \\ H_1^0(\mathbf{r}) &= \sqrt{3/(4\pi)}z, \\ H_1^{-1}(\mathbf{r}) &= \sqrt{3/(4\pi)}(1/\sqrt{2})(x-iy) = \sqrt{3/(8\pi)}(x-iy); \end{aligned} \quad (15.2.28)$$

$$\begin{aligned} H_2^2(\mathbf{r}) &= \sqrt{15/(32\pi)}(x+iy)^2, \\ H_2^1(\mathbf{r}) &= -\sqrt{15/(8\pi)}(x+iy)z, \\ H_2^0(\mathbf{r}) &= \sqrt{5/(16\pi)}(2z^2 - x^2 - y^2), \\ H_2^{-1}(\mathbf{r}) &= \sqrt{15/(8\pi)}(x-iy)z, \\ H_2^{-2}(\mathbf{r}) &= \sqrt{15/(32\pi)}(x-iy)^2. \end{aligned} \quad (15.2.29)$$

---

<sup>3</sup>Note that although we initially work with spherical coordinates, the final result is in the form of power series in *Cartesian* coordinates.

See Appendix U. Because they are defined in terms of the  $Y_\ell^m$  and powers of  $r$ , the harmonic polynomials  $H_\ell^m$  have well-defined properties under the action of the rotation group  $SO(3)$ .<sup>4</sup>

From potential theory we know that any harmonic function analytic at the origin  $\mathbf{r} = 0$  can be expanded in harmonic polynomials. Thus, under the assumption that a harmonic function  $\psi$  is analytic at the origin, it has the expansion

$$\psi(x, y, z) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} g_{\ell m} H_\ell^m(\mathbf{r}) \quad (15.2.30)$$

where the coefficients  $g_{\ell m}$  are arbitrary.

We can also define *real* versions of harmonic polynomials, call them  $H_\ell^{m,\alpha}$  where  $m \geq 0$  and  $\alpha = c$  or  $s$ , by writing

$$\begin{aligned} H_\ell^{m,c}(x, y, z) &= \{[(2\ell + 1)(\ell - m)!]/[4\pi(l + m)!]\}^{1/2} r^\ell P_\ell^m(\cos \theta) \cos(m\phi) \\ \text{with } \ell &= 0, 1, \dots, \infty \text{ and } m = 0, 1, \dots, \ell \end{aligned} \quad (15.2.31)$$

and

$$\begin{aligned} H_\ell^{m,s}(x, y, z) &= \{[(2\ell + 1)(\ell - m)!]/[4\pi(l + m)!]\}^{1/2} r^\ell P_\ell^m(\cos \theta) \sin(m\phi) \\ \text{with } \ell &= 1, 2, \dots, \infty \text{ and } m = 1, 2, \dots, \ell. \end{aligned} \quad (15.2.32)$$

These definitions yield, for example, the results

$$H_0^{0,c} = 1/\sqrt{4\pi}; \quad (15.2.33)$$

$$\begin{aligned} H_1^{1,c} &= -[3/(8\pi)]^{1/2} x, \\ H_1^{0,c} &= [3/(4\pi)]^{1/2} z, \\ H_1^{1,s} &= -[3/(8\pi)]^{1/2} y; \end{aligned} \quad (15.2.34)$$

$$\begin{aligned} H_2^{2,c} &= (1/4)[15/(2\pi)]^{1/2}(x^2 - y^2), \\ H_2^{1,c} &= -[15/(8\pi)]^{1/2} xz, \\ H_2^{0,c} &= (1/2)[5/(4\pi)]^{1/2}(2z^2 - x^2 - y^2), \\ H_2^{2,s} &= (1/2)[15/(2\pi)]^{1/2} xy, \\ H_2^{1,s} &= -[15/(8\pi)]^{1/2} yz; \end{aligned} \quad (15.2.35)$$

$$\begin{aligned} H_3^{3,c} &= -(1/4)[35/(4\pi)]^{1/2}(x^3 - 3xy^2), \\ H_3^{2,c} &= (1/4)[105/(2\pi)]^{1/2}[z(x^2 - y^2)], \\ H_3^{1,c} &= -(1/4)[21/(2\pi)]^{1/2}[x(4z^2 - x^2 - y^2)], \\ H_3^{0,c} &= (1/2)[7/(4\pi)]^{1/2}[2z^3 - 3z(x^2 + y^2)], \\ H_3^{3,s} &= (1/4)[35/(4\pi)]^{1/2}(y^3 - 3x^2 y), \\ H_3^{2,s} &= (1/2)[105/(2\pi)]^{1/2}(xyz), \\ H_3^{1,s} &= -(1/4)[21/(4\pi)]^{1/2}[y(4z^2 - x^2 - y^2)]. \end{aligned} \quad (15.2.36)$$

---

<sup>4</sup>We note that the spherical harmonics  $Y_\ell^m$  could better be called *surface* harmonics. Then we could use the name *spherical harmonics* to refer to the functions  $H_\ell^m$ .

In terms of these polynomials any harmonic function  $\psi$  analytic at the origin has an expansion of the form

$$\psi(x, y, z) = \sum_{\ell=0}^{\infty} \sum_{m=0}^{\ell} g_{\ell,m,c} H_{\ell}^{m,c}(\mathbf{r}) + \sum_{\ell=1}^{\infty} \sum_{m=1}^{\ell} g_{\ell,m,s} H_{\ell}^{m,s}(\mathbf{r}), \quad (15.2.37)$$

where the coefficients  $g_{\ell,m,\alpha}$  are arbitrary.

It is also convenient to define general *spherical polynomials*  $S_{n\ell}^m(\mathbf{r})$  in terms of the spherical harmonics and powers of  $r$  by making the definition

$$S_{n\ell}^m(\mathbf{r}) = r^n Y_{\ell}^m(\theta, \phi) \text{ provided } n \text{ is of the form } n = \ell + 2k. \quad (15.2.38)$$

See Subsection U.2.4. The  $S_{n\ell}^m(\mathbf{r})$  are evidently of degree  $n$ . They form a basis for the space of *all* functions that are analytic at the origin, and harmonic polynomials comprise special cases for which  $\ell = n$ ,

$$H_n^m(\mathbf{r}) = S_{nn}^m(\mathbf{r}). \quad (15.2.39)$$

Therefore, we may also write (2.30) in the form

$$\psi(x, y, z) = \sum_{n=0}^{\infty} \sum_{m=-n}^n g_{nm} S_{nn}^m(\mathbf{r}). \quad (15.2.40)$$

Finally, because they are constructed from the  $Y_{\ell}^m$  and powers of  $r$ , the  $S_{n\ell}^m(\mathbf{r})$  also have well-defined properties under the action of  $SO(3)$ .

### 15.2.4 Spherical Polynomial Vector Fields

We have been working with *scalar* fields such as  $\psi(\mathbf{r})$ . Just as it is possible to construct *scalar* polynomial fields  $S_{n\ell}^m(\mathbf{r})$  having well-defined properties under the action of  $SO(3)$ , it is also possible to construct polynomial *vector* fields that have well-defined properties under the action of  $SO(3)$ . We call such polynomial vector fields *spherical polynomial* vector fields and denote them by the symbols  $\mathbf{S}_{n\ell J}^M(\mathbf{r})$ . These are vector fields whose components are homogenous polynomials of degree  $n$  in the components of  $\mathbf{r}$ . See Subsections U.3.2 and U.3.3 for their definition and some examples.

Any vector field analytic at the origin can be expanded in terms of spherical polynomial vector fields. In particular, both the magnetic field  $\mathbf{B}(\mathbf{r})$  and any associated vector potential  $\mathbf{A}(\mathbf{r})$  can be expanded in terms of spherical polynomial vector fields.

Because of their properties under the action of  $SO(3)$ , there are well-organized relations between spherical polynomials and spherical polynomial vector fields. For example, there is the relation

$$\nabla H_n^m(\mathbf{r}) = \nabla S_{nn}^m(\mathbf{r}) = \sqrt{n(2n+1)} \mathbf{S}_{n-1,n-1,n}^m(\mathbf{r}). \quad (15.2.41)$$

See (U.5.6). Upon combining (2.40) and (2.41) we see that a general field  $\mathbf{B}$  that is source free in some vicinity of  $\mathbf{R}_0$  has a spherical polynomial vector field expansion of the form

$$\mathbf{B}(\mathbf{r}) = -\nabla\psi = -\sum_{n=0}^{\infty} \sum_{m=-n}^n g_{nm} \sqrt{n(2n+1)} \mathbf{S}_{n-1,n-1,n}^m(\mathbf{r}). \quad (15.2.42)$$

Note that the  $n = 0$  term does not actually contribute, as expected.

### 15.2.5 Determination of Minimum Vector Potential: the Poincaré-Coulomb Gauge

Suppose we are given a magnetic field, specified either by a scalar potential presented in the forms (2.30) or (2.40), or equivalently by a spherical polynomial vector field expansion of the form (2.42). And suppose, in order to treat charged-particle motion in this field using a Hamiltonian formulation, we wish to find an associated vector potential. We know that in principle there are many such vector potentials, all of which are related by gauge transformations. Given  $\mathbf{B}(\mathbf{r})$  in some region, the goal of this subsection is to find an associated vector potential  $\mathbf{A}^{\min}(\mathbf{r})$  that is, at least *locally*, as *minimal*/small as possible in the sense that  $\mathbf{A}^{\min}(\mathbf{r})$  is small if  $\mathbf{B}(\mathbf{r})$  is small. The reason for this goal is that, according to (1.5.30), mechanical and canonical momenta differ by the vector potential; and there are situations where we would like this difference to be as small as possible. See Sections 16.7 and 22.8.

Conceptually, our plan is as follows: Make Taylor expansions, with initially unknown coefficients, for the Cartesian components of  $\mathbf{A}^{\min}(\mathbf{r})$ , organize these expansions into homogeneous polynomials, and then further organize them as spherical polynomial vector fields. Then use this representation to compute and organize  $\nabla \times \mathbf{A}^{\min}$  in terms of spherical polynomial vector fields. At the same time parameterize  $\mathbf{B}(\mathbf{r})$  in terms of a scalar potential  $\psi$  expanded in harmonic polynomials. Finally, compare the two expansions for  $\mathbf{B}(\mathbf{r})$  given by  $\mathbf{B} = -\nabla\psi$  and  $\mathbf{B} = \nabla \times \mathbf{A}^{\min}$ , equate coefficients of like terms, and thereby determine the coefficients in the Taylor expansion for the components of  $\mathbf{A}^{\min}$  in terms of the coefficients in the expansion for  $\psi$ . Do all this while keeping the minimal/small goal in mind. For the notation and machinery required for the execution of this plan, again see Appendix U. What lies ahead may seem complicated, but the final result will prove to be remarkably simple.

#### Construction of Minimum Vector Potential

We begin with the harmonic polynomial expansion (2.40) for the scalar potential  $\psi$ , which we rewrite in the form

$$\psi(\mathbf{r}) = \sum_{n=1}^{n_{\max}} \sum_m g_{nm} S_{nn}^m(\mathbf{r}). \quad (15.2.43)$$

Here we assume an expansion through terms of degree  $n_{\max}$ , and omit  $n = 0$  terms since constant terms make no contribution to  $\mathbf{B}$  as given by (2.42).

For the associated vector potential  $\mathbf{A}^{\min}$  we make the spherical polynomial vector field expansion

$$\mathbf{A}^{\min}(\mathbf{r}) = \sum_{n=1}^{n_{\max}} \sum_{\ell} \sum_J \sum_M f_{n\ell JM} \mathbf{S}_{n\ell J}^M(\mathbf{r}). \quad (15.2.44)$$

Again see Appendix U. Given the coefficients  $g_{nm}$ , our task is to use the equality

$$\nabla \times \mathbf{A}^{\min}(\mathbf{r}) = \nabla \times \sum_{n=1}^{n_{\max}} \sum_{\ell} \sum_J \sum_M f_{n\ell JM} \mathbf{S}_{n\ell J}^M(\mathbf{r}) = -\nabla \sum_{n=1}^{n_{\max}} \sum_m g_{nm} S_{nn}^m(\mathbf{r}) = -\nabla\psi(\mathbf{r}) \quad (15.2.45)$$

to find the coefficients  $f_{n\ell JM}$  in terms of the  $g_{nm}$ .

We already know the result of evaluating the right side of (2.45). Use of (2.42) gives the result

$$\mathbf{B}(\mathbf{r}) = -\nabla\psi(\mathbf{r}) = -\nabla \sum_{n=1}^{n_{\max}} \sum_m g_{nm} S_{nn}^m(\mathbf{r}) = -\sum_{n=1}^{n_{\max}} \sum_m g_{nm} \sqrt{n(2n+1)} \mathbf{S}_{n-1,n-1,n}^m(\mathbf{r}). \quad (15.2.46)$$

Next work on evaluating the left side of (2.45). This is a more complicated task. In accord with the range rules (U.3.7) and (U.3.8), we decompose the expansion into the sum of four pieces with each containing a particular kind of term:

- a) All terms for which  $\ell = 0$  and hence  $J = 1$ . Also, therefore,  $n = 2k$  with  $k > 0$ . The associated spherical polynomial vectors are of the form  $\mathbf{S}_{2k,0,1}^M(\mathbf{r})$ .
- b) All terms for which  $\ell > 0$  and  $J = \ell + 1$ . The associated spherical polynomial vectors are of the form  $\mathbf{S}_{n,\ell,\ell+1}^M(\mathbf{r})$ .
- c) All terms for which  $\ell > 0$  and  $J = \ell$ . The associated spherical polynomial vectors are of the form  $\mathbf{S}_{n,\ell,\ell}^M(\mathbf{r})$ .
- d) All terms for which  $\ell > 0$  and  $J = \ell - 1$ . The associated spherical polynomial vectors are of the form  $\mathbf{S}_{n,\ell,\ell-1}^M(\mathbf{r})$ .

Thus, we write

$$\mathbf{A}^{\min} = \mathbf{A}^{\min a} + \mathbf{A}^{\min b} + \mathbf{A}^{\min c} + \mathbf{A}^{\min d} \quad (15.2.47)$$

where

$$\mathbf{A}^{\min a}(\mathbf{r}) = \sum_{k=1}^{k_{\max}} \sum_M f_{2k,0,1,M} \mathbf{S}_{2k,0,1}^M(\mathbf{r}), \quad (15.2.48)$$

$$\mathbf{A}^{\min b}(\mathbf{r}) = \sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M f_{n,\ell,\ell+1,M} \mathbf{S}_{n,\ell,\ell+1}^M(\mathbf{r}), \quad (15.2.49)$$

$$\mathbf{A}^{\min c}(\mathbf{r}) = \sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M f_{n,\ell,\ell,M} \mathbf{S}_{n,\ell,\ell}^M(\mathbf{r}), \quad (15.2.50)$$

$$\mathbf{A}^{\min d}(\mathbf{r}) = \sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M f_{n,\ell,\ell-1,M} \mathbf{S}_{n,\ell,\ell-1}^M(\mathbf{r}). \quad (15.2.51)$$

We are ready to proceed. For the  $\mathbf{A}^{\min a}$  term we find, using (U.5.20), the result

$$\nabla \times \mathbf{A}^{\min a}(\mathbf{r}) = \nabla \times \sum_{k=1}^{k_{\max}} \sum_M f_{2k,0,1,M} \mathbf{S}_{2k,0,1}^M(\mathbf{r}) = \sum_{k=1}^{k_{\max}} \sum_M f_{2k,0,1,M} [i(\sqrt{2/3})(2k)] \mathbf{S}_{2k-1,1,1}^M(\mathbf{r}). \quad (15.2.52)$$

For the  $\mathbf{A}^{\min b}$  term we find, using (U.5.17), the result

$$\begin{aligned}\nabla \times \mathbf{A}^{\min b}(\mathbf{r}) &= \nabla \times \sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M f_{n,\ell,\ell+1,M} \mathbf{S}_{n,\ell,\ell+1}^M(\mathbf{r}) = \\ &\sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M f_{n,\ell,\ell+1,M} [i\sqrt{(\ell+2)/(2\ell+3)}(n-\ell)] \mathbf{S}_{n-1,\ell+1,\ell+1}^M(\mathbf{r}).\end{aligned}\quad (15.2.53)$$

For the  $\mathbf{A}^{\min c}$  term we find, using (U.5.18), the result

$$\begin{aligned}\nabla \times \mathbf{A}^{\min c}(\mathbf{r}) &= \nabla \times \sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M f_{n,\ell,\ell,M} \mathbf{S}_{n,\ell,\ell}^M(\mathbf{r}) = \\ &\sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M f_{n,\ell,\ell,M} [i\sqrt{(\ell+1)/(2\ell+1)}(n+\ell+1)] \mathbf{S}_{n-1,\ell-1,\ell}^M(\mathbf{r}) \\ &+ \sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M f_{n,\ell,\ell,M} [i\sqrt{\ell/(2\ell+1)}(n-\ell)] \mathbf{S}_{n-1,\ell+1,\ell}^M(\mathbf{r}).\end{aligned}\quad (15.2.54)$$

Finally, for the  $\mathbf{A}^{\min d}$  term we find, using (U.5.19), the result

$$\begin{aligned}\nabla \times \mathbf{A}^{\min d}(\mathbf{r}) &= \nabla \times \sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M f_{n,\ell,\ell-1,M} \mathbf{S}_{n,\ell,\ell-1}^M(\mathbf{r}) = \\ &\sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M f_{n,\ell,\ell-1,M} [i\sqrt{(\ell-1)/(2\ell-1)}(n+\ell+1)] \mathbf{S}_{n-1,\ell-1,\ell-1}^M(\mathbf{r}).\end{aligned}\quad (15.2.55)$$

We can now equate coefficients of like terms. Let us begin with the first few corresponding to small values of  $n$ . The first of these, corresponding to  $n = 0$ , is  $\mathbf{S}_{0,0,1}^M$ . For the right side of (2.45) we see from (2.46) that

$$\text{coefficient of } \mathbf{S}_{0,0,1}^M \text{ in } -\nabla\psi = -\sqrt{3} g_{1M}. \quad (15.2.56)$$

Next, for the left side, examine the terms in  $\nabla \times \mathbf{A}^{\min}$ : From (2.52) we see that there are no terms of the desired kind, namely terms involving  $\mathbf{S}_{0,0,1}^M$ , in  $\nabla \times \mathbf{A}^{\min a}$ . From (2.53) we see that there are no terms of the desired kind in  $\nabla \times \mathbf{A}^{\min b}$ . From (2.54) we see that there are terms of the desired kind in  $\nabla \times \mathbf{A}^{\min c}$ , and find the result

$$\text{coefficient of } \mathbf{S}_{0,0,1}^M \text{ in } \nabla \times \mathbf{A}^{\min c} = i\sqrt{6} f_{1,1,1,M}. \quad (15.2.57)$$

Finally, from (2.55) we see that there are no terms of the desired kind in  $\nabla \times \mathbf{A}^{\min d}$ .

Upon comparing (2.56) and (2.57) we conclude that there must be the relation

$$i\sqrt{6} f_{1,1,1,M} = -\sqrt{3} g_{1M}, \quad (15.2.58)$$

and therefore

$$f_{1,1,1,M} = i\sqrt{1/2} g_{1M}. \quad (15.2.59)$$

Note that this relation is consistent with (U.6.39). Moreover, we conclude that the six remaining  $n = 1$  coefficients in  $\mathbf{A}^{\min}$ , namely  $f_{1,1,0,0}$  and the  $f_{1,1,2,M}$ , can be anything since there are the relations (U.6.38) and (U.6.40). In pursuit of our minimal/small goal, we set these coefficients to zero. Then, so far, we have the result

$$\mathbf{A}^{\min}(\mathbf{r}) = \sum_M (i) \sqrt{1/2} g_{1M} \mathbf{S}_{111}^M(\mathbf{r}) + \text{terms of degree } > 1. \quad (15.2.60)$$

In terms of Cartesian components, (2.60) yields the relation

$$\mathbf{A}^{\min}(\mathbf{r}) = -(1/2) \mathbf{r} \times \mathbf{B}(0) + \text{terms of degree } > 1. \quad (15.2.61)$$

Here we have used (2.42), (2.43), and (U.6.25) evaluated for  $n = 1$ .

Let us push on to the case  $n = 2$ ; in which case there are the spherical polynomial vector fields  $\mathbf{S}_{110}^0$ ,  $\mathbf{S}_{111}^M$  with  $-1 \leq M \leq 1$ , and  $\mathbf{S}_{112}^M$  with  $-2 \leq M \leq 2$ . First see where/how they occur in  $-\nabla\psi$ . Examination of (2.46) shows that the only such term in  $-\nabla\psi$  is  $\mathbf{S}_{112}^M$ , and we have the relation

$$\text{coefficient of } \mathbf{S}_{1,1,2}^M \text{ in } -\nabla\psi = -\sqrt{10} g_{2M}. \quad (15.2.62)$$

We next examine the terms in  $\nabla \times \mathbf{A}^{\min}$ : From (2.52) we see that there are no terms of the desired kind, namely terms involving  $\mathbf{S}_{1,1,2}^M$ , in  $\nabla \times \mathbf{A}^{\min a}$ . From (2.53) we see that there are no terms of the desired kind in  $\nabla \times \mathbf{A}^{\min b}$ . From (2.54) we see that there are terms of the desired kind in  $\nabla \times \mathbf{A}^{\min c}$ , and find the relation

$$\text{coefficient of } \mathbf{S}_{1,1,2}^M \text{ in } \nabla \times \mathbf{A}^{\min c} = i\sqrt{15} f_{2,2,2,M}. \quad (15.2.63)$$

Finally, from (2.55) we see that there are no terms of the desired kind in  $\nabla \times \mathbf{A}^{\min d}$ .

Upon comparing (2.62) and (2.63) we conclude that there must be the relation

$$i\sqrt{15} f_{2,2,2,M} = -\sqrt{10} g_{2M}, \quad (15.2.64)$$

and therefore

$$f_{2,2,2,M} = i\sqrt{2/3} g_{2M}. \quad (15.2.65)$$

What can be said about the thirteen remaining  $n = 2$  coefficients in  $\mathbf{A}^{\min}$ , namely the coefficients  $f_{201M}$ ,  $f_{2,2,3,M}$ , and  $f_{2,2,1,M}$ ? It can be shown that  $\nabla \times \mathbf{S}_{223}^M(\mathbf{r}) = 0$ , and therefore the terms with coefficients  $f_{2,2,3,M}$  make no contribution to  $\mathbf{B}(\mathbf{r})$ . See Exercise (U.6.21). In further pursuit of our minimal/small goal, we set these coefficients to zero. It can be shown that terms with the coefficients  $f_{201M}$  and  $f_{2,2,1,M}$  produce terms in  $\mathbf{B}(\mathbf{r})$  having nonzero curl. Again see Exercise (U.6.21). We also set these coefficients to zero to ensure that  $\mathbf{B}(\mathbf{r})$  is curl free. Putting everything together we have learned so far yields the result

$$\mathbf{A}^{\min}(\mathbf{r}) = \sum_M (i) \sqrt{1/2} g_{1M} \mathbf{S}_{111}^M(\mathbf{r}) + \sum_M (i) \sqrt{2/3} g_{2M} \mathbf{S}_{222}^M(\mathbf{r}) + \text{terms of degree } > 2. \quad (15.2.66)$$

The pattern should now be clear. There are the general relations

$$\nabla S_{nn}^M(\mathbf{r}) = \sqrt{n(2n+1)} \mathbf{S}_{n-1,n-1,n}^M(\mathbf{r}) \quad (15.2.67)$$

and

$$\nabla \times \mathbf{S}_{n,n,n}^M(\mathbf{r}) = i\sqrt{(n+1)(2n+1)} \mathbf{S}_{n-1,n-1,n}^M(\mathbf{r}). \quad (15.2.68)$$

Therefore there is the general relation

$$\mathbf{A}^{\min}(\mathbf{r}) = \sum_{n=1}^{n_{\max}} \sum_{M=-n}^n (i) \sqrt{n/(n+1)} g_{nM} \mathbf{S}_{nnn}^M(\mathbf{r}). \quad (15.2.69)$$

We have found a formula for the vector potential  $\mathbf{A}^{\min}(\mathbf{r})$  in terms of the harmonic expansion coefficients for the scalar potential  $\psi(\mathbf{r})$ .

### Properties of Minimum Vector Potential

It can be verified that this particular choice of  $\mathbf{A}^{\min}(\mathbf{r})$  has the property

$$\nabla \cdot \mathbf{A}^{\min}(\mathbf{r}) = 0. \quad (15.2.70)$$

See (U.5.11). Therefore  $\mathbf{A}^{\min}(\mathbf{r})$  is in a *Coulomb/solenoidal* gauge.<sup>5</sup> It also has the property

$$\mathbf{r} \cdot \mathbf{A}^{\min}(\mathbf{r}) = 0. \quad (15.2.71)$$

See (U.6.9). This is the condition that  $\mathbf{A}^{\min}(\mathbf{r})$  be in what is called a *Poincaré* gauge.<sup>6</sup> Taken together, we will say that a vector potential that satisfies both (2.70) and (2.71) is in the *Poincaré-Coulomb* gauge.<sup>7</sup> At this point we observe that any vector potential  $\mathbf{A}(\mathbf{r})$  that obeys the Poincaré gauge condition

$$\mathbf{r} \cdot \mathbf{A}(\mathbf{r}) = 0 \quad (15.2.72)$$

must vanish at the origin,

$$\mathbf{A}(0) = 0. \quad (15.2.73)$$

That is, it has no constant part. Note that, according to (2.3),  $\mathbf{r} = 0$  corresponds to the expansion point  $\mathbf{R} = \mathbf{R}_0$ , and we again assume analyticity so that  $\mathbf{A}(\mathbf{r})$  is analytic at the origin.

To see that  $\mathbf{A}(\mathbf{r})$  has no constant part, expand it in homogeneous polynomials of degree  $n$  by writing

$$\mathbf{A}(\mathbf{r}) = \sum_{n=0}^{\infty} \mathbf{A}^n(\mathbf{r}). \quad (15.2.74)$$

---

<sup>5</sup>Fields that are divergence free are also called *solenoidal*.

<sup>6</sup>A Poincaré gauge is also sometimes called a *multipolar* gauge.

<sup>7</sup>There is a certain “symmetry” here: It is interesting to observe that the two conditions (2.70) and (2.71) have related “Fourier analogs”. In Fourier space the condition (2.70) becomes  $\mathbf{k} \cdot \tilde{\mathbf{A}}^{\min}(\mathbf{k}) = 0$  and the condition (2.71) becomes  $\nabla_{\mathbf{k}} \cdot \tilde{\mathbf{A}}^{\min}(\mathbf{k}) = 0$ . Here a tilde denotes a Fourier transform. See Exercise 22.2.23. Put another way, both  $\mathbf{r}$  and  $\nabla$  may be viewed as vectors that are the Fourier analogs of each other, and (2.70) and (2.71) state that  $\mathbf{A}^{\min}(\mathbf{r})$  is “orthogonal” to both. Note also that  $\nabla \cdot \mathbf{r}$  has the scalar value 3,  $\nabla \cdot \mathbf{r} = 3$ . And, according to Euler’s relation [see (1.5.90)],  $\mathbf{r} \cdot \nabla$  is a “scalar” operator that extracts from any homogeneous three-variable entity its degree, which itself is a scalar.

Combining (2.72) and (2.74) gives the relation

$$\mathbf{r} \cdot \mathbf{A}(\mathbf{r}) = \sum_{n=0}^{\infty} \mathbf{r} \cdot \mathbf{A}^n(\mathbf{r}) = 0, \quad (15.2.75)$$

from which it follows, upon equating terms of like degree, that

$$\mathbf{r} \cdot \mathbf{A}^n(\mathbf{r}) = 0 \quad (15.2.76)$$

for all  $n$ . In particular, there is the result

$$\mathbf{r} \cdot \mathbf{A}^0(\mathbf{r}) = \mathbf{r} \cdot \mathbf{A}(0) = 0 \quad (15.2.77)$$

for all  $\mathbf{r}$ , from which (2.73) follows.

### Evaluation of Work

Have we achieved our goal of finding a *minimal* vector potential? We have, in the following sense: Inspection of (2.46) shows that it provides an expansion of  $\mathbf{B}(\mathbf{r})$  in terms of spherical polynomial vector fields  $\mathbf{S}_{n-1,n-1,n}^m(\mathbf{r})$  with expansion coefficients proportional to the  $g_{nm}$ . Inspection of (2.69) shows that it provides an expansion of  $\mathbf{A}^{\min}(\mathbf{r})$  in terms of spherical polynomial vector fields  $\mathbf{S}_{nnn}^M(\mathbf{r})$  with expansion coefficients again proportional to the  $g_{nM}$ . As shown above, the vector potential  $\mathbf{A}^{\min}(\mathbf{r})$  has no constant part. We also see that its non-constant parts are directly proportional to the coefficients  $g_{nm}$  that describe the constant and non-constant parts of  $\mathbf{B}(\mathbf{r})$ . Moreover, there is an order-by-order relation. Terms of order  $n$  in  $\mathbf{A}^{\min}(\mathbf{r})$  are proportional to terms of order  $n-1$  in  $\mathbf{B}(\mathbf{r})$ . Thus,  $\mathbf{A}^{\min}(\mathbf{r})$  is small if  $\mathbf{B}(\mathbf{r})$  is small. In particular, if high-order terms in  $\mathbf{B}(\mathbf{r})$  are negligible, they will also be negligible in  $\mathbf{A}^{\min}(\mathbf{r})$ .

There is yet another sense in which the vector potential we have found is minimal. Suppose, for example, that we confine our attention to the case of a vector potential that is homogeneous of degree 1, which is the case we need to produce a constant magnetic field. When  $n = 1$  we see from Table U.3.1 that  $\ell = 1$  and  $J = 0, 1, 2$ . Therefore, such a vector potential, call it  $\mathbf{A}^1$ , can be written in the form

$$\mathbf{A}^1(\mathbf{r}) = \sum_J \sum_M f_{11JM} \mathbf{S}_{11J}^M(\mathbf{r}). \quad (15.2.78)$$

Recall (2.44). Let us compute  $\|\mathbf{A}^1(\mathbf{r})\|$ , the *norm* of  $\mathbf{A}^1$ , as defined by the rule

$$\|\mathbf{A}^1(\mathbf{r})\|^2 = < \mathbf{A}^1(\mathbf{r}), \mathbf{A}^1(\mathbf{r}) >. \quad (15.2.79)$$

See Subsection U.4.2. Then, in accord with the work of that subsection, we know there is the result

$$\|\mathbf{A}^1(\mathbf{r})\|^2 = \sum_J \sum_M |f_{11JM}|^2. \quad (15.2.80)$$

See (U.4.21). Also we know the value of  $f_{111M}$  is fixed by (2.59), and we have chosen to set the remaining  $f_{11JM}$  to zero. We now see, since (2.80) is a sum of squares, that doing so *minimizes*  $\|\mathbf{A}^1(\mathbf{r})\|$ . Similar computations may be made for other values of  $n$ . The result is that the choice we have made for  $\mathbf{A}^{\min}$  minimizes  $\|\mathbf{A}^n(\mathbf{r})\|$  for each value of  $n$ .

For yet more explorations of senses in which the vector potential we have found is minimal, see Exercises 2.7 through 2.9 at the end of this section.

## A Further Simplification

We close this subsection by observing that the relation (2.69) can be further manipulated using (U.6.25). Doing so gives the pleasing result

$$\begin{aligned}
 \mathbf{A}^{\min}(\mathbf{r}) &= \sum_{n=1}^{n_{\max}} \sum_{M=-n}^n (i) \sqrt{n/(n+1)} g_{nM} S_{nn}^M(\mathbf{r}) \\
 &= \sum_{n=1}^{n_{\max}} \sum_{M=-n}^n (i) \sqrt{n/(n+1)} g_{nM} [-i/\sqrt{n(n+1)}] [\mathbf{r} \times \nabla S_{nn}^M(\mathbf{r})] \\
 &= \sum_{n=1}^{n_{\max}} \sum_{M=-n}^n [1/(n+1)] g_{nM} [\mathbf{r} \times \nabla S_{nn}^M(\mathbf{r})] \\
 &= \sum_{n=1}^{n_{\max}} \sum_{M=-n}^n [1/(n+1)] g_{nM} [\mathbf{r} \times \nabla H_n^M(\mathbf{r})].
 \end{aligned} \tag{15.2.81}$$

Note that this result has the virtue that none of the extensive machinery of Appendix U is required for its evaluation.

### 15.2.6 Uniqueness of Poincaré-Coulomb Gauge

Is a vector potential in the Poincaré-Coulomb gauge unique? It is. Suppose  $\mathbf{A}$  and  $\mathbf{A}'$  are two vector potentials associated with the same field  $\mathbf{B}$ . Then we know they are related by a gauge transformation of the form

$$\mathbf{A}'(\mathbf{r}) = \mathbf{A}(\mathbf{r}) + \nabla \chi(\mathbf{r}). \tag{15.2.82}$$

If we require that both  $\mathbf{A}$  and  $\mathbf{A}'$  be in the Coulomb gauge, then  $\chi$  must be harmonic: taking the divergence of both sides of (2.82) yields the result

$$\nabla^2 \chi = 0. \tag{15.2.83}$$

See Section 15.6. If we further require that both  $\mathbf{A}$  and  $\mathbf{A}'$  be in the Poincaré-Coulomb gauge, then there must be the additional relations

$$\mathbf{r} \cdot \mathbf{A}(\mathbf{r}) = 0 \text{ and } \mathbf{r} \cdot \mathbf{A}'(\mathbf{r}) = 0. \tag{15.2.84}$$

Requiring (2.84) of (2.82) yields the result that  $\chi$  must also obey the condition

$$\mathbf{r} \cdot \nabla \chi = 0. \tag{15.2.85}$$

Suppose  $\chi$  is decomposed into homogeneous polynomials of degree  $n$  by writing

$$\chi = \sum_{n=0}^{\infty} \chi^n. \tag{15.2.86}$$

Then, by Euler's relation for homogeneous functions, it follows that

$$\mathbf{r} \cdot \nabla \chi = \sum_{n=0}^{\infty} n \chi^n. \tag{15.2.87}$$

Comparison of (2.85) and (2.87) and equating terms of like degree yields the result

$$n\chi^n = 0, \quad (15.2.88)$$

from which it follows that  $\chi^n = 0$  for  $n \neq 0$ . We see that all that is left in the sum (2.86) and in the relation (2.82) is the *constant* term  $\chi^0$ , and this term does not contribute to (2.82). We therefore conclude that  $\mathbf{A}'(\mathbf{r}) = \mathbf{A}(\mathbf{r})$ .

### 15.2.7 Direct Construction of Poincaré-Coulomb Gauge Vector Potential

Subsection 2.5 obtained the final result (2.81) using the machinery of Appendix U. The purpose of this subsection is to proceed in reverse. After some stage setting, we will make an Ansatz that is essentially equivalent to (2.81), and then verify that the vector potential produced by this Ansatz yields  $\mathbf{B}(\mathbf{r})$  as desired, and has other desired/interesting properties.

With reference to (2.37), define scalar fields  $\psi_{\ell,m,\alpha}$  by the rule

$$\psi_{\ell,m,\alpha} = H_\ell^{m,\alpha} \quad (15.2.89)$$

so that we may write

$$\psi(x, y, z) = \sum_{\ell=0}^{\infty} \sum_{m=0}^{\ell} g_{\ell,m,c} \psi_{\ell,m,c} + \sum_{\ell=1}^{\infty} \sum_{m=1}^{\ell} g_{\ell,m,s} \psi_{\ell,m,s}. \quad (15.2.90)$$

Next define related vector fields  $\mathbf{B}^{\ell,m,\alpha}$  by the rule

$$\mathbf{B}^{\ell,m,\alpha} = -\nabla \psi_{\ell,m,\alpha}. \quad (15.2.91)$$

Then, with the aid of (2.90) and (2.91), we may write

$$\mathbf{B} = -\nabla \psi = \sum_{\ell=1}^{\infty} \sum_{m=0}^{\ell} g_{\ell,m,c} \mathbf{B}^{\ell,m,c} + \sum_{\ell=1}^{\infty} \sum_{m=1}^{\ell} g_{\ell,m,s} \mathbf{B}^{\ell,m,s}. \quad (15.2.92)$$

We now seek individual vector potentials  $\mathbf{A}^{\ell,m,\alpha}$  such that

$$\nabla \times \mathbf{A}^{\ell,m,\alpha} = \mathbf{B}^{\ell,m,\alpha}. \quad (15.2.93)$$

Simple calculation shows that, if we can find them, then we may write

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (15.2.94)$$

with

$$\mathbf{A} = \sum_{\ell=1}^{\infty} \sum_{m=0}^{\ell} g_{\ell,m,c} \mathbf{A}^{\ell,m,c} + \sum_{\ell=1}^{\infty} \sum_{m=1}^{\ell} g_{\ell,m,s} \mathbf{A}^{\ell,m,s}. \quad (15.2.95)$$

We claim that a solution to (2.93) is given by the Ansatz

$$\mathbf{A}^{\ell,m,\alpha} = [-1/(\ell+1)][\mathbf{r} \times \mathbf{B}^{\ell,m,\alpha}] = [1/(\ell+1)][\mathbf{r} \times \nabla \psi_{\ell,m,\alpha}]. \quad (15.2.96)$$

Let us check this claim. Recall the vector identity

$$\nabla \times (\mathbf{a} \times \mathbf{b}) = \mathbf{a}(\nabla \cdot \mathbf{b}) - \mathbf{b}(\nabla \cdot \mathbf{a}) + (\mathbf{b} \cdot \nabla)\mathbf{a} - (\mathbf{a} \cdot \nabla)\mathbf{b}. \quad (15.2.97)$$

Then, with

$$\mathbf{a} = \mathbf{r} \quad (15.2.98)$$

and

$$\mathbf{b} = -\nabla\psi_{\ell,m,\alpha} = \mathbf{B}^{\ell,m,\alpha}, \quad (15.2.99)$$

the identity (2.97) yields the result

$$\begin{aligned} \nabla \times (\mathbf{r} \times \mathbf{b}) &= \mathbf{r}(\nabla \cdot \mathbf{b}) - \mathbf{B}^{\ell,m,\alpha}(\nabla \cdot \mathbf{r}) + (\mathbf{B}^{\ell,m,\alpha} \cdot \nabla)\mathbf{r} - (\mathbf{r} \cdot \nabla)\mathbf{B}^{\ell,m,\alpha}. \end{aligned} \quad (15.2.100)$$

Moreover, there are the relations

$$\nabla \cdot \mathbf{b} = -\nabla^2\psi_{\ell,m,\alpha} = 0, \quad (15.2.101)$$

$$\nabla \cdot \mathbf{r} = 3, \quad (15.2.102)$$

$$(\mathbf{B}^{\ell,m,\alpha} \cdot \nabla)\mathbf{r} = \mathbf{B}^{\ell,m,\alpha}, \quad (15.2.103)$$

and

$$(\mathbf{r} \cdot \nabla)\mathbf{B}^{\ell,m,\alpha} = (\ell - 1)\mathbf{B}^{\ell,m,\alpha}. \quad (15.2.104)$$

This last relation follows from the fact that the Cartesian components of  $\mathbf{B}^{\ell,m,\alpha}$  are homogeneous polynomials of degree  $(\ell - 1)$ . We conclude that

$$\nabla \times (\mathbf{r} \times \mathbf{b}) = [-3 + 1 - (\ell - 1)]\mathbf{B}^{\ell,m,\alpha} = [-(\ell + 1)]\mathbf{B}^{\ell,m,\alpha}. \quad (15.2.105)$$

Therefore, the  $\mathbf{A}^{\ell,m,\alpha}$  defined by (2.96) satisfy (2.93). We also note that the Cartesian components of the  $\mathbf{A}^{\ell,m,\alpha}$  are homogeneous polynomials of degree  $\ell$ .

In addition there is the vector identity

$$\nabla \cdot (\mathbf{a} \times \mathbf{b}) = \mathbf{b} \cdot (\nabla \times \mathbf{a}) - \mathbf{a} \cdot (\nabla \times \mathbf{b}). \quad (15.2.106)$$

From this identity and from (2.96), (2.98), and (2.99) it follows that

$$\nabla \cdot \mathbf{A}^{\ell,m,\alpha} = 0. \quad (15.2.107)$$

Thus, the  $\mathbf{A}^{\ell,m,\alpha}$  are in the Coulomb gauge. Moreover, from (2.96), there is the relation

$$\mathbf{r} \cdot \mathbf{A}^{\ell,m,\alpha}(\mathbf{r}) = 0. \quad (15.2.108)$$

Therefore the  $\mathbf{A}^{\ell,m,\alpha}$  are also in the Poincaré gauge, and thus in the Poincaré-Coulomb gauge.

Three final points: First, suppose the magnetic field  $\mathbf{B}(\mathbf{r})$  is decomposed into homogeneous polynomials by writing

$$\mathbf{B}(\mathbf{r}) = \sum_{n=0}^{\infty} \mathbf{B}^n(\mathbf{r}). \quad (15.2.109)$$

The vector potential can also be decomposed into homogeneous polynomials by writing

$$\mathbf{A}(\mathbf{r}) = \sum_{n=1}^{\infty} \mathbf{A}^n(\mathbf{r}). \quad (15.2.110)$$

From (2.96) we see that there is the relation

$$\mathbf{A}^n(\mathbf{r}) = -[1/(n+1)][\mathbf{r} \times \mathbf{B}^{n-1}(\mathbf{r})] \text{ for } n = 1, 2, \dots. \quad (15.2.111)$$

Second, suppose that (2.109) and (2.110) are *truncated* by writing

$$\mathbf{B}^{\text{trunc}}(\mathbf{r}) = \sum_{n=0}^N \mathbf{B}^n(\mathbf{r}), \quad (15.2.112)$$

$$\mathbf{A}^{\text{trunc}}(\mathbf{r}) = \sum_{n=1}^{N+1} \mathbf{A}^n(\mathbf{r}), \quad (15.2.113)$$

with (2.111) continuing to hold for  $n = 1, 2, \dots, N+1$ . It is easy to verify that  $\mathbf{B}^{\text{trunc}}(\mathbf{r})$  is curl and divergence free if  $\mathbf{B}(\mathbf{r})$  is. It is also true that

$$\mathbf{B}^{\text{trunc}}(\mathbf{r}) = \nabla \times \mathbf{A}^{\text{trunc}}(\mathbf{r}). \quad (15.2.114)$$

Thus, truncation by degree does not violate the Maxwell equations.

Finally, we note that the relation (2.111) specifies the vector potential order-by-order in terms of the order-by-order magnetic field. There is an equivalent integral relation that specifies the full vector potential in terms of the full magnetic field. It is given by the relation

$$\mathbf{A}(\mathbf{r}) = -\mathbf{r} \times \int_0^1 d\lambda \lambda \mathbf{B}(\lambda \mathbf{r}). \quad (15.2.115)$$

See Exercise 2.4.

## Exercises

**15.2.1.** The purpose of this exercise is to verify (2.19). In terms of the spherical coordinates  $r, \theta, \phi$  defined in Subsection 15.2.2 there is the result

$$\mathbf{r} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z = r \sin(\theta) \cos(\phi)\mathbf{e}_x + r \sin(\theta) \sin(\phi)\mathbf{e}_y + r \cos(\theta)\mathbf{e}_z. \quad (15.2.116)$$

Verify that there is an associated orthonormal triad  $\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi$  obeying the relations

$$\begin{aligned} \mathbf{e}_r &= [\partial \mathbf{r} / \partial r] / \|[\partial \mathbf{r} / \partial r]\| = \sin(\theta) \cos(\phi)\mathbf{e}_x + \sin(\theta) \sin(\phi)\mathbf{e}_y + \cos(\theta)\mathbf{e}_z \\ &= (1/r)(x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z) = \mathbf{r}/r, \end{aligned} \quad (15.2.117)$$

$$\mathbf{e}_\theta = [\partial \mathbf{r} / \partial \theta] / \|[\partial \mathbf{r} / \partial \theta]\| = \cos(\theta) \cos(\phi)\mathbf{e}_x + \cos(\theta) \sin(\phi)\mathbf{e}_y - \sin(\theta)\mathbf{e}_z, \quad (15.2.118)$$

$$\mathbf{e}_\phi = [\partial \mathbf{r} / \partial \phi] / \|[\partial \mathbf{r} / \partial \phi]\| = -\sin(\phi)\mathbf{e}_x + \cos(\phi)\mathbf{e}_y, \quad (15.2.119)$$

$$\mathbf{r} = r\mathbf{e}_r. \quad (15.2.120)$$

**15.2.2.** The purpose of this exercise is to verify (2.20) and (2.22) through (2.25). In terms of the cylindrical coordinates  $\rho, \phi, z$  defined in Section 15.2 there is the result

$$\mathbf{r} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z = \rho \cos \phi \mathbf{e}_x + \rho \sin \phi \mathbf{e}_y + z\mathbf{e}_z. \quad (15.2.121)$$

Verify that there is an associated orthonormal triad  $\mathbf{e}_\rho, \mathbf{e}_\phi, \mathbf{e}_z$  obeying the relations

$$\mathbf{e}_\rho = [\partial \mathbf{r} / \partial \rho] / \|[\partial \mathbf{r} / \partial \rho]\| = \cos \phi \mathbf{e}_x + \sin \phi \mathbf{e}_y = (1/\rho)(x\mathbf{e}_x + y\mathbf{e}_y), \quad (15.2.122)$$

$$\mathbf{e}_\phi = [\partial \mathbf{r} / \partial \phi] / \|[\partial \mathbf{r} / \partial \phi]\| = -\sin \phi \mathbf{e}_x + \cos \phi \mathbf{e}_y = (1/\rho)(-y\mathbf{e}_x + x\mathbf{e}_y), \quad (15.2.123)$$

$$\mathbf{e}_z = [\partial \mathbf{r} / \partial z] / \|[\partial \mathbf{r} / \partial z]\|, \quad (15.2.124)$$

$$x\mathbf{e}_x + y\mathbf{e}_y = \rho \mathbf{e}_\rho, \quad (15.2.125)$$

$$\mathbf{r} = \rho \mathbf{e}_\rho + z\mathbf{e}_z. \quad (15.2.126)$$

Verify (2.22) through (2.25).

**15.2.3.** Define an operator  $\mathbf{L}$  by the rule

$$\mathbf{L} = \mathbf{r} \times \nabla. \quad (15.2.127)$$

Show that it has the components

$$L_x = y\partial_z - z\partial_y, \quad (15.2.128)$$

$$L_y = z\partial_x - x\partial_z, \quad (15.2.129)$$

$$L_z = x\partial_y - y\partial_x. \quad (15.2.130)$$

Verify that the components of  $\mathbf{L}$  satisfy the commutation rules

$$\{L_x, L_y\} = -L_z, \text{ etc.}, \quad (15.2.131)$$

which are a variant of the commutation rules for  $so(3, \mathbb{R})$ .

Show that (2.96) can be rewritten in the form

$$\mathbf{A}^{\ell, m, \alpha} = [-1/(\ell + 1)][\mathbf{L}\psi_{\ell, m, \alpha}]. \quad (15.2.132)$$

Verify that  $\mathbf{L}$  and  $\nabla^2$  commute, and use this fact to show (as expected, see Section 5) that all the Cartesian components of  $\mathbf{A}^{\ell, m, \alpha}$  are harmonic functions,

$$\nabla^2 \mathbf{A}^{\ell, m, \alpha} = 0. \quad (15.2.133)$$

Verify that  $\mathbf{L}$  and  $r$  commute,

$$\{\mathbf{L}, r\} = 0. \quad (15.2.134)$$

Of course, this result is to be expected since  $r$  is invariant under rotations. Next, in view of (2.31) and (2.32), we may define functions  $h_\ell^{m, \alpha}(\theta, \phi)$  by writing

$$\psi_{\ell, m, \alpha} = H_\ell^{m, \alpha} = r^\ell h_\ell^{m, \alpha}(\theta, \phi). \quad (15.2.135)$$

Show it follows from (2.134) and (2.135) that

$$\mathbf{L}\psi_{\ell, m, \alpha} = r^\ell \mathbf{L}h_\ell^{m, \alpha}(\theta, \phi). \quad (15.2.136)$$

Therefore, if we wish, we can evaluate (2.132) using a raising and lowering operator formalism.

**15.2.4.** The purpose of this exercise is to verify (2.115) thus showing that the relations (2.109) through (2.111) can be written in a more compact form. By the definition of homogeneity, there is the relation

$$\mathbf{B}^n(\lambda \mathbf{r}) = \lambda^n \mathbf{B}^n(\mathbf{r}) \quad (15.2.137)$$

where  $\lambda$  is a scalar. Show from (2.109) that

$$\mathbf{B}(\lambda \mathbf{r}) = \sum_{n=0}^{\infty} \lambda^n \mathbf{B}^n(\mathbf{r}). \quad (15.2.138)$$

Next integrate both sides of (2.138) to demonstrate that

$$\int_0^1 d\lambda \lambda \mathbf{B}(\lambda \mathbf{r}) = \sum_{n=0}^{\infty} [1/(n+2)] \mathbf{B}^n(\mathbf{r}) = \sum_{n=1}^{\infty} [1/(n+1)] \mathbf{B}^{n-1}(\mathbf{r}). \quad (15.2.139)$$

Finally, using (2.110), (2.111), and (2.139), verify that there is the integral relation

$$\mathbf{A}(\mathbf{r}) = -\mathbf{r} \times \int_0^1 d\lambda \lambda \mathbf{B}(\lambda \mathbf{r}). \quad (15.2.140)$$

**15.2.5.** Subsection 2.6 showed that the Poincaré-Coulomb gauge is unique. Accordingly, starting from the requirement  $\mathbf{B} = \nabla \times \mathbf{A}^{\min}$  and the requirements (2.70) and (2.71) and the assumption that  $\mathbf{B}(\mathbf{r})$  is analytic in a neighborhood of  $\mathbf{r} = 0$ , it should be possible to derive the relations (2.111) and (2.115). Do so! Acknowledgement: This exercise was motivated by a suggestion of Sateesh Mane.

**15.2.6.** The relations (2.111) and (2.115) specify the minimum vector potential  $\mathbf{A}^{\min}$  in terms of the magnetic field  $\mathbf{B}$ . The purpose of this exercise is to derive relations that specify the scalar potential  $\psi$  in terms of  $\mathbf{B}$ .

Since  $\mathbf{B}$  is assumed to be analytic and curl free, show that  $\psi$  may be defined by the rule

$$\psi(\mathbf{r}) = - \int_0^{\mathbf{r}} \mathbf{B}(\mathbf{r}') \cdot d\mathbf{r}' \quad (15.2.141)$$

where the integral (because  $\mathbf{B}$  is curl free) may be carried out over any path joining 0 and  $\mathbf{r}$ . Note that with this definition

$$\psi(0) = 0. \quad (15.2.142)$$

Choose the path to be the straight line joining 0 and  $\mathbf{r}$  by making the Ansatz

$$\mathbf{r}' = \lambda \mathbf{r} \text{ with } \lambda \in [0, 1]. \quad (15.2.143)$$

Show that so doing yields the result

$$\psi(\mathbf{r}) = -\mathbf{r} \cdot \int_0^1 d\lambda \mathbf{B}(\lambda \mathbf{r}). \quad (15.2.144)$$

Assume that  $\mathbf{B}(\mathbf{r})$  is decomposed into homogeneous polynomials as in (2.109). Show, using (2.138) and (2.144), that

$$\psi(\mathbf{r}) = -\mathbf{r} \cdot \sum_{n=0}^{\infty} [1/(n+1)] \mathbf{B}^n(\mathbf{r}). \quad (15.2.145)$$

Suppose that  $\psi(\mathbf{r})$  is also decomposed into homogeneous polynomials by writing

$$\psi(\mathbf{r}) = \sum_{n=1}^{\infty} \psi^n(\mathbf{r}). \quad (15.2.146)$$

Show that

$$\psi^n(\mathbf{r}) = -(1/n)\mathbf{r} \cdot \mathbf{B}^{n-1}(\mathbf{r}) \text{ for } n = 1, 2, \dots \quad (15.2.147)$$

Verify directly that

$$-\nabla\psi^n(\mathbf{r}) = \mathbf{B}^{n-1}(\mathbf{r}). \quad (15.2.148)$$

Hint: Use the vector identity

$$\nabla(\mathbf{a} \cdot \mathbf{b}) = (\mathbf{a} \cdot \nabla)\mathbf{b} + (\mathbf{b} \cdot \nabla)\mathbf{a} + \mathbf{a} \times (\nabla \times \mathbf{b}) + \mathbf{b} \times (\nabla \times \mathbf{a}) \quad (15.2.149)$$

with

$$\mathbf{a} = \mathbf{r} \quad (15.2.150)$$

and

$$\mathbf{b} = \mathbf{B}^{n-1}(\mathbf{r}). \quad (15.2.151)$$

**15.2.7.** Review Exercise 1.5.7. There it is found that a uniform vertical magnetic field  $\mathbf{B} = B\mathbf{e}_y$  can be derived from the (Coulomb gauge) vector potential

$$\mathbf{A} = -Bx\mathbf{e}_z. \quad (15.2.152)$$

Note that both the magnetic field and the vector potential (2.152) have no  $z$  dependence so that we may conveniently take any point on the  $z$  axis to be an expansion point  $\mathbf{R}_0$ .

Also, review Subsection U.4.2. In addition recall the relations (2.9) through (2.11) and the definition

$$\int d\Omega = \int_0^\pi \int_0^{2\pi} \sin(\theta) d\theta d\phi = \int_0^\pi \sin(\theta) d\theta \int_0^{2\pi} d\phi. \quad (15.2.153)$$

Show that

$$\int x^2 d\Omega = \int y^2 d\Omega = \int z^2 d\Omega = (4/3)\pi r^2. \quad (15.2.154)$$

Use these results and (U.4.30) to show that

$$\|\mathbf{A}(\mathbf{r})\|^2 = (4/3)\pi B^2. \quad (15.2.155)$$

Let  ${}^{PC}\mathbf{A}$  be the associated vector potential in the Poincaré-Coulomb gauge. Verify that

$${}^{PC}\mathbf{A} = -(B/2)(x\mathbf{e}_z - z\mathbf{e}_x). \quad (15.2.156)$$

Show that

$$\|{}^{PC}\mathbf{A}(\mathbf{r})\|^2 = (2/3)\pi B^2. \quad (15.2.157)$$

Comparison of (2.155) and (2.157) shows that the Poincaré-Coulomb gauge vector potential has a smaller norm, as expected.

Show that

$${}^{PC}\mathbf{A} = \mathbf{A} + \nabla\chi \quad (15.2.158)$$

with

$$\chi = (B/2)xz. \quad (15.2.159)$$

Verify that  $\chi$  is harmonic, as expected because both  ${}^{PC}\mathbf{A}$  and  $\mathbf{A}$  are in the Coulomb gauge.

**15.2.8.** Review Exercise 1.5.9. There it is found that a quadrupole magnetic field with midplane symmetry and infinite extent in the  $z$  direction,

$$\mathbf{B} = Qye_x + Qxe_y, \quad (15.2.160)$$

can be derived from the (Coulomb gauge) vector potential

$$\mathbf{A} = -(Q/2)(x^2 - y^2)\mathbf{e}_z. \quad (15.2.161)$$

Review Exercise 2.7. As in this previous exercise, we may conveniently take any point on the  $z$  axis to be an expansion point  $\mathbf{R}_0$ . Show that

$$\int (x^2 - y^2)^2 d\Omega = (16/15)\pi r^4, \quad (15.2.162)$$

and therefore

$$\|\mathbf{A}(\mathbf{r})\|^2 = (4/15)\pi Q^2. \quad (15.2.163)$$

Let  ${}^{PC}\mathbf{A}$  be the associated vector potential in the Poincaré-Coulomb gauge. Verify that

$$\mathbf{r} \times \mathbf{B} = Q[(x^2 - y^2)\mathbf{e}_z + zye_y - zx\mathbf{e}_x], \quad (15.2.164)$$

and therefore

$${}^{PC}\mathbf{A} = -(Q/3)[-zx\mathbf{e}_x + zy\mathbf{e}_y + (x^2 - y^2)\mathbf{e}_z]. \quad (15.2.165)$$

Show that

$$\int z^2 x^2 d\Omega = \int z^2 y^2 d\Omega = \int x^2 y^2 d\Omega = (4/15)\pi r^4, \quad (15.2.166)$$

$$\int x^4 d\Omega = \int y^4 d\Omega = \int z^4 d\Omega = (4/5)\pi r^4, \quad (15.2.167)$$

and therefore

$$\|{}^{PC}\mathbf{A}\|^2 = (2/3)(4/15)\pi Q^2. \quad (15.2.168)$$

Comparison of (2.163) and (2.168) shows that the Poincaré-Coulomb gauge vector potential has a smaller norm, as expected.

Show that

$${}^{PC}\mathbf{A} = \mathbf{A} + \nabla\chi \quad (15.2.169)$$

with

$$\chi = (Q/6)z(x^2 - y^2). \quad (15.2.170)$$

Verify that  $\chi$  is harmonic, as expected because both  ${}^{PC}\mathbf{A}$  and  $\mathbf{A}$  are in the Coulomb gauge.

**15.2.9.** Let  $\mathbf{B}$  be any source-free magnetic field and select any expansion point  $\mathbf{R}_0$ . Since  $\mathbf{B}$  is assumed to be source free, it can be written as minus the gradient of a scalar potential  $\psi$  and, in terms of the deviation variable  $\mathbf{r}$ ,  $\psi$  will have through some order  $n_{\max}$  a polynomial expansion of the form (2.43). Correspondingly, we have found that  $\mathbf{B}$  can be written in the form

$$\mathbf{B}(\mathbf{r}) = \nabla \times \mathbf{A}^{\min}(\mathbf{r}) \quad (15.2.171)$$

with  $\mathbf{A}^{\min}(\mathbf{r})$ , the unique vector potential that produces  $\mathbf{B}$  and is in the Poincaré-Coulomb gauge, given by

$$\mathbf{A}^{\min}(\mathbf{r}) = \sum_{n=1}^{n_{\max}} \sum_{M=-n}^n (i) \sqrt{n/(n+1)} g_{nM} \mathbf{S}_{nnn}^M(\mathbf{r}). \quad (15.2.172)$$

Recall the work of Subsection 2.5 and view (2.69).

We also claimed that, in the sense of norms,  $\mathbf{A}^{\min}(\mathbf{r})$  is the *smallest* vector potential that produces  $\mathbf{B}$ , and have seen some examples consistent with this claim. Your task is to verify that this claim is true in general. What you will use, in essence, is that if  $u$  and  $v$  are any two nonzero vectors that are orthogonal to each other, then the vector  $u + v$  will be longer (have a greater norm) than either of the vectors  $u$  and  $v$ : This is a mathematical generalization of the fact that, for a right triangle, the hypotenuse will be longer than either of the two other sides.

Let  $\mathbf{A}^{ao}(\mathbf{r})$  be *any other* vector potential that also produces  $\mathbf{B}$ . Then we know that  $\mathbf{A}^{ao}(\mathbf{r})$  and  $\mathbf{A}^{\min}(\mathbf{r})$  are related by a gauge transformation so that we may write

$$\mathbf{A}^{ao}(\mathbf{r}) = \mathbf{A}^{\min}(\mathbf{r}) + \Delta(\mathbf{r}) \quad (15.2.173)$$

where  $\Delta$  is of the form

$$\Delta(\mathbf{r}) = \nabla \chi(\mathbf{r}). \quad (15.2.174)$$

Verify that there is the result

$$\begin{aligned} & < \mathbf{A}^{ao}(\mathbf{r}), \mathbf{A}^{ao}(\mathbf{r}) > = \\ & < \mathbf{A}^{\min}(\mathbf{r}), \mathbf{A}^{\min}(\mathbf{r}) > + < \Delta(\mathbf{r}), \Delta(\mathbf{r}) > \\ & + < \mathbf{A}^{\min}(\mathbf{r}), \Delta(\mathbf{r}) > + < \Delta(\mathbf{r}), \mathbf{A}^{\min}(\mathbf{r}) >. \end{aligned} \quad (15.2.175)$$

You will soon be asked to verify that the last line on the right side of (2.175) vanishes,

$$< \mathbf{A}^{\min}(\mathbf{r}), \Delta(\mathbf{r}) > + < \Delta(\mathbf{r}), \mathbf{A}^{\min}(\mathbf{r}) > = 0. \quad (15.2.176)$$

Assuming this is the case, verify that (2.175) can be rewritten in the form

$$\|\mathbf{A}^{ao}(\mathbf{r})\|^2 = \|\mathbf{A}^{\min}(\mathbf{r})\|^2 + \|\Delta(\mathbf{r})\|^2. \quad (15.2.177)$$

Verify it then follows that there is the valid chain of reasoning

$$\mathbf{A}^{ao}(\mathbf{r}) \neq \mathbf{A}^{\min}(\mathbf{r}) \Rightarrow \Delta(\mathbf{r}) \neq 0 \Rightarrow \|\Delta(\mathbf{r})\| > 0 \Rightarrow \|\mathbf{A}^{ao}(\mathbf{r})\| > \|\mathbf{A}^{\min}(\mathbf{r})\|. \quad (15.2.178)$$

$\mathbf{A}^{\min}(\mathbf{r})$  as defined by (2.172) is indeed the smallest vector potential in the sense that

$$\mathbf{A}^{ao}(\mathbf{r}) \neq \mathbf{A}^{\min}(\mathbf{r}) \Rightarrow \|\mathbf{A}^{\min}(\mathbf{r})\| < \|\mathbf{A}^{ao}(\mathbf{r})\|. \quad (15.2.179)$$

It remains to be shown that (2.176) holds. Examination of (2.172) shows that  $\mathbf{A}^{\min}(\mathbf{r})$  involves *only* the spherical polynomial vector field basis elements for which the three lower indices are the *same*, e.g. the  $\mathbf{S}_{nnn}^M(\mathbf{r})$ . What can be said about  $\Delta(\mathbf{r})$ ? Since the spherical polynomials  $S_{n\ell}^m(\mathbf{r})$  form a basis,  $\chi$  may be expanded in terms of them. Verify from (2.174), (U.5.3), and (U.5.5) that  $\Delta(\mathbf{r})$  involves only the spherical polynomial vector field basis elements for which the three lower indices are *not* all the same. Verify using (U.4.17) that both inner products appearing in (2.176) must vanish.

**15.2.10.** Demonstration that harmonic functions take their extrema on boundaries.

## 15.3 Cylindrical Harmonic Expansion

In the previous section we employed spherical coordinates to find *local* expansions (expansions about a point  $\mathbf{R}_0$ ) for the scalar potential  $\psi$  and the associated magnetic field  $\mathbf{B}$ . We also found a suitable vector potential  $\mathbf{A}^{\min}$ . The goal of this section is to show that it is possible to obtain *semi-global* expansions for  $\psi$  and  $\mathbf{B}$  in the case of a *straight* geometry, the case of straight beam-line elements. By semi-global we mean that an expansion holds all along the vicinity of the beam-line axis (which we take to be the  $z$  axis). That is, while the variables  $x$  and  $y$  are treated as being small, the variable  $z$  need not be small. Again we will want to exploit/ensure the condition that  $\mathbf{B}$  be source free.

For this purpose is convenient to work in cylindrical coordinates  $\rho$ ,  $\phi$ , and  $z$  as given by (2.12) through (2.14). We also note, for future use, that (2.13) and (2.14) can be written in the form

$$x + iy = \rho \exp(i\phi). \quad (15.3.1)$$

From this form it follows that

$$\rho^{2\ell} = (x^2 + y^2)^\ell \quad (15.3.2)$$

and, for  $m \geq 0$ ,

$$\rho^m \cos m\phi = \Re[(x + iy)^m], \quad (15.3.3)$$

$$\rho^m \sin m\phi = \Im[(x + iy)^m]. \quad (15.3.4)$$

We see that *even* powers of  $\rho$  and the combinations  $\rho^m \cos m\phi$  and  $\rho^m \sin m\phi$  are *analytic* (in fact, *polynomial*) functions of  $x$  and  $y$ .

### 15.3.1 Complex Cylindrical Harmonic Expansion

To find the general  $\psi$  in cylindrical coordinates that satisfies Laplace's equation, recall that the functions  $\exp(im\phi)$  form a complete set for the Hilbert space of functions over the interval  $\phi \in [0, 2\pi]$ , and the functions  $\exp(ikz)$  form a complete set for the Hilbert space of functions over the interval  $z \in [-\infty, \infty]$ . Therefore any function  $\psi$  in the product Hilbert space can be written as a superposition of functions of the form  $\Gamma_m(k, \rho) \exp(ikz) \exp(im\phi)$  where the functions  $\Gamma_m(k, \rho)$  are yet to be determined. In cylindrical coordinates the Laplacian has the form

$$\nabla^2 = (1/\rho)(\partial/\partial\rho)(\rho\partial/\partial\rho) + (1/\rho^2)(\partial^2/\partial\phi^2) + \partial^2/\partial z^2. \quad (15.3.5)$$

Thus if the product  $\Gamma_m(k, \rho) \exp(ikz) \exp(im\phi)$  is to satisfy Laplace's equation, the functions  $\Gamma_m(k, \rho)$  must satisfy the modified Bessel equation,

$$(1/\rho)(\partial/\partial\rho)(\rho\partial\Gamma_m/\partial\rho) - (m^2/\rho^2)\Gamma_m - k^2\Gamma_m = 0. \quad (15.3.6)$$

The solutions to this equation (that are regular for small  $\rho$ ) are the modified Bessel functions  $I_m(k\rho)$ . Consequently, in cylindrical coordinates, a general  $\psi$  satisfying Laplace's equation and analytic in  $x, y$  near the  $z$  axis has the expansion

$$\psi(x, y, z) = - \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} dk G_m(k) \exp(ikz) \exp(im\phi) I_m(k\rho) \quad (15.3.7)$$

where the functions  $G_m(k)$  are arbitrary.<sup>8</sup> We remark, for future use, that the modified Bessel functions  $I_m(w)$  have the property

$$I_{-m}(w) = I_m(w). \quad (15.3.8)$$

The representation (3.7) is a *cylindrical harmonic* or *cylindrical multipole* expansion, where  $m$  is related to the order of the multipole. For example,  $m = 0$  for a ‘monopole’ source (including a solenoid),  $m = 1$  for a dipole,  $m = 2$  for a quadrupole, etc. We also remark that these are what we will call *pure* multipoles. For example, a real/physical quadrupole (even one with perfect four-fold symmetry) will have primarily  $m = 2$  components plus smaller higher-order pure multipole components that are not forbidden by symmetry. Both pole/coil shape and rotational symmetry matter. See Subsection 3.5. Finally we should remark that (3.7) could more accurately be called a *circular* cylinder harmonic expansion. In Section 17.4 we will extend this work to include the case of cylinders with *elliptic* cross sections, and in Section 17.5 we will treat the case of cylinders with *rectangular* cross sections.

As stated at the beginning of this chapter, our ultimate goal is a Taylor expansion of the vector potential  $\mathbf{A}$  in the variables  $x, y$ . To do this, we first seek an expansion of  $\psi$  as a Taylor series in the variables  $x, y$  with coefficients that depend on  $z$ . This can be achieved, by using the Taylor expansions for  $I_m(w)$ , as follows: Using (3.7) we may write

$$\begin{aligned} \psi(x, y, z) &= - \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} dk G_m(k) \exp(ikz) \exp(im\phi) I_m(k\rho) \\ &= - \sum_{m=-\infty}^{\infty} \exp(im\phi) \int_{-\infty}^{\infty} dk G_m(k) \exp(ikz) I_m(k\rho) \\ &= \sum_{m=-\infty}^{\infty} \exp(im\phi) \Psi_m(\rho, z) \end{aligned} \quad (15.3.9)$$

where

$$\Psi_m(\rho, z) = - \int_{-\infty}^{\infty} dk G_m(k) \exp(ikz) I_m(k\rho). \quad (15.3.10)$$

---

<sup>8</sup>One may wonder about the minus sign on the right side of (3.7). At this stage it is harmless since the functions  $G_m(k)$  are arbitrary, but will prove to be convenient later on.

(Note that the symbol  $\Psi$  that appears here and in what is to follow should not be confused with the  $\Psi$  in Subsection 2.1. Fortunately it should be clear from the context where it is employed what is actually meant.) The modified Bessel functions have the expansions

$$I_m(w) = (1/2)^{|m|} w^{|m|} \sum_{\ell=0}^{\infty} w^{2\ell} / [2^{2\ell} \ell! (\ell + |m|)!]. \quad (15.3.11)$$

Therefore we may also write

$$\begin{aligned} \Psi_m(\rho, z) &= - \int_{-\infty}^{\infty} dk G_m(k) \exp(ikz) I_m(k\rho) = \\ &= - \int_{-\infty}^{\infty} dk G_m(k) \exp(ikz) (1/2)^{|m|} (k\rho)^{|m|} \sum_{\ell=0}^{\infty} (k\rho)^{2\ell} / [2^{2\ell} \ell! (\ell + |m|)!] = \\ &= - \sum_{\ell=0}^{\infty} \{1/[2^{2\ell} \ell! (\ell + |m|)!]\} \rho^{2\ell+|m|} (1/2)^{|m|} \int_{-\infty}^{\infty} dk k^{2\ell+|m|} G_m(k) \exp(ikz). \end{aligned} \quad (15.3.12)$$

Define functions  $C_m^{[0]}(z)$  by writing

$$C_m^{[0]}(z) \stackrel{\text{def}}{=} (1/2)^{|m|} (1/|m|!) \int_{-\infty}^{\infty} dk k^{|m|} G_m(k) \exp(ikz). \quad (15.3.13)$$

Also, define functions  $C_m^{[n]}(z)$  by writing

$$C_m^{[n]}(z) = (\partial_z)^n C_m^{[0]}(z). \quad (15.3.14)$$

Then, by differentiating under the integral sign, we have the result

$$C_m^{[n]}(z) = (\partial_z)^n C_m^{[0]}(z) = i^n (1/2)^{|m|} (1/|m|!) \int_{-\infty}^{\infty} dk k^{n+|m|} G_m(k) \exp(ikz) \quad (15.3.15)$$

and, in particular,

$$C_m^{[2\ell]}(z) = (-1)^\ell (1/2)^{|m|} (1/|m|!) \int_{-\infty}^{\infty} dk k^{2\ell+|m|} G_m(k) \exp(ikz). \quad (15.3.16)$$

Thus, we may also write the relation

$$(1/2)^{|m|} \int_{-\infty}^{\infty} dk k^{2\ell+|m|} G_m(k) \exp(ikz) = (-1)^\ell |m|! C_m^{[2\ell]}(z). \quad (15.3.17)$$

Putting everything together gives the result

$$\Psi_m(\rho, z) = - \sum_{\ell=0}^{\infty} (-1)^\ell \frac{|m|!}{2^{2\ell} \ell! (\ell + |m|)!} C_m^{[2\ell]}(z) \rho^{2\ell+|m|}. \quad (15.3.18)$$

Consequently,  $\psi(x, y, z)$  has the representation

$$\psi(x, y, z) = - \sum_{m=-\infty}^{\infty} \exp(im\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{|m|!}{2^{2\ell} \ell! (\ell + |m|)!} C_m^{[2\ell]}(z) \rho^{2\ell + |m|}. \quad (15.3.19)$$

Note that, in view of (3.2) through (3.4), the terms appearing on the right side of (3.19) are polynomial in the variables  $x$  and  $y$ .

From (3.18) we see that

$$C_m^{[0]}(z) = - \lim_{\rho \rightarrow 0} (1/\rho^{|m|}) \Psi_m(\rho, z). \quad (15.3.20)$$

For this reason, the functions  $C_m^{[0]}(z)$  are called the *on-axis gradients*.<sup>9</sup> Note that the on-axis gradients depend on the longitudinal variable  $z$ . However we will soon see that, for fields produced by long well-made magnets, the  $z$  dependence will be significant only at the ends.

### 15.3.2 Real Cylindrical Harmonic Expansion in terms of Real On-axis Gradients

So far, for mathematical convenience, we have worked with a possibly complex representation for  $\psi$ . We will now convert our results into equivalent real forms suitable for physical applications. We begin with the relation (3.19). Suppose we require that  $\psi(x, y, z)$  be real. Forming the complex conjugate of (3.19) gives the result

$$\bar{\psi}(x, y, z) = - \sum_{m=-\infty}^{\infty} \exp(-im\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{|m|!}{2^{2\ell} \ell! (\ell + |m|)!} \bar{C}_m^{[2\ell]}(z) \rho^{2\ell + |m|}. \quad (15.3.21)$$

The right side of (3.21) can be rewritten to give the relation

$$\begin{aligned} & - \sum_{m=-\infty}^{\infty} \exp(-im\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{|m|!}{2^{2\ell} \ell! (\ell + |m|)!} \bar{C}_m^{[2\ell]}(z) \rho^{2\ell + |m|} = \\ & - \sum_{m=-\infty}^{\infty} \exp(im\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{|m|!}{2^{2\ell} \ell! (\ell + |m|)!} \bar{C}_{-m}^{[2\ell]}(z) \rho^{2\ell + |m|}. \end{aligned} \quad (15.3.22)$$

Therefore requiring

$$\bar{\psi}(x, y, z) = \psi(x, y, z) \quad (15.3.23)$$

is equivalent to the requirement

$$\bar{C}_{-m}^{[2\ell]}(z) = C_m^{[2\ell]}(z), \quad (15.3.24)$$

or

$$C_{-m}^{[2\ell]}(z) = \bar{C}_m^{[2\ell]}(z). \quad (15.3.25)$$

---

<sup>9</sup>Although (3.20) is mathematically correct, it is not a good way to actually compute the on-axis gradients due to the delicate nature of the limiting process. Indeed, one of the aims of Chapters 17 through 21 is to provide reliable ways of computing the on-axis gradients.

Let us now use this information to rewrite  $\psi$ . From (3.19) we have, in the general case,

$$\begin{aligned}\psi(x, y, z) &= -\sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! \ell!} C_0^{[2\ell]}(z) \rho^{2\ell} \\ &\quad - \sum_{m \neq 0} \cos(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{|m|!}{2^{2\ell} \ell! (\ell + |m|)!} C_m^{[2\ell]}(z) \rho^{2\ell+|m|} \\ &\quad - i \sum_{m \neq 0} \sin(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{|m|!}{2^{2\ell} \ell! (\ell + |m|)!} C_m^{[2\ell]}(z) \rho^{2\ell+|m|}.\end{aligned}\tag{15.3.26}$$

The second sum over  $m$  in (3.26) can be rewritten as

$$\begin{aligned}- \sum_{m \neq 0} \cos(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{|m|!}{2^{2\ell} \ell! (\ell + |m|)!} C_m^{[2\ell]}(z) \rho^{2\ell+|m|} &= \\ - \sum_{m=1}^{\infty} \cos(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell + m)!} [C_m^{[2\ell]}(z) + C_{-m}^{[2\ell]}(z)] \rho^{2\ell+m}.\end{aligned}\tag{15.3.27}$$

Now define functions  $C_{m,c}^{[2\ell]}(z)$  by the rule

$$C_{m,c}^{[2\ell]}(z) = C_m^{[2\ell]}(z) + C_{-m}^{[2\ell]}(z) \text{ for } m \geq 1,\tag{15.3.28}$$

so we may also write

$$\begin{aligned}- \sum_{m \neq 0} \cos(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{|m|!}{2^{2\ell} \ell! (\ell + |m|)!} C_m^{[2\ell]}(z) \rho^{2\ell+|m|} &= \\ - \sum_{m=1}^{\infty} \cos(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell + m)!} C_{m,c}^{[2\ell]}(z) \rho^{2\ell+m}.\end{aligned}\tag{15.3.29}$$

According to (3.25), the functions  $C_{m,c}^{[2\ell]}(z)$  will be real if  $\psi$  is real. The third sum over  $m$  in (3.26) can be rewritten as

$$\begin{aligned}-i \sum_{m \neq 0} \sin(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{|m|!}{2^{2\ell} \ell! (\ell + |m|)!} C_m^{[2\ell]}(z) \rho^{2\ell+|m|} &= \\ -i \sum_{m=1}^{\infty} \sin(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell + m)!} [C_m^{[2\ell]}(z) - C_{-m}^{[2\ell]}(z)] \rho^{2\ell+m}.\end{aligned}\tag{15.3.30}$$

Now define functions  $C_{m,s}^{[2\ell]}(z)$  by the rule

$$C_{m,s}^{[2\ell]}(z) = i[C_m^{[2\ell]}(z) - C_{-m}^{[2\ell]}(z)] \text{ for } m \geq 1,\tag{15.3.31}$$

so we may also write

$$\begin{aligned} -i \sum_{m \neq 0} \sin(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{|m|!}{2^{2\ell} \ell! (\ell + |m|)!} C_m^{[2\ell]}(z) \rho^{2\ell+|m|} = \\ - \sum_{m=1}^{\infty} \sin(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell + m)!} C_{m,s}^{[2\ell]}(z) \rho^{2\ell+m}. \end{aligned} \quad (15.3.32)$$

According to (3.25), the functions  $C_{m,s}^{[2\ell]}(z)$  with  $m \geq 1$  will be real if  $\psi$  is real. Combining the various results so far gives the representation

$$\begin{aligned} \psi(x, y, z) = & - \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! \ell!} C_0^{[2\ell]}(z) \rho^{2\ell} \\ = & \sum_{m=1}^{\infty} \cos(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell + m)!} C_{m,c}^{[2\ell]}(z) \rho^{2\ell+m} \\ = & \sum_{m=1}^{\infty} \sin(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell + m)!} C_{m,s}^{[2\ell]}(z) \rho^{2\ell+m}. \end{aligned} \quad (15.3.33)$$

Observe that, according to (3.25), the  $C_0^{[2\ell]}(z)$  will be real if  $\psi$  is real,

$$C_0^{[0]}(z) = \bar{C}_0^{[0]}(z). \quad (15.3.34)$$

Thus, all the quantities appearing in (3.33) are real if  $\psi$  is real. For future use it will be convenient to extend the definitions (3.28) and (3.31) to the  $m = 0$  case by writing

$$C_{0,c}^{[0]}(z) = C_0^{[0]}(z), \quad (15.3.35)$$

$$C_{0,s}^{[0]}(z) = 0. \quad (15.3.36)$$

We note that all the functions  $C_0^{[0]}(z)$ ,  $C_{m,c}^{[0]}(z)$ , and  $C_{m,s}^{[0]}(z)$  may be chosen independently, and *any* such choice produces a harmonic function when employed in (3.33). See Exercise 3.6. Finally, we observe that all the terms in (3.33) are sums of quantities of the form  $\rho^m \cos(m\phi)$  or  $\rho^m \sin(m\phi)$  multiplied by powers of  $\rho^2$  with  $z$ -dependent coefficients  $C_0^{[2\ell]}(z)$ ,  $C_{m,c}^{[2\ell]}(z)$ , and  $C_{m,s}^{[2\ell]}(z)$ . Thus, in view of (3.2) through (3.4), we have achieved our goal of finding a Taylor expansion for  $\psi(x, y, z)$  in powers of  $x, y$  with coefficients that depend on  $z$ .

We close this subsection by introducing some further notation that will be of future use. Define quantities  $\Psi_0(\rho, z)$ ,  $\Psi_{m,c}(\rho, z)$ , and  $\Psi_{m,s}(\rho, z)$  by the equations

$$\Psi_0(\rho, z) = \Psi_{0,c}(\rho, z) = - \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! \ell!} C_0^{[2\ell]}(z) \rho^{2\ell}, \quad (15.3.37)$$

$$\Psi_{m,c}(\rho, z) = - \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell + m)!} C_{m,c}^{[2\ell]}(z) \rho^{2\ell+m}, \quad (15.3.38)$$

$$\Psi_{m,s}(\rho, z) = - \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,s}^{[2\ell]}(z) \rho^{2\ell+m}, \quad (15.3.39)$$

so that, in view of (3.33), we may write

$$\begin{aligned} \psi(x, y, z) &= \Psi_0(\rho, z) + \sum_{m=1}^{\infty} \Psi_{m,c}(\rho, z) \cos m\phi + \sum_{m=1}^{\infty} \Psi_{m,s}(\rho, z) \sin m\phi \\ &= \psi_0(x, y, z) + \sum_{m=1}^{\infty} \psi_{m,c}(x, y, z) + \sum_{m=1}^{\infty} \psi_{m,s}(x, y, z) \end{aligned} \quad (15.3.40)$$

where

$$\psi_0(x, y, z) = \psi_{0,c}(x, y, z) = \Psi_0(\rho, z), \quad (15.3.41)$$

$$\psi_{m,c}(x, y, z) = \psi_{m,c}(\rho, \phi, z) = \cos(m\phi) \Psi_{m,c}(\rho, z), \quad (15.3.42)$$

$$\psi_{m,s}(x, y, z) = \psi_{m,s}(\rho, \phi, z) = \sin(m\phi) \Psi_{m,s}(\rho, z). \quad (15.3.43)$$

Note it follows from (3.40) that there are the relations

$$\Psi_0(\rho, z) = [1/(2\pi)] \int_{-\pi}^{\pi} d\phi \psi(x, y, z), \quad (15.3.44)$$

$$\Psi_{m,c}(\rho, z) = (1/\pi) \int_{-\pi}^{\pi} d\phi \psi(x, y, z) \cos(m\phi), \quad (15.3.45)$$

$$\Psi_{m,s}(\rho, z) = (1/\pi) \int_{-\pi}^{\pi} d\phi \psi(x, y, z) \sin(m\phi). \quad (15.3.46)$$

And it follows from the definitions (3.37) through (3.39) that there are the on-axis gradient relations

$$C^{[0]}(z) = - \lim_{\rho \rightarrow 0} \Psi_0(\rho, z), \quad (15.3.47)$$

$$C_{m,c}^{[0]}(z) = - \lim_{\rho \rightarrow 0} (1/\rho^m) \Psi_{m,c}(\rho, z), \quad (15.3.48)$$

$$C_{m,s}^{[0]}(z) = - \lim_{\rho \rightarrow 0} (1/\rho^m) \Psi_{m,s}(\rho, z). \quad (15.3.49)$$

These relations are the real counterparts of the complex relations (3.20).

What are the units for the  $C_{m,\alpha}^{[0]}(z)$  (with  $\alpha = c, s$ )? We know that the  $\Psi_{m,\alpha}(\rho, z)$  have units  $BL$ . See (3.45) and (3.46). It follows from (3.48) and (3.49) that the  $C_{m,\alpha}^{[0]}(z)$  have units  $BL/L^m = B/L^{m-1}$ . Correspondingly, the  $C_{m,\alpha}^{[\ell]}(z)$  have units  $B/L^{m+\ell-1}$ .

Finally, there is a variant coefficient relation that will be of subsequent use. Begin by rewriting (3.7) in the form

$$\begin{aligned} \psi(x, y, z) &= - \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} dk G_m(k) \exp(ikz) \cos(m\phi) I_m(k\rho) \\ &\quad - i \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} dk G_m(k) \exp(ikz) \sin(m\phi) I_m(k\rho), \end{aligned} \quad (15.3.50)$$

from which it follows that

$$\begin{aligned}\psi(x, y, z) &= - \int_{-\infty}^{\infty} dk G_0(k) \exp(ikz) I_0(k\rho) \\ &- \sum_{m=1}^{\infty} \int_{-\infty}^{\infty} dk [G_m(k) + G_{-m}(k)] \exp(ikz) \cos(m\phi) I_m(k\rho) \\ &- \sum_{m=1}^{\infty} \int_{-\infty}^{\infty} dk [iG_m(k) - iG_{-m}(k)] \exp(ikz) \sin(m\phi) I_m(k\rho).\end{aligned}\quad (15.3.51)$$

Next, with  $m \geq 1$ , introduce the notation

$$G_{m,c}(k) = G_m(k) + G_{-m}(k), \quad (15.3.52)$$

$$G_{m,s}(k) = iG_m(k) - iG_{-m}(k), \quad (15.3.53)$$

with the conventions

$$G_{0,c}(k) = G_0(k), \quad (15.3.54)$$

$$G_{0,s}(k) = 0. \quad (15.3.55)$$

In terms of this notation,  $\psi$  has the representation

$$\begin{aligned}\psi(x, y, z) &= - \sum_{m=0}^{\infty} \int_{-\infty}^{\infty} dk G_{m,c}(k) \exp(ikz) \cos(m\phi) I_m(k\rho) \\ &- \sum_{m=1}^{\infty} \int_{-\infty}^{\infty} dk G_{m,s}(k) \exp(ikz) \sin(m\phi) I_m(k\rho).\end{aligned}\quad (15.3.56)$$

We also observe, see Exercise 3.9, that the  $C_{m,\alpha}^{[n]}$  can be found in terms of the  $G_{m,\alpha}$  from the relations

$$C_{m,c}^{[n]}(z) = (\partial_z)^n C_{m,c}^{[0]}(z) = i^n (1/2)^m (1/m!) \int_{-\infty}^{\infty} dk k^{n+m} G_{m,c}(k) \exp(ikz), \quad (15.3.57)$$

$$C_{m,s}^{[n]}(z) = (\partial_z)^n C_{m,s}^{[0]}(z) = i^n (1/2)^m (1/m!) \int_{-\infty}^{\infty} dk k^{n+m} G_{m,s}(k) \exp(ikz). \quad (15.3.58)$$

### 15.3.3 Some Simple Examples: $m = 0, 1, 2$

Let us seek a physical interpretation for the functions  $C_0^{[0]}(z)$ ,  $C_{m,c}^{[0]}(z)$ , and  $C_{m,s}^{[0]}(z)$  by computing the associated magnetic fields using (2.6).

#### The Case $m = 0$

If only the  $C_0^{[0]}(z)$  related terms (the  $m = 0$  terms) are present in (3.33),  $\psi$  has the expansion

$$\psi_0(x, y, z) = -C_0^{[0]}(z) + (1/4)(x^2 + y^2)C_0^{[2]}(z) + \dots, \quad (15.3.59)$$

and therefore

$$B_x = -\partial_x \psi_0 = -(1/2)x C_0^{[2]}(z) + \dots, \quad (15.3.60)$$

$$B_y = -\partial_y \psi_0 = -(1/2)y C_0^{[2]}(z) + \dots, \quad (15.3.61)$$

$$B_z = -\partial_z \psi_0 = C_0^{[1]}(z) - (1/4)(x^2 + y^2)C_0^{[3]}(z) + \dots. \quad (15.3.62)$$

We see that  $\mathbf{B}$  is primarily in the  $z$  direction and that  $B_z(0, 0, z)$ , the on-axis  $z$  component of  $\mathbf{B}$ , has a profile given by  $C_0^{[1]}(z)$ . As long as  $C_0^{[1]}(z)$  is nearly constant,  $C_0^{[2]}(z)$  will be small, and therefore the transverse field components  $B_x$  and  $B_y$  will be small. Such would be the case for the field of a solenoid where the field is primarily longitudinal and only has transverse components in the fringe-field regions at each end where  $C_0^{[1]}(z)$  is changing. (See Sections 16.1 and 21.1.) We know that for any solenoid-like element (an element having a nonvanishing  $m = 0$  component in the cylindrical harmonic expansion of its scalar potential and a nonvanishing longitudinal field somewhere on axis)  $C_0^{[1]}(z)$  must depend on  $z$  because this on-axis gradient must be nonzero somewhere for such an element and must vanish far outside any such element because  $\mathbf{B}$  vanishes there. Therefore the functions  $C_0^{[2]}(z)$ ,  $C_0^{[3]}(z)$ , etc., must be nonzero, at least near the end and fringe-field regions of any such element. We conclude that, as a consequence of Maxwell's equations, the scalar potential  $\psi_0$  (and as we will see, the associated vector potential) for any such element must contain terms beyond degree two in the variables  $x, y$ . Correspondingly, the transfer map for any real solenoid must contain nonlinear terms.

Moreover, according to (3.277), the function  $C_0^{[1]}(z)$  is given in terms of the longitudinal on-axis field  $B_z(0, 0, z)$ . We observe that for a long uniform solenoid the on-axis field  $B_z(0, 0, z)$  will be nearly constant in the body of the solenoid, and therefore the quantities  $C_0^{[n]}(z)$  will be small in this region for  $n > 1$ . However, these derivatives may be large in fringe-field regions.

The same set up could also describe some portion of the field due to an off-center dipole (or any other off-center higher-order multipole) since such magnets would also have an on-axis  $B_z$  component somewhere in the fringe-field regions. In all cases we know that  $C_0^{[1]}(z)$  must depend on  $z$  because this on-axis gradient must be nonzero somewhere in or near the element and must vanish far outside the element, again because  $\mathbf{B}$  vanishes there.

### The Case $m = 1$

Next, suppose that only the  $C_{1,s}^{[0]}(z)$  related terms are present in (3.33). In this  $m = 1$  case,  $\psi$  has an expansion of the form

$$\begin{aligned} \psi_{1,s}(x, y, z) &= -\rho \sin(\phi)[C_{1,s}^{[0]}(z) - (1/8)(x^2 + y^2)C_{1,s}^{[2]}(z) + \dots] \\ &= -y[C_{1,s}^{[0]}(z) - (1/8)(x^2 + y^2)C_{1,s}^{[2]}(z) + \dots] \end{aligned} \quad (15.3.63)$$

and therefore

$$B_x = -\partial_x \psi_{1,s} = -(1/4)xy C_{1,s}^{[2]}(z) + \dots, \quad (15.3.64)$$

$$B_y = -\partial_y \psi_{1,s} = C_{1,s}^{[0]}(z) - (1/8)(x^2 + 3y^2)C_{1,s}^{[2]}(z) + \dots, \quad (15.3.65)$$

$$B_z = -\partial_z \psi_{1,s} = y[C_{1,s}^{[1]}(z) - (1/8)(x^2 + y^2)C_{1,s}^{[3]}(z) + \dots]. \quad (15.3.66)$$

We see that  $\mathbf{B}$  is primarily in the  $y$  direction with a profile given by  $B_y(0, 0, z) = C_{1,s}^{[0]}(z)$ . As long as  $C_{1,s}^{[0]}(z)$  is nearly constant,  $C_{1,s}^{[1]}(z)$  and higher derivatives of  $C_{1,s}^{[0]}(z)$  will be small, and therefore the other field components  $B_x$  and  $B_z$  will be small. Such would be the case for the field of a (normal) dipole where the field is primarily vertical and only has  $x$  and  $z$  components in the fringe-field regions at each end where  $C_{1,s}^{[0]}(z)$  is changing. (See Exercise 1.5.7.) However, there will always be nonlinear terms at the ends where  $C_{1,s}^{[0]}(z)$  and higher derivatives cannot be constant. Correspondingly, the transfer map for any real dipole must contain nonlinear terms. We end the discussion of the  $m = 1$  case by remarking that the  $C_{1,c}^{[0]}(z)$  related terms describe the field of a *skew* dipole. See Exercise 4.1.

### The Case $m = 2$

As a last example, suppose that only the  $C_{2,s}^{[0]}(z)$  related terms are present in (3.33). In this  $m = 2$  case,  $\psi$  has an expansion of the form

$$\begin{aligned}\psi_{2,s}(x, y, z) &= -\rho^2 \sin(2\phi)[C_{2,s}^{[0]}(z) - (1/24)(x^2 + y^2)C_{2,s}^{[2]}(z) + \dots] \\ &= -2xy[C_{2,s}^{[0]}(z) - (1/24)(x^2 + y^2)C_{2,s}^{[2]}(z) + \dots]\end{aligned}\quad (15.3.67)$$

and therefore

$$B_x = -\partial_x \psi_{2,s} = 2yC_{2,s}^{[0]}(z) - (1/12)(3x^2y + y^3)C_{2,s}^{[2]}(z) + \dots, \quad (15.3.68)$$

$$B_y = -\partial_y \psi_{2,s} = 2xC_{2,s}^{[0]}(z) - (1/12)(x^3 + 3xy^2)C_{2,s}^{[2]}(z) + \dots, \quad (15.3.69)$$

$$B_z = -\partial_z \psi_{2,s} = 2xy[C_{2,s}^{[1]}(z) - (1/24)(x^2 + y^2)C_{2,s}^{[3]}(z) + \dots]. \quad (15.3.70)$$

We see that  $\mathbf{B}$  is primarily the field of a (normal) quadrupole with a profile given by  $Q(z) = 2C_{2,s}^{[0]}(z)$ . See Exercise 1.5.9. As long as  $C_{2,s}^{[0]}(z)$  is nearly constant,  $C_{2,s}^{[1]}(z)$  and higher derivatives of  $C_{2,s}^{[0]}(z)$  will be small, and therefore the other field components will be small. Such would be the case for the field of a (normal) quadrupole where the field is primarily of the form given by (1.5.62) through (1.5.64) and only has  $z$  components in the fringe-field regions at each end where  $C_{2,s}^{[0]}(z)$  is changing. Again, the transfer map for a real quadrupole must contain nonlinear terms because  $C_{2,s}^{[0]}(z)$  must have nonzero derivatives in the fringe-field regions. We close the discussion of the  $m = 2$  case by remarking that the  $C_{2,c}^{[0]}(z)$  related terms describe the field of a skew quadrupole. See Exercise 4.2.

### 15.3.4 Magnetic Field Expansions for the General Case

#### General Results

Since  $\psi$  is a harmonic function and the operators  $\partial_x, \partial_y, \partial_z$  commute with  $\nabla^2$ , it follows from (2.6) that the (Cartesian) components of  $\mathbf{B}$  must also be harmonic functions. Consequently each component of  $\mathbf{B}$  must have a cylindrical multipole expansion of the form (3.33). Indeed, if  $\psi$  has the expansion (3.33), then it can be shown that the components of  $\mathbf{B}$  have the

expansions

$$\begin{aligned}
B_x = -\partial_x \psi(x, y, z) &= \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! \ell!} C_{1,c}^{[2\ell]}(z) \rho^{2\ell} \\
&+ \cos(\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{2}{2^{2\ell} \ell! (\ell+1)!} [C_{2,c}^{[2\ell]}(z) - (1/4) C_0^{[2\ell+2]}(z)] \rho^{2\ell+1} \\
&+ \sum_{m=2}^{\infty} \cos(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} \{(m+1) C_{m+1,c}^{[2\ell]}(z) - [1/(4m)] C_{m-1,c}^{[2\ell+2]}(z)\} \rho^{2\ell+m} \\
&+ \sin(\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{2}{2^{2\ell} \ell! (\ell+1)!} C_{2,s}^{[2\ell]}(z) \rho^{2\ell+1} \\
&+ \sum_{m=2}^{\infty} \sin(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} \{(m+1) C_{m+1,s}^{[2\ell]}(z) - [1/(4m)] C_{m-1,s}^{[2\ell+2]}(z)\} \rho^{2\ell+m},
\end{aligned} \tag{15.3.71}$$

$$\begin{aligned}
B_y = -\partial_y \psi(x, y, z) &= \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! \ell!} C_{1,s}^{[2\ell]}(z) \rho^{2\ell} \\
&+ \cos(\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{2}{2^{2\ell} \ell! (\ell+1)!} C_{2,s}^{[2\ell]}(z) \rho^{2\ell+1} \\
&+ \sum_{m=2}^{\infty} \cos(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} \{(m+1) C_{m+1,s}^{[2\ell]}(z) + [1/(4m)] C_{m-1,s}^{[2\ell+2]}(z)\} \rho^{2\ell+m} \\
&- \sin(\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{2}{2^{2\ell} \ell! (\ell+1)!} [C_{2,c}^{[2\ell]}(z) + (1/4) C_0^{[2\ell+2]}(z)] \rho^{2\ell+1} \\
&- \sum_{m=2}^{\infty} \sin(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} \{(m+1) C_{m+1,c}^{[2\ell]}(z) + [1/(4m)] C_{m-1,c}^{[2\ell+2]}(z)\} \rho^{2\ell+m},
\end{aligned} \tag{15.3.72}$$

$$\begin{aligned}
B_z = -\partial_z \psi(x, y, z) &= \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! \ell!} C_0^{[2\ell+1]}(z) \rho^{2\ell} \\
&+ \sum_{m=1}^{\infty} \cos(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,c}^{[2\ell+1]}(z) \rho^{2\ell+m} \\
&+ \sum_{m=1}^{\infty} \sin(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,s}^{[2\ell+1]}(z) \rho^{2\ell+m}.
\end{aligned} \tag{15.3.73}$$

See Appendix H.<sup>10</sup>

---

<sup>10</sup>That the components of  $\mathbf{B}$  should depend on the coordinates  $\rho, \phi, z$  and the coefficients  $C_{m,c}^{[n]}, C_{m,s}^{[n]}$  in some fashion is a consequence of (2.6). That they should do so in the particular combinations (3.71) through

### Leading Behavior in Body

Let us compute the leading behavior of the various components of  $\mathbf{B}$ . For the monopole (solenoid) case, for which  $m = 0$ , we assume  $C_0^{[1]}(z)$  is *constant* and all other coefficients are zero. Then we find from (3.71) through (3.73) the results

$$B_x^0 = 0, \quad (15.3.74)$$

$$B_y^0 = 0, \quad (15.3.75)$$

$$B_z^0 = C_0^{[1]}. \quad (15.3.76)$$

For the dipole ( $m = 1$ ) cases, assuming  $C_{1,c}^{[0]}(z)$  or  $C_{1,s}^{[0]}(z)$  is constant and all other coefficients are zero, we find the results

$$B_x^{1,c} = C_{1,c}^{[0]}, \quad (15.3.77)$$

$$B_y^{1,c} = 0, \quad (15.3.78)$$

$$B_z^{1,c} = 0; \quad (15.3.79)$$

$$B_x^{1,s} = 0, \quad (15.3.80)$$

$$B_y^{1,s} = C_{1,s}^{[0]}, \quad (15.3.81)$$

$$B_z^{1,s} = 0. \quad (15.3.82)$$

For the  $m \geq 2$  cases, assuming  $C_{m,c}^{[0]}(z)$  or  $C_{m,s}^{[0]}(z)$  is constant and therefore all other coefficients  $C_{m,c}^{[n]}(z)$  and  $C_{m,s}^{[n]}(z)$  vanish for  $n > 0$ , we find the results

$$B_x^{m,c} = m \cos[(m-1)\phi] \rho^{m-1} C_{m,c}^{[0]} = m C_{m,c}^{[0]} \Re[(x+iy)^{m-1}], \quad (15.3.83)$$

$$B_y^{m,c} = -m \sin[(m-1)\phi] \rho^{m-1} C_{m,c}^{[0]} = -m C_{m,c}^{[0]} \Im[(x+iy)^{m-1}], \quad (15.3.84)$$

$$B_z^{m,c} = 0; \quad (15.3.85)$$

$$B_x^{m,s} = m \sin[(m-1)\phi] \rho^{m-1} C_{m,s}^{[0]} = m C_{m,s}^{[0]} \Im[(x+iy)^{m-1}], \quad (15.3.86)$$

$$B_y^{m,s} = m \cos[(m-1)\phi] \rho^{m-1} C_{m,s}^{[0]} = m C_{m,s}^{[0]} \Re[(x+iy)^{m-1}], \quad (15.3.87)$$

$$B_z^{m,s} = 0. \quad (15.3.88)$$

Here we have used (3.3) and (3.4). Note that, if the relations (3.83) through (3.88) are evaluated for  $m = 1$ , they reproduce the results (3.77) through (3.82). They therefore actually hold for all  $m > 0$ .

When both  $C_{m,c}^{[0]}(z)$  and  $C_{m,s}^{[0]}(z)$  are present and constant, and all other coefficients are zero, we may write

$$\mathbf{B}^m = \mathbf{B}^{m,c} + \mathbf{B}^{m,s}. \quad (15.3.89)$$

---

(3.73) is in part a consequence of the components of  $\mathbf{B}$  being harmonic. See Exercise 3.4.

With this notation, the relations (3.77), (3.78), (3.80), (3.81), (3.83), (3.84), (3.86), and (3.87) can be rewritten in the compact (but complex) form

$$B_y^m + iB_x^m = m(C_{m,s}^{[0]} + iC_{m,c}^{[0]})(x + iy)^{m-1}. \quad (15.3.90)$$

It must be emphasized, however, that (3.90) holds only in the *body* of a pure multipole magnet, and not in the fringe-field regions at the ends. Moreover, fringe fields and their nonlinear contributions to transfer maps are inescapable consequences of Maxwell's equations.

### 15.3.5 Symmetry and Allowed and Forbidden Multipoles

There are restrictions on multipole content dictated by symmetry conditions. As a first example, consider a rectangular bending magnet such as that shown in Figures 1.6.1 and 1.6.2. Suppose the magnet is rotated by  $180^\circ$  about the  $z$  axis, and simultaneously the strength of the magnet is reversed in sign (so that  $\psi$  is replaced by  $-\psi$ ). Assuming perfect symmetry, doing so should produce the same magnetic field as before. Correspondingly, the scalar potential  $\psi$  should remain unchanged. Suppose  $\psi$  as given by (3.33) is regarded as a function of  $\rho$ ,  $\phi$ , and  $z$ . Then we demand that

$$-\psi(\rho, \phi - \pi, z) = \psi(\rho, \phi, z). \quad (15.3.91)$$

Inspection of (3.33) shows that the requirement (3.91) forces all the coefficients  $C_{m,c}^{[2\ell]}$  and  $C_{m,s}^{[2\ell]}$  to be zero save those for which  $m = 1, 3, 5, \dots$ .

Next consider a quadrupole magnet. Again assuming perfect symmetry, its magnet field should be unchanged if it is rotated by  $90^\circ$  about the  $z$  axis, and simultaneously the strength of the magnet is reversed in sign. In this case we demand that

$$-\psi(\rho, \phi - \pi/2, z) = \psi(\rho, \phi, z). \quad (15.3.92)$$

Now inspection of (3.33) shows that the requirement (3.92) forces all the coefficients  $C_{m,c}^{[2\ell]}$  and  $C_{m,s}^{[2\ell]}$  to be zero save those for which  $m = 2, 6, 10, \dots$ .

Finally, consider a perfectly symmetric  $2n$ -pole magnet for  $n = 1, 2, 3, \dots$ . In this case a rotation by  $(360/2n)^\circ$  and reversing the strength should leave the field unchanged. Now we conclude all multipole coefficients must vanish save possibly those for which

$$m = n(2j + 1) \text{ with } j = 0, 1, 2, 3, \dots \quad (15.3.93)$$

In addition to rotational symmetry, there is the consideration of *midplane* symmetry in which one observes what happens when  $y \rightarrow -y$ , or equivalently, in view of (2.13) and (2.14),  $\phi \rightarrow -\phi$ . From (3.2) and (3.37) through (3.39) we see that the functions  $\Psi_0$ ,  $\Psi_{m,c}$ , and  $\Psi_{m,s}$  are invariant under this operation. It follows from (3.41) through (3.43) that there are the relations

$$\psi_0(x, -y, z) = \psi_0(x, y, z), \quad (15.3.94)$$

$$\psi_{m,c}(x, -y, z) = \psi_{m,c}(x, y, z), \quad (15.3.95)$$

$$\psi_{m,s}(x, -y, z) = -\psi_{m,s}(x, y, z). \quad (15.3.96)$$

We will say that a magnetic field  $\mathbf{B}$  has *midplane* symmetry if it arises from a scalar potential that only has terms of the form  $\psi_{m,s}$ ,

$$\mathbf{B} = -\nabla \sum_{m=1}^{\infty} \psi_{m,s}. \quad (15.3.97)$$

Such a field is also said to be *normal* or produced by normal multipoles. Correspondingly, fields of the form

$$\mathbf{B} = -\nabla \sum_{m=1}^{\infty} \psi_{m,c} \quad (15.3.98)$$

are said to be *skew* or produced by skew multipoles. From (2.6) we see that normal fields have the symmetry property

$$\mathbf{B}_x^{\text{normal}}(x, -y, z) = -\mathbf{B}_x^{\text{normal}}(x, y, z), \quad (15.3.99)$$

$$\mathbf{B}_y^{\text{normal}}(x, -y, z) = \mathbf{B}_y^{\text{normal}}(x, y, z), \quad (15.3.100)$$

$$\mathbf{B}_z^{\text{normal}}(x, -y, z) = -\mathbf{B}_z^{\text{normal}}(x, y, z); \quad (15.3.101)$$

and skew fields have the property

$$\mathbf{B}_x^{\text{skew}}(x, -y, z) = \mathbf{B}_x^{\text{skew}}(x, y, z), \quad (15.3.102)$$

$$\mathbf{B}_y^{\text{skew}}(x, -y, z) = -\mathbf{B}_y^{\text{skew}}(x, y, z), \quad (15.3.103)$$

$$\mathbf{B}_z^{\text{skew}}(x, -y, z) = \mathbf{B}_z^{\text{skew}}(x, y, z). \quad (15.3.104)$$

The same conclusions can be drawn from (3.71) through (3.73). Observe that, by these definitions, the field arising from  $\psi_0$ , for example the field of a solenoid, is also skew.

We note that, multipole by multipole, skew elements are related to normal elements and vice versa by rotations about the  $z$  axis. From (3.42) and (3.43) we find the relations

$$\psi_{m,c}[\rho, \phi - \pi/(2m), z] = \sin(m\phi)\Psi_{m,c}(\rho, z), \quad (15.3.105)$$

$$\psi_{m,s}[\rho, \phi - \pi/(2m), z] = -\cos(m\phi)\Psi_{m,s}(\rho, z). \quad (15.3.106)$$

We see that a skew element is converted into a normal element, but with on-axis gradients  $C_{m,c}^{[2\ell]}(z)$ ; and a normal element is converted into a skew element, but with on-axis gradients  $-C_{m,s}^{[2\ell]}(z)$ .

Finally, we observe that a similar discussion could be given to the possible symmetry operation  $x \rightarrow -x$ , for which the properties of  $\psi_{m,c}$  and  $\psi_{m,s}$  are interchanged.

### 15.3.6 Relation between Harmonic Polynomials in Spherical and Cylindrical Coordinates

Equation (2.26) defined Harmonic polynomials in terms of a radius  $r$  and the spherical harmonics  $Y_\ell^m(\theta, \phi)$ , thereby providing a description in terms of spherical coordinates; and the relations (2.27) through (2.29) illustrate that the results of this definition are indeed

polynomials in the Cartesian coordinates  $x, y, z$ . Suppose these polynomials are re-expressed in terms of cylindrical coordinates. This could be done using the  $z$  axis as the axis of the cylinder as in the relations (2.13) and (2.14), or with some other axis taken to be the axis of the cylinder as in Exercise 1.5.4. What would be the appearance of such expansions? And how are such expansions related to the cylindrical harmonic expansions found in (3.19) and (3.33)? The answers to these questions are the subject of this section.

Suppose the substitutions (2.13) and (2.14) are made in the relations (2.27) through (2.29). So doing and employing (3.1) yields the results

$$H_0^0(\mathbf{r}) = 1/\sqrt{4\pi}; \quad (15.3.107)$$

$$\begin{aligned} H_1^1(\mathbf{r}) &= -\sqrt{3/(8\pi)}\rho \exp(i\phi), \\ H_1^0(\mathbf{r}) &= \sqrt{3/(4\pi)}z, \\ H_1^{-1}(\mathbf{r}) &= \sqrt{3/(8\pi)}\rho \exp(-i\phi); \end{aligned} \quad (15.3.108)$$

$$\begin{aligned} H_2^2(\mathbf{r}) &= \sqrt{15/(32\pi)}\rho^2 \exp(2i\phi), \\ H_2^1(\mathbf{r}) &= -\sqrt{15/(8\pi)}z\rho \exp(i\phi), \\ H_2^0(\mathbf{r}) &= \sqrt{5/(16\pi)}(2z^2 - \rho^2), \\ H_2^{-1}(\mathbf{r}) &= \sqrt{15/(8\pi)}z\rho \exp(-i\phi), \\ H_2^{-2}(\mathbf{r}) &= \sqrt{15/(32\pi)}\rho^2 \exp(-2i\phi). \end{aligned} \quad (15.3.109)$$

How are these results related to cylindrical harmonic expansions? For the expansion (3.19) consider the special case in which  $C_m^0(z) \neq 0$  for only one value of  $m$ , and suppose for this value of  $m$  that  $C_m^{[0]}(z)$  has the special form

$$C_m^{[0]}(z) = a_n^m f_n(z) \quad (15.3.110)$$

with

$$f_n(z) = z^n. \quad (15.3.111)$$

Call the result  $\psi_n^m(x, y, z)$ . That is, make the Ansatz

$$\psi_n^m(x, y, z) = -a_n^m \rho^{|m|} \exp(im\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{|m|!}{2^{2\ell} \ell! (\ell + |m|)!} f_n^{[2\ell]}(z) \rho^{2\ell}. \quad (15.3.112)$$

What are the properties of this Ansatz?

We begin by observing that the combination  $f_n^{[2\ell]}(z) \rho^{2\ell+|m|}$  is a monomial in the variables  $z$  and  $\rho$  for each value of  $\ell$ , with  $2\ell \leq n$ , and all these monomials are of degree  $n + |m|$ . For example, there are the relations

$$f_0^{[0]} = 1, \quad (15.3.113)$$

$$f_0^{[2]} = 0; \quad (15.3.114)$$

$$f_1^{[0]} = z, \quad (15.3.115)$$

$$f_1^{[2]} = 0; \quad (15.3.116)$$

$$f_2^{[0]} = z^2, \quad (15.3.117)$$

$$f_2^{[2]} = 2, \quad (15.3.118)$$

$$f_2^{[4]} = 0; \quad (15.3.119)$$

$$f_3^{[0]} = z^3, \quad (15.3.120)$$

$$f_3^{[2]} = 6z, \quad (15.3.121)$$

$$f_3^{[4]} = 0, \text{ etc.} \quad (15.3.122)$$

It follows that there are the results

$$f_0^{[2\ell]}(z)\rho^{2\ell} = 1 \text{ when } \ell = 0; \quad (15.3.123)$$

$$f_1^{[2\ell]}(z)\rho^{2\ell} = z \text{ when } \ell = 0; \quad (15.3.124)$$

$$f_2^{[2\ell]}(z)\rho^{2\ell} = z^2 \text{ when } \ell = 0, \quad (15.3.125)$$

$$f_2^{[2\ell]}(z)\rho^{2\ell} = 2\rho^2 \text{ when } \ell = 1; \quad (15.3.126)$$

$$f_3^{[2\ell]}(z)\rho^{2\ell} = z^3 \text{ when } \ell = 0, \quad (15.3.127)$$

$$f_3^{[2\ell]}(z)\rho^{2\ell} = 6z\rho^2 \text{ when } \ell = 1, \text{ etc.} \quad (15.3.128)$$

Now let us use these results to work out the first few  $\psi_n^m(x, y, z)$ . So doing gives the relations

$$\psi_0^0(x, y, z) = -a_0^0 \propto H_0^0(\mathbf{r}); \quad (15.3.129)$$

$$\psi_0^1(x, y, z) = -a_0^1 \rho \exp(i\phi) \propto H_1^1(\mathbf{r}), \quad (15.3.130)$$

$$\psi_1^0(x, y, z) = -a_1^0 z \propto H_1^0(\mathbf{r}), \quad (15.3.131)$$

$$\psi_0^{-1}(x, y, z) = -a_0^{-1} \rho \exp(-i\phi) \propto H_1^{-1}(\mathbf{r}); \quad (15.3.132)$$

$$\psi_0^2(x, y, z) = -a_0^2 \rho^2 \exp(2i\phi) \propto H_2^2(\mathbf{r}), \quad (15.3.133)$$

$$\psi_1^1(x, y, z) = -a_1^1 z \rho \exp(i\phi) \propto H_2^1(\mathbf{r}), \quad (15.3.134)$$

$$\psi_2^0(x, y, z) = -a_2^0 (z^2 - \rho^2/2) \propto H_2^0(\mathbf{r}), \quad (15.3.135)$$

$$\psi_1^{-1}(x, y, z) = -a_1^{-1} z \rho \exp(-i\phi) \propto H_2^{-1}(\mathbf{r}), \quad (15.3.136)$$

$$\psi_0^{-2}(x, y, z) = -a_0^{-2} \rho^2 \exp(-2i\phi) \propto H_2^{-2}(\mathbf{r}). \quad (15.3.137)$$

These examples illustrate that the Ansatz specified by (3.111) and (3.112) produces the harmonic polynomials expressed in cylindrical coordinates. There is the general relation

$$\psi_n^m(x, y, z) \propto H_{n+|m|}^m(\mathbf{r}) \quad (15.3.138)$$

where both sides of (3.138) are to be expressed in terms of cylindrical coordinates. Moreover, we note that the functions (3.110) with (3.111) provide a basis for the set of functions  $C_m^{[0]}(z)$ . Therefore, as we already know from other arguments, harmonic polynomials form a basis for the set of harmonic functions.

## Exercises

**15.3.1.** Verify (3.50) and (3.51).

**15.3.2.** Verify (3.52) through (3.56).

**15.3.3.** Given (3.71) through (3.73), verify (3.74) through (3.90).

**15.3.4.** Suppose that some beam line element is described by the magnetic scalar potential  $\psi(x, y, z) = \psi(\rho, \phi, z)$  and suppose this element is rotated by angle  $\phi'$  about the  $z$  axis. With regard to sign convention, look *down* the  $z$  axis in the direction of decreasing  $z$  (so that the arrow pointing along the positive  $z$  axis is aimed at your face) and suppose the rotation is made in the counterclockwise direction by an angle  $\phi'$  when  $\phi'$  is positive. (See, for illustrations, Figures 1.6.1 and 16.2.1.) Let  $\hat{\psi}$  be the magnetic scalar potential for this rotated element. Show that there is the relation

$$\hat{\psi}(\rho, \phi, z) = \psi(\rho, \phi - \phi', z). \quad (15.3.139)$$

Suppose that  $\psi$  has the expansion given by the first line of (3.40) and that  $\hat{\psi}$  has an expansion of the form

$$\hat{\psi}(x, y, z) = \hat{\Psi}_0(\rho, z) + \sum_{m=1}^{\infty} \hat{\Psi}_{m,c}(\rho, z) \cos m\phi + \sum_{m=1}^{\infty} \hat{\Psi}_{m,s}(\rho, z) \sin m\phi. \quad (15.3.140)$$

Show that

$$\hat{\Psi}_0(\rho, z) = \Psi_0(\rho, z), \quad (15.3.141)$$

$$\hat{\Psi}_{m,c}(\rho, z) = \cos(m\phi') \Psi_{m,c}(\rho, z) - \sin(m\phi') \Psi_{m,s}(\rho, z), \quad (15.3.142)$$

$$\hat{\Psi}_{m,s}(\rho, z) = \sin(m\phi') \Psi_{m,c}(\rho, z) + \cos(m\phi') \Psi_{m,s}(\rho, z). \quad (15.3.143)$$

With regard to on-axis gradients, suppose the original on-axis gradients are the functions  $C_{m,\alpha}^{[n]}(z)$ . See (3.33). Suppose that the on-axis gradients for the rotated element are the functions  $\hat{C}_{m,\alpha}^{[n]}(z)$ . Show that there the relations

$$\hat{C}_0^{[n]}(z) = C_0^{[n]}(z), \quad (15.3.144)$$

$$\hat{C}_{m,c}^{[n]}(z) = \cos(m\phi') C_{m,c}^{[n]}(z) - \sin(m\phi') C_{m,s}^{[n]}(z), \quad (15.3.145)$$

$$\hat{C}_{m,s}^{[n]}(z) = \sin(m\phi') C_{m,c}^{[n]}(z) + \cos(m\phi') C_{m,s}^{[n]}(z). \quad (15.3.146)$$

**15.3.5.** Show that the definitions (3.28) and (3.31) can be inverted to give the relations

$$C_m^{[0]}(z) = (1/2)[C_{m,c}^{[0]}(z) - iC_{m,s}^{[0]}(z)] \text{ for } m \geq 1, \quad (15.3.147)$$

$$C_{-m}^{[0]}(z) = (1/2)[C_{m,c}^{[0]}(z) + iC_{m,s}^{[0]}(z)] \text{ for } m \geq 1. \quad (15.3.148)$$

**15.3.6.** The relation (3.5) displays the Laplacian in cylindrical coordinates. Verify that one may write

$$\nabla^2 = \nabla_{\perp}^2 + \partial^2/\partial z^2 \quad (15.3.149)$$

where

$$\begin{aligned} \nabla_{\perp}^2 &= \partial^2/\partial x^2 + \partial^2/\partial y^2 = (1/\rho)(\partial/\partial\rho)(\rho\partial/\partial\rho) + (1/\rho^2)(\partial^2/\partial\phi^2) \\ &= \partial^2/\partial\rho^2 + (1/\rho)\partial/\partial\rho + (1/\rho^2)(\partial^2/\partial\phi^2). \end{aligned} \quad (15.3.150)$$

Define functions  $\chi_c$  and  $\chi_s$  by the rules

$$\chi_c = \rho^{2\ell+m} \cos m\phi, \quad (15.3.151)$$

$$\chi_s = \rho^{2\ell+m} \sin m\phi. \quad (15.3.152)$$

Show that they are homogeneous polynomials of order  $(2\ell + m)$  in the variables  $x$  and  $y$ , and have the property

$$\nabla_{\perp}^2 \chi_{\alpha} = 4\ell(\ell + m)\chi_{\alpha}/\rho^2 \quad (15.3.153)$$

where  $\alpha = c, s$ . Use this property to show that  $\psi$  as given by (3.33) satisfies the Laplace equation (2.7).

**15.3.7.** Review Exercise 3.4. Next, note the resemblance between the functional forms of  $\psi$  as given by (3.33) and the Cartesian components of the negative gradient of  $\psi$  as given by (3.71) through (3.73). Looking forward, observe that the Cartesian components of the associated vector potential in a Coulomb gauge, see (5.89) through (5.94), also have analogous functional forms. Why should this be?

**15.3.8.** Suppose the Fourier coefficient  $G_m(k)$  appearing in (3.7) has the form

$$G_m(k) = \lambda\delta_{m,m'}\delta(k)/k^{|m|}. \quad (15.3.154)$$

Show that in this case there is the relation

$$\int_{-\infty}^{\infty} dk k^{n+|m|} G_m(k) \exp(ikz) = \lambda\delta_{m,m'}\delta_{n,0}. \quad (15.3.155)$$

Correspondingly, show that use of (3.15) in this case yields the result

$$C_m^{[n]}(z) = i^n(1/2)^{|m|}(1/|m|!)\lambda\delta_{m,m'}\delta_{n,0}. \quad (15.3.156)$$

Finally, employ (3.156) in (3.19) to show that in this case

$$\psi(x, y, z) = -[\lambda/(2^{|m'|}|m'|!)]\rho^{|m'|} \exp(im'\phi). \quad (15.3.157)$$

Verify, by direct calculation, that  $\psi$  as given by (3.157) is harmonic.

**15.3.9.** Your task in this exercise is to derive (3.57) and (3.58).

**15.3.10.** Show that the scalar potential  $\psi_{1,c}$  produces a skew dipole magnetic field that is primarily in the  $x$  direction. 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. Compare your results with those of Exercise 1.5.7.

**15.3.11.** Show that the scalar potential  $\psi_{2,c}$  produces a skew quadrupole magnetic field. 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. Compare your results with those of Exercise 1.5.9.

**15.3.12.** Review Equations (3.1) through (3.4). Define functions  $u_m$  and  $v_m$  by the rules

$$u_m(x, y) = \Re[(x + iy)^m], \quad (15.3.158)$$

$$v_m(x, y) = \Im[(x + iy)^m], \quad (15.3.159)$$

so that

$$(x + iy)^m = u_m + iv_m. \quad (15.3.160)$$

Since  $z^m = (x + iy)^m$  is an *analytic* function of the complex variable  $z$ , verify that each pair  $u_m, v_m$  satisfies the *Cauchy-Riemann* equations

$$\partial_x u_m = \partial_y v_m \text{ and } \partial_y u_m = -\partial_x v_m. \quad (15.3.161)$$

Using (3.161) verify that both  $u_m$  and  $v_m$  are harmonic functions,

$$(\partial_x^2 + \partial_y^2)u_m = (\partial_x^2 + \partial_y^2)v_m = 0. \quad (15.3.162)$$

Verify that both  $u_m$  and  $v_m$  are *polynomials*, and work/write out the first few. In fact, as a point of culture, they are called *Šiljac polynomials*. See the reference to his book at the end of this chapter.

## 15.4 Determination of the Vector Potential: Azimuthal-Free Gauge

Although the description of magnetic fields  $\mathbf{B}$  in terms of the scalar potential  $\psi$  is convenient, it is not what we ultimately need. What we need, if we wish to exploit the symplectic structure of Hamiltonian dynamics, is a description of  $\mathbf{B}$  in terms of a vector potential  $\mathbf{A}$  such that

$$\mathbf{B} = \nabla \times \mathbf{A}. \quad (15.4.1)$$

We recall that in cylindrical coordinates the radial and azimuthal components of a vector, in this case the vector potential  $\mathbf{A}$ , are related to the transverse Cartesian components by equations (2.22) through (2.25). Since there is gauge freedom in the choice of a vector

potential, it is sometimes convenient, if possible, to work in a gauge for which the azimuthal component vanishes,

$$A_\phi = 0. \quad (15.4.2)$$

According to (2.24) and (2.25), in this gauge we have the relations

$$A_x = A_\rho \cos \phi, \quad (15.4.3)$$

$$A_y = A_\rho \sin \phi. \quad (15.4.4)$$

We call this gauge the *azimuthal-free* gauge. We will see that it is possible to find a vector potential in the azimuthal-free gauge for the magnetic field of any multipole save for  $m = 0$ . (In subsequent sections, we will find a vector potential for the  $m = 0$  case in a Coulomb gauge.)

### 15.4.1 Derivation

We will employ the notation (3.38) and (3.39). With this notation in mind, define vector potentials  $\mathbf{A}^{m,c}$  and  $\mathbf{A}^{m,s}$  by the rules

$$A_\rho^{m,c} = \frac{\sin(m\phi)}{m} \rho \frac{\partial}{\partial z} \Psi_{m,c}, \quad (15.4.5)$$

$$A_\phi^{m,c} = 0, \quad (15.4.6)$$

$$A_z^{m,c} = -\frac{\sin(m\phi)}{m} \rho \frac{\partial}{\partial \rho} \Psi_{m,c}; \quad (15.4.7)$$

$$A_\rho^{m,s} = -\frac{\cos(m\phi)}{m} \rho \frac{\partial}{\partial z} \Psi_{m,s}, \quad (15.4.8)$$

$$A_\phi^{m,s} = 0, \quad (15.4.9)$$

$$A_z^{m,s} = \frac{\cos(m\phi)}{m} \rho \frac{\partial}{\partial \rho} \Psi_{m,s}. \quad (15.4.10)$$

(Note that these definitions fail for the  $m = 0$  case. See Exercise 4.5.) Then, it is easily verified that

$$\nabla \times \mathbf{A}^{m,c} = -\nabla \psi_{m,c}, \quad (15.4.11)$$

$$\nabla \times \mathbf{A}^{m,s} = -\nabla \psi_{m,s}. \quad (15.4.12)$$

See Exercise 4.6. Correspondingly, if we define  $\mathbf{A}$  by the sum

$$\mathbf{A} = \sum_{m=1}^{\infty} \mathbf{A}^{m,c} + \sum_{m=1}^{\infty} \mathbf{A}^{m,s}, \quad (15.4.13)$$

we have, by linearity and again omitting the  $m = 0$  term, the result,

$$\nabla \times \mathbf{A} = -\nabla \psi = \mathbf{B}. \quad (15.4.14)$$

At this point we make an important observation. We know that  $\mathbf{B}$  falls to zero for large  $|z|$  because for large  $|z|$  the observation point must be well outside the beam-line element

in question. From (3.71) through (3.73) and the definitions of  $\Psi_{m,c}$  and  $\Psi_{m,s}$ , we see that these  $\Psi$  must also fall to zero for large  $|z|$ . Correspondingly, from (4.5) through (4.7), we see that  $\mathbf{A}^{m,c}$  and  $\mathbf{A}^{m,s}$  must fall to zero for large  $|z|$ . This behavior is important because it guarantees that, for the azimuthal-free gauge, the canonical and mechanical momenta will be *equal* far outside any beam-line element. See (1.5.30).

We close this subsection by presenting explicit formulas for the cylindrical and Cartesian components of  $\mathbf{A}^{m,c}$  and  $\mathbf{A}^{m,s}$  for general  $m \geq 1$ . From (4.5) through (4.10) and the expansions (3.38) and (3.39) we find the results

$$A_\rho^{m,c} = -\frac{\sin(m\phi)}{m} \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,c}^{[2\ell+1]}(z) \rho^{2\ell+m+1}, \quad (15.4.15)$$

$$A_\phi^{m,c} = 0, \quad (15.4.16)$$

$$A_z^{m,c} = \frac{\sin(m\phi)}{m} \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(2\ell+m)(m!)}{2^{2\ell} \ell! (\ell+m)!} C_{m,c}^{[2\ell]}(z) \rho^{2\ell+m}; \quad (15.4.17)$$

$$A_\rho^{m,s} = \frac{\cos(m\phi)}{m} \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,s}^{[2\ell+1]}(z) \rho^{2\ell+m+1}, \quad (15.4.18)$$

$$A_\phi^{m,s} = 0, \quad (15.4.19)$$

$$A_z^{m,s} = -\frac{\cos(m\phi)}{m} \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(2\ell+m)(m!)}{2^{2\ell} \ell! (\ell+m)!} C_{m,s}^{[2\ell]}(z) \rho^{2\ell+m}. \quad (15.4.20)$$

From (4.3), (4.4), and (4.15) through (4.17) we find the results

$$A_x^{m,c} = -(1/m)x \Im[(x+iy)^m] \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,c}^{[2\ell+1]}(z) (x^2 + y^2)^\ell, \quad (15.4.21)$$

$$A_y^{m,c} = -(1/m)y \Im[(x+iy)^m] \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,c}^{[2\ell+1]}(z) (x^2 + y^2)^\ell, \quad (15.4.22)$$

$$A_z^{m,c} = -(1/m)\Im[(x+iy)^m] \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(2\ell+m)(m!)}{2^{2\ell} \ell! (\ell+m)!} C_{m,c}^{[2\ell]}(z) (x^2 + y^2)^\ell. \quad (15.4.23)$$

From (4.3), (4.4), and (4.18) through (4.20) we find the results

$$A_x^{m,s} = -(1/m)x \Re[(x+iy)^m] \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,s}^{[2\ell+1]}(z) (x^2 + y^2)^\ell, \quad (15.4.24)$$

$$A_y^{m,s} = -(1/m)y \Re[(x+iy)^m] \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,s}^{[2\ell+1]}(z) (x^2 + y^2)^\ell, \quad (15.4.25)$$

$$A_z^{m,s} = -(1/m)\Re[(x+iy)^m] \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(2\ell+m)(m!)}{2^{2\ell}\ell!(\ell+m)!} C_{m,s}^{[n]}(z)(x^2+y^2)^\ell. \quad (15.4.26)$$

Note that (4.21) through (4.26) provide expansions of the vector potential in terms of homogeneous polynomials in the variables  $x, y$  with  $z$ -dependent coefficients  $C_{m,\alpha}^{[n]}(z)$ , and that the *minimum* degree of these polynomials is  $m$ . Therefore, if the design orbit is on the  $z$  axis, as it will be for all beam-line elements not having  $m = 1$  (dipole) content, only a finite number of  $m$  and  $\ell$  values are required to compute the transfer map through some finite order in the Lie generators.

### 15.4.2 Some Simple Examples: $m = 1, 2$

As a first example of a vector potential in the azimuthal-free gauge, suppose all terms in (3.33) vanish save for the (normal) dipole terms  $C_{1,s}^{[n]}(z)$ . Then, using (4.24) through (4.26), we find through terms of degree five that  $\mathbf{A}^{1,s}$  has the expansion

$$A_x^{1,s} = x^2 C_{1,s}^{[1]}(z) - (1/8)(x^4 + x^2y^2)C_{1,s}^{[3]}(z) + \dots, \quad (15.4.27)$$

$$A_y^{1,s} = xy C_{1,s}^{[1]}(z) - (1/8)(x^3y + xy^3)C_{1,s}^{[3]}(z) + \dots, \quad (15.4.28)$$

$$\begin{aligned} A_z^{1,s} &= -xC_{1,s}^{[0]}(z) + (3/8)(x^3 + xy^2)C_{1,s}^{[2]}(z) \\ &\quad - (5/192)(x^5 + 2x^3y^2 + xy^4)C_{1,s}^{[4]}(z) + \dots. \end{aligned} \quad (15.4.29)$$

Note that the lowest order terms in (4.27) through (4.29) agree with (1.5.77) if we make the identification  $B = C_{1,s}^{[0]}$ . However, we know that  $C_{1,s}^{[0]}(z)$  must depend on  $z$  because the on-axis gradients must vanish far outside any magnet. Therefore the functions  $C_{1,s}^{[1]}(z)$ ,  $C_{1,s}^{[2]}(z)$ ,  $C_{1,s}^{[3]}(z)$ , etc. must be nonzero, at least near the end and fringe-field regions of any (rectangular) dipole magnet. We conclude that, as a consequence of Maxwell's equations, the vector potential must contain terms beyond degree two in the variables  $x, y$ . Correspondingly, as already stated earlier, the transfer map for any real dipole must contain nonlinear terms.

Analogous results hold for the skew case corresponding to nonvanishing  $C_{1,c}^{[n]}(z)$ .

As a second example of a vector potential in the azimuthal-free gauge, suppose all terms in (3.33) vanish save for the (normal) quadrupole terms  $C_{2,s}^{[n]}(z)$ . Then, again using (4.24) through (4.26), we find through terms of degree four that  $\mathbf{A}^{2,s}$  has the expansion

$$A_x^{2,s} = (1/2)(x^3 - xy^2)C_{2,s}^{[1]}(z) + \dots, \quad (15.4.30)$$

$$A_y^{2,s} = -(1/2)(y^3 - yx^2)C_{2,s}^{[1]}(z) + \dots, \quad (15.4.31)$$

$$A_z^{2,s} = -(x^2 - y^2)C_{2,s}^{[0]}(z) + (1/6)(x^4 - y^4)C_{2,s}^{[2]}(z) + \dots. \quad (15.4.32)$$

Note that the lowest order terms in (4.30) through (4.32) agree with (1.5.83) if we make the identification  $Q/2 = C_{2,s}^{[0]}$ . However, we know that  $C_{2,s}^{[0]}(z)$  must depend on  $z$  because the on-axis gradients must vanish far outside any magnet. Therefore the functions  $C_{2,s}^{[1]}(z)$ ,  $C_{2,s}^{[2]}(z)$ , etc. must be nonzero, at least near the end and fringe-field regions of any quadrupole magnet.

We conclude again that, as a consequence of Maxwell's equations, the vector potential must contain terms beyond degree two in the variables  $x, y$ . Correspondingly, the transfer map for any real quadrupole must contain nonlinear terms.

Analogous results hold for the skew case corresponding to nonvanishing  $C_{2,c}^{[n]}(z)$ .

## Exercises

**15.4.1.** Consider, as a model, the field of an iron-dominated dipole with very wide (in the  $x$  direction) pole faces. See Figures 1.6.1 and 1.6.2. Based on symmetry one might imagine that the field of such a magnet would have no  $B_x$  component and, correspondingly,  $\psi$  for such a magnet would have no  $x$  dependence. Let us make the Ansatz

$$\begin{aligned}\psi(y, z) &= \sum_{n=0}^{\infty} (-1)^n [1/(2n+1)!] y^{2n+1} O^{[2n]}(z) \\ &= y O^{[0]}(z) - (1/6)y^3 O^{[2]}(z) + (1/120)y^5 O^{[4]}(z) + \dots\end{aligned}\quad (15.4.33)$$

where  $O^{[0]}(z)$  is, in principle, an arbitrary function, but required in our case to go to zero as  $|z| \rightarrow \infty$ . Show that this  $\psi$  is harmonic. Hint: See Appendix H.

Next, show that this  $\psi$  will produce a magnetic field  $\mathbf{B}^{\text{iwd}}$ , the field of an *infinite-width dipole*, with components

$$B_x^{\text{iwd}} = 0, \quad (15.4.34)$$

$$\begin{aligned}B_y^{\text{iwd}} &= \sum_{n=0}^{\infty} (-1)^n [1/(2n)!] y^{2n} O^{[2n]}(z) \\ &= O^{[0]}(z) - (1/2)y^2 O^{[2]}(z) + (1/24)y^4 O^{[4]}(z) + \dots,\end{aligned}\quad (15.4.35)$$

$$\begin{aligned}B_z^{\text{iwd}} &= \sum_{n=0}^{\infty} (-1)^n [1/(2n+1)!] y^{2n+1} O^{[2n+1]}(z) \\ &= y O^{[1]}(z) - (1/6)y^3 O^{[3]}(z) + (1/120)y^5 O^{[5]}(z) + \dots.\end{aligned}\quad (15.4.36)$$

Thus,  $\mathbf{B}^{\text{iwd}}$  is primarily in the  $y$  direction, has no  $x$  component, but does have a  $z$  component in the fringe-field regions where  $O^{[n]}(z) \neq 0$  for  $n > 0$ .

How is the expansion (4.33) related to a cylindrical multipole expansion? Note the identities

$$y = \rho \sin \phi, \quad (15.4.37)$$

$$y^3 = \rho^3 \sin^3 \phi = \rho^3 [(3/4) \sin \phi - (1/4) \sin 3\phi], \quad (15.4.38)$$

$$y^5 = \rho^5 \sin^5 \phi = \rho^5 [(10/16) \sin \phi - (5/16) \sin 3\phi + (1/16) \sin 5\phi], \text{ etc.} \quad (15.4.39)$$

Show from (3.43) that there are the relations

$$\Psi_{1,s}(\rho, z) = -O^{[0]}(z)\rho + (1/8)O^{[2]}(z)\rho^3 - (1/192)O^{[4]}(z)\rho^5 + \dots, \quad (15.4.40)$$

$$\Psi_{3,s}(\rho, z) = -(1/24)O^{[2]}(z)\rho^3 + (1/384)O^{[4]}(z)\rho^5 + \dots, \quad (15.4.41)$$

$$\Psi_{5,s}(\rho, z) = -(1/1920)O^{[4]}(z)\rho^5 + \dots, \text{ etc}, \quad (15.4.42)$$

Using (3.39), show that there are the expansions

$$\Psi_{1,s}(\rho, z) = -C_{1,s}^{[0]}(z)\rho + (1/8)C_{1,s}^{[2]}(z)\rho^3 - (1/192)C_{1,s}^{[4]}\rho^5 + \dots, \quad (15.4.43)$$

$$\Psi_{3,s}(\rho, z) = -C_{3,s}^{[0]}(z)\rho^3 + (1/16)C_{3,s}^{[2]}(z)\rho^5 + \dots, \quad (15.4.44)$$

$$\Psi_{5,s}(\rho, z) = -C_{5,s}^{[0]}(z)\rho^5 + \dots. \quad (15.4.45)$$

Now derive the relations

$$C_{1,s}^{[n]}(z) = O^{[n]}(z), \quad (15.4.46)$$

$$C_{3,s}^{[n]}(z) = (1/24)O^{[n+2]}(z), \quad (15.4.47)$$

$$C_{5,s}^{[n]}(z) = (1/1920)O^{[n+4]}(z), \text{ etc.} \quad (15.4.48)$$

Thus the infinite-width dipole has a major contribution from  $\psi_{1,s}(x, y, z)$ , and further contributions, in the fringe-field regions, from all the  $\psi_{m,s}(x, y, z)$  with  $m = 3, 5, \dots$ ; and all the relevant  $C_{m,s}^{[n]}(z)$  are determined by the  $O^{[n+m-1]}(z)$ .

What is the azimuthal-free vector potential for this model field? The vector potential for  $\psi_{1,s}$  has already been found. It is given by (4.27) through (4.29). Show that an analogous calculation for  $\psi_{3,s}$  gives the result

$$A_x^{3,s} = (1/3)x(x^3 - 3xy^2)C_{3,s}^{[1]}(z) + \dots, \quad (15.4.49)$$

$$A_y^{3,s} = (1/3)y(x^3 - 3xy^2)C_{3,s}^{[1]}(z) + \dots, \quad (15.4.50)$$

$$A_z^{3,s} = -(x^3 - 3xy^2)C_{3,s}^{[0]}(z) + (5/48)(x^5 - 2x^3y^2 - 3xy^4)C_{3,s}^{[2]}(z) + \dots; \quad (15.4.51)$$

and an analogous calculation for  $\psi_{5,s}$  gives the result

$$A_x^{5,s} = (1/5)x(x^5 - 10x^3y^2 + 5xy^4)C_{5,s}^{[1]}(z) + \dots, \quad (15.4.52)$$

$$A_y^{5,s} = (1/5)y(x^5 - 10x^3y^2 + 5xy^4)C_{5,s}^{[1]}(z) + \dots, \quad (15.4.53)$$

$$A_z^{5,s} = -(x^5 - 10x^3y^2 + 5xy^4)C_{5,s}^{[0]}(z) + \dots. \quad (15.4.54)$$

Add all these vector potentials together and use (4.46) through (4.48) to show that the azimuthal-free vector potential for the model field is given by the relations

$$A_x = x^2O^{[1]}(z) - (1/18)(2x^4 + 3x^2y^2)O^{[3]}(z) + \dots, \quad (15.4.55)$$

$$A_y = xyO^{[1]}(z) - (1/18)(2x^3y + 3xy^3)O^{[3]}(z) + \dots, \quad (15.4.56)$$

$$\begin{aligned} A_z = & -xO^{[0]}(z) + (1/6)(2x^3 + 3xy^2)O^{[2]}(z) \\ & - (1/360)(8x^5 + 20x^3y^2 + 15xy^4)O^{[4]}(z) + \dots. \end{aligned} \quad (15.4.57)$$

We have learned, as inspection of (4.55) through (4.57) illustrates, that in the azimuthal-free gauge the vector potential associated with even a fairly simple magnetic field, such as

that of our infinite-width dipole model, is quite complicated. Perhaps other gauges would give simpler results? Indeed, consider the infinite-width-dipole vector potential  $\mathbf{A}^{\text{iwd}}$  defined by the equations

$$A_x^{\text{iwd}} = \sum_{n=1}^{\infty} (-1)^n [1/(2n)!] y^{2n} O^{[2n-1]}(z) = -(1/2)y^2 O^{[1]}(z) + (1/24)y^4 O^{[3]}(z) + \dots, \quad (15.4.58)$$

$$A_y^{\text{iwd}} = 0, \quad (15.4.59)$$

$$A_z^{\text{iwd}} = -x O^{[0]}(z). \quad (15.4.60)$$

Show that use of this much simpler vector potential also yields the field  $\mathbf{B}^{\text{iwd}}$  given by (4.34) through (4.36). The question of other gauges will be explored in Sections 15.5 and 15.6. Also, see Exercise 6.2.

Finally note that, for the vector potentials given either by (4.55) through (4.57) or by (4.58) through (4.60), the primary component is in the  $z$  direction with additional components in some transverse direction significant only in the fringe-field regions. This feature is advantageous when using  $z$  as the independent variable, see Exercise 1.6.1, because then the  $z$  component of the vector potential appears only outside the square root that is ubiquitous in all canonical and relativistic formulations.

**15.4.2.** Let  $\mathcal{C}$  be a circle in some plane of constant  $z$  and centered on  $x = y = 0$ . Show that for any vector potential  $\mathbf{A}$  in the azimuthal-free gauge there is the result

$$\int_{\mathcal{C}} \mathbf{A} \cdot d\mathbf{r} = 0. \quad (15.4.61)$$

But, by Stokes' theorem, there is also the result

$$\int_{\mathcal{C}} \mathbf{A} \cdot d\mathbf{r} = \int_{\mathcal{D}} (\nabla \times \mathbf{A}) \cdot d\mathbf{S} = \int_{\mathcal{D}} \mathbf{B} \cdot d\mathbf{S} = \int_{\mathcal{D}} B_z dS \quad (15.4.62)$$

where  $\mathcal{D}$  is the disc in the constant  $z$  plane surrounded by  $\mathcal{C}$ . For the monopole case ( $m = 0$ ) we know that the magnetic field is predominantly in the  $z$  direction when  $x, y$  are near zero. Therefore the integral (4.62) cannot vanish. Correspondingly, an  $m = 0$  magnetic field cannot be derived from an azimuthal-free vector potential.

**15.4.3.** The goal of this exercise is to verify (4.11) and (4.12). Begin with the fact that, by construction,  $\psi_{m,c}$  and  $\psi_{m,s}$  are harmonic. Show, using (3.5), that

$$[(1/\rho)(\partial/\partial\rho)(\rho\partial/\partial\rho) + \partial^2/\partial z^2]\Psi_{m,c} = (m/\rho)^2\Psi_{m,c}, \quad (15.4.63)$$

with an analogous result for  $\Psi_{m,s}$ . Using these results, verify (4.11) and (4.12) by employing the formulas for  $\nabla \times$  and  $\nabla$  in cylindrical coordinates.

**15.4.4.** Assume that (3.77) through (3.83) hold in the *body* of a pure multipole. Let

$$\mathbf{B}^m = \nabla \times \mathbf{A}^m \quad (15.4.64)$$

with

$$\mathbf{A}^m = \mathbf{A}^{m,c} + \mathbf{A}^{m,s}. \quad (15.4.65)$$

Show from (4.21) through (4.26) that in this case (the azimuthal-free gauge case)

$$A_x^m = A_y^m = 0, \quad (15.4.66)$$

and

$$\begin{aligned} A_z^m &= C_{m,c}^{[0]} \Im[(x+iy)^m] - C_{m,s}^{[0]} \Re[(x+iy)^m] \\ &= -\Re[(C_{m,s}^{[0]} + iC_{m,c}^{[0]})(x+iy)^m] = \Im[(C_{m,c}^{[0]} - iC_{m,s}^{[0]})(x+iy)^m]. \end{aligned} \quad (15.4.67)$$

Here the quantities  $C_{m,\alpha}^{[0]}$  are assumed to be *constant* ( $z$  independent).

## 15.5 Determination of the Vector Potential: Symmetric Coulomb Gauge

Sometimes it is convenient to work in a *Coulomb* gauge rather than the azimuthal-free gauge. In this section we will find such a vector potential which, for reasons that will become apparent, we will call the *symmetric* Coulomb gauge vector potential. Before doing so, there is an important fact to be noted about Coulomb-gauge vector potentials for source-free magnetic fields. Suppose that  $\hat{\mathbf{A}}$  is a vector potential for  $\mathbf{B}$  that satisfies the Coulomb gauge condition

$$\nabla \cdot \hat{\mathbf{A}} = 0. \quad (15.5.1)$$

We know by construction that

$$\nabla \times (\nabla \times \hat{\mathbf{A}}) = \nabla \times \mathbf{B} = 0. \quad (15.5.2)$$

(Here we have assumed that  $\mathbf{B}$  is source free.) But there is also the vector identity

$$\nabla \times (\nabla \times \hat{\mathbf{A}}) = \nabla(\nabla \cdot \hat{\mathbf{A}}) - \nabla^2 \hat{\mathbf{A}}. \quad (15.5.3)$$

It follows from (5.1) through (5.3) that there is the relation

$$\nabla^2 \hat{\mathbf{A}} = 0. \quad (15.5.4)$$

That is, each *Cartesian* component of a Coulomb-gauge vector potential is a harmonic function. (It need not be true of other components such as spherical and cylindrical components.) This fact will be useful in the next section.

### 15.5.1 The $m = 0$ Case

We now turn to the task of computing vector potentials in a Coulomb gauge. We begin with the  $m = 0$  case, for which there is no azimuthal-free gauge vector potential. In the  $m = 0$  case  $\psi$  takes the form

$$\psi_0(x, y, z) = - \sum_{\ell=0}^{\infty} (-1)^\ell \frac{1}{2^{2\ell} \ell! \ell!} C_0^{[2\ell]}(z) \rho^{2\ell}. \quad (15.5.5)$$

See (3.33). Define a function  $U(\rho, z)$  by the rule

$$U(\rho, z) = (1/2) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! (\ell+1)!} C_0^{[2\ell+1]}(z) \rho^{2\ell}. \quad (15.5.6)$$

Now define a vector potential  $\hat{\mathbf{A}}^0$  by the Ansatz

$$\hat{A}_x^0 = -yU, \quad (15.5.7)$$

$$\hat{A}_y^0 = xU, \quad (15.5.8)$$

$$\hat{A}_z^0 = 0. \quad (15.5.9)$$

We will soon see that this vector potential produces  $\mathbf{B}$  and is in a Coulomb gauge. Because the two transverse components of this vector potential involve the same master function  $U$  in analogous ways, we will refer to this vector potential as the *symmetric*  $m = 0$  Coulomb gauge vector potential.

Let us verify that this vector potential produces  $\mathbf{B}$ . First we have to answer the question

$$(\nabla \times \hat{\mathbf{A}}^0)_z = \partial_x \hat{A}_y^0 - \partial_y \hat{A}_x^0 = -\partial_z \psi_0 = B_z? \quad (15.5.10)$$

From (5.8) we find the result

$$\partial_x \hat{A}_y^0 = U + x \partial_x U. \quad (15.5.11)$$

But, by the chain rule we have the result

$$\partial_x U = (\partial U / \partial \rho)(\partial \rho / \partial x) = (x/\rho)(\partial U / \partial \rho). \quad (15.5.12)$$

Here we have used the relation

$$\partial \rho / \partial x = x/\rho. \quad (15.5.13)$$

See Appendix H. From (5.11) and (5.12) we conclude that

$$\partial_x \hat{A}_y^0 = U + (x^2/\rho)(\partial U / \partial \rho). \quad (15.5.14)$$

Similarly, we find that

$$-\partial_y \hat{A}_x^0 = U + (y^2/\rho)(\partial U / \partial \rho), \quad (15.5.15)$$

and therefore

$$\partial_x \hat{A}_y^0 - \partial_y \hat{A}_x^0 = 2U + \rho(\partial U / \partial \rho). \quad (15.5.16)$$

But, from the definition (5.6), we see that

$$\rho(\partial U / \partial \rho) = -(1/2) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{2\ell}{2^{2\ell} \ell! (\ell+1)!} C_0^{[2\ell+1]}(z) \rho^{2\ell} \quad (15.5.17)$$

and consequently,

$$\begin{aligned} \partial_x \hat{A}_y^0 - \partial_y \hat{A}_x^0 &= 2U + \rho(\partial U / \partial \rho) = (1/2) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{2\ell+2}{2^{2\ell} \ell! (\ell+1)!} C_0^{[2\ell+1]}(z) \rho^{2\ell} \\ &= \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! (\ell+1)!} C_0^{[2\ell+1]}(z) \rho^{2\ell} = -\partial_z \psi_0 = B_z. \end{aligned} \quad (15.5.18)$$

Next examine the question

$$(\nabla \times \hat{\mathbf{A}}^0)_x = \partial_y \hat{A}_z^0 - \partial_z \hat{A}_y^0 = -\partial_z \hat{A}_y^0 = -\partial_x \psi_0 = B_x? \quad (15.5.19)$$

From (5.8) we see that

$$-\partial_z \hat{A}_y^0 = -x(\partial U/\partial z) = -(x/2) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! (\ell+1)!} C_0^{[2\ell+2]}(z) \rho^{2\ell}. \quad (15.5.20)$$

But we also see from the form (5.5) for  $\psi_0$  that

$$\partial_x \psi_0 = (\partial \psi_0 / \partial \rho)(\partial \rho / \partial x) = (x/\rho)(\partial \psi_0 / \partial \rho). \quad (15.5.21)$$

Similarly, and for future use, there is the relation

$$\partial_y \psi_0 = (\partial \psi_0 / \partial \rho)(\partial \rho / \partial y) = (y/\rho)(\partial \psi_0 / \partial \rho). \quad (15.5.22)$$

But from the representation (5.5) we see that

$$\begin{aligned} (1/\rho)(\partial \psi_0 / \partial \rho) &= - \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{2\ell}{2^{2\ell} \ell! \ell!} C_0^{[2\ell]}(z) \rho^{2\ell-2} \\ &= - \sum_{\ell=1}^{\infty} (-1)^{\ell} \frac{2\ell}{2^{2\ell} \ell! \ell!} C_0^{[2\ell]}(z) \rho^{2\ell-2} = - \sum_{n=0}^{\infty} (-1)^{n+1} \frac{(2n+2)}{2^{(2n+2)} (n+1)! (n+1)!} C_0^{[2n+2]}(z) \rho^{2n} \\ &= (1/2) \sum_{n=0}^{\infty} (-1)^n \frac{1}{2^{2n} n! (n+1)!} C_0^{[2n+2]}(z) \rho^{2n}, \end{aligned} \quad (15.5.23)$$

and consequently

$$\partial_x \psi_0 = (x/\rho)(\partial \psi_0 / \partial \rho) = (x/2) \sum_{n=0}^{\infty} (-1)^n \frac{1}{2^{2n} n! (n+1)!} C_0^{[2n+2]}(z) \rho^{2n}. \quad (15.5.24)$$

Comparison of (5.20) and (5.24) shows that (5.19) is satisfied.

The last question to examine is

$$(\nabla \times \hat{\mathbf{A}}^0)_y = \partial_z \hat{A}_x^0 - \partial_x \hat{A}_z^0 = \partial_z \hat{A}_x^0 = -\partial_y \psi_0 = B_y? \quad (15.5.25)$$

From (5.7) we see that

$$\partial_z \hat{A}_x^0 = -y(\partial U/\partial z) = -(y/2) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! (\ell+1)!} C_0^{[2\ell+2]}(z) \rho^{2\ell}. \quad (15.5.26)$$

But from (5.22) and (5.23) we have

$$\partial_y \psi_0 = (y/\rho)(\partial \psi_0 / \partial \rho) = (-y/2) \sum_{n=0}^{\infty} (-1)^n \frac{1}{2^{2n} n! (n+1)!} C_0^{[2n+2]}(z) \rho^{2n}. \quad (15.5.27)$$

Comparison of (5.26) and (5.27) shows that (5.25) is satisfied.

We can also check that  $\hat{\mathbf{A}}^0$  is divergence free. From (5.7) through (5.9) we see that

$$\nabla \cdot \hat{\mathbf{A}}^0 = \partial_x \hat{A}_x^0 + \partial_y \hat{A}_y^0 + \partial_z \hat{A}_z^0 = -y\partial_x U + x\partial_y U. \quad (15.5.28)$$

Now use (5.12) and its  $y$  analog to find the result

$$-y\partial_x U + x\partial_y U = [(-yx/\rho) + (xy/\rho)](\partial U/\partial\rho) = 0. \quad (15.5.29)$$

From (5.6) through (5.8), we see that  $\hat{A}_x^0$  and  $\hat{A}_y^0$  can be written as

$$\hat{A}_x^0 = -\sin(\phi)(1/2) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell}\ell!(\ell+1)!} C_0^{[2\ell+1]}(z)\rho^{2\ell+1} \quad (15.5.30)$$

and

$$\hat{A}_y^0 = \cos(\phi)(1/2) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell}\ell!(\ell+1)!} C_0^{[2\ell+1]}(z)\rho^{2\ell+1}. \quad (15.5.31)$$

Comparison of these expressions with (3.33) shows that both  $\hat{A}_x^0$  and  $\hat{A}_y^0$  are harmonic functions, as expected.

Inserting (2.13) and (2.14) into (5.30) and (5.31) gives the even more explicit results

$$\begin{aligned} \hat{A}_x^0 &= -(y/2) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell}\ell!(\ell+1)!} C_0^{[2\ell+1]}(z)(x^2+y^2)^{\ell} \\ &= -(y/2)[C_0^{[1]} - (1/8)C_0^{[3]}(x^2+y^2) + \dots], \end{aligned} \quad (15.5.32)$$

$$\begin{aligned} \hat{A}_y^0 &= (x/2) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell}\ell!(\ell+1)!} C_0^{[2\ell+1]}(z)(x^2+y^2)^{\ell} \\ &= (x/2)[C_0^{[1]} - (1/8)C_0^{[3]}(x^2+y^2) + \dots], \end{aligned} \quad (15.5.33)$$

$$\hat{A}_z^0 = 0. \quad (15.5.34)$$

From the relation

$$B_z = -\partial_z \psi_0 = \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell}\ell!\ell!} C_0^{[2\ell+1]}(z)\rho^{2\ell}, \quad (15.5.35)$$

we see that

$$C_0^{[1]}(z) = B_z(0,0,z). \quad (15.5.36)$$

Since we assume that  $B_z(0,0,z)$  falls off for large  $|z|$ , the same will be true for  $C_0^{[1]}(z)$  and for the  $C_0^{[2\ell+1]}(z)$ , and hence, according to (5.32) through (5.34), also for  $\hat{\mathbf{A}}^0$ .

For future work it is also useful to have expressions for  $\hat{\mathbf{A}}^0$  in cylindrical components. Using (2.22), (2.23), and (5.7) through (5.9) gives the results

$$\hat{A}_{\rho}^0 = 0, \quad (15.5.37)$$

$$\hat{A}_\phi^0 = \rho U(\rho, z) = (1/2) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{1}{2^{2\ell} \ell! (\ell+1)!} C_0^{[2\ell+1]}(z) \rho^{2\ell+1}, \quad (15.5.38)$$

$$\hat{A}_z^0 = 0. \quad (15.5.39)$$

Note that, in contrast to the azimuthal-free gauge of Section 15.4, this vector potential has *only* an azimuthal component.

### 15.5.2 The $m \geq 1$ Cases

#### Derivation

Let us begin by recalling some notation: Since we will often be dealing with both the skew and normal cases simultaneously, we will use as before the symbol  $\alpha$  to denote either  $c$  or  $s$ . For example, we will use  $\psi_{m,\alpha}$  to denote either  $\psi_{m,c}$  or  $\psi_{m,s}$ . With this convention in mind, the purpose of the present subsection is to find vector potentials  $\hat{\mathbf{A}}^{m,\alpha}$  that are in a Coulomb gauge,

$$\nabla \cdot \hat{\mathbf{A}}^{m,\alpha} = 0, \quad (15.5.40)$$

and also produce the  $\mathbf{B}$  fields associated with the  $\psi_{m,\alpha}$ ,

$$\nabla \times \hat{\mathbf{A}}^{m,\alpha} = -\nabla \psi_{m,\alpha}. \quad (15.5.41)$$

The requirements (5.41), when combined with the relations (4.11) and (4.12), yield the conditions

$$\nabla \times (\hat{\mathbf{A}}^{m,\alpha} - \mathbf{A}^{m,\alpha}) = 0, \quad (15.5.42)$$

from which it follows that there are functions  $\chi_{m,\alpha}$  such that

$$\hat{\mathbf{A}}^{m,\alpha} = \mathbf{A}^{m,\alpha} + \nabla \chi_{m,\alpha}. \quad (15.5.43)$$

Of course, the relations (5.43) are simply gauge transformations. Our strategy will be to find the functions  $\chi_{m,\alpha}$  and then use (5.43) to yield the desired vector potentials.

Upon taking the divergence of both sides of (5.43), and using the Coulomb conditions (5.40), we find that the  $\chi_{m,\alpha}$  must satisfy the equations

$$\nabla^2 \chi_{m,\alpha} = -\nabla \cdot \mathbf{A}^{m,\alpha}. \quad (15.5.44)$$

For an azimuthal-free  $\mathbf{A}$ , see (4.2), we have the relation

$$\nabla \cdot \mathbf{A} = (1/\rho)(\partial/\partial\rho)(\rho A_\rho) + (\partial/\partial z)A_z. \quad (15.5.45)$$

Using the representations (4.5) through (4.10) for the right sides of (5.44) gives the results

$$\nabla \cdot \mathbf{A}^{m,\alpha} = (2/\rho)A_\rho^{m,\alpha}. \quad (15.5.46)$$

Consequently, the  $\chi_{m,\alpha}$  must satisfy the relations

$$\nabla^2 \chi_{m,\alpha} = -(2/\rho)A_\rho^{m,\alpha}. \quad (15.5.47)$$

To find the  $\chi_{m,\alpha}$ , let us make the Ansätze

$$\chi_{m,c} = -(\sin m\phi)d_{m,c}(\rho, z), \quad (15.5.48)$$

$$\chi_{m,s} = (\cos m\phi)d_{m,s}(\rho, z), \quad (15.5.49)$$

where the functions  $d_{m,\alpha}(\rho, z)$  are yet to be determined. From the representations (5.48) and (5.49) we find the relations

$$\nabla^2 \chi_{m,c} = -(\sin m\phi)[(1/\rho)(\partial/\partial\rho)(\rho\partial/\partial\rho) - m^2/\rho^2 + (\partial/\partial z)^2]d_{m,s}(\rho, z), \quad (15.5.50)$$

$$\nabla^2 \chi_{m,s} = (\cos m\phi)[(1/\rho)(\partial/\partial\rho)(\rho\partial/\partial\rho) - m^2/\rho^2 + (\partial/\partial z)^2]d_{m,s}(\rho, z). \quad (15.5.51)$$

See (3.5). Upon using (5.47) and comparing (5.50) and (5.51) with (4.5) and (4.8), we find the relations

$$[(1/\rho)(\partial/\partial\rho)(\rho\partial/\partial\rho) - m^2/\rho^2 + (\partial/\partial z)^2]d_{m,\alpha}(\rho, z) = (-2/m)(\partial/\partial z)\Psi_{m,\alpha}(\rho, z). \quad (15.5.52)$$

Next assume that each  $d_{m,\alpha}(\rho, z)$  has an expansion of the form

$$d_{m,\alpha}(\rho, z) = \sum_{\ell=0}^{\infty} D_{m,\alpha}^{2\ell}(z)\rho^{2\ell+m+2} \quad (15.5.53)$$

where the functions  $D_{m,\alpha}^{2\ell}(z)$  are yet to be determined. It easily verified that there is the relation

$$[(1/\rho)(\partial/\partial\rho)(\rho\partial/\partial\rho) - m^2/\rho^2]\rho^n = (n^2 - m^2)\rho^{n-2}. \quad (15.5.54)$$

It follows, by using the expansion (5.53), that there is the relation

$$\begin{aligned} & [(1/\rho)(\partial/\partial\rho)(\rho\partial/\partial\rho) - m^2/\rho^2 + (\partial/\partial z)^2]d_{m,\alpha}(\rho, z) \\ &= \sum_{\ell=0}^{\infty} [(2\ell + m + 2)^2 - m^2]D_{m,\alpha}^{2\ell}(z)\rho^{2\ell+m} + \sum_{\ell=0}^{\infty} [(\partial/\partial z)^2 D_{m,\alpha}^{2\ell}(z)]\rho^{2\ell+m+2}. \end{aligned} \quad (15.5.55)$$

The sum consisting of the second set of terms on the right side of (5.55) can be rewritten in the form

$$\sum_{\ell=0}^{\infty} [(\partial/\partial z)^2 D_{m,\alpha}^{2\ell}(z)]\rho^{2\ell+m+2} = \sum_{n=1}^{\infty} [(\partial/\partial z)^2 D_{m,\alpha}^{2n-2}(z)]\rho^{2n+m} \quad (15.5.56)$$

or, equivalently, in the form

$$\sum_{\ell=0}^{\infty} [(\partial/\partial z)^2 D_{m,\alpha}^{2\ell}(z)]\rho^{2\ell+m+2} = \sum_{\ell=0}^{\infty} [(\partial/\partial z)^2 D_{m,\alpha}^{2\ell-2}(z)]\rho^{2\ell+m} \quad (15.5.57)$$

with the understanding that

$$D_{m,\alpha}^{-2} = 0. \quad (15.5.58)$$

Consequently, we also have the relation

$$\begin{aligned} & [(1/\rho)(\partial/\partial\rho)(\rho\partial/\partial\rho) - m^2/\rho^2 + (\partial/\partial z)^2]d_{m,\alpha}(\rho, z) \\ = & \sum_{\ell=0}^{\infty} \{[(2\ell+m+2)^2 - m^2]D_{m,\alpha}^{2\ell}(z) + [(\partial/\partial z)^2 D_{m,\alpha}^{2\ell-2}(z)]\}\rho^{2\ell+m}. \end{aligned} \quad (15.5.59)$$

We have found an expansion in powers of  $\rho$  for the left side of (5.52). From (3.38) and (3.39) we already have such an expansion for the right side of (5.52), which we write in the form

$$(-2/m)(\partial/\partial z)\Psi_{m,\alpha}(\rho, z) = \sum_{\ell=0}^{\infty} r(\ell, m)C_{m,\alpha}^{[2\ell+1]}(z)\rho^{2\ell+m} \quad (15.5.60)$$

where

$$r(\ell, m) = -2(-1)^\ell(m!)/[m2^{2\ell}\ell!(l+m)!]. \quad (15.5.61)$$

Equating like powers of  $\rho$  on both sides of (5.52) gives the relation

$$[(2\ell+m+2)^2 - m^2]D_{m,\alpha}^{2\ell}(z) + (\partial/\partial z)^2 D_{m,\alpha}^{2\ell-2}(z) = r(\ell, m)C_{m,\alpha}^{[2\ell+1]}(z), \quad (15.5.62)$$

which can be rewritten as the recursion relation

$$D_{m,\alpha}^{2\ell}(z) = s(\ell, m)C_{m,\alpha}^{[2\ell+1]}(z) + t(\ell, m)(\partial/\partial z)^2 D_{m,\alpha}^{2\ell-2}(z) \quad (15.5.63)$$

where  $s(\ell, m)$  and  $t(\ell, m)$  are the coefficients

$$s(\ell, m) = r(\ell, m)/[(2\ell+m+2)^2 - m^2], \quad (15.5.64)$$

$$t(\ell, m) = -1/[(2\ell+m+2)^2 - m^2]. \quad (15.5.65)$$

We find, for the first few terms, the results

$$D_{m,\alpha}^0(z) = s(0, m)C_{m,\alpha}^{[1]}(z) = -\{1/[2m(m+1)]\}C_{m,\alpha}^{[1]}(z), \quad (15.5.66)$$

$$\begin{aligned} D_{m,\alpha}^2(z) &= s(1, m)C_{m,\alpha}^{[3]}(z) + t(1, m)(\partial/\partial z)^2 D_{m,\alpha}^0(z) \\ &= s(1, m)C_{m,\alpha}^{[3]}(z) + t(1, m)s(0, m)C_{m,\alpha}^{[3]}(z) \\ &= [s(1, m) + t(1, m)s(0, m)]C_{m,\alpha}^{[3]}(z) \\ &= \{1/[8m(m+1)(m+2)]\}C_{m,\alpha}^{[3]}(z), \end{aligned} \quad (15.5.67)$$

$$\begin{aligned} D_{m,\alpha}^4(z) &= s(2, m)C_{m,\alpha}^{[5]}(z) + t(2, m)(\partial/\partial z)^2 D_{m,\alpha}^2(z) \\ &= \{s(2, m) + t(2, m)[s(1, m) + t(1, m)s(0, m)]\}C_{m,\alpha}^{[5]}(z), \\ &= -\{1/[64m(m+1)(m+2)(m+3)]\}C_{m,\alpha}^{[5]}(z), \text{ etc.} \end{aligned} \quad (15.5.68)$$

We conclude that the  $D_{m,\alpha}^{2\ell}(z)$  are completely specified in terms of the  $C_{m,\alpha}^{[2\ell+1]}(z)$ . Indeed, we have by induction the relation

$$D_{m,\alpha}^{2\ell}(z) = u(\ell, m)C_{m,\alpha}^{[2\ell+1]}(z) \quad (15.5.69)$$

where the coefficients  $u(\ell, m)$  are given by the recursion relation

$$u(\ell, m) = s(\ell, m) + t(\ell, m)u(l - 1, m) \quad (15.5.70)$$

with

$$u(-1, m) = 0. \quad (15.5.71)$$

This recursion relation has the solution

$$u(\ell, m) = -(-1)^\ell c_\ell \{[(m - 1)!]/[(m + \ell + 1)!]\} \quad (15.5.72)$$

where

$$c_\ell = 1/[(2)(2^{2\ell})(\ell!)]. \quad (15.5.73)$$

We are now able to write explicit series expansions for the  $d_{m,\alpha}(\rho, z)$  and the  $\chi_{m,\alpha}$  in terms of the  $C_{m,\alpha}^{[n]}(z)$ . Upon combining (5.53), (5.69), (5.72), and (5.73), we find the result

$$d_{m,\alpha}(\rho, z) = -(1/2) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(m - 1)!}{2^{2\ell} \ell! (\ell + m + 1)!} C_{m,\alpha}^{[2\ell+1]}(z) \rho^{2\ell+m+2}. \quad (15.5.74)$$

With the use of (5.48) and (5.49), we find for the  $\chi_{m,\alpha}$  the expansions

$$\begin{aligned} \chi_{m,c} &= (1/2)(\sin m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(m - 1)!}{2^{2\ell} \ell! (\ell + m + 1)!} C_{m,c}^{[2\ell+1]}(z) \rho^{2\ell+m+2} \\ &= (1/2)\rho^m (\sin m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(m - 1)!}{2^{2\ell} \ell! (\ell + m + 1)!} C_{m,c}^{[2\ell+1]}(z) \rho^{2\ell+2} \\ &= (1/2)\Im[(x + iy)^m] \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(m - 1)!}{2^{2\ell} \ell! (\ell + m + 1)!} C_{m,c}^{[2\ell+1]}(z) \rho^{2\ell+2}, \end{aligned} \quad (15.5.75)$$

$$\begin{aligned} \chi_{m,s} &= -(1/2)(\cos m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(m - 1)!}{2^{2\ell} \ell! (\ell + m + 1)!} C_{m,s}^{[2\ell+1]}(z) \rho^{2\ell+m+2} \\ &= -(1/2)\rho^m (\cos m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(m - 1)!}{2^{2\ell} \ell! (\ell + m + 1)!} C_{m,s}^{[2\ell+1]}(z) \rho^{2\ell+2} \\ &= -(1/2)\Re[(x + iy)^m] \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(m - 1)!}{2^{2\ell} \ell! (\ell + m + 1)!} C_{m,s}^{[2\ell+1]}(z) \rho^{2\ell+2}. \end{aligned} \quad (15.5.76)$$

Here, for future use and employing (3.3) and (3.4), we have written results in both cylindrical and Cartesian variables.

Next, we can compute series expansions for the  $\nabla \chi_{m,\alpha}$ . It is convenient to work out the gradients in cylindrical coordinates. We find from (5.75) and (5.76) the results

$$\begin{aligned} (\nabla \chi_{m,c})_\rho &= \partial_\rho \chi_{m,c} \\ &= (1/2)(\sin m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(m-1)!(2\ell+m+2)}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(z)\rho^{2\ell+m+1}, \end{aligned} \quad (15.5.77)$$

$$\begin{aligned} (\nabla \chi_{m,c})_\phi &= (1/\rho)\partial_\phi \chi_{m,c} \\ &= (1/2)(\cos m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(m)(m-1)!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(z)\rho^{2\ell+m+1} \\ &= (1/2)(\cos m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(z)\rho^{2\ell+m+1}, \end{aligned} \quad (15.5.78)$$

$$\begin{aligned} (\nabla \chi_{m,c})_z &= \partial_z \chi_{m,c} \\ &= (1/2)(\sin m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(m-1)!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+2]}(z)\rho^{2\ell+m+2}; \end{aligned} \quad (15.5.79)$$

$$\begin{aligned} (\nabla \chi_{m,s})_\rho &= \partial_\rho \chi_{m,s} \\ &= -(1/2)(\cos m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(m-1)!(2\ell+m+2)}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z)\rho^{2\ell+m+1}, \end{aligned} \quad (15.5.80)$$

$$\begin{aligned} (\nabla \chi_{m,s})_\phi &= (1/\rho)\partial_\phi \chi_{m,s} \\ &= (1/2)(\sin m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(m)(m-1)!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z)\rho^{2\ell+m+1} \\ &= (1/2)(\sin m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z)\rho^{2\ell+m+1}, \end{aligned} \quad (15.5.81)$$

$$\begin{aligned} (\nabla \chi_{m,s})_z &= \partial_z \chi_{m,s} \\ &= -(1/2)(\cos m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(m-1)!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,s}^{[2\ell+2]}(z)\rho^{2\ell+m+2}. \end{aligned} \quad (15.5.82)$$

We have all the ingredients at hand to compute the  $\hat{\mathbf{A}}^{m,\alpha}$ . We find, in cylindrical coordinates and using (4.15) through (4.20), (5.43), and (5.77) through (5.82), the results

$$\hat{A}_\rho^{m,c} = -(1/2)(\sin m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(z)\rho^{2\ell+m+1}, \quad (15.5.83)$$

$$\hat{A}_\phi^{m,c} = (1/2)(\cos m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(z)\rho^{2\ell+m+1}, \quad (15.5.84)$$

$$\hat{A}_z^{m,c} = (\sin m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m)!} C_{m,c}^{[2\ell]}(z)\rho^{2\ell+m}; \quad (15.5.85)$$

$$\hat{A}_\rho^{m,s} = (1/2)(\cos m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z)\rho^{2\ell+m+1}, \quad (15.5.86)$$

$$\hat{A}_\phi^{m,s} = (1/2)(\sin m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z)\rho^{2\ell+m+1}, \quad (15.5.87)$$

$$\hat{A}_z^{m,s} = -(\cos m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m)!} C_{m,s}^{[2\ell]}(z)\rho^{2\ell+m}. \quad (15.5.88)$$

Correspondingly, using (2.24) and (2.25), we find for the Cartesian components of the  $\hat{\mathbf{A}}^{m,\alpha}$  the results

$$\begin{aligned} \hat{A}_x^{m,c} &= -(1/2)[(\cos \phi)(\sin m\phi) + (\sin \phi)(\cos m\phi)] \times \\ &\quad \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(z)\rho^{2\ell+m+1} \\ &= -(1/2)[\sin(m+1)\phi] \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(z)\rho^{2\ell+m+1}, \\ &= -(1/2)\Im[(x+iy)^{m+1}] \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(z)(x^2+y^2)^\ell, \end{aligned} \quad (15.5.89)$$

$$\begin{aligned} \hat{A}_y^{m,c} &= (1/2)[-(\sin \phi)(\sin m\phi) + (\cos \phi)(\cos m\phi)] \times \\ &\quad \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(z)\rho^{2\ell+m+1} \\ &= (1/2)[\cos(m+1)\phi] \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(z)\rho^{2\ell+m+1}, \\ &= (1/2)\Re[(x+iy)^{m+1}] \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(z)(x^2+y^2)^\ell, \end{aligned} \quad (15.5.90)$$

$$\begin{aligned}
\hat{A}_z^{m,c} &= (\sin m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,c}^{[2\ell]}(z) \rho^{2\ell+m} \\
&= \Im[(x+iy)^m] \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,c}^{[2\ell]}(z) (x^2+y^2)^{\ell};
\end{aligned} \tag{15.5.91}$$

$$\begin{aligned}
\hat{A}_x^{m,s} &= (1/2)[(\cos \phi)(\cos m\phi) - (\sin \phi)(\sin m\phi)] \times \\
&\quad \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z) \rho^{2\ell+m+1} \\
&= (1/2)[\cos(m+1)\phi] \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z) \rho^{2\ell+m+1} \\
&= (1/2)\Re[(x+iy)^{m+1}] \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z) (x^2+y^2)^{\ell},
\end{aligned} \tag{15.5.92}$$

$$\begin{aligned}
\hat{A}_y^{m,s} &= (1/2)[(\sin \phi)(\cos m\phi) + (\cos \phi)(\sin m\phi)] \times \\
&\quad \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z) \rho^{2\ell+m+1} \\
&= (1/2)[\sin(m+1)\phi] \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z) \rho^{2\ell+m+1} \\
&= (1/2)\Im[(x+iy)^{m+1}] \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z) (x^2+y^2)^{\ell},
\end{aligned} \tag{15.5.93}$$

$$\begin{aligned}
\hat{A}_z^{m,s} &= -(\cos m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,s}^{[2\ell]}(z) \rho^{2\ell+m} \\
&= -\Re[(x+iy)^m] \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,s}^{[2\ell]}(z) (x^2+y^2)^{\ell}.
\end{aligned} \tag{15.5.94}$$

Note four important facts: First, we see that the relations (5.83) through (5.88) for  $\hat{\mathbf{A}}^{m,\alpha}$  are also defined for  $m = 0$ . When so evaluated they produce, in the case  $\alpha = c$ , a result that agrees with the  $\hat{\mathbf{A}}^0$  given by (5.37) through (5.39); and [recalling (3.36)] they produce, in the case  $\alpha = s$ , the zero vector. That is, we may make the definitions

$$\hat{\mathbf{A}}^{0,c} = \hat{\mathbf{A}}^0, \tag{15.5.95}$$

$$\hat{\mathbf{A}}^{0,s} = 0. \quad (15.5.96)$$

Second, from their form in (5.89) through (5.94), it is evident that the Cartesian components of  $\hat{\mathbf{A}}^{m,\alpha}$  are all harmonic functions, as expected. Third, we observe that  $\hat{A}_x^{m,c}$  and  $\hat{A}_y^{m,c}$  as given by (5.89) and (5.90) involve complementary trigonometric functions multiplying the same master function. The same is true of  $\hat{A}_x^{m,s}$  and  $\hat{A}_y^{m,s}$  as given by (5.92) and (5.93). See Exercise 5.6. For this reason we refer to the Coulomb gauge vector potential we have found as being the *symmetric* Coulomb gauge vector potential. Finally, note again that (5.89) through (5.94) provide expansions of the vector potential in terms of homogeneous polynomials in the variables  $x, y$  with  $z$ -dependent coefficients  $C_{m,\alpha}^{[n]}(z)$ , and that the *minimum* degree of these polynomials is  $m$ . In summary, we have found formulas for the  $\chi_{m,\alpha}$  and the  $\hat{\mathbf{A}}^{m,\alpha}$  in terms of the  $C_{m,\alpha}^{[n]}(z)$ .

### Symmetric Coulomb Gauge Examples for $m = 1, 2$

As an example of the use of these relations, let us compute  $\hat{\mathbf{A}}^{1,s}$  for the (normal) dipole case  $m = 1$ . As before, suppose all terms in (3.33) vanish save for the dipole terms  $C_{1,s}^{[n]}(z)$ . Using (5.92) through (5.94) we then find, through terms of degree five, that  $\hat{\mathbf{A}}^{1,s}$  has the expansion

$$\hat{A}_x^{1,s} = (1/4)(x^2 - y^2)C_{1,s}^{[1]}(z) - (1/48)(x^4 - y^4)C_{1,s}^{[3]}(z) + \dots, \quad (15.5.97)$$

$$\hat{A}_y^{1,s} = (1/2)xyC_{1,s}^{[1]}(z) - (1/24)(x^3y + xy^3)C_{1,s}^{[3]}(z) + \dots, \quad (15.5.98)$$

$$\hat{A}_z^{1,s} = -xC_{1,s}^{[0]}(z) + (1/8)(x^3 + xy^2)C_{1,s}^{[2]}(z) - (1/192)(x^5 + 2x^3y^2 + xy^4)C_{1,s}^{[4]}(z) + \dots. \quad (15.5.99)$$

This expansion should be compared with the azimuthal-free gauge expansion given by (4.27) through (4.29). Direct calculation verifies that  $\hat{\mathbf{A}}^{1,s}$  satisfies (5.1) and (5.4) through terms of degree four, which is what is expected based on the order of the terms that have been retained in the expansion. Finally, we remark that analogous results can be found for the skew case  $\hat{\mathbf{A}}^{1,c}$ .

As a second example of the use of these relations, let us compute  $\hat{\mathbf{A}}^{2,s}$  for the (normal) quadrupole case  $m = 2$ . As before, suppose all terms in (3.33) vanish save for the quadrupole terms  $C_{2,s}^{[n]}(z)$ . Then, again using (5.92) through (5.94), we find, through terms of degree four, that  $\hat{\mathbf{A}}^{2,s}$  has the expansion

$$\hat{A}_x^{2,s} = (1/6)(x^3 - 3xy^2)C_{2,s}^{[1]}(z) + \dots, \quad (15.5.100)$$

$$\hat{A}_y^{2,s} = -(1/6)(y^3 - 3x^2y)C_{2,s}^{[1]}(z) + \dots, \quad (15.5.101)$$

$$\hat{A}_z^{2,s} = -(x^2 - y^2)C_{2,s}^{[0]}(z) + (1/12)(x^4 - y^4)C_{2,s}^{[2]}(z) + \dots. \quad (15.5.102)$$

This expansion should be compared with the azimuthal-free gauge expansion given by (4.30) through (4.32). See Exercise 5.7. Direct calculation again verifies that (5.1) and (5.4) are satisfied by  $\hat{\mathbf{A}}^{2,s}$  through the order of the terms that have been retained in the expansion. Finally, analogous results can be found for the skew case  $\hat{\mathbf{A}}^{2,c}$ .

## A More Elaborate Notation for Future Use

We have been working with vector potentials in the azimuthal-free and symmetric Coulomb gauges, and the connections between them. Soon we will also be introducing vector potentials in the Poincaré Coulomb gauge. It is therefore sometimes useful to introduce a more elaborate notation. We have been using the notation  $\mathbf{A}^{m,\alpha}$  to denote the vector potential in the *azimuthal-free* gauge. In future we may sometimes use the notation  ${}^{AF}\mathbf{A}^{m,\alpha}$  for this type of vector potential,

$${}^{AF}\mathbf{A}^{m,\alpha} = \mathbf{A}^{m,\alpha}. \quad (15.5.103)$$

We have also been using the notation  $\hat{\mathbf{A}}^{m,\alpha}$  to denote the vector potential in the *symmetric Coulomb* gauge. In future we may sometimes use the notation  ${}^{SC}\mathbf{A}^{m,\alpha}$  for this type of vector potential,

$${}^{SC}\mathbf{A}^{m,\alpha} = \hat{\mathbf{A}}^{m,\alpha}. \quad (15.5.104)$$

Also, we have been using the notation  $\chi_{m,\alpha}$  to denote the gauge function that connects the vector potentials in the azimuthal-free and symmetric Coulomb gauges. In future we may use the notation  ${}^{AF}\vec{\chi}_{m,\alpha}^{SC}(x, y, z)$  for this gauge function,

$${}^{AF}\vec{\chi}_{m,\alpha}^{SC}(x, y, z) = \chi_{m,\alpha}(x, y, z). \quad (15.5.105)$$

When this more elaborate notation is employed, the gauge transformation relation (5.43) takes the form

$${}^{SC}\mathbf{A}^{m,\alpha}(x, y, z) = {}^{AF}\mathbf{A}^{m,\alpha}(x, y, z) + \nabla \cdot {}^{AF}\vec{\chi}_{m,\alpha}^{SC}(x, y, z). \quad (15.5.106)$$

Here the *arrow* over  $\chi_{m,\alpha}$  does not denote a vector, but rather that a vector potential in the *AF* gauge is sent to a vector potential in the *SC* gauge.

## Exercises

**15.5.1.** For  $\hat{\mathbf{A}}^0$  given by (5.37) through (5.39), compute the curl and divergence of  $\hat{\mathbf{A}}^0$  in cylindrical coordinates.

**15.5.2.** Verify that (5.72) and (5.73) satisfy the recursion relation (5.70) with the initial condition (5.71).

**15.5.3.** Verify the expansions (5.83) through (5.88) and verify that  $\nabla \cdot \hat{\mathbf{A}}^{m,\alpha} = 0$

**15.5.4.** Consider the case of straight beam-line elements, such as solenoids, quadrupoles, sextupoles, octupoles, etc., for which the design orbit lies on the  $z$  axis. Suppose we wish to retain, in the expansion of the Hamiltonian  $H$  appearing in (1.3), homogeneous polynomials in  $x$  and  $y$  through degree 4. This would be required if we wished to make a Lie factorization of  $\mathcal{M}$  that retained all Lie generators of degree 4 and lower,

$$\mathcal{M} = \mathcal{R} \exp(: f_3 :) \exp(: f_4 :). \quad (15.5.107)$$

Assuming no particular field symmetries, and working in the symmetric Coulomb gauge of this section, show that the following on-axis gradients and their derivatives would then be required:

$$\begin{aligned}
 & C_0^{[0]}(z), \quad C_0^{[1]}(z), \quad C_0^{[2]}(z), \quad C_0^{[3]}(z); \\
 & C_{1,\alpha}^{[0]}(z), \quad C_{1,\alpha}^{[1]}(z), \quad C_{1,\alpha}^{[2]}(z); \\
 & C_{2,\alpha}^{[0]}(z), \quad C_{2,\alpha}^{[1]}(z), \quad C_{2,\alpha}^{[2]}(z); \\
 & C_{3,\alpha}^{[0]}(z); \\
 & C_{4,\alpha}^{[0]}(z).
 \end{aligned} \tag{15.5.108}$$

Verify that in the  $m = 0$  case the  $C_m^{[n]}$  with  $n$  even are actually not needed. See Subsection 5.1. Also, strictly speaking, the dipole terms, the terms in the second row of (5.104), should actually vanish in order for the design orbit to lie on the  $z$  axis. A possible exception could be the case of a wiggler/undulator where the  $C_{1,\alpha}^{[n]}(z)$  oscillate in  $z$  and nearly average to zero in such a way that the design orbit does not depart significantly from the  $z$  axis.

Suppose, instead, we wish to retain homogeneous polynomials through degree 8. This would be required if we wished to make a Lie factorization of  $\mathcal{M}$  that retained all Lie generators of degree 8 and lower,

$$\mathcal{M} = \mathcal{R} \exp(:f_3:) \exp(:f_4:) \exp(:f_5:) \exp(:f_6:) \exp(:f_7:) \exp(:f_8:). \tag{15.5.109}$$

Assuming no particular field symmetries, and working in the symmetric Coulomb gauge of this section, show that the following on-axis gradients and their derivatives would then be required:

$$\begin{aligned}
 & C_0^{[0]}(z), \quad C_0^{[1]}(z), \quad C_0^{[2]}(z), \quad C_0^{[3]}(z), \quad C_0^{[4]}(z), \quad C_0^{[5]}(z), \quad C_0^{[6]}(z), \quad C_0^{[7]}(z); \\
 & C_{1,\alpha}^{[0]}(z), \quad C_{1,\alpha}^{[1]}(z), \quad C_{1,\alpha}^{[2]}(z), \quad C_{1,\alpha}^{[3]}(z), \quad C_{1,\alpha}^{[4]}(z), \quad C_{1,\alpha}^{[5]}(z), \quad C_{1,\alpha}^{[6]}(z); \\
 & C_{2,\alpha}^{[0]}(z), \quad C_{2,\alpha}^{[1]}(z), \quad C_{2,\alpha}^{[2]}(z), \quad C_{2,\alpha}^{[3]}(z), \quad C_{2,\alpha}^{[4]}(z), \quad C_{2,\alpha}^{[5]}(z), \quad C_{2,\alpha}^{[6]}(z); \\
 & C_{3,\alpha}^{[0]}(z), \quad C_{3,\alpha}^{[1]}(z), \quad C_{3,\alpha}^{[2]}(z), \quad C_{3,\alpha}^{[3]}(z), \quad C_{3,\alpha}^{[4]}(z); \\
 & C_{4,\alpha}^{[0]}(z), \quad C_{4,\alpha}^{[1]}(z), \quad C_{4,\alpha}^{[2]}(z), \quad C_{4,\alpha}^{[3]}(z), \quad C_{4,\alpha}^{[4]}(z); \\
 & C_{5,\alpha}^{[0]}(z), \quad C_{5,\alpha}^{[1]}(z), \quad C_{5,\alpha}^{[2]}(z); \\
 & C_{6,\alpha}^{[0]}(z), \quad C_{6,\alpha}^{[1]}(z), \quad C_{6,\alpha}^{[2]}(z); \\
 & C_{7,\alpha}^{[0]}(z); \\
 & C_{8,\alpha}^{[0]}(z).
 \end{aligned} \tag{15.5.110}$$

Again verify that in the  $m = 0$  case the  $C_m^{[n]}$  with  $n$  even are actually not needed. And again, with the possible exception of a wiggler/undulator, the dipole terms, the terms in the second row of (5.110), should actually vanish in order for the design orbit to lie on the  $z$  axis.

**15.5.5.** Assume that (3.77) through (3.83) hold in the *body* of a pure multipole (with  $m > 0$ ). Let

$$\mathbf{B}^m = \nabla \times \hat{\mathbf{A}}^m \quad (15.5.111)$$

with

$$\hat{\mathbf{A}}^m = \hat{\mathbf{A}}^{m,c} + \hat{\mathbf{A}}^{m,s}. \quad (15.5.112)$$

Show from (5.89) through (5.94) that in this case (the symmetric Coulomb gauge case) there are the relations

$$\hat{A}_x^m = \hat{A}_y^m = 0, \quad (15.5.113)$$

and

$$\begin{aligned} \hat{A}_z^m &= C_{m,c}^{[0]} \Im[(x+iy)^m] - C_{m,s}^{[0]} \Re[(x+iy)^m] \\ &= -\Re[(C_{m,s}^{[0]} + iC_{m,c}^{[0]}) (x+iy)^m] = \Im[(C_{m,c}^{[0]} - iC_{m,s}^{[0]}) (x+iy)^m]. \end{aligned} \quad (15.5.114)$$

Here the quantities  $C_{m,\alpha}^{[0]}$  are assumed to be *constant* ( $z$  independent).

Note that the results (5.109) and (5.110) agree with (4.66) and (4.67). The azimuthal-free and symmetric Coulomb gauges give the *same* result in the body for all terms with  $m > 0$ . The difference between the two gauges occurs only in the fringe-field regions.

**15.5.6.** As a consequence of the symmetry present in the symmetric Coulomb gauge, verify that the two real relations (5.89) and (5.90) can be combined to produce the single complex relation

$$\hat{A}_x^{m,c} + i\hat{A}_y^{m,c} = (i/2)(x+iy)^{m+1} \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell} \ell! (\ell+m+1)!} C_{m,c}^{[2\ell+1]}(z) (x^2+y^2)^\ell. \quad (15.5.115)$$

Also verify that the two real relations (5.92) and (5.93) can be combined to produce the single complex relation

$$\hat{A}_x^{m,s} + i\hat{A}_y^{m,s} = (1/2)(x+iy)^{m+1} \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell} \ell! (\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z) (x^2+y^2)^\ell. \quad (15.5.116)$$

**15.5.7.** The purpose of this exercise is to study and compare the azimuthal-free and symmetric Coulomb gauge vector potentials for the case of a normal quadrupole, the case  $m = 2$  and  $\alpha = s$ . Let  ${}^{AF}\mathbf{A}$  and  ${}^{SC}\mathbf{A}$  be the vector potentials for a normal quadrupole in the *azimuthal-free* and *symmetric Coulomb* gauges, respectively. Then, from (4.30) through (4.32), we find for a normal quadrupole the results

$${}^{AF}A_x = (1/2)(x^3 - xy^2)C_{2,s}^{[1]}(z) + \dots, \quad (15.5.117)$$

$${}^{AF}A_y = -(1/2)(y^3 - yx^2)C_{2,s}^{[1]}(z) + \dots, \quad (15.5.118)$$

$${}^{AF}A_z = -(x^2 - y^2)C_{2,s}^{[0]}(z) + (1/6)(x^4 - y^4)C_{2,s}^{[2]}(z) + \dots. \quad (15.5.119)$$

And, from (5.100) through (5.102), we find the results

$${}^{SC}A_x = (1/6)(x^3 - 3xy^2)C_{2,s}^{[1]}(z) + \dots, \quad (15.5.120)$$

$${}^{SC}A_y = -(1/6)(y^3 - 3x^2y)C_{2,s}^{[1]}(z) + \dots, \quad (15.5.121)$$

$${}^{SC}A_z = -(x^2 - y^2)C_{2,s}^{[0]}(z) + (1/12)(x^4 - y^4)C_{2,s}^{[2]}(z) + \dots. \quad (15.5.122)$$

Observe that the two vector potentials agree in the body of the quadrupole, and differ in the fringe-field regions. In particular, in the fringe-field regions, the nonlinear terms are *smaller* for the symmetric Coulomb gauge.

Verify that  $\nabla \cdot {}^{SC}\mathbf{A} = 0$  as expected. Verify, again as expected, that  ${}^{AF}\mathbf{A}$  and  ${}^{SC}\mathbf{A}$  produce the same magnetic field and that this field agrees with (3.62) through (3.64).

**15.5.8.** Why is the Coulomb gauge named after Coulomb?

## 15.6 Nonuniqueness of Coulomb Gauge

There still remains the question of uniqueness. We will see that there are other Coulomb gauge vector potentials beyond the symmetric one already found.

### 15.6.1 The General Case

Suppose  $\lambda(x, y, z)$  is any *harmonic* function,

$$\nabla^2 \lambda = 0. \quad (15.6.1)$$

If we add  $\nabla \lambda$  to  $\hat{\mathbf{A}}$  to produce a vector potential  $\tilde{\mathbf{A}}$ , it is easily verified that the result

$$\tilde{\mathbf{A}} = \hat{\mathbf{A}} + \nabla \lambda \quad (15.6.2)$$

also satisfies the desired relations

$$\nabla \times \tilde{\mathbf{A}} = \mathbf{B} \quad (15.6.3)$$

and

$$\nabla \cdot \tilde{\mathbf{A}} = 0. \quad (15.6.4)$$

Conversely, if we require that the Ansatz (6.2) also yield a vector potential in the Coulomb gauge, then  $\lambda$  must be harmonic.

We next observe that, by construction,  $\hat{\mathbf{A}}$  falls to zero as  $|z| \rightarrow \infty$  so that we should require the same of  $\lambda$  in order for  $\tilde{\mathbf{A}}$  to have the same asymptotic behavior.<sup>11</sup> Thanks to the work already done, it is easy to describe all  $\lambda$  satisfying (6.1) that have this property. Namely, by repeating the arguments leading to the representation of  $\psi$  in terms of on-axis gradients, we may write

$$\lambda = \sum_{m=0}^{\infty} \Lambda_{m,c}(\rho, z) \cos m\phi + \sum_{m=1}^{\infty} \Lambda_{m,s}(\rho, z) \sin m\phi \quad (15.6.5)$$

---

<sup>11</sup>As stated earlier, this asymptotic behavior is desirable in order that the canonical and mechanical momenta be asymptotically the same. See (1.5.30). If we are working with  $z$  as the independent variable, in which case  $p_z$  is not a dynamical variable and does not appear in the Hamiltonian, we will at least want the  $x$  and  $y$  components of  $\tilde{\mathbf{A}}$  to vanish for large  $|z|$ .

and set

$$\Lambda_{m,\alpha}(\rho, z) = \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} L_{m,\alpha}^{[2\ell]}(z) \rho^{2\ell+m}. \quad (15.6.6)$$

Here the functions  $L_{m,\alpha}^{[0]}(z)$  may be specified at will save for the condition that they fall to zero for large  $|z|$ .

We know that  $\hat{A}_x$  and  $\hat{A}_y$  are harmonic functions because we may write

$$\hat{A}_x = \sum_{m=0}^{\infty} \hat{A}_x^{m,c} + \sum_{m=1}^{\infty} \hat{A}_x^{m,s}, \text{ etc.} \quad (15.6.7)$$

and we have already seen that each term in the above sums is harmonic. It is a remarkable fact that if  $\sigma(x, y, z)$  is a harmonic function that falls off for large  $|z|$  and thus can be written in a form analogous to (6.5), then there is another harmonic function  $\lambda$  with the same properties such that

$$\partial_y \lambda = \sigma. \quad (15.6.8)$$

Or, if one prefers, there is a  $\lambda'$  such that

$$\partial_x \lambda' = \sigma. \quad (15.6.9)$$

See Appendix H. Let us apply this result, for example, to the case

$$\sigma = -\hat{A}_y \quad (15.6.10)$$

and then use  $\lambda$  to make the gauge transformation (6.2). Doing so, we find the results

$$\tilde{A}_x = \hat{A}_x + \partial_x \lambda, \quad (15.6.11)$$

$$\tilde{A}_y = \hat{A}_y + \partial_y \lambda = \hat{A}_y + \sigma = 0, \quad (15.6.12)$$

$$\tilde{A}_z = \hat{A}_z + \partial_z \lambda. \quad (15.6.13)$$

That is, we have found a gauge which is both Coulomb and for which the  $y$  component of the vector potential is zero. We call this the *vertical-free* Coulomb gauge. Similarly, by using  $\lambda'$ , one can find a *horizontal-free* Coulomb gauge for which the  $x$  component of the vector potential is zero.

Even a bit more can be accomplished. Suppose  $\tilde{\mathbf{A}}$  is a vector potential in the vertical-free Coulomb gauge so that  $\tilde{A}_y = 0$ . Let  $\tau(x, z)$  be a harmonic function that depends *only* on  $x$  and  $z$ . Such a function can be written in the form

$$\begin{aligned} \tau(x, z) &= \sum_{n=0}^{\infty} (-1)^n [1/(2n+1)!] x^{2n+1} O^{[2n]}(z) + \sum_{n=0}^{\infty} (-1)^n [1/(2n)!] x^{2n} E^{[2n]}(z) \\ &= [xO^{[0]}(z) - (1/6)x^3 O^{[2]}(z) + (1/120)x^5 O^{[4]}(z) + \dots] \\ &\quad + [E^{[0]}(z) - (1/2)x^2 E^{[2]}(z) + (1/24)x^4 E^{[4]}(z) + \dots] \end{aligned} \quad (15.6.14)$$

where  $O^{[0]}(z)$  and  $E^{[0]}(z)$  are arbitrary functions of  $z$  save that they fall off for large  $|z|$ . See Appendix H. Now use  $\tau$  to make a further gauge transformation that sends  $\tilde{\mathbf{A}}$  to  $\check{\mathbf{A}}$ ,

$$\check{\mathbf{A}} = \tilde{\mathbf{A}} + \nabla \tau. \quad (15.6.15)$$

By construction,

$$\partial_y \tau = 0 \quad (15.6.16)$$

so that  $\check{\mathbf{A}}$  is also vertical free,

$$\check{A}_y = 0. \quad (15.6.17)$$

And for the  $x$  and  $z$  components of  $\check{\mathbf{A}}$  we find the results

$$\begin{aligned} \check{A}_x &= \hat{A}_x + [O^{[0]}(z) - (1/2)x^2 O^{[2]}(z) + (1/24)x^4 O^{[4]}(z) + \dots] \\ &\quad + [-xE^{[2]}(z) + (1/6)x^3 E^{[4]}(z) + \dots], \end{aligned} \quad (15.6.18)$$

$$\begin{aligned} \check{A}_z &= \hat{A}_z + [xO^{[1]}(z) - (1/6)x^3 O^{[3]}(z) + (1/120)x^5 O^{[5]}(z) + \dots] \\ &\quad + [E^{[1]}(z) - (1/2)x^2 E^{[3]}(z) + (1/24)x^4 E^{[5]}(z) + \dots]. \end{aligned} \quad (15.6.19)$$

Evidently, by a suitable choice of  $O^{[0]}(z)$  and  $E^{[0]}(z)$ , we are able to make some further adjustments to  $\check{A}_x$  and  $\check{A}_z$  while keeping the gauge Coulomb and vertical free.

### 15.6.2 Normal Dipole Example

As an example of the further gauge freedom just described, let us consider the case of a normal dipole whose Coulomb-gauge vector potential  $\hat{\mathbf{A}}^{1,s}$  is given by (5.95) through (5.97). We will first perform a succession of gauge transformations to make the vector potential vertical free while maintaining its Coulomb nature. Then we will adjust its  $x$  component.

To begin, use a  $\lambda$ , which we will denote by the symbols  $\lambda^3$ , for which all the  $L_{m,\alpha}^{[0]}(z)$  are zero save for  $L_{3,c}^{[0]}(z)$ . Then, from (6.6), we find through terms of degree six the result

$$\Lambda_{3,c}(\rho, z) = L_{3,c}^{[0]}(z)\rho^3 - (1/16)L_{3,c}^{[2]}(z)\rho^5 + \dots, \quad (15.6.20)$$

from which it follows, using (6.5), that

$$\begin{aligned} \lambda^3 &= (\cos 3\phi)\Lambda_{3,c}(\rho, z) = (\cos 3\phi)[L_{3,c}^{[0]}(z)\rho^3 - (1/16)L_{3,c}^{[2]}(z)\rho^5 + \dots] \\ &= (x^3 - 3xy^2)L_{3,c}^{[0]}(z) - (1/16)(x^5 - 2x^3y^2 - 3xy^4)L_{3,c}^{[2]}(z) + \dots. \end{aligned} \quad (15.6.21)$$

This  $\lambda^3$  has the gradients

$$(\nabla \lambda^3)_x = 3(x^2 - y^2)L_{3,c}^{[0]}(z) - (1/16)(5x^4 - 6x^2y^2 - 3y^4)L_{3,c}^{[2]}(z) + \dots, \quad (15.6.22)$$

$$(\nabla \lambda^3)_y = -6xyL_{3,c}^{[0]}(z) + (1/16)(4x^3y + 12xy^3)L_{3,c}^{[2]}(z) + \dots, \quad (15.6.23)$$

$$(\nabla \lambda^3)_z = (x^3 - 3xy^2)L_{3,c}^{[1]}(z) - (1/16)(x^5 - 2x^3y^2 - 3xy^4)L_{3,c}^{[3]}(z) + \dots. \quad (15.6.24)$$

Let us employ  $\lambda^3$  to produce a transformed vector potential  $\mathbf{A}'$  using the relation

$$\mathbf{A}' = \hat{\mathbf{A}}^{1,s} + \nabla \lambda^3. \quad (15.6.25)$$

Then, from (5.95) through (5.97) and (6.22) through (6.25), we find the results

$$\begin{aligned} A'_x &= (x^2 - y^2)[(1/4)C_{1,s}^{[1]}(z) + 3L_{3,c}^{[0]}(z)] \\ &\quad - (1/48)(x^4 - y^4)C_{1,s}^{[3]}(z) - (1/16)(5x^4 - 6x^2y^2 - 3y^4)L_{3,c}^{[2]}(z) + \dots, \end{aligned} \quad (15.6.26)$$

$$\begin{aligned} A'_y &= xy[(1/2)C_{1,s}^{[1]}(z) - 6L_{3,c}^{[0]}(z)] \\ &\quad - (1/24)(x^3y + xy^3)C_{1,s}^{[3]}(z) + (1/16)(4x^3y + 12xy^3)L_{3,c}^{[2]}(z) + \dots, \end{aligned} \quad (15.6.27)$$

$$\begin{aligned} A'_z &= -xC_{1,s}^{[0]}(z) + (1/8)(x^3 + xy^2)C_{1,s}^{[2]}(z) - (1/192)(x^5 + 2x^3y^2 + xy^4)C_{1,s}^{[4]}(z) \\ &\quad + (x^3 - 3xy^2)L_{3,c}^{[1]}(z) - (1/16)(x^5 - 2x^3y^2 - 3xy^4)L_{3,c}^{[3]}(z) + \dots. \end{aligned} \quad (15.6.28)$$

Observe that we can make the leading term of  $A'_x$  vanish by setting

$$[(1/4)C_{1,s}^{[1]}(z) + 3L_{3,c}^{[0]}(z)] = 0. \quad (15.6.29)$$

Or, we can make the leading term of  $A'_y$  vanish by setting

$$[(1/2)C_{1,s}^{[1]}(z) - 6L_{3,c}^{[0]}(z)] = 0. \quad (15.6.30)$$

Suppose we decide to make the leading term of  $A'_y$  vanish. Then we have the relation

$$L_{3,c}^{[0]}(z) = (1/12)C_{1,s}^{[1]}(z), \quad (15.6.31)$$

from which it follows that

$$L_{3,c}^{[1]}(z) = (1/12)C_{1,s}^{[2]}(z), \quad (15.6.32)$$

$$L_{3,c}^{[2]}(z) = (1/12)C_{1,s}^{[3]}(z), \quad (15.6.33)$$

$$L_{3,c}^{[3]}(z) = (1/12)C_{1,s}^{[4]}(z), \text{ etc.} \quad (15.6.34)$$

When this is done,  $\mathbf{A}'$  takes the form

$$A'_x = (1/2)(x^2 - y^2)C_{1,s}^{[1]}(z) - (1/192)(9x^4 - 6x^2y^2 - 7y^4)C_{1,s}^{[3]}(z) + \dots, \quad (15.6.35)$$

$$A'_y = -(1/48)(x^3y - xy^3)C_{1,s}^{[3]}(z) + \dots, \quad (15.6.36)$$

$$A'_z = -xC_{1,s}^{[0]}(z) + (1/24)(5x^3 - 3xy^2)C_{1,s}^{[2]}(z) - (1/96)(x^5 - xy^4)C_{1,s}^{[4]}(z) + \dots. \quad (15.6.37)$$

At this point, as a sanity check on the algebra used to yield (6.35) through (6.37), the reader should verify (through the order of the terms retained) that  $\mathbf{A}'$  is still Coulombic and its components are still harmonic.

Let us see if we can make the next term in  $A'_y$  vanish by performing an additional gauge transformation. Suppose we make a further gauge transformation using a  $\lambda$ , which we will call  $\lambda^5$ , for which all the  $L_{m,\alpha}^{[0]}(z)$  are zero save for  $L_{5,c}^{[0]}(z)$ . Then we find through terms of degree six the result

$$\Lambda_{5,c}(\rho, z) = L_{5,c}^{[0]}(z)\rho^5 + \dots, \quad (15.6.38)$$

from which it follows that

$$\begin{aligned} \lambda^5 &= (\cos 5\phi)\Lambda_{5,c}(\rho, z) = (\cos 5\phi)[L_{5,c}^{[0]}(z)\rho^5 + \dots] \\ &= (x^5 - 10x^3y^2 + 5xy^4)L_{5,c}^{[0]}(z) + \dots. \end{aligned} \quad (15.6.39)$$

This  $\lambda$  has the gradients

$$(\nabla \lambda^5)_x = (5x^4 - 30x^2y^2 + 5y^4)L_{5,c}^{[0]}(z) + \dots, \quad (15.6.40)$$

$$(\nabla \lambda^5)_y = -20(x^3y - xy^3)L_{5,c}^{[0]}(z) + \dots, \quad (15.6.41)$$

$$(\nabla \lambda^5)_z = (x^5 - 10x^3y^2 + 5xy^4)L_{5,c}^{[1]}(z). \quad (15.6.42)$$

Correspondingly, we will define a further transformed vector potential  $\mathbf{A}''$  by writing

$$\mathbf{A}'' = \mathbf{A}' + \nabla \lambda^5. \quad (15.6.43)$$

Then, using (6.35) through (6.37) and (6.40) through (6.43),  $\mathbf{A}''$  takes the form

$$\begin{aligned} A''_x &= (1/2)(x^2 - y^2)C_{1,s}^{[1]}(z) - (1/192)(9x^4 - 6x^2y^2 - 7y^4)C_{1,s}^{[3]}(z) \\ &\quad + (5x^4 - 30x^2y^2 + 5y^4)L_{5,c}^{[0]}(z) + \dots, \end{aligned} \quad (15.6.44)$$

$$A''_y = -(1/48)(x^3y - xy^3)C_{1,s}^{[3]}(z) - 20(x^3y - xy^3)L_{5,c}^{[0]}(z) + \dots, \quad (15.6.45)$$

$$\begin{aligned} A''_z &= -xC_{1,s}^{[0]}(z) + (1/24)(5x^3 - 3xy^2)C_{1,s}^{[2]}(z) - (1/96)(x^5 - xy^4)C_{1,s}^{[4]}(z) \\ &\quad + (x^5 - 10x^3y^2 + 5xy^4)L_{5,c}^{[1]}(z) + \dots. \end{aligned} \quad (15.6.46)$$

We see that  $A''_y$  will vanish through terms of degree four provided  $L_{5,c}^{[0]}(z)$  is selected to satisfy the relation

$$L_{5,c}^{[0]}(z) = -(1/960)C_{1,s}^{[3]}(z), \quad (15.6.47)$$

from which it follows that

$$L_{5,c}^{[1]}(z) = -(1/960)C_{1,s}^{[4]}(z), \text{ etc.} \quad (15.6.48)$$

When this condition is met,  $\mathbf{A}''$  takes the form

$$A''_x = (1/2)(x^2 - y^2)C_{1,s}^{[1]}(z) - (1/96)(5x^4 - 6x^2y^2 - 3y^4)C_{1,s}^{[3]}(z) + \dots, \quad (15.6.49)$$

$$A_y'' = 0 + \dots , \quad (15.6.50)$$

$$\begin{aligned} A_z'' &= -x C_{1,s}^{[0]}(z) + (1/24)(5x^3 - 3xy^2)C_{1,s}^{[2]}(z) \\ &\quad - (1/960)(11x^5 - 10x^3y^2 - 5xy^4)C_{1,s}^{[4]}(z) + \dots . \end{aligned} \quad (15.6.51)$$

(Here the reader should again perform Coulombic and harmonic sanity checks.) We have achieved, through terms of degree four, a vertical-free Coulomb gauge vector potential for the normal dipole.

There is still the possibility of adjusting the  $x$  (and correspondingly the  $z$  component) of  $\mathbf{A}''$  by making yet another gauge transformation using the harmonic function  $\tau(x, z)$  given by (6.14). We define a still further transformed vector potential  $\mathbf{A}'''$  by writing

$$\mathbf{A}''' = \mathbf{A}'' + \nabla\tau. \quad (15.6.52)$$

So doing gives the result

$$\begin{aligned} A_x''' &= O^{[0]}(z) - xE^{[2]}(z) + (1/2)(x^2 - y^2)C_{1,s}^{[1]}(z) - (1/2)x^2O^{[2]}(z) \\ &\quad + (1/6)x^3E^{[4]}(z) - (1/96)(5x^4 - 6x^2y^2 - 3y^4)C_{1,s}^{[3]}(z) + (1/24)x^4O^{[4]}(z) + \dots , \end{aligned} \quad (15.6.53)$$

$$A_y''' = 0 + \dots , \quad (15.6.54)$$

$$\begin{aligned} A_z''' &= E^{[1]}(z) - xC_{1,s}^{[0]}(z) + xO^{[1]}(z) - (1/2)x^2E^{[3]}(z) \\ &\quad + (1/24)(5x^3 - 3xy^2)C_{1,s}^{[2]}(z) - (1/6)x^3O^{[3]}(z) + (1/24)x^4E^{[5]}(z) \\ &\quad - (1/960)(11x^5 - 10x^3y^2 - 5xy^4)C_{1,s}^{[4]}(z) + (1/120)x^5O^{[5]}(z) + \dots . \end{aligned} \quad (15.6.55)$$

Here the functions  $O^{[0]}(z)$  and  $E^{[1]}(z)$  are arbitrary except that they must vanish as  $|z| \rightarrow \infty$ .

## Exercises

**15.6.1.** Verify that the vector potential  $\mathbf{A}''$  given by (6.49) through (6.51) yields the magnetic field  $\mathbf{B}$  given by (3.58) through (3.60), and is Coulombic and harmonic.

**15.6.2.** Review Exercise 4.4 and, in particular, the vector potential given by (4.58) through (4.60). Show that this vector potential is in neither the azimuthal-free nor the Coulomb gauge. Let  $\chi$  be the function

$$\begin{aligned} \chi(x, z) &= -\sum_{n=1}^{\infty}(-1)^n[1/(2n+1)!]x^{2n+1}O^{[2n-1]}(z) \\ &= (1/6)x^3O^{[1]}(z) - (1/120)x^5O^{[3]}(z) + \dots . \end{aligned} \quad (15.6.56)$$

Define a vector potential  $\hat{\mathbf{A}}^{\text{iwd}}$  by the making the gauge transformation

$$\hat{\mathbf{A}}^{\text{iwd}} = \mathbf{A}^{\text{iwd}} + \nabla\chi. \quad (15.6.57)$$

Show that

$$\begin{aligned}\hat{A}_x^{\text{iwd}} &= \sum_{n=1}^{\infty} (-1)^n [1/(2n)!] y^{2n} O^{[2n-1]}(z) - \sum_{n=1}^{\infty} (-1)^n [1/(2n)!] x^{2n} O^{[2n-1]}(z) \\ &= -(1/2)(y^2 - x^2) O^{[1]}(z) + (1/24)(y^4 - x^4) O^{[3]}(z) + \dots,\end{aligned} \quad (15.6.58)$$

$$\hat{A}_y^{\text{iwd}} = 0, \quad (15.6.59)$$

$$\begin{aligned}\hat{A}_z^{\text{iwd}} &= - \sum_{n=0}^{\infty} (-1)^n [1/(2n+1)!] x^{2n+1} O^{[2n]}(z) \\ &= -x O^{[0]}(z) + (1/6)x^3 O^{[2]}(z) - (1/120)x^5 O^{[4]}(z) + \dots.\end{aligned} \quad (15.6.60)$$

Thus, the vector potential  $\hat{\mathbf{A}}^{\text{iwd}}$  is vertical free. Show that  $\hat{\mathbf{A}}^{\text{iwd}}$  is also in the Coulomb gauge, and that all its Cartesian components are harmonic functions.

Equations (6.49) through (6.51) give the vector potential in the vertical-free Coulomb gauge corresponding to the  $C_{1,s}^{[n]}$ . Change notation to call this result  $\bar{\mathbf{A}}^{1,s}$ ,

$$\bar{A}_x^{1,s} = (1/2)(x^2 - y^2) C_{1,s}^{[1]}(z) - (1/96)(5x^4 - 6x^2y^2 - 3y^4) C_{1,s}^{[3]}(z) + \dots, \quad (15.6.61)$$

$$\bar{A}_y^{1,s} = 0 + \dots, \quad (15.6.62)$$

$$\begin{aligned}\bar{A}_z^{1,s} &= -x C_{1,s}^{[0]}(z) + (1/24)(5x^3 - 3xy^2) C_{1,s}^{[2]}(z) \\ &\quad - (1/960)(11x^5 - 10x^3y^2 - 5xy^4) C_{1,s}^{[4]}(z) + \dots.\end{aligned} \quad (15.6.63)$$

Show, by analogous calculations, that the vector potential in the vertical-free Coulomb gauge corresponding to the  $C_{3,s}^{[n]}$  is given by the relations

$$\bar{A}_x^{3,s} = (1/4)(x^4 - 6x^2y^2 + y^4) C_{3,s}^{[1]}(z) + \dots, \quad (15.6.64)$$

$$\bar{A}_y^{3,s} = 0, \quad (15.6.65)$$

$$\bar{A}_z^{3,s} = -(x^3 - 3xy^2) C_{3,s}^{[0]}(z) + (1/80)(7x^5 - 30x^3y^2 - 5xy^4) C_{3,s}^{[2]}(z) + \dots. \quad (15.6.66)$$

Use these results to find, through terms of degree four, the Coulombic and vertical-free vector potential for the infinite-width dipole, and show that your results agree with (6.58) through (6.60).

Note that again, as was the case for the vector potentials found in Exercise 3.4, that the vector potential is primarily in the  $z$  direction.

Study Appendix H.3.3, which finds a Coulombic and horizontal-free vector potential for the infinite-width dipole.

**15.6.3.** Review Exercise 4.7. Show that, under the same assumptions, (4.66) and (4.67) also hold in the Coulomb gauge.

**15.6.4.** The relations (5.32) through (5.34) provide expansions for the components of  $\hat{\mathbf{A}}^0$ , the  $m = 0$  vector potential in the symmetric Coulomb gauge. Find the first few terms in the expansions for the components of  $\tilde{\mathbf{A}}^0$ , the vector potential for the  $m = 0$  case in the vertical-free Coulomb gauge. Recall (6.11) through (6.13).

## 15.7 Determination of the Vector Potential: Poincaré-Coulomb Gauge

Assuming  $m \neq 0$ , the relations (4.15) through (4.26) provide, through all orders, formulas for the azimuthal-free gauge vector potential in terms of the on-axis gradients. And, for all values of  $m$ , the relations (5.83) through (5.94) provide, again through all orders, formulas for the symmetric Coulomb gauge vector potential in terms of the on-axis gradients. For general  $m$  and through all orders, are there relations that provide formulas for the Poincaré-Coulomb gauge vector potential in terms of the on-axis gradients? The purpose of this section is to explore this question.

At this point it is necessary to be precise about what we wish to accomplish. Recall the vectors  $\mathbf{R}$ ,  $\mathbf{R}_0$ , and  $\mathbf{r}$  introduced in Subsection 2.1 by writing (2.3). Since the axis of any of the straight beam-line elements we are considering is supposed to lie along the  $z$  axis, we stipulate that  $\mathbf{R}_0$  be of the form

$$\mathbf{R}_0 = (0, 0, Z_0). \quad (15.7.1)$$

Correspondingly,  $\mathbf{R}$  then takes the form

$$\mathbf{R} = (x, y, Z_0 + z) \quad (15.7.2)$$

where  $x$  and  $y$  are assumed to be small, and  $z$  may or may not also be small depending on the choice of  $Z_0$ .

Let  $\hat{\mathbf{A}}^{m,\alpha}$  be the symmetric Coulomb gauge vector potential of Section 5. In view of (7.2) we may write

$$\hat{\mathbf{A}}^{m,\alpha}(\mathbf{R}) = \hat{\mathbf{A}}^{m,\alpha}(x, y, Z_0 + z). \quad (15.7.3)$$

Assuming it exists, let us introduce the symbols  ${}^{PC}\mathbf{A}^{m,\alpha}(x, y, z; Z_0)$  to denote the Poincaré-Coulomb gauge counterpart to the symmetric Coulomb gauge vector potential  $\hat{\mathbf{A}}^{m,\alpha}$ . Since we require that these vector potentials produce the same magnetic field,

$$\nabla \times \hat{\mathbf{A}}^{m,\alpha}(x, y, Z_0 + z) = \nabla \times {}^{PC}\mathbf{A}^{m,\alpha}(x, y, z; Z_0), \quad (15.7.4)$$

they must be related by a gauge transformation of the form

$${}^{PC}\mathbf{A}^{m,\alpha}(x, y, z; Z_0) = \hat{\mathbf{A}}^{m,\alpha}(x, y, Z_0 + z) + \nabla \hat{\chi}_{m,\alpha} \quad (15.7.5)$$

described by the gauge function  $\hat{\chi}_{m,\alpha}(x, y, z; Z_0)$ . Here

$$\nabla = (\partial/\partial x, \partial/\partial y, \partial/\partial z), \quad (15.7.6)$$

and

$$\nabla \hat{\chi}_{m,\alpha} = \nabla \hat{\chi}_{m,\alpha}(x, y, z; Z_0). \quad (15.7.7)$$

Since both  ${}^{PC}\mathbf{A}^{m,\alpha}$  and  $\hat{\mathbf{A}}^{m,\alpha}$  are supposed to be in the Coulomb gauge, i.e. divergence free, we see from (7.5) that the function  $\hat{\chi}_{m,\alpha}$  must be harmonic,

$$\nabla^2 \hat{\chi}_{m,\alpha}(x, y, z; Z_0) = 0. \quad (15.7.8)$$

Finally if  ${}^{PC}\mathbf{A}^{m,\alpha}$  is to be in the Poincaré-Coulomb gauge, then from (7.5) we see that  $\hat{\chi}_{m,\alpha}$  must satisfy the further condition

$$\mathbf{r} \cdot \nabla \hat{\chi}_{m,\alpha} = -\mathbf{r} \cdot \hat{\mathbf{A}}^{m,\alpha}. \quad (15.7.9)$$

Our task now is to find the functions  $\hat{\chi}_{m,\alpha}(x, y, z; Z_0)$  in terms of the on-axis gradients.

### 15.7.1 The $m = 0$ Case

Observe that the  $m = 0$  symmetric Coulomb gauge vector potential  $\hat{\mathbf{A}}^0$  given by (5.37) through (5.39) has only a  $\phi$  component. It follows, from the fact that the  $\mathbf{e}_\rho$ ,  $\mathbf{e}_\phi$ ,  $\mathbf{e}_z$  form an orthonormal triad and (2.20), that there is the relation

$$\mathbf{r} \cdot \hat{\mathbf{A}}^0(\mathbf{R}) = 0. \quad (15.7.10)$$

Alternatively, this same relation follows from (5.32) through (5.34) and (2.5). Either way, we conclude that  $\hat{\mathbf{A}}^0$ , when evaluated with respect to any origin on the  $z$  axis, is in the Poincaré-Coulomb gauge.

Also,  $\hat{\mathbf{A}}^0$  can be expressed in terms of on-axis gradients. Indeed, in terms of the variables employed in this section, the relations (5.32) through (5.34) and (5.36) take the form

$$\begin{aligned} \hat{A}_x^0(x, y, Z_0 + z) &= -(y/2) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{1}{2^{2\ell} \ell! (\ell+1)!} C_0^{[2\ell+1]}(Z_0 + z) (x^2 + y^2)^\ell \\ &= -(y/2)[C_0^{[1]}(Z_0 + z) - (1/8)C_0^{[3]}(Z_0 + z)(x^2 + y^2) + \dots], \end{aligned} \quad (15.7.11)$$

$$\begin{aligned} \hat{A}_y^0(x, y, Z_0 + z) &= (x/2) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{1}{2^{2\ell} \ell! (\ell+1)!} C_0^{[2\ell+1]}(Z_0 + z) (x^2 + y^2)^\ell \\ &= (x/2)[C_0^{[1]}(Z_0 + z) - (1/8)C_0^{[3]}(Z_0 + z)(x^2 + y^2) + \dots], \end{aligned} \quad (15.7.12)$$

$$\hat{A}_z^0(x, y, Z_0 + z) = 0, \quad (15.7.13)$$

where

$$C_0^{[1]}(Z_0 + z) = B_z(0, 0, Z_0 + z). \quad (15.7.14)$$

### 15.7.2 The $m \geq 1$ Cases

We have seen that for the case  $m = 0$  the symmetric Coulomb gauge vector potential  $\hat{\mathbf{A}}^0$  is in the Poincaré-Coulomb gauge. What can be said about the cases  $m \geq 1$ ? Can we construct, from the on-axis gradients, a vector potential in the Poincaré-Coulomb gauge for these cases? Here we assume that the expansion point  $\mathbf{R}_0$  has been selected such that  $\mathbf{r}$  is small, at least initially. Subsequently we will require that  $x$  and  $y$  remain small, but may allow  $z$  to become large.

In cylindrical coordinates the gradient operator takes the form

$$\nabla = \mathbf{e}_\rho(\partial/\partial\rho) + \mathbf{e}_\phi(1/\rho)(\partial/\partial\phi) + \mathbf{e}_z(\partial/\partial z). \quad (15.7.15)$$

It follows from (2.20) and (7.15) that

$$\mathbf{r} \cdot \nabla = \rho(\partial/\partial\rho) + z(\partial/\partial z). \quad (15.7.16)$$

Also, we conclude from (2.20) and (2.21) that

$$\mathbf{r} \cdot \hat{\mathbf{A}}^{m,\alpha} = \rho\hat{A}_\rho^{m,\alpha} + z\hat{A}_z^{m,\alpha}. \quad (15.7.17)$$

Upon combining (7.9), (7.16), and (7.17) we see that  $\hat{\chi}_{m,\alpha}$  must satisfy the equation

$$[\rho(\partial/\partial\rho) + z(\partial/\partial z)]\hat{\chi}_{m,\alpha} = -(\rho\hat{A}_\rho^{m,\alpha} + z\hat{A}_z^{m,\alpha}). \quad (15.7.18)$$

Since the  $\hat{\chi}_{m,\alpha}$  are known to be harmonic, and in view of (3.33), let us make the Ansätze

$$\begin{aligned} \hat{\chi}_{m,c}(x, y, z; Z_0) &= -\sin(m\phi) \sum_{k=0}^{\infty} (-1)^k \frac{m!}{2^{2k} k!(k+m)!} D_{m,c}^{[2k]}(z; Z_0) \rho^{2k+m} \\ &= -\rho^m \sin(m\phi) \sum_{k=0}^{\infty} (-1)^k \frac{m!}{2^{2k} k!(k+m)!} D_{m,c}^{[2k]}(z; Z_0) \rho^{2k} \\ &= -\Im[(x+iy)^m] \sum_{k=0}^{\infty} (-1)^k \frac{m!}{2^{2k} k!(k+m)!} D_{m,c}^{[2k]}(z; Z_0) \rho^{2k}, \end{aligned} \quad (15.7.19)$$

$$\begin{aligned} \hat{\chi}_{m,s}(x, y, z; Z_0) &= \cos(m\phi) \sum_{k=0}^{\infty} (-1)^k \frac{m!}{2^{2k} k!(k+m)!} D_{m,s}^{[2k]}(z; Z_0) \rho^{2k+m} \\ &= \rho^m \cos(m\phi) \sum_{k=0}^{\infty} (-1)^k \frac{m!}{2^{2k} k!(k+m)!} D_{m,s}^{[2k]}(z; Z_0) \rho^{2k} \\ &= \Re[(x+iy)^m] \sum_{k=0}^{\infty} (-1)^k \frac{m!}{2^{2k} k!(k+m)!} D_{m,s}^{[2k]}(z; Z_0) \rho^{2k}, \end{aligned} \quad (15.7.20)$$

where the functions  $D_{m,\alpha}^{[0]}(z; Z_0)$  are yet to be determined. Note that these  $D$  functions are not to be confused with those appearing in (5.53). Observe also that, for future use, (7.19) and (7.20) are expressed in both cylindrical and Cartesian variables.

Recall that in Subsection 15.4.2 we found the Coulomb gauge results

$$\hat{A}_\rho^{m,c} = -(1/2)(\sin m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(Z_0 + z)\rho^{2\ell+m+1}, \quad (15.7.21)$$

$$\hat{A}_z^{m,c} = (\sin m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m)!} C_{m,c}^{[2\ell]}(Z_0 + z)\rho^{2\ell+m}; \quad (15.7.22)$$

$$\hat{A}_\rho^{m,s} = (1/2)(\cos m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,s}^{[2\ell+1]}(Z_0 + z)\rho^{2\ell+m+1}, \quad (15.7.23)$$

$$\hat{A}_z^{m,s} = -(\cos m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m)!} C_{m,s}^{[2\ell]}(Z_0 + z)\rho^{2\ell+m}. \quad (15.7.24)$$

[Here we have again employed the variables of this section in writing (7.21) through (7.24).] We observe that the operator appearing on the left side of (7.18) does not involve the variable  $\phi$ . Therefore, we may cancel like trigonometric factors appearing on the right and left sides of (7.18) to find, in the case  $\alpha = c$ , the requirement

$$\begin{aligned} & [\rho(\partial/\partial\rho) + z(\partial/\partial z)] \left[ \sum_{k=0}^{\infty} (-1)^k \frac{m!}{2^{2k}k!(k+m)!} D_{m,c}^{[2k]}(z; Z_0)\rho^{2k+m} \right] = \\ & -(\rho/2) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(Z_0 + z)\rho^{2\ell+m+1} \\ & + z \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m)!} C_{m,c}^{[2\ell]}(Z_0 + z)\rho^{2\ell+m}; \end{aligned} \quad (15.7.25)$$

and, in the case  $\alpha = s$ , the requirement

$$\begin{aligned} & [\rho(\partial/\partial\rho) + z(\partial/\partial z)] \left[ \sum_{k=0}^{\infty} (-1)^k \frac{m!}{2^{2k}k!(k+m)!} D_{m,s}^{[2k]}(z; Z_0)\rho^{2k+m} \right] = \\ & -(\rho/2) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,s}^{[2\ell+1]}(Z_0 + z)\rho^{2\ell+m+1} \\ & + z \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m)!} C_{m,s}^{[2\ell]}(Z_0 + z)\rho^{2\ell+m}. \end{aligned} \quad (15.7.26)$$

Note the pleasant fact that the requirements (7.25) and (7.26) are identical in form. The operator on the left sides of (7.25) and (7.26) may be moved under the summation sign and allowed to work its will. For example, there is the general result

$$[\rho(\partial/\partial\rho) + z(\partial/\partial z)][D_{m,\alpha}^{[2k]}(z; Z_0)\rho^{2k+m}] = [(2k+m)D_{m,\alpha}^{[2k]}(z; Z_0) + zD_{m,\alpha}^{[2k+1]}(z; Z_0)]\rho^{2k+m}. \quad (15.7.27)$$

Also, the indicated multiplications on the right sides of (7.25) and (7.26) can be carried out. The net result of these two manipulations is the requirement

$$\begin{aligned} & \sum_{k=0}^{\infty} [(-1)^k \frac{m!}{2^{2k} k!(k+m)!}] [(2k+m) D_{m,\alpha}^{[2k]}(z; Z_0) + z D_{m,\alpha}^{[2k+1]}(z; Z_0)] \rho^{2k+m} = \\ & \sum_{\ell=0}^{\infty} (-1)^{\ell+1} \frac{m!}{2^{2\ell+1} \ell!(\ell+m+1)!} C_{m,\alpha}^{[2\ell+1]}(Z_0 + z) \rho^{2\ell+m+2} \\ & + \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell!(\ell+m)!} z C_{m,\alpha}^{[2\ell]}(Z_0 + z) \rho^{2\ell+m}. \end{aligned} \quad (15.7.28)$$

Let us equate the coefficients of powers of  $\rho$  on both sides of (7.28) to obtain, we hope, relations that will specify the  $D_{m,\alpha}^{[2k]}(z; Z_0)$  in terms of the  $C_{m,\alpha}^{[2\ell]}(Z_0 + z)$ . The lowest power of  $\rho$  on the left side of (7.28) occurs for  $k = 0$ , and is  $\rho^m$ . Its coefficient is

$$\text{Coefficient of } \rho^m \text{ on left side} = m D_{m,\alpha}^{[0]}(z; Z_0) + z D_{m,\alpha}^{[1]}(z; Z_0). \quad (15.7.29)$$

The lowest power of  $\rho$  on the right side of (7.28) occurs for  $\ell = 0$ , and is also  $\rho^m$ . Its coefficient is

$$\text{Coefficient of } \rho^m \text{ on right side} = z C_{m,\alpha}^{[0]}(Z_0 + z). \quad (15.7.30)$$

We conclude, so far, that there is the requirement that  $D_{m,\alpha}^{[0]}(z; Z_0)$  must satisfy the differential equation

$$z D_{m,\alpha}^{[1]}(z; Z_0) + m D_{m,\alpha}^{[0]}(z; Z_0) = z C_{m,\alpha}^{[0]}(Z_0 + z). \quad (15.7.31)$$

We now seek to solve (7.31). Begin by multiplying both sides of (7.31) by  $z^{m-1}$  to yield the result

$$z^m D_{m,\alpha}^{[1]}(z; Z_0) + z^{m-1} m D_{m,\alpha}^{[0]}(z; Z_0) = z^m C_{m,\alpha}^{[0]}(Z_0 + z). \quad (15.7.32)$$

Observe that

$$z^m D_{m,\alpha}^{[1]}(z; Z_0) + z^{m-1} m D_{m,\alpha}^{[0]}(z; Z_0) = (d/dz)[z^m D_{m,\alpha}^{[0]}(z; Z_0)], \quad (15.7.33)$$

and therefore  $z^{m-1}$  is an integrating factor for (7.31). It follows that (7.31) can be rewritten in the form

$$(d/dz)[z^m D_{m,\alpha}^{[0]}(z; Z_0)] = z^m C_{m,\alpha}^{[0]}(Z_0 + z) \quad (15.7.34)$$

with the immediate general solution

$$z^m D_{m,\alpha}^{[0]}(z; Z_0) = \text{constant} + \int_0^z dz' (z')^m C_{m,\alpha}^{[0]}(Z_0 + z'), \quad (15.7.35)$$

or

$$D_{m,\alpha}^{[0]}(z; Z_0) = \text{constant} \times z^{-m} + \int_0^z dz' (z'/z)^m C_{m,\alpha}^{[0]}(Z_0 + z'). \quad (15.7.36)$$

If we seek a particular solution that is analytic in  $z$ , then we must set the constant term to zero. Also, suppose we introduce a new variable of integration  $\lambda$  by writing

$$\lambda = z'/z. \quad (15.7.37)$$

When these steps are made, (7.36) takes the final form

$$D_{m,\alpha}^{[0]}(z; Z_0) = z \int_0^1 d\lambda \lambda^m C_{m,\alpha}^{[0]}(Z_0 + \lambda z). \quad (15.7.38)$$

We arrived at the requirement (7.31) by equating the coefficients of the lowest power of  $\rho^m$  in (7.28). What happens if we equate the coefficients of the higher powers? We hope *nothing* new since  $D_{m,\alpha}^{[0]}(z; Z_0)$  is already specified by (7.38). The next highest power of  $\rho$  appearing on the left side of (7.28) is  $\rho^{m+2}$ , and occurs for  $k = 1$ . Its coefficient is

$$\text{Coefficient of } \rho^{m+2} \text{ on left side} = -[\frac{m!}{4(m+1)!}] [(m+2)D_{m,\alpha}^{[2]}(z; Z_0) + zD_{m,\alpha}^{[3]}(z; Z_0)]. \quad (15.7.39)$$

The next highest power of  $\rho$  appearing on the right side of (7.28) is also  $\rho^{m+2}$ , and occurs for  $\ell = 0$  in the first term on the right and  $\ell = 1$  in the second term. Its coefficient is

$$\text{Coefficient of } \rho^{m+2} \text{ on right side} = -[\frac{m!}{2(m+1)!}] C_{m,\alpha}^{[1]}(Z_0 + z) - [\frac{m!}{4(m+1)!}] zC_{m,\alpha}^{[2]}(Z_0 + z). \quad (15.7.40)$$

Equating these two coefficients yields the result

$$(m+2)D_{m,\alpha}^{[2]}(z; Z_0) + zD_{m,\alpha}^{[3]}(z; Z_0) = 2C_{m,\alpha}^{[1]}(Z_0 + z) + zC_{m,\alpha}^{[2]}(Z_0 + z). \quad (15.7.41)$$

Is this result new? It is not. Differentiating both sides of the previous result (7.31) yields the relation

$$\partial_z [zD_{m,\alpha}^{[1]}(z; Z_0) + mD_{m,\alpha}^{[0]}(z; Z_0)] = \partial_z [zC_{m,\alpha}^{[0]}(Z_0 + z)] \quad (15.7.42)$$

which, upon expansion, yields the result

$$zD_{m,\alpha}^{[2]}(z; Z_0) + (1+m)D_{m,\alpha}^{[1]}(z; Z_0) = zC_{m,\alpha}^{[1]}(Z_0 + z) + C_{m,\alpha}^{[0]}(Z_0 + z). \quad (15.7.43)$$

Next, differentiating both sides of (7.43) yields the relation

$$\partial_z [zD_{m,\alpha}^{[2]}(z; Z_0) + (1+m)D_{m,\alpha}^{[1]}(z; Z_0)] = \partial_z [zC_{m,\alpha}^{[1]}(Z_0 + z) + C_{m,\alpha}^{[0]}(Z_0 + z)] \quad (15.7.44)$$

which, upon expansion, yields the result

$$zD_{m,\alpha}^{[3]}(z; Z_0) + (2+m)D_{m,\alpha}^{[2]}(z; Z_0) = zC_{m,\alpha}^{[2]}(Z_0 + z) + 2C_{m,\alpha}^{[1]}(Z_0 + z). \quad (15.7.45)$$

We see that this result agrees with (7.41). Further calculation shows that all the results obtained by equating the coefficients of like powers of  $\rho$  on both sides of (7.28) are consistent with the relation (7.31) and are identical to results that flow from it upon differentiation.

In summary, the  $\hat{\chi}_{m,\alpha}$  are specified by (7.19) and (7.20) in terms of the  $D_{m,\alpha}^{[0]}(z; Z_0)$ , and the  $D_{m,\alpha}^{[0]}(z; Z_0)$  are specified by (7.38) in terms of the  $C_{m,\alpha}^{[0]}(Z_0 + z)$ . What remains, according to (7.5), is to compute  $\nabla \hat{\chi}_{m,\alpha}$ . Introduce the notation

$$\Delta \mathbf{A}^{m,\alpha} = {}^P C \mathbf{A}^{m,\alpha} - \hat{\mathbf{A}}^{m,\alpha} = \nabla \hat{\chi}_{m,\alpha}. \quad (15.7.46)$$

Then, from (7.15), (7.19), and (7.20), we have the relations

$$\Delta A_{\rho}^{m,c} = (\partial/\partial\rho)\hat{\chi}_{m,c} = -\sin(m\phi) \sum_{k=0}^{\infty} (-1)^k \frac{m!(2k+m)}{2^{2k}k!(k+m)!} D_{m,c}^{[2k]}(z; Z_0) \rho^{2k+m-1}, \quad (15.7.47)$$

$$\Delta A_{\phi}^{m,c} = (1/\rho)(\partial/\partial\phi)\hat{\chi}_{m,c} = -m \cos(m\phi) \sum_{k=0}^{\infty} (-1)^k \frac{m!}{2^{2k}k!(k+m)!} D_{m,c}^{[2k]}(z; Z_0) \rho^{2k+m-1}, \quad (15.7.48)$$

$$\Delta A_z^{m,c} = (\partial/\partial z)\hat{\chi}_{m,c} = -\sin(m\phi) \sum_{k=0}^{\infty} (-1)^k \frac{m!}{2^{2k}k!(k+m)!} D_{m,c}^{[2k+1]}(z; Z_0) \rho^{2k+m}; \quad (15.7.49)$$

$$\Delta A_{\rho}^{m,s} = (\partial/\partial\rho)\hat{\chi}_{m,s} = \cos(m\phi) \sum_{k=0}^{\infty} (-1)^k \frac{m!(2k+m)}{2^{2k}k!(k+m)!} D_{m,s}^{[2k]}(z; Z_0) \rho^{2k+m-1}, \quad (15.7.50)$$

$$\Delta A_{\phi}^{m,s} = (1/\rho)(\partial/\partial\phi)\hat{\chi}_{m,s} = -m \sin(m\phi) \sum_{k=0}^{\infty} (-1)^k \frac{m!}{2^{2k}k!(k+m)!} D_{m,s}^{[2k]}(z; Z_0) \rho^{2k+m-1}, \quad (15.7.51)$$

$$\Delta A_z^{m,s} = (\partial/\partial z)\hat{\chi}_{m,s} = \cos(m\phi) \sum_{k=0}^{\infty} (-1)^k \frac{m!}{2^{2k}k!(k+m)!} D_{m,s}^{[2k+1]}(z; Z_0) \rho^{2k+m}. \quad (15.7.52)$$

These are the results in cylindrical coordinates. Suppose results in Cartesian coordinates are desired. The relations (7.49) and (7.52) are already in Cartesian form. The remaining Cartesian-form results may be found using the relations

$$\Delta A_x^{m,\alpha} = \cos\phi \Delta A_{\rho}^{m,\alpha} - \sin\phi \Delta A_{\phi}^{m,\alpha}, \quad (15.7.53)$$

$$\Delta A_y^{m,\alpha} = \sin\phi \Delta A_{\rho}^{m,\alpha} + \cos\phi \Delta A_{\phi}^{m,\alpha}. \quad (15.7.54)$$

Recall (2.24) and (2.25).

There are two other points that require discussion. The first, a matter of consistency, is this: We have found formulas for the  $\Delta \mathbf{A}^{m,\alpha}$  in the case  $m > 0$ . What happens when these formulas are evaluated for  $m = 0$ ? From (7.47) through (7.49) we immediately see that

$$\Delta \mathbf{A}^{0,c} = 0, \quad (15.7.55)$$

and therefore

$${}^{PC}\mathbf{A}^{0,c} = \hat{\mathbf{A}}^{0,c}. \quad (15.7.56)$$

This result is consistent with previous results because we know that  $\hat{\mathbf{A}}^{0,c} = \hat{\mathbf{A}}^0$  is already in the Poincaré-Coulomb gauge, and this gauge is unique. What about  $\Delta \mathbf{A}^{0,s}$ ? From (3.36) we recall that  $C_{0,s}^{[0]}(Z_0 + z)$  vanishes, and therefore according to (7.38) all the  $D_{0,s}^{[n]}$  vanish. Consequently, according to (7.50) through (7.52),  $\Delta \mathbf{A}^{0,s}$  also vanishes,

$$\Delta \mathbf{A}^{0,s} = 0. \quad (15.7.57)$$

Therefore, in view of (5.96), there is the result

$${}^{PC}\mathbf{A}^{0,s} = 0. \quad (15.7.58)$$

The second point has to do with the nature of the relation between the Poincaré-Coulomb vector potential we have found (which we know is unique from the work of Subsection 2.6) and the on-axis gradients. The relations (4.15) through (4.26) and (5.83) through (5.94) provide, at any point  $z$ , formulas for the vector potential in the azimuthal-free and Coulomb gauges in terms of the on-axis gradients  $C_{m,\alpha}^{[0]}(z)$  and their first few derivatives at the *same* point  $z$ . In particular, if expansions in powers of  $x$  and  $y$  are required only through some finite order (as is the case), then only a finite number of derivatives of the  $C_{m,\alpha}^{[0]}(z)$  are required. In this sense, we may say that these vector potentials depend *locally* on the  $C_{m,\alpha}^{[0]}(z)$ . By contrast, according to (7.38), it appears that computation of the  $D_{m,\alpha}^{[0]}(z; Z_0)$ , and therefore of the vector potential in the Poincaré-Coulomb gauge at this value of  $z$ , requires a knowledge of the  $C_{m,\alpha}^{[0]}(Z_0 + z')$  over the full interval  $z' \in [0, z]$ . Thus the  $z$  dependence of the vector potential in the Poincaré-Coulomb gauge appears to be *nonlocal* in the  $C_{m,\alpha}^{[0]}(Z_0 + z')$ . This conclusion is correct if  $z$  is large, as it may well be. However if  $z$  is small, which will be the case in the vicinity of the expansion point  $\mathbf{R}_0$ , and if we are content with a polynomial expansion in powers of  $z$ , which is all that is required to find a polynomial expansion of the Poincaré-Coulomb gauge vector potential in the vicinity of  $\mathbf{R}_0$ , then we can do better. By Taylor's theorem we may write

$$\begin{aligned} C_{m,\alpha}^{[0]}(Z_0 + \lambda z) = & C_{m,\alpha}^{[0]}(Z_0) + C_{m,\alpha}^{[1]}(Z_0)(\lambda z) + (1/2!)C_{m,\alpha}^{[2]}(Z_0)(\lambda z)^2 \\ & + (1/3!)C_{m,\alpha}^{[3]}(Z_0)(\lambda z)^3 + \dots \end{aligned} \quad (15.7.59)$$

It follows from (7.38) that

$$\begin{aligned} D_{m,\alpha}^{[0]}(z; Z_0) = & \{[1/(m+1)]C_{m,\alpha}^{[0]}(Z_0)\}z + \{[1/(m+2)]C_{m,\alpha}^{[1]}(Z_0)\}z^2 \\ & + \{[1/(m+3)][1/(2!)]C_{m,\alpha}^{[2]}(Z_0)\}z^3 + \{[1/(m+4)][1/(3!)]C_{m,\alpha}^{[3]}(Z_0)\}z^4 + \dots \end{aligned} \quad (15.7.60)$$

Correspondingly we conclude that, in order to obtain a polynomial expansion of the Poincaré-Coulomb gauge vector potential in the vicinity of  $\mathbf{R}_0$ , only a finite number of the  $C_{m,\alpha}^{[n]}(Z_0)$  are required.

The fact that the relation between the Poincaré-Coulomb gauge vector potential and the  $C_{m,\alpha}^{[0]}(Z_0)$  depends on the expansion point  $\mathbf{R}_0$  should not be entirely surprising. The defining requirements (4.2) and (5.1) for the azimuthal-free and Coulomb-gauge vector potentials are required to hold for all  $z$ , and are thus  $z$  *independent*. By contrast, the defining requirement (2.72) for the Poincaré-Coulomb gauge involves  $\mathbf{r}$ , which in turn involves the expansion point.

## 15.8 Relations Between Vector Potentials in Various Gauges and Associated Gauge Functions

We have found three general vector potentials specified in terms of on-axis gradients, namely the azimuthal free, symmetric Coulomb, and Poincaré-Coulomb gauge vector potentials. The purpose of this section is to find, in terms of on-axis gradients, the gauge transformation

functions that interrelate these vector potentials. And, once these gauge transformation functions are known, there are associated symplectic maps given by relations of the form (6.2.79). See Exercises 6.2.8 and 6.5.3. These results will be of subsequent use. See, for example, Subsection 16.1.3.

### 15.8.1 Transformation Between Azimuthal-Free Gauge and Symmetric Coulomb Gauge

The desired results for this subsection have already been found. The azimuthal free gauge and symmetric Coulomb gauge vector potentials are related by the gauge transformations (5.43) employing the gauge functions  $\chi_{m,\alpha}$ . And the gauge functions  $\chi_{m,\alpha}$  are given in terms of on-axis gradients by the relations (5.75) and (5.76).

As an example, let us see how this machinery works in the case of a normal quadrupole ( $m = 2$  and  $\alpha = s$ ). For this purpose it is convenient to employ the more elaborate notation introduced at the end of Subsection 5.2. Let  ${}^{AF}\mathbf{A}^{2,s}$  and  ${}^{SC}\mathbf{A}^{2,s}$  be the vector potentials in the *azimuthal-free* and *symmetric Coulomb* gauges, respectively. Then, from (4.30) through (4.32), we find the results

$${}^{AF}A_x^{2,s} = (1/2)(x^3 - xy^2)C_{2,s}^{[1]}(z) + \dots, \quad (15.8.1)$$

$${}^{AF}A_y^{2,s} = -(1/2)(y^3 - yx^2)C_{2,s}^{[1]}(z) + \dots, \quad (15.8.2)$$

$${}^{AF}A_z^{2,s} = -(x^2 - y^2)C_{2,s}^{[0]}(z) + (1/6)(x^4 - y^4)C_{2,s}^{[2]}(z) + \dots. \quad (15.8.3)$$

And, from (5.100) through (5.102), we find the results

$${}^{SC}A_x^{2,s} = (1/6)(x^3 - 3xy^2)C_{2,s}^{[1]}(z) + \dots, \quad (15.8.4)$$

$${}^{SC}A_y^{2,s} = -(1/6)(y^3 - 3x^2y)C_{2,s}^{[1]}(z) + \dots, \quad (15.8.5)$$

$${}^{SC}A_z^{2,s} = -(x^2 - y^2)C_{2,s}^{[0]}(z) + (1/12)(x^4 - y^4)C_{2,s}^{[2]}(z) + \dots. \quad (15.8.6)$$

Review Exercise 5.7.

According to (5.43) we expect the relation

$${}^{SC}\mathbf{A}^{m,s} = {}^{AF}\mathbf{A}^{m,s} + \nabla\chi_{m,s} \quad (15.8.7)$$

with, according to (5.76),  $\chi_{m,s}$  given by

$$\chi_{m,s} = -(1/2)\Re[(x+iy)^m] \sum_{\ell=0}^{\infty} (-1)^\ell \frac{(m-1)!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z)\rho^{2\ell+2}. \quad (15.8.8)$$

As a side comment we note that, if we wish, we may employ the more elaborate notation introduced at the end of Subsection 5.2 to rewrite (8.7) in the form

$${}^{SC}\mathbf{A}^{m,s} = {}^{AF}\mathbf{A}^{m,s} + \nabla {}^{AF}\vec{\chi}_{m,s}^{SC}(x,y,z). \quad (15.8.9)$$

where

$${}^{AF}\vec{\chi}_{m,s}^{SC}(xyz) = \chi_{m,s}(x,y,z). \quad (15.8.10)$$

Let us evaluate the result (8.8) for  $m = 2$  and then list the first few terms in the series. So doing gives the result

$$\begin{aligned}\chi_{2,s} &= -(1/2)\Re[(x+iy)^2]\{(1/6)C_{2,s}^{[1]}(z)\rho^2 - (1/96)C_{2,s}^{[3]}(z)\rho^4 + \dots\} \\ &= -(1/2)(x^2 - y^2)(1/6)C_{2,s}^{[1]}(z)(x^2 + y^2) + \dots \\ &= -(1/12)(x^4 - y^4)C_{2,s}^{[1]}(z) + \dots.\end{aligned}\quad (15.8.11)$$

It is easily checked that  $\chi_{2,s}$  as given by (8.11) satisfies (8.7) for  $m = 2$  and  $\alpha = s$ .

### 15.8.2 Transformation Between Symmetric Coulomb Gauge and Poincaré-Coulomb Gauge

The necessary ingredients for completing this subsection have also already been found. The symmetric Coulomb gauge vector potentials and the Poincaré-Coulomb gauge vector potentials are related by the gauge transformations (7.5) employing the gauge functions  $\hat{\chi}_{m,\alpha}$ . In turn, the gauge functions  $\hat{\chi}_{m,\alpha}$  are given in terms of the functions  $D_{m,\alpha}^{[2k]}$  by the relations (7.19) and (7.20). Finally, the functions  $D_{m,\alpha}^{[0]}$  are given in terms of the on-axis gradients by the relations (7.38) or, equivalently, (7.60).

Let us make the relations between the  $\hat{\chi}_{m,\alpha}$  and the on-axis gradients explicit. To do so it is convenient to introduce a *master* function that we will call  $F_{m,\alpha}(\rho, z; Z_0)$  and is defined by the rule

$$F_{m,\alpha}(\rho, z; Z_0) = \sum_{k=0}^{\infty} (-1)^k \frac{m!}{2^{2k} k!(k+m)!} D_{m,\alpha}^{[2k]}(z; Z_0) \rho^{2k}. \quad (15.8.12)$$

Look again at (7.19) and (7.20). Evidently, when Cartesian coordinates are employed, the  $\hat{\chi}_{m,\alpha}$  are given in terms of  $F_{m,\alpha}(\rho, z; Z_0)$  by the relations

$$\hat{\chi}_{m,c}(x, y, z; Z_0) = -\Im[(x+iy)^m] F_{m,c}(\rho, z; Z_0), \quad (15.8.13)$$

$$\hat{\chi}_{m,s}(x, y, z; Z_0) = \Re[(x+iy)^m] F_{m,s}(\rho, z; Z_0). \quad (15.8.14)$$

Let us rewrite the relation between the  $D_{m,\alpha}^{[0]}(z; Z_0)$  and the on-axis gradients as given by (7.60),

$$\begin{aligned}D_{m,\alpha}^{[0]}(z; Z_0) &= \{[1/(m+1)]C_{m,\alpha}^{[0]}(Z_0)\}z + \{[1/(m+2)]C_{m,\alpha}^{[1]}(Z_0)\}z^2 \\ &\quad + \{[1/(m+3)](1/2!)C_{m,\alpha}^{[2]}(Z_0)\}z^3 + \{[1/(m+4)](1/3!)C_{m,\alpha}^{[3]}(Z_0)\}z^4 + \dots.\end{aligned}\quad (15.8.15)$$

From (8.4) we find, for examples, the relations

$$\begin{aligned}D_{m,\alpha}^{[2]}(z; Z_0) &= \{[2/(m+2)]C_{m,\alpha}^{[1]}(Z_0)\} + \{[1/(m+3)](6/2!)C_{m,\alpha}^{[2]}(Z_0)\}z \\ &\quad + \{[1/(m+4)](12/3!)C_{m,\alpha}^{[3]}(Z_0)\}z^2 + \dots,\end{aligned}\quad (15.8.16)$$

$$D_{m,\alpha}^{[4]}(z; Z_0) = \{[1/(m+4)](24/3!)C_{m,\alpha}^{[3]}(Z_0)\} + \dots . \quad (15.8.17)$$

Upon employing (8.4) through (8.6) in (8.1) we see that  $F_{m,\alpha}(\rho, z; Z_0)$  is given in terms of on-axis gradients by an expansion of the form

$$\begin{aligned} F_{m,\alpha}(\rho, z; Z_0) &= C_{m,\alpha}^{[0]}(Z_0)[1/(m+1)]z \\ &+ C_{m,\alpha}^{[1]}(Z_0)\{[1/(m+2)]z^2 - [2(m+2)(m+1)]^{-1}\rho^2\} \\ &+ C_{m,\alpha}^{[2]}(Z_0)\{[1/(m+3)](1/2!)z^3 + *z\rho^2\} \\ &+ C_{m,\alpha}^{[3]}(Z_0)\{[1/(m+4)](1/3!)z^4 + *z^2\rho^2 + *\rho^4\} \\ &+ \dots . \end{aligned} \quad (15.8.18)$$

As an example, let us apply this result to the case of a normal quadrupole, the case  $\alpha = s$  and  $m = 2$ . Inserting these values into (8.16) yields for  $F_{2,s}(\rho, z; Z_0)$  the expansion

$$\begin{aligned} F_{2,s}(\rho, z; Z_0) &= C_{2,s}^{[0]}(Z_0)[1/(2+1)]z \\ &+ C_{2,s}^{[1]}(Z_0)\{[1/(2+2)]z^2 - [2(2+2)(2+1)]^{-1}\rho^2\} + \dots \\ &= C_{2,s}^{[0]}(Z_0)(1/3)z + C_{2,s}^{[1]}(Z_0)[(1/4)z^2 - (1/24)\rho^2] + \dots . \end{aligned} \quad (15.8.19)$$

Correspondingly, according to (8.14),  $\hat{\chi}_{2,s}$  has the expansion

$$\begin{aligned} \hat{\chi}_{2,s}(x, y, z; Z_0) &= (1/3)(x^2 - y^2)zC_{2,s}^{[0]}(Z_0) \\ &+ (x^2 - y^2)[(1/4)z^2 - (1/24)\rho^2]C_{2,s}^{[1]}(Z_0) + \dots \\ &= (1/3)[(x^2 - y^2)z]C_{2,s}^{[0]}(Z_0) \\ &+ [(1/4)(x^2 - y^2)z^2 - (1/24)(x^4 - y^4)]C_{2,s}^{[1]}(Z_0) + \dots . \end{aligned} \quad (15.8.20)$$

We can use this result to find  ${}^{PC}\mathbf{A}^{2,s}(x, y, z; Z_0)$ . For the case under consideration (7.5) takes the form

$${}^{PC}\mathbf{A}^{2,s}(x, y, z; Z_0) = {}^{SC}\mathbf{A}^{2,s}(x, y, Z_0 + z) + \nabla \hat{\chi}_{2,s}(x, y, z; Z_0) \quad (15.8.21)$$

or, in our more elaborate notation,

$${}^{PC}\mathbf{A}^{2,s}(x, y, z; Z_0) = {}^{SC}\mathbf{A}^{2,s}(x, y, Z_0 + z) + \nabla {}^{SC}\vec{\chi}_{2,s}^{PC}(x, y, z; Z_0). \quad (15.8.22)$$

where

$${}^{SC}\vec{\chi}_{2,s}^{PC}(x, y, z; Z_0) = \hat{\chi}_{2,s}(x, y, z; Z_0). \quad (15.8.23)$$

The first step is to find  ${}^{SC}\mathbf{A}^{2,s}(x, y, Z_0 + z)$  in the form of an expansion in  $x$ ,  $y$ , and  $z$ . That is, we wish to write

$${}^{SC}\mathbf{A}^{2,s}(x, y, Z_0 + z) = {}^{SC}\mathbf{A}^{2,s}(x, y, z; Z_0). \quad (15.8.24)$$

Observe that the results (8.4) through (8.6) hold for all  $z$  and that the expansion is being made in the transverse variables  $x$  and  $y$ . Therefore we may also write

$${}^{SC}A_x^{2,s}(x, y, Z) = (1/6)(x^3 - 3xy^2)C_{2,s}^{[1]}(Z) + \dots, \quad (15.8.25)$$

$${}^{SC}A_y^{2,s}(x, y, Z) = -(1/6)(y^3 - 3x^2y)C_{2,s}^{[1]}(Z) + \dots, \quad (15.8.26)$$

$${}^{SC}A_z^{2,s}(x, y, Z) = -(x^2 - y^2)C_{2,s}^{[0]}(Z) + (1/12)(x^4 - y^4)C_{2,s}^{[2]}(Z) + \dots. \quad (15.8.27)$$

At this point make the substitution  $Z = Z_0 + z$  where it is understood that  $z$  now plays the role of a *deviation* variable about  $Z_0$ . Then we may make the expansions

$$C_{2,s}^{[0]}(Z_0 + z) = C_{2,s}^{[0]}(Z_0) + zC_{2,s}^{[1]}(Z_0) + (1/2)z^2C_{2,s}^{[2]}(Z_0) + \dots, \quad (15.8.28)$$

$$C_{2,s}^{[1]}(Z_0 + z) = C_{2,s}^{[1]}(Z_0) + zC_{2,s}^{[2]}(Z_0) + \dots, \quad (15.8.29)$$

$$C_{2,s}^{[2]}(Z_0 + z) = C_{2,s}^{[2]}(Z_0) + \dots. \quad (15.8.30)$$

Now employ, for example, (8.28) through (8.30) in (8.25) to find the result

$$\begin{aligned} {}^{SC}A_x^{2,s}(x, y, z; Z_0) &= {}^{SC}A_x^{2,s}(x, y, Z_0 + z) = (1/6)(x^3 - 3xy^2)C_{2,s}^{[1]}(Z_0 + z) + \dots \\ &= (1/6)(x^3 - 3xy^2)[C_{2,s}^{[1]}(Z_0) + zC_{2,s}^{[2]}(Z_0)] + \dots \\ &= (1/6)(x^3 - 3xy^2)C_{2,s}^{[1]}(Z_0) + (1/6)[(x^3 - 3xy^2)z]C_{2,s}^{[2]}(Z_0) + \dots. \end{aligned} \quad (15.8.31)$$

Similarly, one finds

$${}^{SC}A_y^{2,s}(x, y, z; Z_0) = (1/6)(y^3 - 3x^2y)C_{2,s}^{[1]}(Z_0) + (1/6)[(y^3 - 3x^2y)z]C_{2,s}^{[2]}(Z_0) + \dots, \quad (15.8.32)$$

$$\begin{aligned} {}^{SC}A_z^{2,s}(x, y, z; Z_0) &= -(x^2 - y^2)C_{2,s}^{[0]}(Z_0) - [(x^2 - y^2)z]C_{2,s}^{[1]}(Z_0) \\ &\quad + [-(1/2)(x^2 - y^2)z^2 + (1/12)(x^4 - y^4)]C_{2,s}^{[2]}(Z_0) + \dots. \end{aligned} \quad (15.8.33)$$

The next step is to find the gradient of  $\hat{\chi}_{2,s}(x, y, z; Z_0)$ . From the expansion (8.20) one finds for the gradients the results

$$\partial_x \hat{\chi}_{2,s}(x, y, z; Z_0) = [(2/3)xz]C_{2,s}^{[0]}(Z_0) + [(1/2)xz^2 - (1/6)x^3]C_{2,s}^{[1]}(Z_0) + \dots, \quad (15.8.34)$$

$$\partial_y \hat{\chi}_{2,s}(x, y, z; Z_0) = [-(2/3)yz]C_{2,s}^{[0]}(Z_0) + [-(1/2)yz^2 + (1/6)y^3]C_{2,s}^{[1]}(Z_0) + \dots, \quad (15.8.35)$$

$$\partial_z \hat{\chi}_{2,s}(x, y, z; Z_0) = [(1/3)(x^2 - y^2)]C_{2,s}^{[0]}(Z_0) + [(1/2)z(x^2 - y^2)]C_{2,s}^{[1]}(Z_0) + \dots. \quad (15.8.36)$$

Finally we are ready to use (8.21), now written in the form

$${}^{PC}\mathbf{A}^{2,s}(x, y, z; Z_0) = {}^{SC}\mathbf{A}^{2,s}(x, y, z; Z_0) + \nabla \hat{\chi}_{2,s}(x, y, z; Z_0) \quad (15.8.37)$$

[Recall (8.24).] Adding the results (8.31) through (8.33) for  ${}^{SC}\mathbf{A}^{2,s}(x, y, z; Z_0)$  to the results (8.34) through (8.36) for  $\nabla\hat{\chi}_{2,s}(x, y, z; Z_0)$  gives the final results

$$\begin{aligned} {}^{PC}\mathbf{A}_x^{2,s}(x, y, z; Z_0) &= [(2/3)xz]C_{2,s}^{[0]}(Z_0) + [*]C_{2,s}^{[1]}(Z_0) \\ &\quad + [*]C_{2,s}^{[2]}(Z_0) + \dots, \end{aligned} \quad (15.8.38)$$

$$\begin{aligned} {}^{PC}\mathbf{A}_y^{2,s}(x, y, z; Z_0) &= -[(2/3)yz]C_{2,s}^{[0]}(Z_0) + [*]C_{2,s}^{[1]}(Z_0) \\ &\quad + [*]C_{2,s}^{[2]}(Z_0) + \dots, \end{aligned} \quad (15.8.39)$$

$$\begin{aligned} {}^{PC}\mathbf{A}_z^{2,s}(x, y, z; Z_0) &= -[(2/3)(x^2 - y^2)]C_{2,s}^{[0]}(Z_0) + [*]C_{2,s}^{[1]}(Z_0) \\ &\quad + [*]C_{2,s}^{[2]}(Z_0) + \dots. \end{aligned} \quad (15.8.40)$$

We have found, in terms of on-axis gradients and through terms of degree four, the Poincaré-Coulomb gauge vector potential for a normal quadrupole.

After all this algebra let us make a few sanity checks. Note that, in view of the relation  $Q = 2C_{2,s}^{[0]}$ , the quadratic part of  ${}^{PC}\mathbf{A}^{2,s}(x, y, z; Z_0)$  agrees with (2.165), which is a comfort. Also, it is easy to check that each Cartesian component of  ${}^{PC}\mathbf{A}^{2,s}(x, y, z; Z_0)$  is an harmonic function. Finally is is easily checked that  ${}^{PC}\mathbf{A}^{2,s}(x, y, z; Z_0)$  satisfies the Poincaré-Coulomb gauge conditions

### 15.8.3 Transformation Between Azimuthal-Free Gauge and Poincaré-Coulomb Gauge

Suppose (5.43) is added to (7.5). Doing so gives the result

$${}^{PC}\mathbf{A}^{m,\alpha}(x, y, z; Z_0) = \mathbf{A}^{m,\alpha}(x, y, Z_0 + z) + \nabla\chi_{m,\alpha} + \nabla\hat{\chi}_{m,\alpha}. \quad (15.8.41)$$

We see that there is the gauge transformation relation

$${}^{PC}\mathbf{A}^{m,\alpha}(x, y, z; Z_0) = \mathbf{A}^{m,\alpha}(x, y, Z_0 + z) + \nabla\hat{\chi}_{m,\alpha} \quad (15.8.42)$$

where  $\hat{\chi}_{m,\alpha}$  is the gauge function

$$\hat{\chi}_{m,\alpha}(x, y, z; Z_0) = \chi_{m,\alpha}(x, y, Z_0 + z) + \hat{\chi}_{m,\alpha}(x, y, z; Z_0). \quad (15.8.43)$$

Let us apply this result to the case of a normal quadrupole, the case  $\alpha = s$  and  $m = 2$ . Doing so in (8.42) yields the result

$${}^{PC}\mathbf{A}^{2,s}(x, y, z; Z_0) = \mathbf{A}^{2,s}(x, y, Z_0 + z) + \nabla\hat{\chi}_{2,s} \quad (15.8.44)$$

where  $\hat{\chi}_{2,s}$  is the gauge function

$$\hat{\chi}_{2,s}(x, y, z; Z_0) = \chi_{2,s}(x, y, Z_0 + z) + \hat{\chi}_{2,s}(x, y, z; Z_0). \quad (15.8.45)$$

If we employ the notation  ${}^{AF}\mathbf{A}$ , then (8.44) can be rewritten in the form

$${}^{PC}\mathbf{A}^{2,s}(x, y, z; Z_0) = {}^{AF}\mathbf{A}^{2,s}(x, y, Z_0 + z) + \nabla\hat{\chi}_{2,s}. \quad (15.8.46)$$

## Exercises

**15.8.1.** Verify that  $\chi_{2,s}$  as given by (8.11) satisfies (8.7) for  $m = 2$  and  $\alpha = s$ .

**15.8.2.** Verify the derivation of (8.18).

**15.8.3.** Verify that  $\hat{\chi}_{2,s}(x, y, z; Z_0)$ , as given by (8.20) is harmonic, as expected.

**15.8.4.** Exercises comparing vector potential norms in various gauges.

## 15.9 Scalar Potentials from Sources

We have examined at some length the representation of scalar potentials in terms of cylindrical harmonic expansions and on-axis gradients. The purpose of this section is to describe a complementary representation of scalar potentials in terms of sources.

### 15.9.1 Preliminaries

Define the function  $G(\mathbf{r}, \mathbf{r}')$  by the rule

$$G(\mathbf{r}, \mathbf{r}') = [1/(4\pi)][1/\|\mathbf{r} - \mathbf{r}'\|] \quad (15.9.1)$$

where

$$\|\mathbf{r} - \mathbf{r}'\| = [(\mathbf{r} - \mathbf{r}') \cdot (\mathbf{r} - \mathbf{r}')]^{1/2}. \quad (15.9.2)$$

Let  $\psi(\mathbf{r}, \mathbf{r}')_{mon}$  be the scalar potential at the location  $\mathbf{r}$  arising from a point (*monopole*) source at the location  $\mathbf{r}'$ . In this section we take it to be given by the relation

$$\psi(\mathbf{r}, \mathbf{r}')_{mon} = qG(\mathbf{r}, \mathbf{r}') = [1/(4\pi)][q/\|\mathbf{r} - \mathbf{r}'\|] \quad (15.9.3)$$

where  $q$  is the source strength. For future use recall the mathematical relations

$$\nabla_{\mathbf{r}} G(\mathbf{r}, \mathbf{r}') = -\nabla_{\mathbf{r}'} G(\mathbf{r}, \mathbf{r}') = -[1/(4\pi)](\mathbf{r} - \mathbf{r}')/\|\mathbf{r} - \mathbf{r}'\|^3, \quad (15.9.4)$$

$$\nabla_{\mathbf{r}}^2 G(\mathbf{r}, \mathbf{r}') = \nabla_{\mathbf{r}'}^2 G(\mathbf{r}, \mathbf{r}') = -\delta_3(\mathbf{r} - \mathbf{r}'). \quad (15.9.5)$$

### 15.9.2 Monopole Volume Distributions

Let  $V$  be any volume. Consider the potential  $\psi(\mathbf{r})_{vol}$  given by the *volume* monopole distribution

$$\psi(\mathbf{r})_{vol} = \int_V d^3\mathbf{r}' \rho(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') \quad (15.9.6)$$

where  $\rho(\mathbf{r}')$  is the volume source density. In view of (9.5),  $\psi(\mathbf{r})_{vol}$  satisfies the Poisson equation

$$\nabla_{\mathbf{r}}^2 \psi(\mathbf{r})_{vol} = -\rho(\mathbf{r}) \quad (15.9.7)$$

for any location  $\mathbf{r} \in V$ . Note that  $G$  has units of  $1/L$ . Also, since  $\psi$  has units of  $BL$ ,  $q$  has units of  $BL^2$  and  $\rho$  has units of  $BL/L^2 = B/L$ .

### 15.9.3 Monopole Surface Distributions

We will also need to consider potentials given by *surface* monopole distributions. Let  $S$  be some two-dimensional surface. Define the potential  $\psi(\mathbf{r})_{sur}$  by the integral

$$\psi(\mathbf{r})_{sur} = \psi(\mathbf{r})_{sl} = \int_S dS' \sigma(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') \quad (15.9.8)$$

where  $\sigma(\mathbf{r}')$  is the surface monopole source density. [The meaning of the notation  $\psi(\mathbf{r})_{sl}$  will be explained shortly.] In view of (9.5),  $\psi(\mathbf{r})_{sur}$  is harmonic at locations  $\mathbf{r}$  that are not on the surface  $S$ ,

$$\nabla_{\mathbf{r}}^2 \psi(\mathbf{r})_{sur} = 0 \text{ for } \mathbf{r} \notin S. \quad (15.9.9)$$

### 15.9.4 Dipole Surface Distributions

Next, let  $\psi(\mathbf{r}, \mathbf{r}')_{dip}$  be the scalar potential at the location  $\mathbf{r}$  arising from a point *dipole* source at the location  $\mathbf{r}'$ . It is given by the relation

$$\psi(\mathbf{r}, \mathbf{r}')_{dip} = [1/(4\pi)][\mathbf{p} \cdot (\mathbf{r} - \mathbf{r}')]/||\mathbf{r} - \mathbf{r}'||^3 = -[1/(4\pi)](\mathbf{p} \cdot \nabla_{\mathbf{r}})[1/||\mathbf{r} - \mathbf{r}'||] \quad (15.9.10)$$

where  $\mathbf{p}$  is the dipole moment. Here we have used (9.4). Since the operators  $\nabla_{\mathbf{r}}^2$  and  $\nabla_{\mathbf{r}}$  commute, and in view of (9.4) and (9.5), it follows that  $\psi(\mathbf{r}, \mathbf{r}')_{dip}$  is an harmonic function of  $\mathbf{r}$  at all locations  $\mathbf{r}$  different from  $\mathbf{r}'$ ,

$$\nabla_{\mathbf{r}}^2 \psi(\mathbf{r}, \mathbf{r}')_{dip} = 0 \text{ for } \mathbf{r} \neq \mathbf{r}'. \quad (15.9.11)$$

We may also consider surface dipole distributions. Define the function  $\psi(\mathbf{r})_{surdip}$  by the rule

$$\psi(\mathbf{r})_{surdip} = \int_S dS' [1/(4\pi)][\mathbf{T}(\mathbf{r}') \cdot (\mathbf{r} - \mathbf{r}')]/||\mathbf{r} - \mathbf{r}'||^3 \quad (15.9.12)$$

where  $\mathbf{T}(\mathbf{r}')$  is a surface dipole source density. In view of (9.11),  $\psi(\mathbf{r})_{surdip}$  is harmonic at locations  $\mathbf{r}$  that are not on the surface  $S$ ,

$$\nabla_{\mathbf{r}}^2 \psi(\mathbf{r})_{surdip} = 0 \text{ for } \mathbf{r} \notin S. \quad (15.9.13)$$

There is a special case of (9.12) that is of particular significance for our purposes, namely the case where  $\mathbf{T}(\mathbf{r}')$  points along or against the direction of  $d\mathbf{S}'$ . In this case we speak of a *double layer* source distribution and make the definition

$$\psi(\mathbf{r})_{dl} = [1/(4\pi)] \int_S d\mathbf{S}' \cdot (\mathbf{r} - \mathbf{r}') \tau(\mathbf{r}')/||\mathbf{r} - \mathbf{r}'||^3 \quad (15.9.14)$$

where  $\tau(\mathbf{r}')$  is the double layer source density. Because of (9.13), the function  $\psi(\mathbf{r})_{dl}$  is also harmonic at locations  $\mathbf{r}$  that are not on the surface,

$$\nabla_{\mathbf{r}}^2 \psi(\mathbf{r})_{dl} = 0 \text{ for } \mathbf{r} \notin S. \quad (15.9.15)$$

A dipole may be viewed as the limiting case of a monopole doublet where the spacing between the poles tends to zero and the strengths of the two oppositely charged poles tend

to  $\pm$  infinity. If a distribution of dipoles on some surface  $S$  is arranged to have all dipoles oriented *perpendicular* to the surface, then we may regard this distribution as being a double layer so that we may speak of a double layer source distribution with density  $\tau(\mathbf{r}')$ . In the same vein, we may view the surface monopole distribution appearing in (9.8) as being a *single layer* distribution. That is the intent of the notation  $\psi(\mathbf{r})_{sl}$  in (9.8).

### 15.9.5 Scalar Potential in Terms of Volume and Single and Double Layer Source Distributions

We are now prepared to state and prove a remarkable mathematical result: Suppose  $\psi(\mathbf{r})$  is any function defined in a volume  $V$  surrounded by a surface  $S$ . Then there are functions  $\rho(\mathbf{r}')$ ,  $\sigma(\mathbf{r}')$ , and  $\tau(\mathbf{r}')$  such that

$$\begin{aligned}\psi(\mathbf{r}) &= \int_V d^3\mathbf{r}' \rho(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') \\ &+ \int_S dS' \sigma(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') \\ &+ [1/(4\pi)] \int_S d\mathbf{S}' \cdot (\mathbf{r} - \mathbf{r}') \tau(\mathbf{r}') / ||\mathbf{r} - \mathbf{r}'||^3.\end{aligned}\quad (15.9.16)$$

Note that the three terms appearing on the right side of (9.16) are of the form  $\psi(\mathbf{r})_{vol}$ ,  $\psi(\mathbf{r})_{sl}$ , and  $\psi(\mathbf{r})_{dl}$ , respectively.

Assuming (9.16) holds, from (9.7), (9.9), and (9.15) it follows that there is the Poisson relation

$$\nabla_{\mathbf{r}}^2 \psi(\mathbf{r}) = -\rho(\mathbf{r}) \text{ for } \mathbf{r} \in V. \quad (15.9.17)$$

Consequently, if  $\psi$  is *harmonic* in  $V$  but otherwise arbitrary, then there is the corollary relation

$$\psi(\mathbf{r}) = \int_S dS' \sigma(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') + [1/(4\pi)] \int_S d\mathbf{S}' \cdot (\mathbf{r} - \mathbf{r}') \tau(\mathbf{r}') / ||\mathbf{r} - \mathbf{r}'||^3. \quad (15.9.18)$$

Any solution of Laplace's equation in a volume  $V$  can be written in terms of single and double layer distributions on a surrounding surface  $S$ .

At this point it is convenient to pause to think about units. Suppose the magnetic field strength  $B$  is measured in Tesla. Then, according to (2.6), the scalar potential  $\psi$  has units of Tesla-Meter. Next, from (9.25), it follows that  $\rho$  has units of Tesla/Meter. And, from (9.30),  $\sigma$  has units of Tesla. Finally, from (9.28),  $\tau$  has units of Tesla-Meter. Note that these last three units can also all be derived from looking at (9.32). In summary, letting the symbol  $L$  stand for *Length*, we have the dimensional relations

$$\psi \sim BL, \quad \rho \sim B/L, \quad \sigma \sim B, \quad \tau \sim BL. \quad (15.9.19)$$

Let us continue. To prove (9.16), we will make use of a relation between two particular volume and surface integrals that is referred to by some authors as *Green's theorem*.<sup>12</sup> We begin by stating and proving Green's theorem, and then apply it to the problem at hand.

---

<sup>12</sup>These authors also refer to the relation between closed-path line integrals and related surface integrals as *Stokes' theorem*. Frequently other authors refer to a special case of Stokes' theorem as Green's theorem, namely the case where the closed path lies in a plane and the related surface lies in the same plane.

### 15.9.6 Green's Theorem

Suppose that  $\psi(\mathbf{r}')$  and  $\phi(\mathbf{r}')$  are any two functions sufficiently smooth to justify the mathematical manipulations about to be performed. It is easily verified that there is the algebraic identity

$$\begin{aligned}\psi(\mathbf{r}')\nabla_{\mathbf{r}'}^2\phi(\mathbf{r}') - \phi(\mathbf{r}')\nabla_{\mathbf{r}'}^2\psi(\mathbf{r}') = \\ \nabla_{\mathbf{r}'} \cdot [\psi(\mathbf{r}')\nabla_{\mathbf{r}'}\phi(\mathbf{r}') - \phi(\mathbf{r}')\nabla_{\mathbf{r}'}\psi(\mathbf{r}')].\end{aligned}\quad (15.9.20)$$

Integrate both sides of (9.19) over a volume  $V$  to obtain the result

$$\begin{aligned}\int_V d^3\mathbf{r}' [\psi(\mathbf{r}')\nabla_{\mathbf{r}'}^2\phi(\mathbf{r}') - \phi(\mathbf{r}')\nabla_{\mathbf{r}'}^2\psi(\mathbf{r}')] = \\ \int_V d^3\mathbf{r}' \nabla_{\mathbf{r}'} \cdot [\psi(\mathbf{r}')\nabla_{\mathbf{r}'}\phi(\mathbf{r}') - \phi(\mathbf{r}')\nabla_{\mathbf{r}'}\psi(\mathbf{r}')].\end{aligned}\quad (15.9.21)$$

The volume integral on the right side of (9.20) can be transformed into a surface integral with the aid of Gauss' divergence theorem,

$$\begin{aligned}\int_V d^3\mathbf{r}' \nabla_{\mathbf{r}'} \cdot [\psi(\mathbf{r}')\nabla_{\mathbf{r}'}\phi(\mathbf{r}') - \phi(\mathbf{r}')\nabla_{\mathbf{r}'}\psi(\mathbf{r}')] = \\ \int_S d\mathbf{S}' \cdot [\psi(\mathbf{r}')\nabla_{\mathbf{r}'}\phi(\mathbf{r}') - \phi(\mathbf{r}')\nabla_{\mathbf{r}'}\psi(\mathbf{r}')].\end{aligned}\quad (15.9.22)$$

Therefore we may also write

$$\begin{aligned}\int_V d^3\mathbf{r}' [\psi(\mathbf{r}')\nabla_{\mathbf{r}'}^2\phi(\mathbf{r}') - \phi(\mathbf{r}')\nabla_{\mathbf{r}'}^2\psi(\mathbf{r}')] = \\ \int_S d\mathbf{S}' \cdot [\psi(\mathbf{r}')\nabla_{\mathbf{r}'}\phi(\mathbf{r}') - \phi(\mathbf{r}')\nabla_{\mathbf{r}'}\psi(\mathbf{r}')].\end{aligned}\quad (15.9.23)$$

This result is Green's theorem.

### 15.9.7 Application of Green's Theorem

Green's theorem applies to any pair of sufficiently smooth functions  $\psi$  and  $\phi$ . Now consider the case where  $\psi(\mathbf{r}')$  is any function defined for  $\mathbf{r}' \in V \cup S$ , and sufficiently smooth to justify the mathematical manipulations about to be performed, and  $\phi(\mathbf{r}')$  is the function

$$\phi(\mathbf{r}') = G(\mathbf{r}, \mathbf{r}'). \quad (15.9.24)$$

In this case Green's theorem takes the form

$$\begin{aligned}\int_V d^3\mathbf{r}' [\psi(\mathbf{r}')\nabla_{\mathbf{r}'}^2G(\mathbf{r}, \mathbf{r}') - G(\mathbf{r}, \mathbf{r}')\nabla_{\mathbf{r}'}^2\psi(\mathbf{r}')] = \\ \int_S d\mathbf{S}' \cdot [\psi(\mathbf{r}')\nabla_{\mathbf{r}'}G(\mathbf{r}, \mathbf{r}') - G(\mathbf{r}, \mathbf{r}')\nabla_{\mathbf{r}'}\psi(\mathbf{r}')].\end{aligned}\quad (15.9.25)$$

We will now work to evaluate both sides of (9.24).

Begin with the left side. Specify a function  $\rho(\mathbf{r}')$  by the rule

$$\rho(\mathbf{r}') = -\nabla_{\mathbf{r}'}^2 \psi(\mathbf{r}'). \quad (15.9.26)$$

Also, employ (9.5). Then evaluation of the left side of (9.24) yields the net result

$$\begin{aligned} \int_V d^3 \mathbf{r}' [\psi(\mathbf{r}') \nabla_{\mathbf{r}'}^2 G(\mathbf{r}, \mathbf{r}') - G(\mathbf{r}, \mathbf{r}') \nabla_{\mathbf{r}'}^2 \psi(\mathbf{r}')] &= \\ - \int_V d^3 \mathbf{r}' \psi(\mathbf{r}') \delta_3(\mathbf{r} - \mathbf{r}') + \int_V d^3 \mathbf{r}' \rho(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') &= \\ -\psi(\mathbf{r}) + \int_V d^3 \mathbf{r}' \rho(\mathbf{r}') G(\mathbf{r}, \mathbf{r}'). \end{aligned} \quad (15.9.27)$$

We have evaluated the left side of (9.24). Now work on evaluating the right side of (9.24). By using (9.4), the first term on the right side of (9.24) can be rewritten in the form

$$\begin{aligned} \int_S d\mathbf{S}' \cdot [\psi(\mathbf{r}') \nabla_{\mathbf{r}'} G(\mathbf{r}, \mathbf{r}')] &= \\ [1/(4\pi)] \int_S d\mathbf{S}' \cdot (\mathbf{r} - \mathbf{r}') \psi(\mathbf{r}') / \|\mathbf{r} - \mathbf{r}'\|^3 &= \\ -[1/(4\pi)] \int_S d\mathbf{S}' \cdot (\mathbf{r} - \mathbf{r}') \tau(\mathbf{r}') / \|\mathbf{r} - \mathbf{r}'\|^3 \end{aligned} \quad (15.9.28)$$

where

$$\tau(\mathbf{r}') = -\psi(\mathbf{r}') \text{ for } \mathbf{r}' \in S. \quad (15.9.29)$$

The second term on the right side of (9.24) can be rewritten in the form

$$-\int_S d\mathbf{S}' \cdot G(\mathbf{r}, \mathbf{r}') \nabla_{\mathbf{r}'} \psi(\mathbf{r}') = -\int_S d\mathbf{S}' \sigma(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') \quad (15.9.30)$$

where

$$\sigma(\mathbf{r}') = \mathbf{n}(\mathbf{r}') \cdot \nabla_{\mathbf{r}'} \psi(\mathbf{r}') \text{ for } \mathbf{r}' \in S \quad (15.9.31)$$

and  $\mathbf{n}(\mathbf{r}')$  is the outward unit normal vector to  $S$  at the location  $\mathbf{r}'$ . Combining (9.29) and (9.27) gives for the right side of (9.24) the net result

$$\begin{aligned} \int_S d\mathbf{S}' \cdot [\psi(\mathbf{r}') \nabla_{\mathbf{r}'} G(\mathbf{r}, \mathbf{r}') - G(\mathbf{r}, \mathbf{r}') \nabla_{\mathbf{r}'} \psi(\mathbf{r}')] &= \\ - \int_S d\mathbf{S}' \sigma(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') - [1/(4\pi)] \int_S d\mathbf{S}' \cdot (\mathbf{r} - \mathbf{r}') \tau(\mathbf{r}') / \|\mathbf{r} - \mathbf{r}'\|^3. \end{aligned} \quad (15.9.32)$$

Finally, it follows from (9.24) that the two net results (9.26) and (9.31) must be equal. Therefore we may write

$$\begin{aligned} -\psi(\mathbf{r}) + \int_V d^3 \mathbf{r}' \rho(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') &= \\ - \int_S d\mathbf{S}' \sigma(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') - [1/(4\pi)] \int_S d\mathbf{S}' \cdot (\mathbf{r} - \mathbf{r}') \tau(\mathbf{r}') / \|\mathbf{r} - \mathbf{r}'\|^3, \end{aligned} \quad (15.9.33)$$

which is a result equivalent to the assertion (9.16). We also note that (9.25), (9.30), and (9.28) provide, in terms of  $\psi$ , formulas for the source densities  $\rho$ ,  $\sigma$ , and  $\tau$ .

## Exercises

**15.9.1.** Verify the identity (9.19).

**15.9.2.** Suppose that  $\psi(\mathbf{r})$  is given by the relation

$$\psi(\mathbf{r}) = \mathbf{r} \cdot \mathbf{r}, \quad (15.9.34)$$

that  $S$  is the surface of a sphere of radius  $R$  centered on the origin, and that  $V$  is the interior of the sphere. Find the source densities  $\rho$ ,  $\sigma$ , and  $\tau$ , and verify that (9.16) holds by evaluating all relevant integrals.

## 15.10 Normal Magnetic Monopole Doublet Example and Applications

To validate the numerical methods to be presented in Chapters 17 through 20 and Chapters 23 and 24, it will be useful to have a test problem. One such problem is that of the field of a *normal monopole doublet*. (The justification for the adjective *normal* as opposed to *skew* will become apparent subsequently.) This field has rapid spatial field variations, thereby posing a challenge to numerical methods, and is also exactly computable in analytic form.

The normal monopole-doublet problem will also illustrate the methods of Sections 2 and 3 of this chapter. Finally, the work of this section and \* subsequent sections will prove useful for discovering, in Sections \* and \*, relations between on-axis gradients and single and double layer sources on cylindrical surfaces.

### 15.10.1 Magnetic Scalar Potential and Magnetic Field

In analogy with the case of electrostatics, we suppose there is a magnetic charge density  $\rho_{mag}(\mathbf{r})$  and that the scalar potential  $\psi$  obeys a Poisson equation of the form

$$\nabla^2 \psi(\mathbf{r}) = -\rho_{mag}(\mathbf{r}). \quad (15.10.1)$$

(This relation is identical to that for the electrostatics case in MKS units except that we have set  $\epsilon_0 = 1$ . So doing just defines the units we will be using for magnetic charge.) We then know from electrostatic theory, which in this case is just a mathematical result, that a solution of (10.1) is given by the relation

$$\psi(\mathbf{r}) = [1/(4\pi)] \int_{\text{all space}} d^3 \mathbf{r}' \rho_{mag}(\mathbf{r}') / \|\mathbf{r} - \mathbf{r}'\|. \quad (15.10.2)$$

As a specific application of (10.2), suppose two magnetic monopoles having strengths  $\pm 4\pi g$  are placed at the  $(x, y, z)$  locations

$$\mathbf{r}^+ = (0, a, 0), \quad (15.10.3)$$

and

$$\mathbf{r}^- = (0, -a, 0). \quad (15.10.4)$$

See Figure 10.1, which also shows a circular cylinder with radius  $R$ . The magnetic charge density in this case is given by the relation

$$\rho_{mag}(\mathbf{r}) = 4\pi g\delta^3(\mathbf{r} - \mathbf{r}^+) - 4\pi g\delta^3(\mathbf{r} - \mathbf{r}^-), \quad (15.10.5)$$

and according to (10.2) the associated scalar potential  $\psi(x, y, z)$  is given by the relation

$$\begin{aligned} \psi(x, y, z) &= g[x^2 + (y - a)^2 + z^2]^{-1/2} - g[x^2 + (y + a)^2 + z^2]^{-1/2} \\ &= \psi^+(x, y, z) + \psi^-(x, y, z). \end{aligned} \quad (15.10.6)$$

[Here the notation is such that  $\psi^+$  is singular at  $\mathbf{r}^+$  and  $\psi^-$  is singular at  $\mathbf{r}^-$ . We have also introduced a factor of  $4\pi$  in the specification of the monopole strengths so that subsequent formulas such as (10.6) will be free of  $4\pi$  factors.] Correspondingly, these monopoles produce a magnetic field  $\mathbf{B} = -\nabla\psi$  having the components

$$B_x = gx[x^2 + (y - a)^2 + z^2]^{-3/2} - gx[x^2 + (y + a)^2 + z^2]^{-3/2}, \quad (15.10.7)$$

$$B_y = g(y - a)[x^2 + (y - a)^2 + z^2]^{-3/2} - g(y + a)[x^2 + (y + a)^2 + z^2]^{-3/2}, \quad (15.10.8)$$

$$B_z = gz[x^2 + (y - a)^2 + z^2]^{-3/2} - gz[x^2 + (y + a)^2 + z^2]^{-3/2}. \quad (15.10.9)$$

This field is sketched in Figure 10.2. Note that the magnetic field lines appear to come out of the pole with positive magnetic charge, and into the pole with negative magnetic charge, just as in the electrostatic analog. We know that the gradient of a scalar field points in the direction of greatest change. Because of the minus sign in  $\mathbf{B} = -\nabla\psi$ ,  $\mathbf{B}$  must point in the “down hill” direction. It follows that  $\psi$  becomes evermore positive as one approaches a positive pole, and ever more negative as one approaches a negative pole. Inspection of the mathematical expression (10.6) shows that this is indeed the case for the monopole doublet example. And inspection of (10.2) shows that this is to be expected. Suppose, for example, that  $\rho_{mag}(\mathbf{r}')$  is very positive for  $\mathbf{r}'$  in some region  $R$ . Then, for  $\mathbf{r}$  in  $R$ , the integral (10.2) will be dominated by the value of  $\rho_{mag}(\mathbf{r}')$  in  $R$  because there  $1/\|\mathbf{r} - \mathbf{r}'\|$  will be very large, and correspondingly  $\psi$  will be positive. Although  $1/\|\mathbf{r}\|$  is not as singular as  $\delta^3(\mathbf{r})$ ,  $\psi$  does tend to “track”  $\rho_{mag}$ . [Note that, according to (10.1), what do track exactly are the functions  $-\nabla^2\psi$  and  $\rho_{mag}$ .] Finally, in magnetostatic theory, magnetic field lines are supposed to come out of north poles and go into south poles. Therefore, we identify poles with positive magnetic charge as being north poles and poles with negative magnetic charge as being south poles.

To provide further insight, Figure 10.3 shows the on-axis field component  $B_y(x = 0, y = 0, z)$ , and Figures 10.4 and 10.5 show the off-axis field components  $B_x(\rho = 1/2, \phi = \pi/4, z)$  and  $B_z(\rho = 1/2, \phi = \pi/4, z)$ . In Cartesian coordinates, the field components  $B_x$  and  $B_z$  are shown along the line  $x = y = \sqrt{2}/4$  cm  $\simeq .353$  cm,  $z \in [-\infty, \infty]$ . Note that the on-axis field component  $B_y(x = 0, y = 0, z)$  falls off as  $1/|z|^3$  for large  $|z|$ , as expected for a doublet composed of monopoles of opposite strengths.

At this point the reader might object that this field is unphysical since to this time no magnetic monopoles are known to exist. However, as far as an observer inside an interior cylinder of the kind shown in Figure 10.1 is concerned, the field he/she sees is perfectly possible because within the cylinder it obeys  $\nabla \cdot \mathbf{B} = 0$  and  $\nabla \times \mathbf{B} = 0$ .

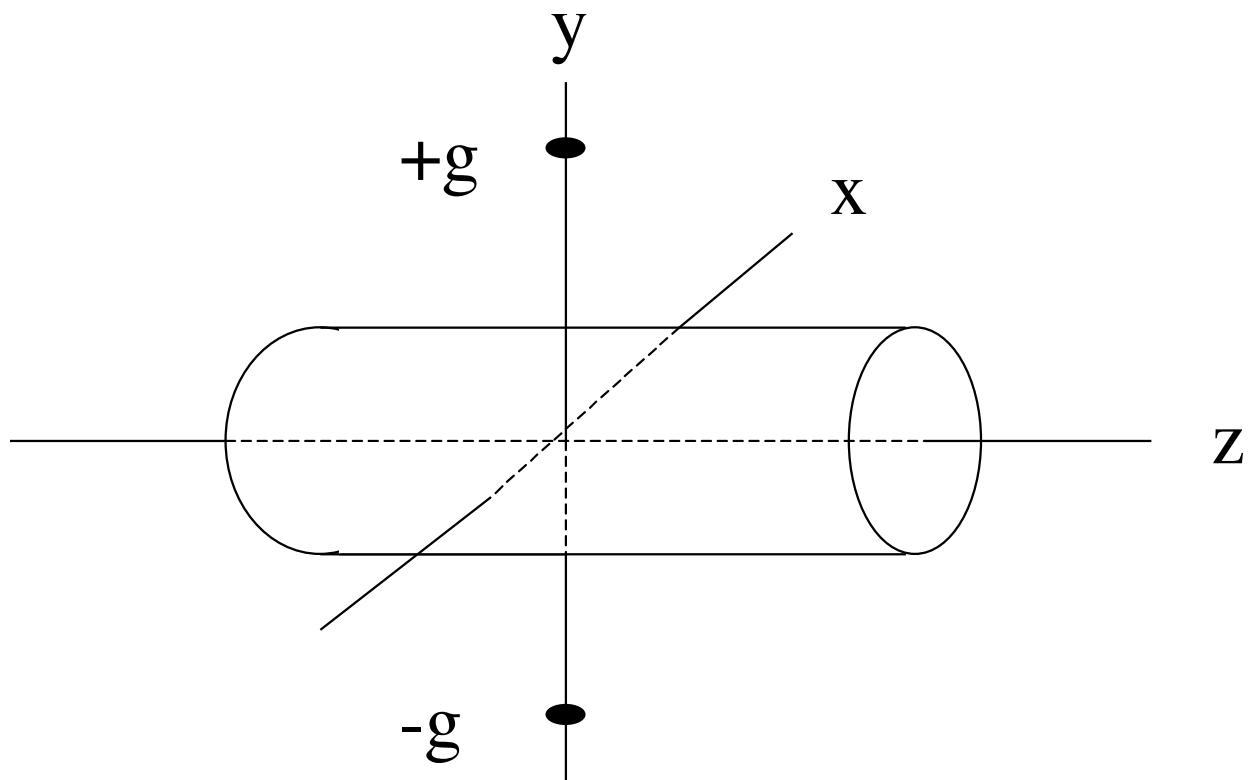


Figure 15.10.1: A normal monopole doublet consisting of two magnetic monopoles of equal and opposite sign placed on the  $y$  axis and centered on the origin. Also shown, for future reference, is a cylinder with circular cross section placed in the interior field.

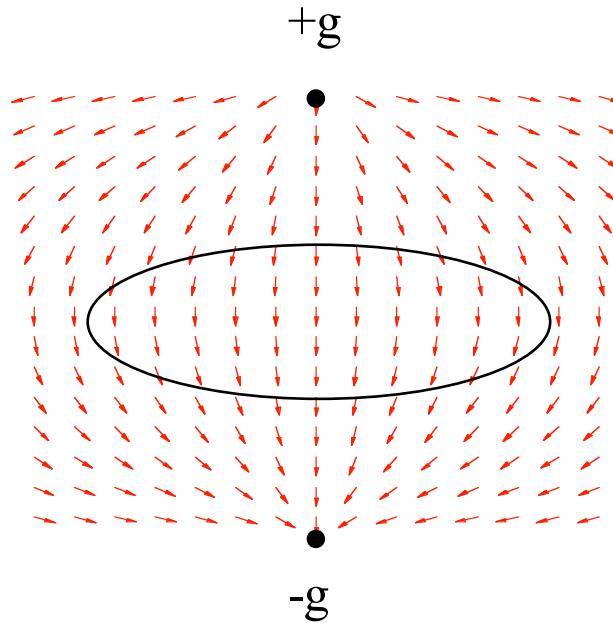


Figure 15.10.2: The interior field of a normal monopole doublet in the  $z = 0$  plane. Also shown is an ellipse whose purpose will become clear in Sections 17.4 and 19.2.

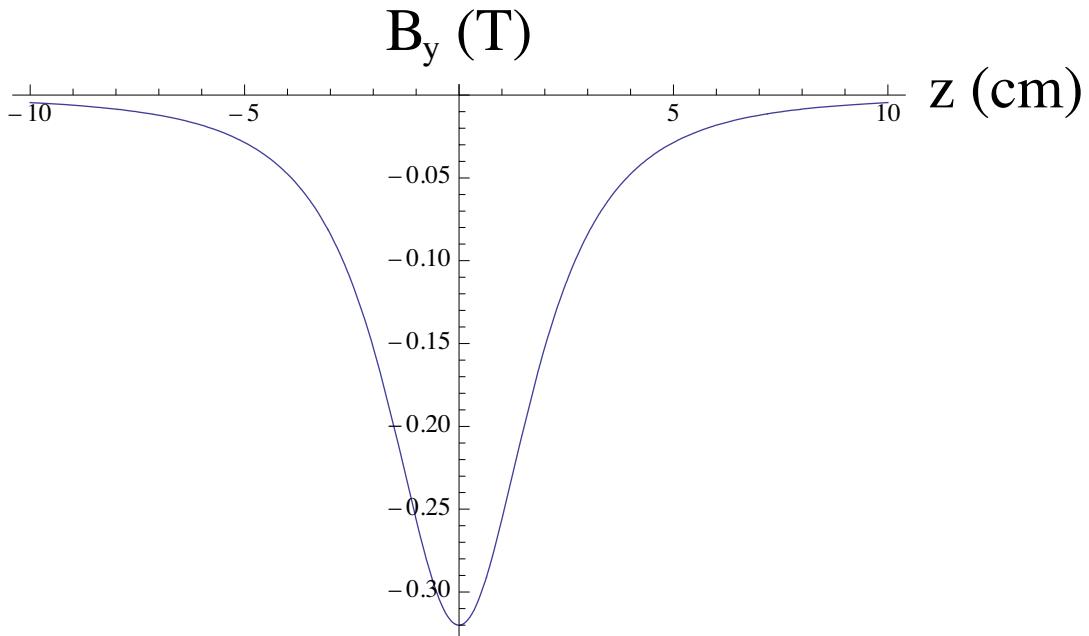


Figure 15.10.3: The on-axis field component  $B_y(x = 0, y = 0, z)$  for the normal monopole doublet in the case that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm) $^2$ . The coordinate  $z$  is given in centimeters.

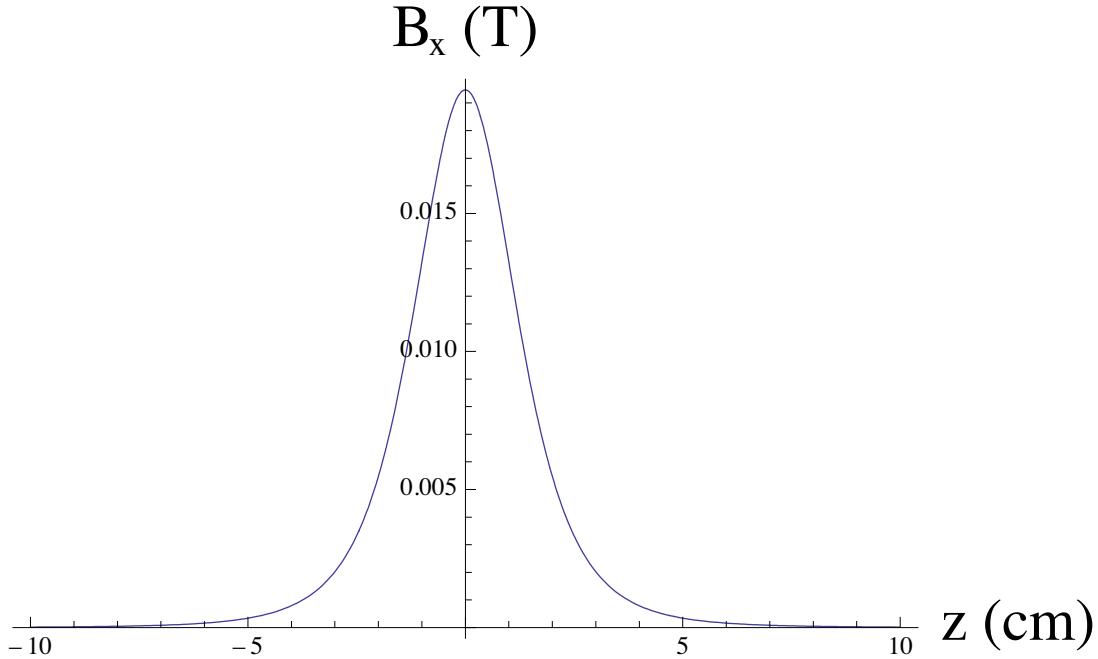


Figure 15.10.4: The field component  $B_x$  on the line  $\rho = 1/2$  cm,  $\phi = \pi/4$ ,  $z \in [-\infty, \infty]$  for the normal monopole doublet in the case that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm) $^2$ . In Cartesian coordinates, this is the line  $x = y \simeq .353$  cm,  $z \in [-\infty, \infty]$ . The coordinate  $z$  is given in centimeters.

The radial component  $B_\rho(\rho, \phi, z)$  of  $\mathbf{B}$  is defined by the relation

$$B_\rho(\rho, \phi, z) = (\cos \phi) B_x + (\sin \phi) B_y. \quad (15.10.10)$$

Recall (2.22). Consequently, using (10.7) and (10.8), we find on the surface  $\rho = R$  the result

$$\begin{aligned} B_\rho(R, \phi, z) &= gR\{[z^2 + R^2 + a^2 - 2aR \sin \phi]^{-3/2} - [z^2 + R^2 + a^2 + 2aR \sin \phi]^{-3/2}\} \\ &\quad - ga \sin \phi \{[z^2 + R^2 + a^2 - 2aR \sin \phi]^{-3/2} + [z^2 + R^2 + a^2 + 2aR \sin \phi]^{-3/2}\}. \end{aligned} \quad (15.10.11)$$

To provide a feel for the behavior of  $B_\rho(R, \phi, z)$ , Figure 10.6 displays  $B_\rho(R = 2, \phi, z = 0)$  as a function of  $\phi$ , and Figure 10.7 shows  $B_\rho(R = 2, \phi = \pi/2, z)$  as a function of  $z$ . We see that the surface field is rather singular. By contrast the fields shown in Figures 10.3 through 10.5, which are those at locations interior to this surface, are less singular. This is to be expected because harmonic functions take their extrema on boundaries.

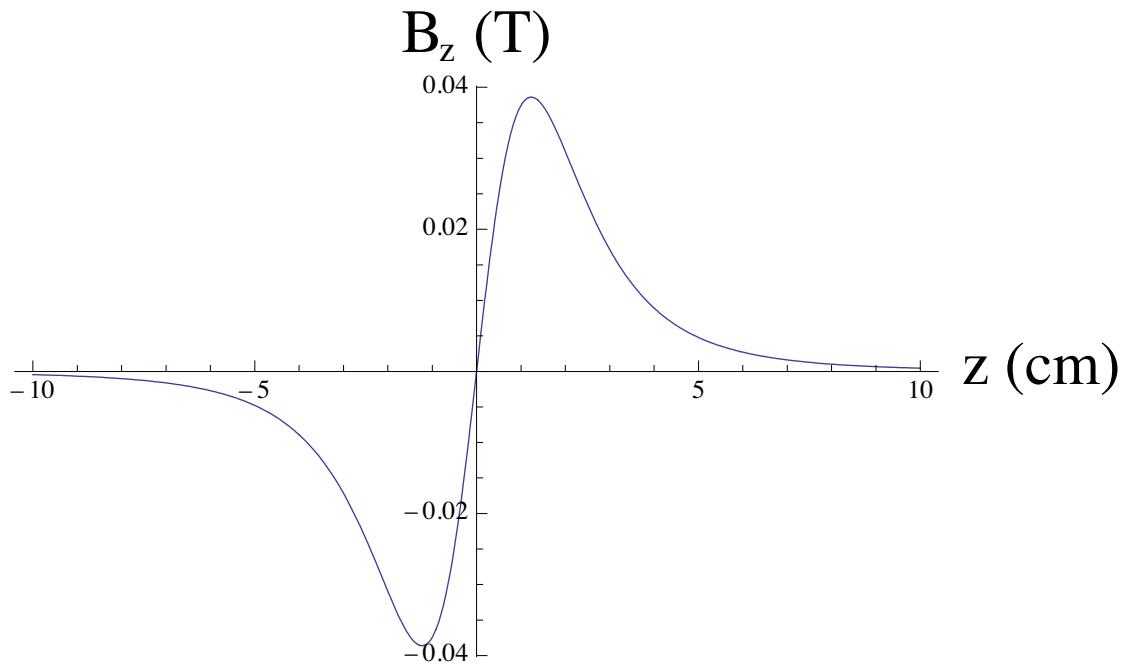


Figure 15.10.5: The field component  $B_z$  on the line  $\rho = 1/2$  cm,  $\phi = \pi/4$ ,  $z \in [-\infty, \infty]$  for the normal monopole doublet in the case that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm) $^2$ . In Cartesian coordinates, this is the line  $x = y \simeq .353$  cm,  $z \in [-\infty, \infty]$ . The coordinate  $z$  is given in centimeters.

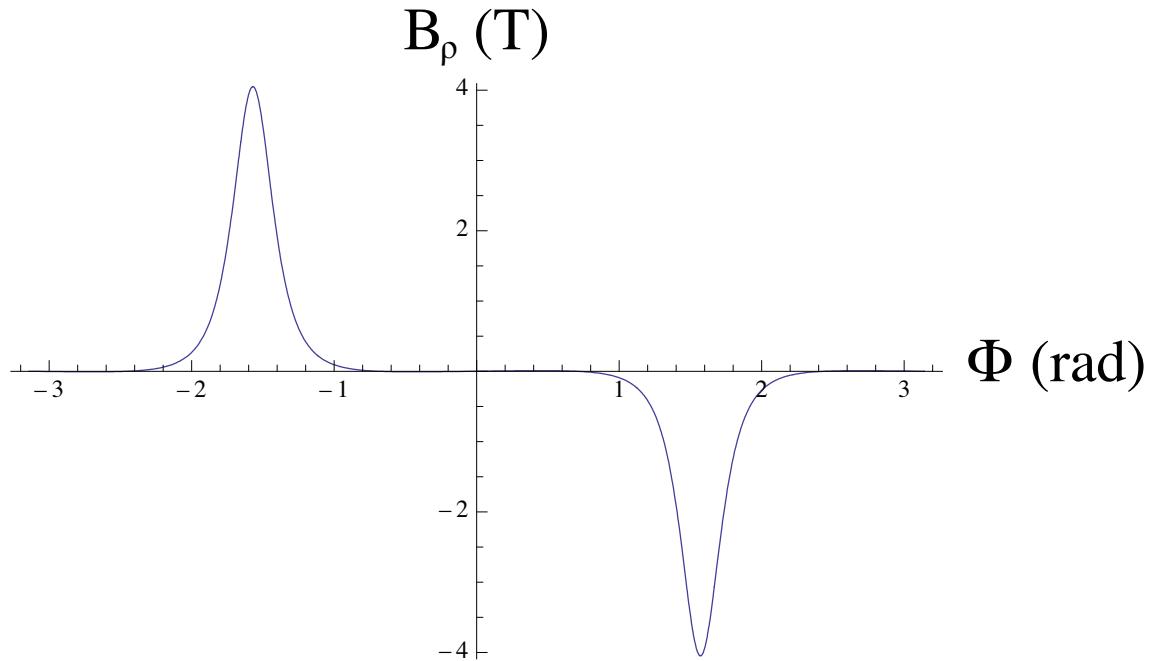


Figure 15.10.6: The quantity  $B_\rho(R, \phi, z = 0)$  for the normal monopole doublet in the case that  $R = 2$  cm,  $a = 2.5$  cm, and  $g = 1$  Tesla-(cm) $^2$ .

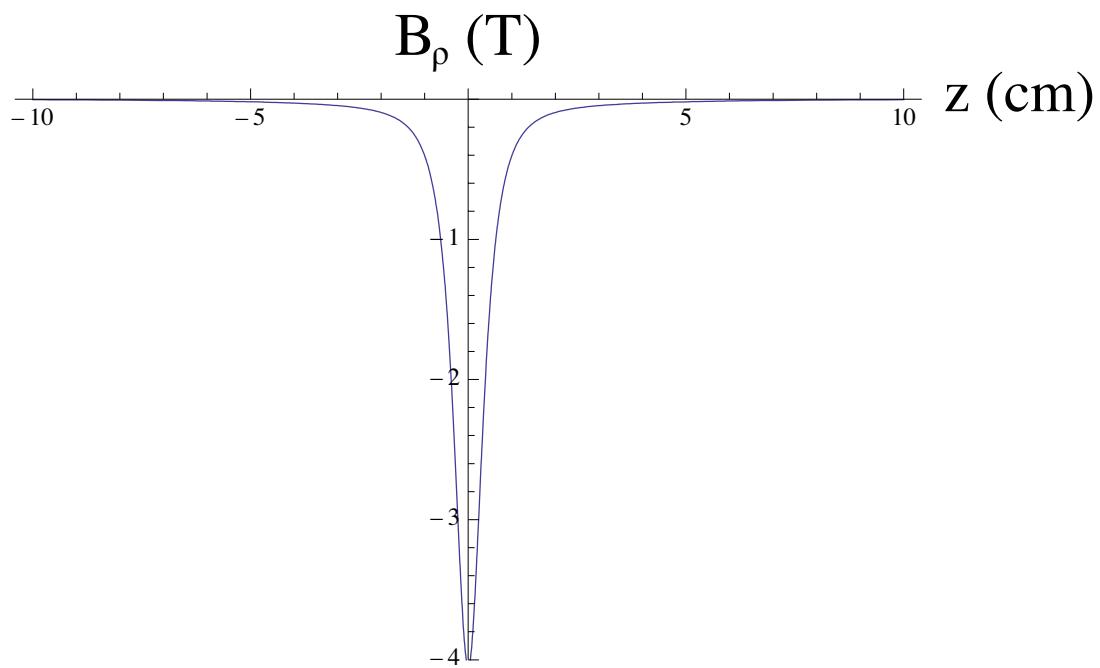


Figure 15.10.7: The quantity  $B_\rho(R, \phi = \pi/2, z)$  for the normal monopole doublet in the case that  $R = 2$  cm,  $a = 2.5$  cm, and  $g = 1$  Tesla-(cm) $^2$ . The coordinate  $z$  is given in centimeters.

### 15.10.2 Analytic On-Axis Gradients for the Normal Monopole Doublet

In this subsection we will find analytic expressions for the on-axis gradients for the normal monopole doublet. In Chapter 19, numerical results for these gradients will be compared against these analytic results.

In view of the form of the expansion (3.33) for  $\psi$ , let us seek power series expansions for  $\psi^\pm(x, y, z)$  in the variable  $\rho$ . In the case of  $\psi^+$ , for example, we may write

$$\begin{aligned}\psi^+(x, y, z) &= g[x^2 + (y - a)^2 + z^2]^{-1/2} = g[(\rho \cos \phi)^2 + (\rho \sin \phi - a)^2 + z^2]^{-1/2} \\ &= g[a^2 + z^2 - 2a\rho \sin \phi + \rho^2]^{-1/2} \\ &= g[a^2 + z^2]^{-1/2} \{1 - [2a\rho/(a^2 + z^2)] \sin \phi + \rho^2/(a^2 + z^2)\} \\ &= g[a^2 + z^2]^{-1/2} [1 - 2wh + h^2]^{-1/2}\end{aligned}\quad (15.10.12)$$

where

$$h = \rho/(a^2 + z^2)^{1/2} \quad (15.10.13)$$

and

$$w = [a/(a^2 + z^2)^{1/2}] \sin \phi. \quad (15.10.14)$$

To proceed further, we employ the mathematical fact that for Legendre polynomials  $P_\ell(w)$  there is the *generating function* expansion

$$[1 - 2wh + h^2]^{-1/2} = \sum_{\ell=0}^{\infty} h^\ell P_\ell(w). \quad (15.10.15)$$

Combining (10.12) through (10.15) gives the result

$$\begin{aligned}\psi^+(x, y, z) &= g[x^2 + (y - a)^2 + z^2]^{-1/2} \\ &= g[a^2 + z^2]^{-1/2} \sum_{\ell=0}^{\infty} [\rho/(a^2 + z^2)^{1/2}]^\ell P_\ell(w).\end{aligned}\quad (15.10.16)$$

Similarly, there is the result

$$\begin{aligned}\psi^-(x, y, z) &= -g[x^2 + (y + a)^2 + z^2]^{-1/2} \\ &= -g[a^2 + z^2]^{-1/2} \sum_{\ell=0}^{\infty} [\rho/(a^2 + z^2)^{1/2}]^\ell P_\ell(-w).\end{aligned}\quad (15.10.17)$$

Introduce the notation  $\psi^d(x, y, z) = \psi^d(\rho, \phi, z)$  for the scalar potential associated with the magnetic monopole doublet we have been studying. Here the superscript  $d$  stands for *doublet*. Employing this notation, we may write

$$\psi^d(\rho, \phi, z) = \psi^+ + \psi^-. \quad (15.10.18)$$

It follows, taking into account the parity of the the Legendre polynomials, that  $\psi^d$  has the expansion

$$\psi^d(\rho, \phi, z) = 2g[a^2 + z^2]^{-1/2} \sum_{n=0}^{\infty} [\rho/(a^2 + z^2)^{1/2}]^{2n+1} P_{2n+1}(w). \quad (15.10.19)$$

Suppose we wish to write  $\psi^d$  in the general form (3.33). From (3.37) through (3.43) we know there are the general relations

$$\Psi_0(\rho, z) = [1/(2\pi)] \int_{-\pi}^{\pi} d\phi \psi(x, y, x), \quad (15.10.20)$$

$$\Psi_{m,c}(\rho, z) = (1/\pi) \int_{-\pi}^{\pi} d\phi \psi(x, y, x) \cos m\phi, \quad (15.10.21)$$

$$\Psi_{m,s}(\rho, z) = (1/\pi) \int_{-\pi}^{\pi} d\phi \psi(x, y, x) \sin m\phi. \quad (15.10.22)$$

In the case of  $\psi = \psi^d$  they take the form

$$\Psi_0^d(\rho, z) = [1/(2\pi)] \int_{-\pi}^{\pi} d\phi \psi^d(x, y, x), \quad (15.10.23)$$

$$\Psi_{m,c}^d(\rho, z) = (1/\pi) \int_{-\pi}^{\pi} d\phi \psi^d(x, y, x) \cos m\phi, \quad (15.10.24)$$

$$\Psi_{m,s}^d(\rho, z) = (1/\pi) \int_{-\pi}^{\pi} d\phi \psi^d(x, y, x) \sin m\phi. \quad (15.10.25)$$

From (10.14) and (10.19) we see that  $\psi^d$  is an *odd* function of  $\phi$ . Therefore, for the normal monopole doublet, the integrals (10.23) and (10.24) vanish by symmetry so that  $\Psi_0^d(\rho, z) = 0$  and  $\Psi_{m,c}^d(\rho, z) = 0$ . Consequently, for the normal monopole doublet, we see from (3.44) and (3.45) that

$$\begin{aligned} C_0^{[0]}(z) &= 0, \\ C_{m,c}^{[0]}(z) &= 0, \end{aligned} \quad (15.10.26)$$

thus justifying use of the adjective *normal*.

With regard to  $\Psi_{m,s}^d(\rho, z)$  we find the result

$$\Psi_{m,s}^d(\rho, z) = (2g/\pi)[a^2 + z^2]^{-1/2} \sum_{n=0}^{\infty} [\rho/(a^2 + z^2)^{1/2}]^{2n+1} \int_{-\pi}^{\pi} d\phi P_{2n+1}(w) \sin m\phi. \quad (15.10.27)$$

To analyze the integral that occurs on the right side of (10.27), introduce the notation

$$\beta(z, a) = a/(a^2 + z^2)^{1/2} \quad (15.10.28)$$

so that

$$w = \beta \sin \phi. \quad (15.10.29)$$

With this notation, we must study integrals of the form

$$c_{m',m}(z) = \int_{-\pi}^{\pi} d\phi P_{m'}(\beta \sin \phi) \sin m\phi \quad (15.10.30)$$

with  $m'$  odd. Indeed, with the definition (10.30) we may rewrite (10.27) in the form

$$\Psi_{m,s}^d(\rho, z) = (2g/\pi)[a^2 + z^2]^{-1/2} \sum_{m' \text{ odd}} c_{m',m}(z) [\rho/(a^2 + z^2)^{1/2}]^{m'}. \quad (15.10.31)$$

We claim, and will verify shortly, that there are the relations

$$c_{m',m} = 0 \text{ for } m' \text{ odd and } m \text{ even,} \quad (15.10.32)$$

$$c_{m',m} = 0 \text{ for } m' < m. \quad (15.10.33)$$

Assuming (10.32) is correct, it follows from (10.31) that

$$\Psi_{m,s}^d(\rho, z) = 0 \text{ for } m \text{ even.} \quad (15.10.34)$$

And, for  $m$  odd, employing (10.33) in (10.31) yields the result

$$\begin{aligned} \Psi_{m,s}^d(\rho, z) &= (2g/\pi)[a^2 + z^2]^{-1/2} c_{m,m}(z) [\rho/(a^2 + z^2)^{1/2}]^m \\ &\quad + (2g/\pi)[a^2 + z^2]^{-1/2} \sum_{m' \text{ odd and } m' > m} c_{m',m}(z) [\rho/(a^2 + z^2)^{1/2}]^{m'}. \end{aligned} \quad (15.10.35)$$

Let us make use of (10.34) and (10.35). Recall from (3.46) that

$$C_{m,s}^{[0]}(z) = -\lim_{\rho \rightarrow 0} (1/\rho^m) \Psi_{m,s}(\rho, z). \quad (15.10.36)$$

An immediate conclusion from (10.31) and (10.32) (consistent with symmetry considerations, see Subsection 3.5) is that  $\Psi_{m,s}^d(\rho, z) = 0$  for even  $m$ ; and hence, by (10.36),

$$C_{m,s}^{[0]}(z) = 0 \text{ for } m \text{ even.} \quad (15.10.37)$$

And from (10.35) and (10.36) we see that

$$C_{m,s}^{[0]}(z) = -(2g/\pi)[a^2 + z^2]^{-1/2} c_{m,m}(z) [1/(a^2 + z^2)^{1/2}]^m \text{ for } m \text{ odd.} \quad (15.10.38)$$

In the limit  $\rho \rightarrow 0$  only the first term on the right side of (10.35) contributes to  $C_{m,s}^{[0]}$  for odd  $m$ .

What remains is to verify (10.32) and (10.33) and to evaluate  $c_{m,m}(z)$  for odd  $m$ . We will use the relations

$$\sin(\ell\phi) \sin(\ell'\phi) = (1/2)\{\cos[(\ell - \ell')\phi] - \cos[(\ell + \ell')\phi]\} \quad (15.10.39)$$

and therefore

$$\int_{-\pi}^{\pi} d\phi \sin(\ell\phi) \sin(\ell'\phi) = \pi \delta_{\ell,\ell'}. \quad (15.10.40)$$

Here, in writing (10.40), we assume that neither  $\ell$  nor  $\ell'$  vanish. If one or both do, then the integral (10.40) obviously vanishes.

To continue, we know from the Taylor expansions for Legendre polynomials that (for odd  $m'$ ) there is the result

$$P_{m'}(w) = \{[(2m')!]/[2^{m'}(m'!)^2]\} w^{m'} + \text{lower (in decrements of two) odd powers of } w. \quad (15.10.41)$$

Consequently, there is the result

$$\begin{aligned} P_{m'}(\beta \sin \phi) &= \{[(2m')!]/[2^{m'}(m'!)^2]\}\beta^{m'}(\sin \phi)^{m'} \\ &\quad + \text{lower (in decrements of two) odd powers of } (\beta \sin \phi). \end{aligned} \quad (15.10.42)$$

We also know (again for odd  $m'$ ) that

$$(\sin \phi)^{m'} = (-1)^{(m'-1)/2}(1/2)^{m'-1} \sin m'\phi + \text{lower odd angular frequency sinusoidal terms.} \quad (15.10.43)$$

It follows from (10.42) and (10.43) that  $P_{m'}(\beta \sin \phi)$  contains terms with highest angular frequency  $m'$  plus lower order odd angular frequency terms. Now suppose  $m$  is even. Then the angular frequencies in  $P_{m'}(\beta \sin \phi)$ , being odd, will differ from the even angular frequency  $m$  in  $\sin m\phi$ ; and therefore by (10.40) the assertion (10.32) holds. Also, since the highest angular frequency in  $P_{m'}(\beta \sin \phi)$  is  $m'$ , it follows from (10.40) that (10.33) is true.

Finally, (with  $m$  odd), we find from (10.30), (10.40), (10.42), and (10.43) that

$$\begin{aligned} c_{m,m} &= (-1)^{(m-1)/2} \{[(2m)!]/[2^m(m!)^2]\} (1/2)^{m-1} \beta^m \int_{-\pi}^{\pi} d\phi \sin^2(m\phi) \\ &= (-1)^{(m-1)/2} \pi \{[(2m)!]/[2^{2m-1}(m!)^2]\} \beta^m. \end{aligned} \quad (15.10.44)$$

We conclude from (10.38) that (for  $m$  odd)

$$\begin{aligned} C_{m,s}^{[0]}(z) &= -(2g/\pi)(a^2 + z^2)^{-1/2}(a^2 + z^2)^{-m/2}c_{m,m} \\ &= -(2g/\pi)(a^2 + z^2)^{-(m+1)/2}(-1)^{(m-1)/2}\pi \{[(2m)!]/[2^{2m-1}(m!)^2]\} \beta^m \\ &= -(2g/\pi)(a^2 + z^2)^{-(m+1)/2}(-1)^{(m-1)/2}\pi \{[(2m)!]/[2^{2m-1}(m!)^2]\} [a/(a^2 + z^2)^{1/2}]^m \\ &= -(2g)a^m(a^2 + z^2)^{-(2m+1)/2}(-1)^{(m-1)/2} \{[(2m)!]/[2^{2m-1}(m!)^2]\} \\ &= -(4g)a^m(a^2 + z^2)^{-(2m+1)/2}(-1)^{(m-1)/2} \{[(2m)!]/[2^{2m}(m!)^2]\}. \\ &= -(4g)(-1)^{(m-1)/2} \{[(2m)!]/[2^{2m}(m!)^2]\} a^m(a^2 + z^2)^{-(2m+1)/2}. \end{aligned} \quad (15.10.45)$$

Taken together, (10.26), (10.37), and (10.45) provide the on-axis gradients for the normal monopole doublet. Note that the  $C_{m,s}^{[0]}(z)$  have the asymptotic fall off

$$|C_{m,s}^{[0]}(z)| \sim (a/|z|)^{2m+1} \quad (15.10.46)$$

for large  $|z|$ . Let us summarize our findings. When (10.19), (10.32), and (10.35) are employed in the general expansion (3.33) we have the result

$$\psi^d(\rho, \phi, z) = - \sum_{m \text{ odd}} \sin(m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,s}^{[2\ell]}(z) \rho^{2\ell+m} \quad (15.10.47)$$

where the needed on-axis gradients  $C_{m,s}^{[2\ell]}(z)$  are specified by (10.45) and its derivatives with respect to  $z$ .

As an application of this result suppose we wish to retain, in the expansion of the Hamiltonian  $H$  appearing in (1.3), homogeneous polynomials through degree 8. Then, as

we see from (1.4), we must retain homogeneous polynomials in the variables  $x, y$  through degree 7 in the expansions of  $A_x$  and  $A_y$ , and homogeneous polynomials in the variables  $x, y$  through degree 8 in the expansion of  $A_z$ . Inspection of (4.21) through (4.26) or (5.89) through (5.94) shows that for the cases  $m = 0$  or  $m$  odd we then need the  $C_{m,\alpha}^{[n]}(z)$  with  $(m + n) \leq 7$ . And for the cases of even  $m$  we need the  $C_{m,\alpha}^{[n]}(z)$  with  $(m + n) \leq 8$ . In particular, for the case of the monopole doublet, (for which only the on-axis gradients with  $\alpha = s$  and  $m$  odd are nonzero) we need the following functions:

$$\begin{aligned} & C_{1,s}^{[0]}(z), C_{1,s}^{[1]}(z), C_{1,s}^{[2]}(z), C_{1,s}^{[3]}(z), C_{1,s}^{[4]}(z), C_{1,s}^{[5]}(z), C_{1,s}^{[6]}(z); \\ & C_{3,s}^{[0]}(z), C_{3,s}^{[1]}(z), C_{3,s}^{[2]}(z), C_{3,s}^{[3]}(z), C_{3,s}^{[4]}(z); \\ & C_{5,s}^{[0]}(z), C_{5,s}^{[1]}(z), C_{5,s}^{[2]}(z); \\ & C_{7,s}^{[0]}(z). \end{aligned} \quad (15.10.48)$$

(See Exercise 7.1.) Graphs of a selected few of these functions, for the monopole doublet in the case that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm) $^2$ , are shown in Figures 10.8 through 10.15. In these plots  $z$  has units of centimeters. Evidently the  $C_{m,s}^{[0]}$  become ever more highly peaked with increasing  $m$ . Fortunately, when working through some fixed degree, we need fewer derivatives with increasing  $m$ . Note that we expect that the function  $C_{m,s}^{[n]}(z)$  should have  $n$  zeroes. This is indeed the case, but some of these zeroes can be hidden in the tails. Figure 10.10 is an enlargement of Figure 10.9 showing a hidden zero for the case of  $C_{1,s}^{[6]}(z)$ .

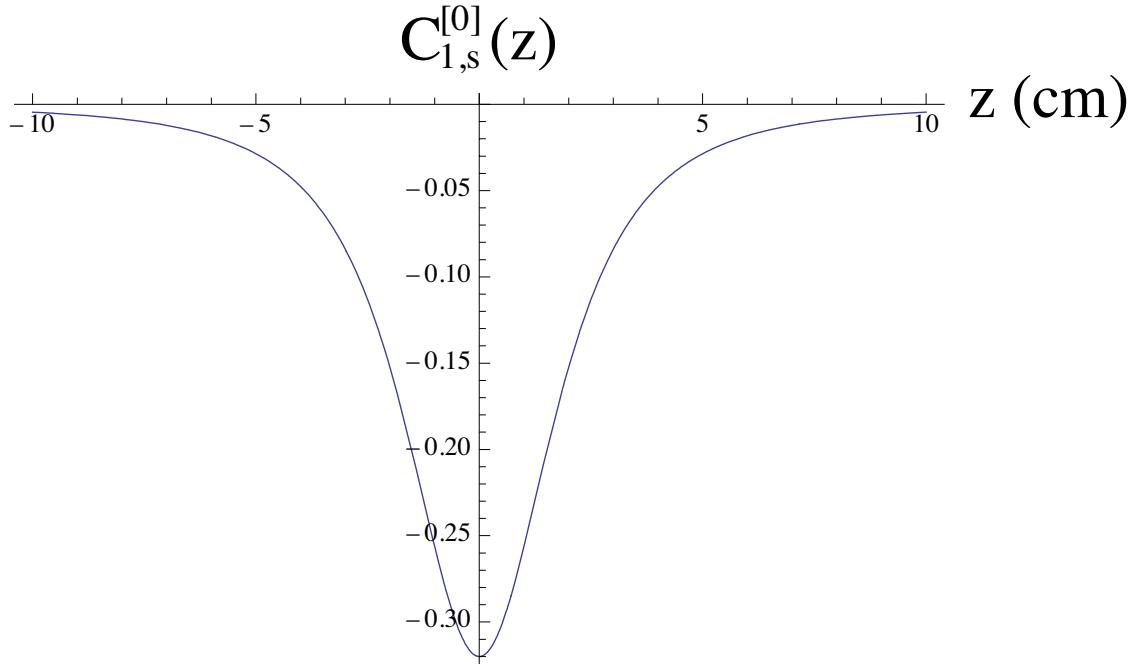


Figure 15.10.8: The on-axis gradient function  $C_{1,s}^{[0]}$  for the normal monopole doublet in the case that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm) $^2$ .

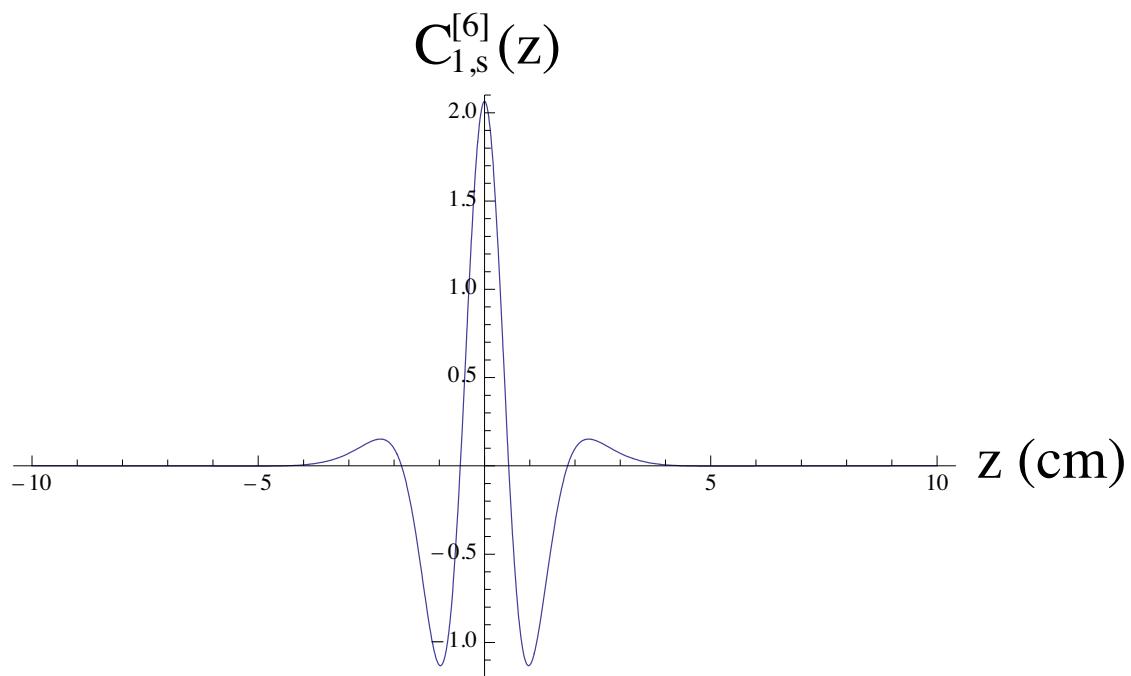


Figure 15.10.9: The on-axis gradient function  $C_{1,s}^{[6]}$  for the normal monopole doublet in the case that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm) $^2$ .

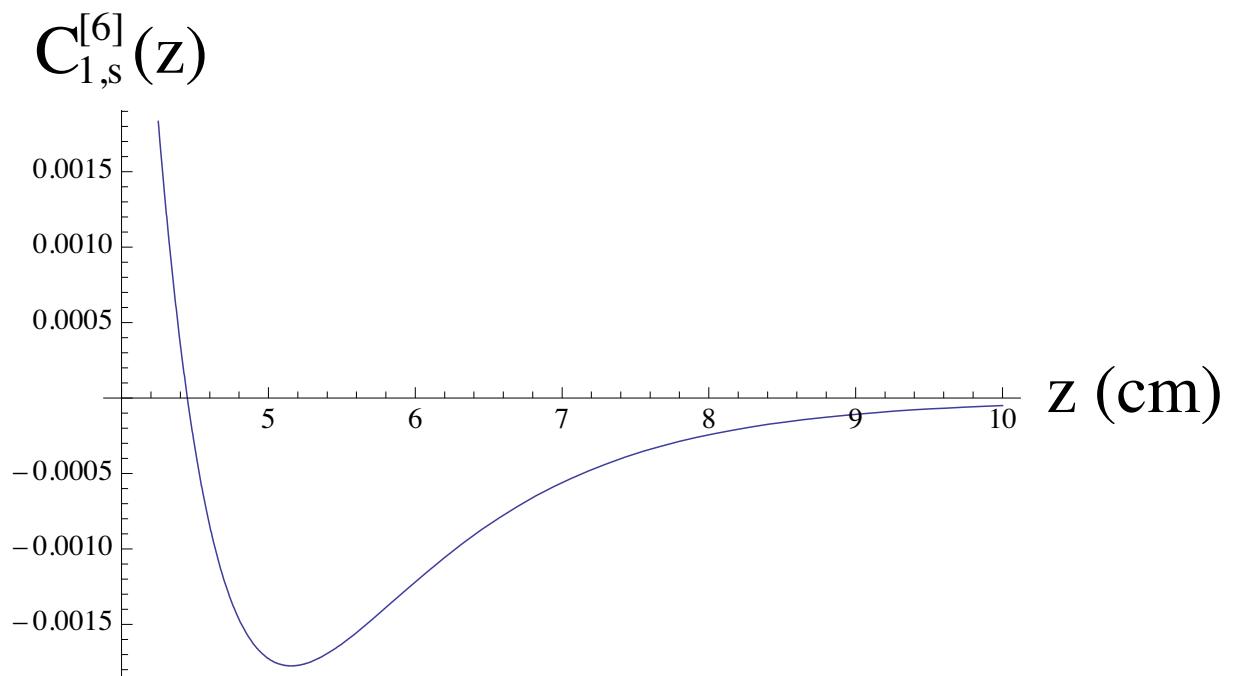


Figure 15.10.10: An enlargement of a portion of Figure 10.9 showing a zero hidden in a tail.

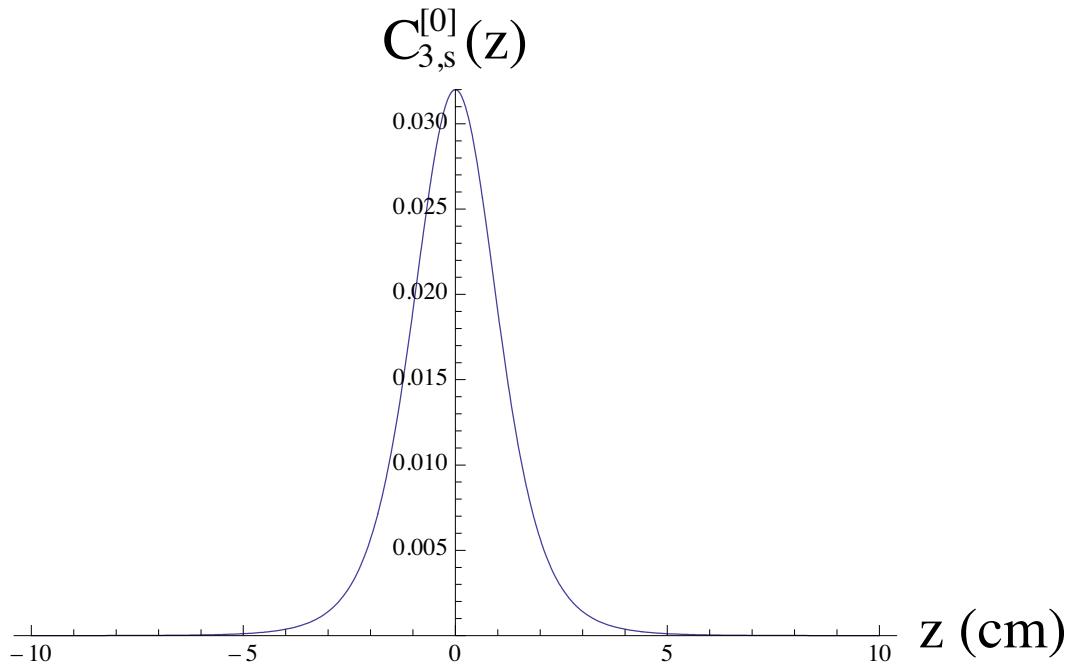


Figure 15.10.11: The on-axis gradient function  $C_{3,s}^{[0]}$  for the normal monopole doublet in the case that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm) $^2$ .

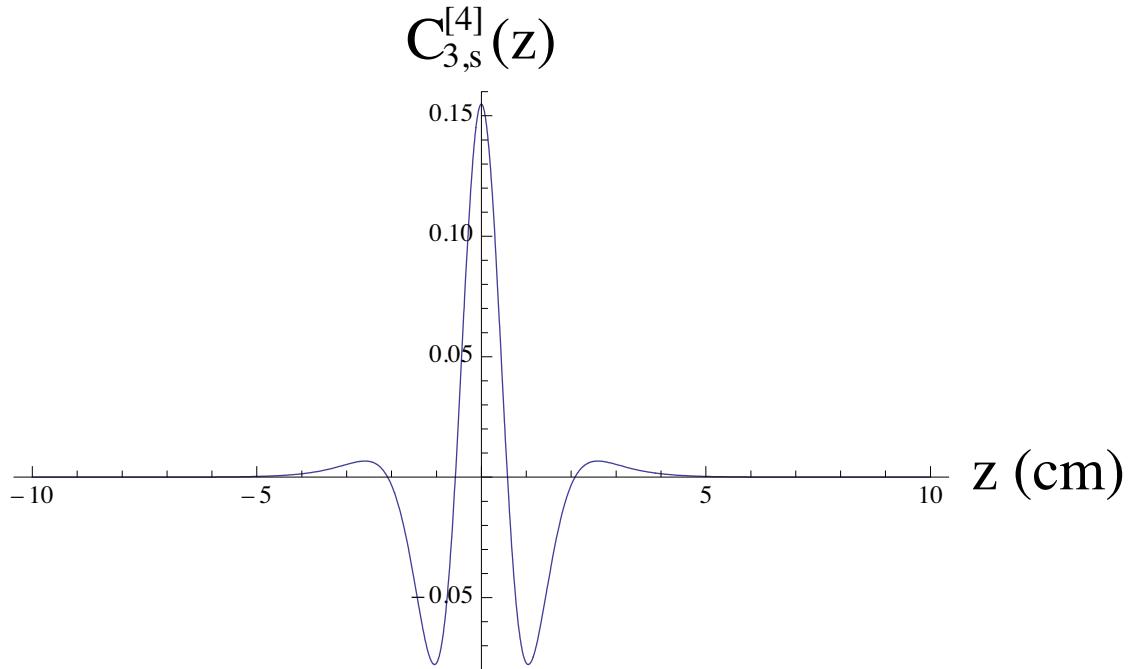


Figure 15.10.12: The on-axis gradient function  $C_{3,s}^{[4]}$  for the normal monopole doublet in the case that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm) $^2$ .

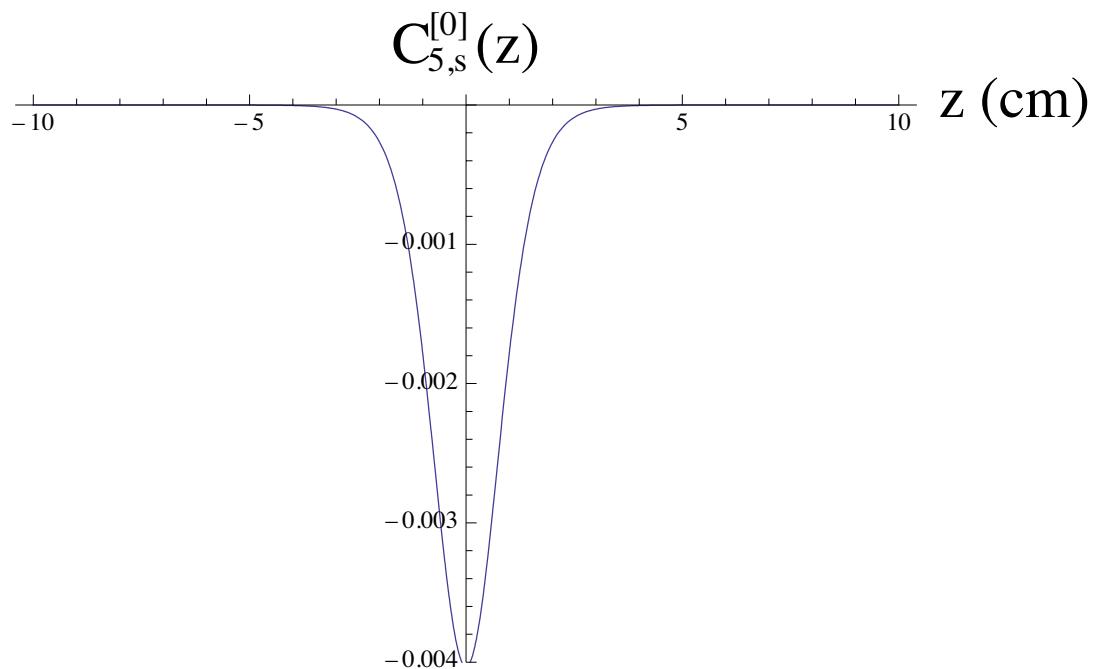


Figure 15.10.13: The on-axis gradient function  $C_{5,s}^{[0]}$  for the normal monopole doublet in the case that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm) $^2$ .

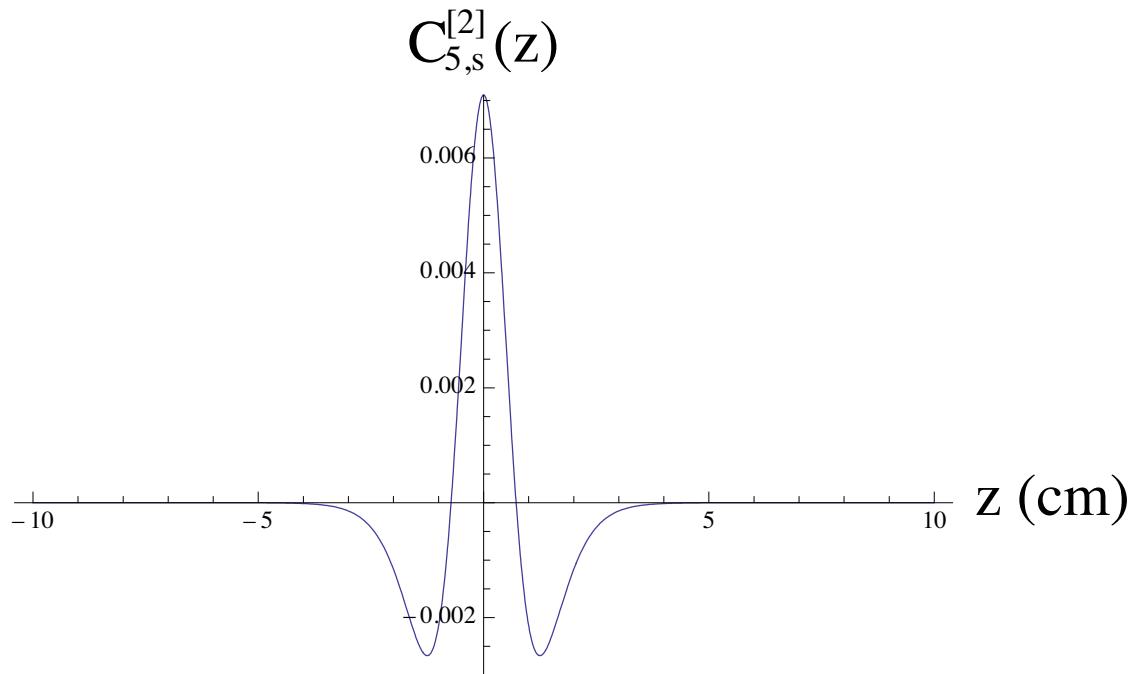


Figure 15.10.14: The on-axis gradient function  $C_{5,s}^{[2]}$  for the normal monopole doublet in the case that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm) $^2$ .

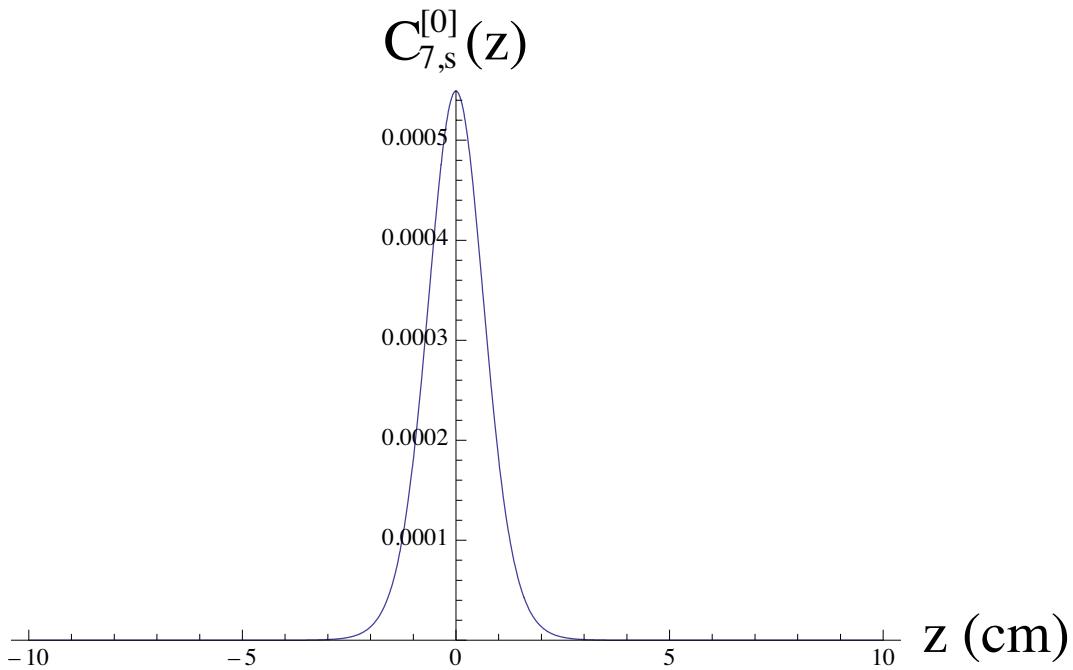


Figure 15.10.15: The on-axis gradient function  $C_{7,s}^{[0]}$  for the normal monopole doublet in the case that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm) $^2$ .

## Exercises

**15.10.1.** Explain why Figures 10.3 and 10.8 should be the same.

## 15.11 Minimum Vector Potential for Normal Magnetic Monopole Doublet

The purpose of this section is to find the first few terms in the expansion of the minimum (Poincaré-Coulomb gauge) vector potential for a normal magnetic monopole doublet. We will first find the minimum vector potential in terms of the scalar potential and its associated magnetic field using the results of Subsection 2.7. Then we will find the minimum vector potential in terms of the on-axis gradients using the results of Section 7.

In particular, we will be interested in expansions for the fringe-field regions and in the midplane. Suppose the doublet is located at the origin  $\mathbf{R} = (0, 0, 0)$  as in Subsection 10.1, and we seek an expansion about the mid-plane point  $\mathbf{R}_0 = (X_0, 0, Z_0)$ . If  $Z_0 \ll 0$ , we will obtain an expansion in the leading fringe-field region, and if  $Z_0 \gg 0$ , we will obtain an expansion in the trailing fringe-field region. Moreover, if  $X_0 = 0$ , the expansion point will be on axis; and setting  $X_0 \neq 0$  allows for expansion about a point on the (curved) design orbit. See Section 23.3 and Figures 23.3.1 and 24.1.1.

### 15.11.1 Computation from the Scalar Potential and Associated Magnetic Field

We begin by specifying the scalar potential  $\Psi(\mathbf{R})$ . According to (10.3), it is given by the relation

$$\Psi(X, Y, Z) = -g[X^2 + (Y - a)^2 + Z^2]^{-1/2} + g[X^2 + (Y + a)^2 + Z^2]^{-1/2}. \quad (15.11.1)$$

Next, according to (2.4),  $\psi(\mathbf{r})$  is given by the relation

$$\begin{aligned} \psi(x, y, z) &= \Psi(\mathbf{R}_0 + \mathbf{r}) \\ &= -g[(X_0 + x)^2 + (y - a)^2 + (Z_0 + z)^2]^{-1/2} \\ &\quad + g[(X_0 + x)^2 + (y + a)^2 + (Z_0 + z)^2]^{-1/2}. \end{aligned} \quad (15.11.2)$$

The right side of (11.2) can now be expanded in powers of the components of  $\mathbf{r}$ . Doing so yields, for the first few terms, the result

$$\begin{aligned} \psi(\mathbf{r}; X_0, Z_0) &= [-2ga/(X_0^2 + Z_0^2 + a^2)^{3/2}]y \\ &\quad + [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}][y(X_0x + Z_0z)] \\ &\quad + \text{terms of order 3 and higher.} \end{aligned} \quad (15.11.3)$$

Note that  $\Psi(X_0, 0, Z_0)$  vanishes so that there is no constant term in the expansion (11.3). We observe that the first term in (11.3) falls off as  $(1/|X_0|)^3$  or  $(1/|Z_0|)^3$  for large  $|X_0|$  or  $|Z_0|$ , and the second falls off as  $(1/|X_0|)^4$  or  $(1/|Z_0|)^4$ . In general, successive terms fall off with ever increasing powers of  $(1/|X_0|)$  or  $(1/|Z_0|)$ .

Let us compute the magnetic field  $\mathbf{B}$  associated with the first two terms in (11.3). We find the result

$$\begin{aligned} \mathbf{B}(\mathbf{r}; X_0, Z_0) &= \nabla_{\mathbf{r}}\psi(\mathbf{r}; X_0, Y_0) \\ &= -[2ga/(X_0^2 + Z_0^2 + a^2)^{3/2}]\mathbf{e}_y \\ &\quad + [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}](X_0x + Z_0z)\mathbf{e}_y \\ &\quad + [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}][y(X_0\mathbf{e}_x + Z_0\mathbf{e}_z)]. \end{aligned} \quad (15.11.4)$$

Next let us find the minimum vector potential  $\mathbf{A}^{\min}$  associated with the first two terms in (11.3). Begin by decomposing  $\mathbf{B}$  into homogeneous polynomials by rewriting (11.4) in the form (2.109) with

$$\mathbf{B}^0(\mathbf{r}; X_0, Z_0) = -[2ga/(X_0^2 + Z_0^2 + a^2)^{3/2}]\mathbf{e}_y \quad (15.11.5)$$

and

$$\mathbf{B}^1(\mathbf{r}; X_0, Z_0) = [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}][(X_0x + Z_0z)\mathbf{e}_y + y(X_0\mathbf{e}_x + Z_0\mathbf{e}_z)]. \quad (15.11.6)$$

The minimum vector potential associated with this magnetic field is given by the relations (2.109) through (2.111). Working out the indicated cross products yields the results

$$\mathbf{A}^{\min 1}(\mathbf{r}; X_0, Z_0) = [ga/(X_0^2 + Z_0^2 + a^2)^{3/2}](-z\mathbf{e}_x + x\mathbf{e}_z), \quad (15.11.7)$$

$$\begin{aligned} \mathbf{A}^{\min 2}(\mathbf{r}; X_0, Z_0) &= [-2ga/(X_0^2 + Z_0^2 + a^2)^{5/2}] \times \\ &[(Z_0y^2 - Z_0z^2 - X_0xz)\mathbf{e}_x + (X_0yz - Z_0xy)\mathbf{e}_y + (X_0x^2 + Z_0xz - X_0y^2)\mathbf{e}_z]. \end{aligned} \quad (15.11.8)$$

Simple calculation verifies that there are the relations

$$\nabla \times \mathbf{A}^{\min 1}(\mathbf{r}; X_0, Z_0) = \mathbf{B}^0(\mathbf{r}; X_0, Z_0), \quad (15.11.9)$$

$$\nabla \times \mathbf{A}^{\min 2}(\mathbf{r}; X_0, Z_0) = \mathbf{B}^1(\mathbf{r}; X_0, Z_0), \quad (15.11.10)$$

as desired. We note that  $\mathbf{A}^{\min 1}$  falls off as  $(1/|X_0|)^3$  or  $(1/|Z_0|)^3$  for large  $|X_0|$  or  $|Z_0|$ , and  $\mathbf{A}^{\min 2}$  falls off as  $(1/|X_0|)^4$  or  $(1/|Z_0|)^4$ . In general, successive  $\mathbf{A}^{\min n}$  fall off with ever increasing powers of  $(1/|X_0|)$  or  $(1/|Z_0|)$ .

### 15.11.2 Computation from the On-Axis Gradients

The  $m = 1$  and  $\alpha = s$  vector potential in the symmetric Coulomb gauge of Section 5 is given in terms of on-axis gradients by the relation

$$\hat{A}_x^{1,s}(x, y, Z_0 + z) = (1/4)(x^2 - y^2)C_{1,s}^{[1]}(Z_0 + z) - (1/48)(x^4 - y^4)C_{1,s}^{[3]}(Z_0 + z) + \dots, \quad (15.11.11)$$

$$\hat{A}_y^{1,s}(x, y, Z_0 + z) = (1/2)xyC_{1,s}^{[1]}(Z_0 + z) - (1/24)(x^3y + xy^3)C_{1,s}^{[3]}(Z_0 + z) + \dots, \quad (15.11.12)$$

$$\begin{aligned} \hat{A}_z^{1,s}(x, y, Z_0 + z) &= -xC_{1,s}^{[0]}(Z_0 + z) + (1/8)(x^3 + xy^2)C_{1,s}^{[2]}(Z_0 + z) \\ &\quad - (1/192)(x^5 + 2x^3y^2 + xy^4)C_{1,s}^{[4]}(Z_0 + z) + \dots. \end{aligned} \quad (15.11.13)$$

See (5.97) through (5.99). If we expand  $\hat{\mathbf{A}}^{1,s}(x, y, Z_0 + z)$  in powers of  $z$ , organize the results into homogeneous polynomials, and retain only terms of degree less than 3, we find the results

$$\hat{A}_x^{1,s}(x, y, Z_0 + z) = (1/4)(x^2 - y^2)C_{1,s}^{[1]}(Z_0) + \dots, \quad (15.11.14)$$

$$\hat{A}_y^{1,s}(x, y, Z_0 + z) = (1/2)xyC_{1,s}^{[1]}(Z_0) + \dots, \quad (15.11.15)$$

$$\hat{A}_z^{1,s}(x, y, Z_0 + z) = -xC_{1,s}^{[0]}(Z_0) - xzC_{1,s}^{[1]}(Z_0) + \dots. \quad (15.11.16)$$

The gauge function  $\hat{\chi}_{1,s}$  that relates the symmetric Coulomb gauge vector potential  $\hat{\mathbf{A}}^{1,s}$  and the Poincaré-Coulomb gauge vector potential  ${}^P\mathbf{C}^{1,s}$  is given by the relation

$$\begin{aligned} \hat{\chi}_{1,s}(x, y, z; Z_0) &= \cos(\phi) \sum_{k=0}^{\infty} (-1)^k \frac{1!}{2^{2k} k!(k+1)!} D_{1,s}^{[2k]}(z; Z_0) \rho^{2k+1} \\ &= \rho \cos(\phi) [D_{1,s}^{[0]}(z; Z_0) - (1/8)\rho^2 D_{1,s}^{[2]}(z; Z_0) + (*)\rho^4 D_{1,s}^{[4]}(z; Z_0) + \dots] \\ &= x[D_{1,s}^{[0]}(z; Z_0) - (1/8)\rho^2 D_{1,s}^{[2]}(z; Z_0) + (*)\rho^4 D_{1,s}^{[4]}(z; Z_0) + \dots]. \end{aligned} \quad (15.11.17)$$

See (7.5) and (7.20). We have also found, see (7.60), the relation

$$\begin{aligned} D_{1,s}^{[0]}(z; Z_0) &= \{[1/(1+1)]C_{1,s}^{[0]}(Z_0)\}z + \{[1/(1+2)]C_{1,s}^{[1]}(Z_0)\}z^2 \\ &\quad + \{[1/(1+3)](1/2!)C_{1,s}^{[2]}(0)\}z^3 + \{[1/(1+4)](1/3!)C_{1,s}^{[3]}(Z_0)\}z^4 + \dots \\ &= (1/2)C_{1,s}^{[0]}(Z_0)z + (1/3)C_{1,s}^{[1]}(Z_0)z^2 + (1/8)C_{1,s}^{[2]}(Z_0)z^3 + (1/30)C_{1,s}^{[3]}(Z_0)z^4 + \dots, \end{aligned} \quad (15.11.18)$$

from which it follows that

$$D_{1,s}^{[2]}(z; Z_0) = (2/3)C_{1,s}^{[1]}(Z_0) + (3/4)C_{1,s}^{[2]}(Z_0)z + (2/5)C_{1,s}^{[3]}(Z_0)z^2 + \dots, \quad (15.11.19)$$

$$D_{1,s}^{[4]}(z; Z_0) = (4/5)C_{1,s}^{[3]}(Z_0) + \dots. \quad (15.11.20)$$

Inserting these results into (11.17) and collecting terms of like degree give the final homogeneous polynomial expansion

$$\begin{aligned} \hat{\chi}_{1,s}(x, y, z; Z_0) &= x[(1/2)C_{1,s}^{[0]}(Z_0)z + (1/3)C_{1,s}^{[1]}(Z_0)z^2 + (1/8)C_{1,s}^{[2]}(Z_0)z^3 + \dots] \\ &\quad - (1/8)x\rho^2[(2/3)C_{1,s}^{[1]}(Z_0) + (3/4)C_{1,s}^{[2]}(Z_0)z + (2/5)C_{1,s}^{[3]}(Z_0)z^2 + \dots] \\ &\quad + (*)x\rho^4[(4/5)C_{1,s}^{[3]}(Z_0) + \dots] + \dots \\ &= \{[xz(1/2)]C_{1,s}^{[0]}(Z_0)\} + \{[xz^2(1/3) - (1/8)x\rho^2(2/3)]C_{1,s}^{[1]}(Z_0)\} + \dots \\ &= \{[(1/2)xz]C_{1,s}^{[0]}(Z_0)\} + \{[(1/3)xz^2 - (1/12)x\rho^2]C_{1,s}^{[1]}(Z_0)\} + \dots. \end{aligned} \quad (15.11.21)$$

The next step is to compute the  $\Delta A_{1,s}^{1,s}$  defined by (7.46). From (11.21) we find the results

$$\begin{aligned} \Delta A_x^{1,s} &= (\partial/\partial x)\hat{\chi}_{1,s}(x, y, z; Z_0) \\ &= \{[(1/2)z]C_{1,s}^{[0]}(Z_0)\} + \{[(1/3)z^2 - (1/12)(3x^2 + y^2)]C_{1,s}^{[1]}(Z_0)\} + \dots, \end{aligned} \quad (15.11.22)$$

$$\begin{aligned} \Delta A_y^{1,s} &= (\partial/\partial y)\hat{\chi}_{1,s}(x, y, z; Z_0) \\ &= +\{[-(1/6)xy]C_{1,s}^{[1]}(Z_0)\} + \dots, \end{aligned} \quad (15.11.23)$$

$$\begin{aligned} \Delta A_z^{1,s} &= (\partial/\partial z)\hat{\chi}_{1,s}(x, y, z; Z_0) \\ &= \{[(1/2)x]C_{1,s}^{[0]}(Z_0)\} + \{[(2/3)xz]C_{1,s}^{[1]}(Z_0)\} + \dots. \end{aligned} \quad (15.11.24)$$

Finally, we may obtain  ${}^{PC}\mathbf{A}^{1,s}$  with the aid of (7.46) and (11.14) through (11.16). Doing so gives the results

$$\begin{aligned} {}^{PC}A_x^{1,s}(x, y, z; Z_0) &= (1/4)(x^2 - y^2)C_{1,s}^{[1]}(Z_0) + \{[(1/2)z]C_{1,s}^{[0]}(Z_0)\} \\ &\quad + \{[(1/3)z^2 - (1/12)(3x^2 + y^2)]C_{1,s}^{[1]}(Z_0)\} + \dots \\ &= \{[(1/2)z]C_{1,s}^{[0]}(Z_0)\} + \{[(1/3)z^2 - (1/3)y^2]C_{1,s}^{[1]}(Z_0)\} + \dots, \end{aligned} \quad (15.11.25)$$

$$\begin{aligned} {}^{PC}A_y^{1,s}(x, y, z; Z_0) &= (1/2)xyC_{1,s}^{[1]}(Z_0) + \{[-(1/6)xy]C_{1,s}^{[1]}(Z_0)\} + \cdots \\ &= \{[(1/3)xy]C_{1,s}^{[1]}(Z_0)\} + \cdots, \end{aligned} \quad (15.11.26)$$

$$\begin{aligned} {}^{PC}A_z^{1,s}(x, y, z; Z_0) &= -xC_{1,s}^{[0]}(Z_0) - xzC_{1,s}^{[1]}(Z_0) + \{(1/2)x]C_{1,s}^{[0]}(Z_0)\} \\ &\quad + \{[(2/3)xz]C_{1,s}^{[1]}(Z_0)\} \cdots \\ &= \{[-(1/2)x]C_{1,s}^{[0]}(Z_0)\} - \{[(1/3)xz]C_{1,s}^{[1]}(Z_0)\} + \cdots. \end{aligned} \quad (15.11.27)$$

How do the results (11.25) through (11.27) compare with the results (11.7) and (11.8) found in the previous subsection? Suppose  $\mathbf{A}^{\min 1}(\mathbf{r}; X_0, Z_0)$  and  $\mathbf{A}^{\min 2}(\mathbf{r}; X_0, Z_0)$  as given by (11.7) and (11.8) are evaluated at  $X_0 = 0$ . So doing gives the results

$$\mathbf{A}^{\min 1}(\mathbf{r}; X_0, Z_0)|_{X_0=0} = [ga/(a^2 + Z_0^2)^{3/2}](-ze_x + xe_z), \quad (15.11.28)$$

$$\begin{aligned} \mathbf{A}^{\min 2}(\mathbf{r}; X_0, Z_0)|_{X_0=0} &= [-2ga/(a^2 + Z_0^2)^{5/2}] \times \\ &\quad [(Z_0y^2 - Z_0z^2)\mathbf{e}_x + (-Z_0xy)\mathbf{e}_y + (Z_0xz)\mathbf{e}_z]. \end{aligned} \quad (15.11.29)$$

Also, from (8.21) and (8.33), we find the results

$$C_{1,s}^{[0]}(z) = -g[(2!)/(a^2)]\beta^3(z) = -2ga/(a^2 + z^2)^{3/2}, \quad (15.11.30)$$

from which it follows that

$$C_{1,s}^{[0]}(Z_0) = -2ga/(a^2 + Z_0^2)^{3/2}, \quad (15.11.31)$$

$$C_{1,s}^{[1]}(Z_0) = 6gaZ_0/(a^2 + Z_0^2)^{5/2}. \quad (15.11.32)$$

Consequently, (11.28) and (11.29) can be rewritten in the form

$$\mathbf{A}^{\min 1}(\mathbf{r}; X_0, Z_0)|_{X_0=0} = -(1/2)C_{1,s}^{[0]}(Z_0)(-ze_x + xe_z), \quad (15.11.33)$$

$$\mathbf{A}^{\min 2}(\mathbf{r}; X_0, Z_0)|_{X_0=0} = (-1/3)C_{1,s}^{[1]}(Z_0)[(y^2 - z^2)\mathbf{e}_x + (-xy)\mathbf{e}_y + (xz)\mathbf{e}_z]. \quad (15.11.34)$$

Comparison of (11.33) and (11.34) with (11.25) through (11.27) reveals that (11.33) agrees with the first-degree terms in (11.25) through (11.27), and (11.34) agrees with the second-degree terms in (11.25) through (11.27). Therefore the on-axis minimum vector potential expansion computed from the scalar potential and associated magnetic field agrees with the on-axis minimum vector potential expansion computed from the on-axis gradients, as desired and required.

Finally we remind the reader that, although we have been considering the case of a monopole doublet field, the relations (11.11) through (11.13) and (11.21) through (11.27) hold for *any*  $m = 1$  and  $\alpha = s$  magnetic field no matter what its source. The same is true for the relation

$$C_{1,s}^{[0]}(z) = B_y(0, 0, z). \quad (15.11.35)$$

Recall (3.59).

## 15.12 Scalar Potentials Produced by Single-Layer Multipole Ring Sources

Section 10 treated the normal magnetic monopole doublet and found its scalar field, magnetic field, and on-axis gradients. The purpose of this section is to find the results produced by hypothetical single-layer rings of magnetic charge having sinusoidal and co-sinusoidal magnetic charge distributions. Initially we will determine scalar potentials, and eventually in Section 13, corresponding on-axis gradients. These results will be useful for subsequent magnetic field modeling.

### 15.12.1 Normal Multipole Case

#### Defining integral relation for potential due to single-layer ring source

Begin by considering a single-layer pure normal multipole ring source. Specifically, suppose this source lies on a *ring* that is in the  $z = 0$  plane, is centered on the  $z$  axis, and has radius  $a$ . Let  $\psi^r(\mathbf{r}; m, s)$  be the potential produced by this source. Then, using (9.1) and the first term in (9.18), we may write

$$\psi^r(\mathbf{r}; m, s) = [1/(4\pi)] \int_S dS' \sigma_m(\mathbf{r}') / \|\mathbf{r} - \mathbf{r}'\|. \quad (15.12.1)$$

Our task is to evaluate the right side of (12.1) with  $\sigma_m(\mathbf{r}')$  of the form

$$\sigma_m(\mathbf{r}') = (4\pi)B_0 a \delta(z') \sin m\phi' \quad (15.12.2)$$

where  $B_0$  is a strength factor that has units of  $B$ . Note that, since  $\delta(z')$  has units of  $1/L$  and  $a$  has units of  $L$ ,  $\sigma_m$  has units of  $B$  and  $\psi^r$  has units of  $BL$ , as desired. See (9.19). Moreover, since the weight function  $\sin m\phi'$  “peaks” at  $\phi' = \pi/(2m)$ ,  $\sigma_m$  as given by (12.2) is a *normal*  $2m$ -pole source.

For the further ingredients in the integral we employ cylindrical coordinates and write

$$\mathbf{r} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z; \quad (15.12.3)$$

$$\mathbf{r}' = x'\mathbf{e}_x + y'\mathbf{e}_y + z'\mathbf{e}_z; \quad (15.12.4)$$

$$x = \rho \cos \phi, \quad x' = a \cos \phi'; \quad (15.12.5)$$

$$y = \rho \sin \phi, \quad y' = a \sin \phi'; \quad (15.12.6)$$

$$dS' = ad\phi' dz'; \quad (15.12.7)$$

$$\begin{aligned} \|\mathbf{r} - \mathbf{r}'\|^2|_{z'=0} &= (x - x')^2 + (y - y')^2 + z^2 \\ &= x^2 + y^2 - 2(xx' + yy') + (x')^2 + (y')^2 + z^2 \\ &= \rho^2 - 2a\rho(\cos \phi \cos \phi' + \sin \phi \sin \phi') + a^2 + z^2 \\ &= \rho^2 - 2a\rho \cos(\phi - \phi') + a^2 + z^2. \end{aligned} \quad (15.12.8)$$

Upon putting everything together we find the result

$$\begin{aligned} \psi^r(\rho, \phi, z; m, s) &= \psi^r(\mathbf{r}; m, s) = \\ B_0 a^2 \int_{-\pi}^{\pi} d\phi' \sin m\phi' [\rho^2 - 2a\rho \cos(\phi - \phi') + a^2 + z^2]^{-1/2}. \end{aligned} \quad (15.12.9)$$

### Standard form integral relation

At this point there is more that can be said which will prove useful for further work. The conversation begins with (12.9). What we will do, by algebraic manipulation, is transform the integral (12.9) to what might be viewed as a kind of “*standard form*”.

Since the cosine function is even, we begin by rewriting (12.9) in the form

$$\begin{aligned}\psi^r(\rho, \phi, z; m, s, a) &= \psi^r(\mathbf{r}; m, s, a) = \\ B_0 a^2 \int_{-\pi}^{\pi} d\phi' \sin m\phi' [\rho^2 - 2a\rho \cos(\phi' - \phi) + a^2 + z^2]^{-1/2}.\end{aligned}\quad (15.12.10)$$

Here, in our notation for  $\psi^r$ , we have made its dependence on  $a$  more explicit. Next write

$$\phi' = (\phi' - \phi) + \phi = \theta + \phi \quad (15.12.11)$$

where

$$\theta = \phi' - \phi. \quad (15.12.12)$$

It follows that

$$\sin(m\phi') = \sin(m\theta + m\phi) = \sin(m\theta) \cos(m\phi) + \cos(m\theta) \sin(m\phi). \quad (15.12.13)$$

Now use (12.11) to make a change of integration variables so that

$$d\phi' = d\theta \quad (15.12.14)$$

and (12.10) becomes

$$\begin{aligned}\psi^r(\rho, \phi, z; m, s, a) &= \\ B_0 a^2 \int_{-\pi}^{\pi} d\theta [\sin(m\theta) \cos(m\phi) + \cos(m\theta) \sin(m\phi)] [\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-1/2}.\end{aligned}\quad (15.12.15)$$

One might wonder if in writing (12.15) we have forgotten to change the limits of integration. We haven't. Since the integrand is periodic with period  $2\pi$ , all that matters is that the integration be over a full period.

Evidently (12.15) can be written as the sum of two terms. Doing so gives the result

$$\begin{aligned}\psi^r(\rho, \phi, z; m, s, a) &= \\ B_0 a^2 \cos m\phi \int_{-\pi}^{\pi} d\theta \sin(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-1/2} \\ &+ B_0 a^2 \sin m\phi \int_{-\pi}^{\pi} d\theta \cos(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-1/2}.\end{aligned}\quad (15.12.16)$$

Observe that the integrand for the first integral on the right side of (12.16) is an *odd* function of  $\theta$ , and therefore this integral vanishes. We are left with what we call the “*standard form*” result

$$\psi^r(\rho, \phi, z; m, s, a) = B_0 a^2 \sin m\phi \int_{-\pi}^{\pi} d\theta \cos(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-1/2}, \quad (15.12.17)$$

which isolates the  $\phi$  dependence of  $\psi^r$ . [Note we have found that the Ansatz of a  $\sin m\phi'$  factor in the source term (12.2) entails a prefactor of  $\sin m\phi$  in the resulting potential (12.17). The potential has the same symmetry as the source.] The result (12.17) will be employed in Section 15 to find a relation between the scalar potentials produced by analogous single- and double-layer sources.

### 15.12.2 Skew Multipole Case

The skew multipole case may be treated analogously. Now consider a single-layer pure *skew* multipole ring source again lying on a ring that is in the  $z = 0$  plane, is centered on the  $z$  axis, and has radius  $a$ . Let  $\psi^r(\mathbf{r}; m, c)$  be the potential produced by this source. Then we may write

$$\psi^r(\mathbf{r}; m, c) = [1/(4\pi)] \int_S dS' \sigma_m(\mathbf{r}') / \|\mathbf{r} - \mathbf{r}'\|. \quad (15.12.18)$$

The only difference from the normal multipole case is that our task now is to evaluate the right side of (12.18) with  $\sigma_m(\mathbf{r}')$  of the form

$$\sigma_m(\mathbf{r}') = (4\pi)B_0 a \delta(z') \cos m\phi'. \quad (15.12.19)$$

For the further ingredients in the integral we again employ (12.3) through (12.8). Upon putting everything together we now find the result

$$\begin{aligned} \psi^r(\rho, \phi, z; m, c) &= \psi^r(\mathbf{r}; m, c) = \\ B_0 a^2 \int_{-\pi}^{\pi} d\phi' \cos m\phi' [\rho^2 - 2a\rho \cos(\phi - \phi') + a^2 + z^2]^{-1/2}. \end{aligned} \quad (15.12.20)$$

#### Standard form integral relation

By algebraic manipulation, let us work to transform the integral (12.20) to standard form. Again use the even property of the cosine function to rewrite (12.20) in the form

$$\begin{aligned} \psi^r(\rho, \phi, z; m, c, a) &= \psi^r(\mathbf{r}; m, c, a) = \\ B_0 a^2 \int_{-\pi}^{\pi} d\phi' \cos m\phi' [\rho^2 - 2a\rho \cos(\phi' - \phi) + a^2 + z^2]^{-1/2}. \end{aligned} \quad (15.12.21)$$

Also, again define a variable  $\theta$  using (12.11) and (12.12). It follows that now

$$\cos(m\phi') = \cos(m\theta + m\phi) = \cos(m\theta) \cos(m\phi) - \sin(m\theta) \sin(m\phi). \quad (15.12.22)$$

Again use (12.14) to make a change of integration variables so that (12.21) becomes

$$\begin{aligned} \psi^r(\rho, \phi, z; m, c, a) &= \\ B_0 a^2 \int_{-\pi}^{\pi} d\theta [\cos(m\theta) \cos(m\phi) - \sin(m\theta) \sin(m\phi)] [\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-1/2}. \end{aligned} \quad (15.12.23)$$

Evidently (12.23) can be written as the sum of two terms. Doing so gives the result

$$\begin{aligned}\psi^r(\rho, \phi, z; m, c, a) = & \\ B_0 a^2 \cos m\phi \int_{-\pi}^{\pi} d\theta \cos(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-1/2} \\ - B_0 a^2 \sin m\phi \int_{-\pi}^{\pi} d\theta \sin(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-1/2}.\end{aligned}\tag{15.12.24}$$

The integrand for the second integral on the right side of (12.24) is an *odd* function of  $\theta$ , and therefore this integral vanishes. We are left with the standard form result

$$\psi^r(\rho, \phi, z; m, c, a) = B_0 a^2 \cos m\phi \int_{-\pi}^{\pi} d\theta \cos(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-1/2}.\tag{15.12.25}$$

[Note we have found that the Ansatz of a  $\cos m\phi'$  factor in the source term (12.19) entails a prefactor of  $\cos m\phi$  in the resulting potential (12.25). Again the potential has the same symmetry as the source.] Observe, by comparing (12.17) and (12.25), that  $\psi^r(\rho, \phi, z; m, s, a)$  and  $\psi^r(\rho, \phi, z; m, c, a)$  differ *only* by their prefactors  $\sin(m\phi)$  and  $\cos(m\phi)$ , respectively.

Also note that the functions  $\sin(m\phi)$  and  $\cos(m\phi)$  form a complete set. Therefore all possible single-layer ring-source results can be found by superposition of the potentials  $\psi^r(\rho, \phi, z; m, s, a)$  and  $\psi^r(\rho, \phi, z; m, c, a)$ .

## Exercises

**15.12.1.** Sketch, on two circles of radius  $\rho = a$ , the magnetic monopole distributions associated with Subsections 12.1 and 12.2, for various values of  $m$ . Verify from their definitions that these distributions have vanishing net “magnetic” charge provided  $m \geq 1$ . Sketch the expected magnetic field lines from these distributions to see that they are indeed consistent with the field lines for pure normal and skew multipoles.

## 15.13 On-Axis Gradients for these Scalar Potentials

In view of (12.17) and (12.25) and the general cylindrical harmonic expansion result (3.33), we know that we may write relations of the form

$$\psi^r(\rho, \phi, z; m, s, a) = -\sin(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} \hat{C}_{m,s}^{[2\ell]}(z) \rho^{2\ell+m}\tag{15.13.1}$$

and

$$\psi^r(\rho, \phi, z; m, c, a) = -\cos(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} \hat{C}_{m,c}^{[2\ell]}(z) \rho^{2\ell+m}.\tag{15.13.2}$$

The goal of this section is to find explicit results for the on-axis gradient functions  $\hat{C}_{m,\alpha}^{[2\ell]}(z)$ .

Define a function  $F(m, \rho, z, a)$  by the rule

$$F(m, \rho, z, a) = \int_{-\pi}^{\pi} d\theta \cos(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-1/2}. \quad (15.13.3)$$

It appears as a common factor in the relations (12.17) and (12.25) for  $\psi^r(\rho, \phi, z; m, s, a)$  and  $\psi^r(\rho, \phi, z; m, c, a)$ . Comparison of (12.17) and (12.25) with (13.1) and (13.2) yields the results

$$\sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} \hat{C}_{m,\alpha}^{[2\ell]}(z) \rho^{2\ell+m} = -B_0 a^2 F(m, \rho, z, a), \quad (15.13.4)$$

from which it follows that

$$\hat{C}_{m,\alpha}^{[0]}(z) = -B_0 a^2 \lim_{\rho \rightarrow 0} (1/\rho)^m F(m, \rho, z, a). \quad (15.13.5)$$

Note that, according to (13.5), the on-axis gradients  $\hat{C}_{m,\alpha}^{[0]}(z)$  are, in fact, *independent* of  $\alpha$ . This happy circumstance occurs because the source functions  $\sigma_m$  appearing in (12.2) and (12.19) differ only in their prefactors  $\sin(m\phi')$  and  $\cos(m\phi')$ , respectively. And, as calculation showed, the unprimed analogs of these prefactors are the same as the prefactors that appear in (12.17) and (12.25), respectively.

Let us work on evaluating (13.5). Our approach will be similar to what was done in Subsection 10.2 for the case of the magnetic monopole doublet. Write

$$[\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-1/2} = [a^2 + z^2]^{-1/2} [1 - 2wh + h^2]^{-1/2} \quad (15.13.6)$$

where

$$h = \rho/(a^2 + z^2)^{1/2} \quad (15.13.7)$$

and

$$w = [a/(a^2 + z^2)^{1/2}] \cos \theta. \quad (15.13.8)$$

Recall the Legendre polynomial generating function expansion (10.15), which we rewrite below,

$$[1 - 2wh + h^2]^{-1/2} = \sum_{\ell=0}^{\infty} h^{\ell} P_{\ell}(w). \quad (15.13.9)$$

Combining (13.6) through (13.9) gives the result

$$[\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-1/2} = [a^2 + z^2]^{-1/2} \sum_{n=0}^{\infty} [\rho/(a^2 + z^2)^{1/2}]^n P_n(w). \quad (15.13.10)$$

Further combining (13.3) and (13.10) produces the relation

$$F(m, \rho, z, a) = [a^2 + z^2]^{-1/2} \sum_{n=0}^{\infty} [\rho/(a^2 + z^2)^{1/2}]^n \int_{-\pi}^{\pi} d\theta \cos(m\theta) P_n(w). \quad (15.13.11)$$

To analyze the integral that occurs on the right side of (13.11) introduce, as before, the notation

$$\beta(z, a) = a/(a^2 + z^2)^{1/2} \quad (15.13.12)$$

so that  $w$  as given by (13.8) takes the form

$$w = \beta \cos \theta. \quad (15.13.13)$$

With this notation, we must study integrals of the form

$$c_{n,m} = \int_{-\pi}^{\pi} d\theta P_n(\beta \cos \theta) \cos m\theta \quad (15.13.14)$$

for general  $m$  and  $n$ .

To begin this study, we know from the Taylor expansions for Legendre polynomials that

$$P_n(w) = \{[(2n)!]/[2^n(n!)^2]\}w^n + \text{lower powers of } w. \quad (15.13.15)$$

We also know that

$$(\cos \theta)^m = (1/2)^{m-1} \cos m\theta + \text{lower angular frequency cosinusoidal terms.} \quad (15.13.16)$$

It follows that

$$c_{n,m} = 0 \text{ for } n < m, \quad (15.13.17)$$

and

$$\begin{aligned} c_{m,m} &= \{[(2m)!]/[2^m(m!)^2]\}(1/2)^{m-1}\beta^m \int_{-\pi}^{\pi} d\theta \cos^2(m\theta) \\ &= \pi\{[(2m)!]/[2^{2m-1}(m!)^2]\}\beta^m. \end{aligned} \quad (15.13.18)$$

Now insert the results obtained so far into (13.11). By the definition (13.14) the relation (13.11) takes the form

$$F(m, \rho, z, a) = [a^2 + z^2]^{-1/2} \sum_{n=0}^{\infty} [\rho/(a^2 + z^2)^{1/2}]^n c_{n,m}. \quad (15.13.19)$$

And making use of (13.17) brings (13.19) to the form

$$\begin{aligned} F(m, \rho, z, a) &= [a^2 + z^2]^{-1/2} [\rho/(a^2 + z^2)^{1/2}]^m c_{m,m} \\ &\quad + [a^2 + z^2]^{-1/2} \sum_{n>m}^{\infty} [\rho/(a^2 + z^2)^{1/2}]^n c_{n,m}. \end{aligned} \quad (15.13.20)$$

We observe from (13.14) that the coefficients  $c_{n,m}$  do not depend on  $\rho$ . Also, we know that we need the limit (13.5). We see that, in the limit  $\rho \rightarrow 0$ , only the first term on the right side of (13.20) contributes to (13.5). We conclude that the on-axis gradients are given by the relations

$$\begin{aligned} \hat{C}_{m,\alpha}^{[0]}(z) &= -(B_0 a^2)(a^2 + z^2)^{-1/2} (a^2 + z^2)^{-m/2} c_{m,m} \\ &= -(B_0 a^2)(a^2 + z^2)^{-1/2} (a^2 + z^2)^{-m/2} \pi\{[(2m)!]/[2^{2m-1}(m!)^2]\} \beta^m \\ &= -(B_0 a^2)\{[(2m)!]/[2^{2m-2}(m!)^2 a^{m+1}]\} \beta^{2m+1}(z). \end{aligned} \quad (15.13.21)$$

We have found exploit results for the on-axis gradients for all the normal and skew multipole potentials produced by single-layer ring sources.

Define, for future use, functions  $\delta_m(z, a)$  by the rules

$$\delta_m(z, a) = d_m(1/a)\beta^{2m+1} = d_m(1/a)a^{2m+1}/(z^2 + a^2)^{(2m+1)/2} \quad (15.13.22)$$

where

$$d_1 = 1/2 \quad (15.13.23)$$

and the  $d_m$  satisfy the recursion relation

$$d_{m+1} = [(2m+1)/(2m)]d_m. \quad (15.13.24)$$

Then it can be verified that  $\delta_m(z, a)$  is an *approximating* delta function in the sense that

$$\lim_{a \rightarrow 0} \delta_m(z, a) = \delta(z). \quad (15.13.25)$$

See Subsection 14.1. With these definitions, we may write

$$\begin{aligned} \hat{C}_{m,\alpha}^{[0]}(z, a) &= -(B_0 a^2) \{[(2m)!]/[2^{2m-2}(m!)^2 a^{m+1}]\} \beta^{2m+1}(z, a) \\ &= -(B_0 a^2) \{[(2m)!]/[2^{2m-2}(m!)^2 a^m]\} (1/a) \beta^{2m+1}(z, a) \\ &= -(B_0 a^2/d_m) \{[(2m)!]/[2^{2m-2}(m!)^2 a^m]\} \delta_m(z, a). \end{aligned} \quad (15.13.26)$$

See (13.21) and (13.22).

At this point we also observe that the  $\hat{C}_{m,\alpha}^{[0]}(z)$  have the asymptotic fall off

$$|\hat{C}_{m,\alpha}^{[0]}(z)| \sim (a/|z|)^{2m+1} \quad (15.13.27)$$

for large  $|z|$ .

Finally we remark that the integral (13.3) can be evaluated and the series (13.4), and also related series, can be summed, to yield expressions involving Legendre functions of the second kind. See Exercise 15.1.

## 15.14 Approximating Delta, Signum, and Bump Functions

Eventually we will want to describe the on-axis gradient functions  $C_{m,\alpha}^{[n]}(z)$  produced by superpositions of single- and double-layer ring sources. When the time comes we will find it convenient to employ what we call approximating delta, signum, and bump functions. Much of the required background for their use is presented in this section.

### 15.14.1 Approximating Delta Functions

Section 13 defined approximating delta functions functions  $\delta_m(z, a)$  by the rules

$$\delta_m(z, a) = d_m(1/a)\beta^{2m+1} = d_m(1/a)(a)^{2m+1}/(z^2 + a^2)^{[(2m+1)/2]} \quad (15.14.1)$$

where

$$d_1 = 1/2 \quad (15.14.2)$$

and the  $d_m$  satisfy the recursion relation

$$d_{m+1} = [(2m+1)/(2m)]d_m. \quad (15.14.3)$$

Recall (13.22) through (13.24). Thus the first few coefficients  $d_m$  have the values

$$d_1 = 1/2, d_2 = 3/4, d_3 = 15/16, d_4 = 35/32, d_5 = *, d_6 = *, d_7 = *, d_8 = *. \quad (15.14.4)$$

And Section 13 claimed, see (13.25), that each  $\delta_m(z, a)$  is an *approximating* delta function in the sense that

$$\lim_{a \rightarrow 0} \delta_m(z, a) = \delta(z). \quad (15.14.5)$$

The purpose of this subsection is to verify these claims and to explore along the way various features of the  $\delta_m(z, a)$  that will be useful for future work.

Easy calculations show that

$$\delta_m(-z, a) = \delta_m(z, a), \quad (15.14.6)$$

$$\delta_m(0, a) = d_m(1/a), \quad (15.14.7)$$

$$\lim_{a \rightarrow 0} \delta_m(z, a) = 0 \text{ for } z \neq 0, \quad (15.14.8)$$

$$\delta_m(z, a) = (d_m/a)\{(a/|z|)^{2m+1} + O[(a/|z|)^{2m+3}]\} \text{ as } |z| \rightarrow \infty. \quad (15.14.9)$$

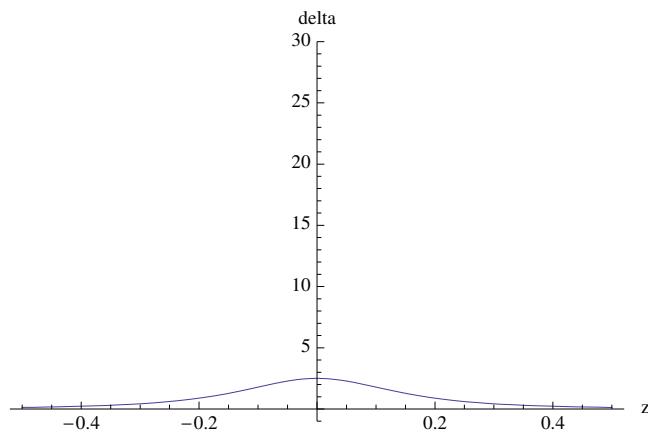
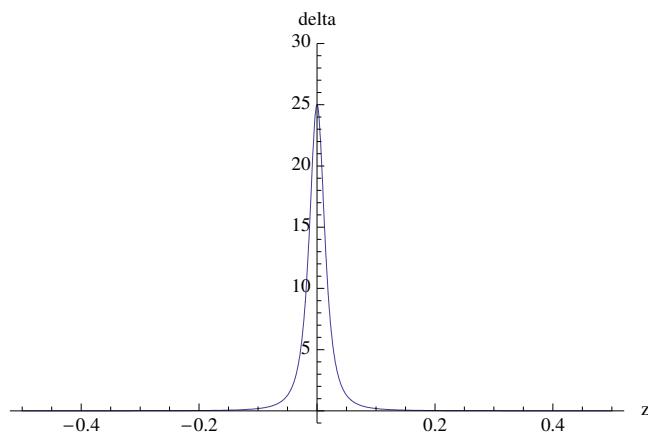
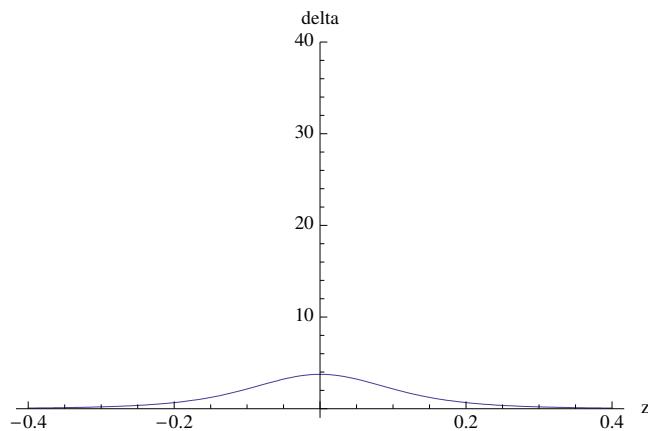
Figures 14.1 through 14.4 below illustrate these results for the case  $m = 1$ , for which

$$\delta_1(z, a) = (1/2)(1/a)(a^3)/(z^2 + a^2)^{3/2} = (1/2)(a^2)/(z^2 + a^2)^{3/2}, \quad (15.14.10)$$

and the case  $m = 2$ , for which

$$\delta_2(z, a) = (3/4)(1/a)(a^5)/(z^2 + a^2)^{5/2} = (3/4)(a^4)/(z^2 + a^2)^{5/2}. \quad (15.14.11)$$

Observe that the peaks becomes higher and narrower as  $a \rightarrow 0$ . And, as specified by (14.9),  $a$  and  $m$  control the fall-off rate. All these results are consistent with the claim (14.5). Moreover, the peaks also become higher and narrower as  $m \rightarrow \infty$ .

Figure 15.14.1: The approximating delta function (14.10) when  $a = .2$ .Figure 15.14.2: The approximating delta function (14.10) when  $a = .02$ .Figure 15.14.3: The approximating delta function (14.11) when  $a = .2$ .

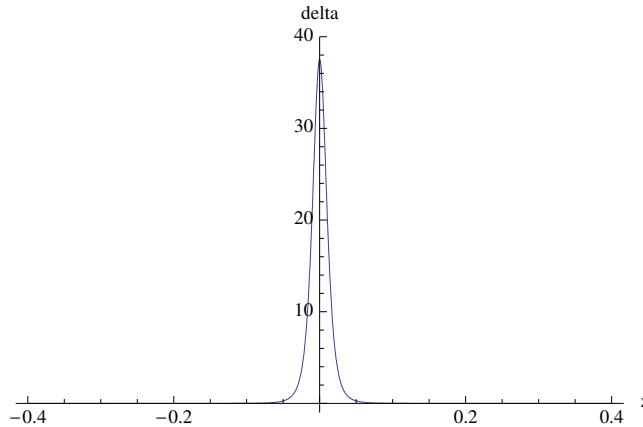


Figure 15.14.4: The approximating delta function (14.11) when  $a = .02$ .

What remains is to verify that

$$\int_{-\infty}^{\infty} dz \delta_m(z, a) = 1, \quad (15.14.12)$$

which will complete the proof of (12.54), and to verify the relations (12.51) and (12.52).

In the case  $m = 1$  we find from (12.59) and an integral table that

$$\begin{aligned} \int_{-\infty}^{\infty} dz \delta_1(z, a) &= (1/2)a^2 \int_{-\infty}^{\infty} dz (z^2 + a^2)^{3/2} \\ &= (1/2)z/(z^2 + a^2)^{1/2}|_{-\infty}^{\infty} \\ &= (1/2)(1 + 1) = 1, \end{aligned} \quad (15.14.13)$$

as desired.

To proceed further, make the induction hypothesis that (12.50) and (12.61) hold for some value of  $m$ , in which case we have the relation

$$(z^2 + a^2)^{[-(2m+1)/2]} = (1/d_m)a^{-2m}\delta_m(z, a) \quad (15.14.14)$$

so that

$$\int_{-\infty}^{\infty} dz (z^2 + a^2)^{[-(2m+1)/2]} = (1/d_m)a^{-2m}. \quad (15.14.15)$$

Now differentiate both sides of (12.64) with respect to  $a$ . For the left side of (12.64) we find the result

$$\begin{aligned} \partial_a \int_{-\infty}^{\infty} dz (z^2 + a^2)^{[-(2m+1)/2]} &= [-(2m+1)/2](2a) \int_{-\infty}^{\infty} dz (z^2 + a^2)^{[-(2m+1)/2]-1} \\ &= -(2m+1)(1/d_{m+1})a^{-(2m+1)} \int_{-\infty}^{\infty} dz \delta_{m+1}(z, a). \end{aligned} \quad (15.14.16)$$

For the right side of (12.64) we find the result

$$\partial_a[(1/d_m)a^{-2m}] = -(2m)(1/d_m)a^{-(2m+1)}. \quad (15.14.17)$$

Since the right sides of (12.65) and (12.66) must be equal, we find that

$$\int_{-\infty}^{\infty} dz \delta_{m+1}(z, a) = 1 \quad (15.14.18)$$

provided

$$(2m + 1)(1/d_{m+1}) = (2m)(1/d_m), \quad (15.14.19)$$

which is equivalent to (12.52).

## 15.14.2 Approximating Signum Functions

The previous subsection defined and studied approximating delta functions  $\delta_m(z, a)$ . The purpose of this subsection and the next is to define related *signum* and *bump* functions.

What we will call the *true* signum function has the definition

$$\begin{aligned} \text{sgn}(z) &= 1 \text{ if } z > 0, \\ \text{sgn}(z) &= 0 \text{ if } z = 0, \\ \text{sgn}(z) &= -1 \text{ if } z < 0. \end{aligned} \quad (15.14.20)$$

Note that  $\text{sgn}(z)$  is an odd function,

$$\text{sgn}(-z) = -\text{sgn}(z). \quad (15.14.21)$$

From the definition (4.20) it follows that, in the context of *distribution theory*, there is the relation

$$\partial_z \text{sgn}(z) = 2\delta(z). \quad (15.14.22)$$

Correspondingly, there is the integral relation

$$\text{sgn}(z) = \int_0^z dz' 2\delta(z'). \quad (15.14.23)$$

With this background in mind, we are prepared to define *approximating* signum functions  $\text{sgn}_m(z, a)$ . Let  $\delta_m(z, a)$  be an approximating delta function. In analogy to (14.23), we make the definition

$$\text{sgn}_m(z, a) = \int_0^z dz' 2\delta_m(z', a). \quad (15.14.24)$$

With this definition we see that there is immediately the relation

$$\partial_z \text{sgn}_m(z, a) = 2\delta_m(z, a), \quad (15.14.25)$$

which is the analog of (14.22). Correspondingly, there will be the relation

$$\lim_{a \rightarrow 0} \text{sgn}_m(z, a) = \text{sgn}(z). \quad (15.14.26)$$

Also, it follows from (14.6) and (14.24) that  $\text{sgn}_m(z, a)$ , like  $\text{sgn}(z)$ , is an odd function,

$$\text{sgn}_m(-z, a) = -\text{sgn}_m(z, a). \quad (15.14.27)$$

Finally, in view of (14.6), (14.12), (14.27) and the definition (14.24), there will be the relations

$$\lim_{z \rightarrow \pm\infty} \operatorname{sgn}_m(z, a) = \pm 1. \quad (15.14.28)$$

Let us examine the first few  $\operatorname{sgn}_m(z, a)$  in detail. To simplify notation, and in analogy with (2.17), make the definition

$$r = (z^2 + a^2)^{1/2}. \quad (15.14.29)$$

With this definition we may write

$$\begin{aligned} \delta_m(z, a) &= d_m(1/a)\beta^{2m+1} = d_m(1/a)(a)^{2m+1}/(z^2 + a^2)^{[(2m+1)/2]} \\ &= d_m(a)^{2m}(1/r)^{2m+1}. \end{aligned} \quad (15.14.30)$$

Correspondingly, following (14.24), the approximating signum functions are given by the relations

$$\operatorname{sgn}_m(z, a) = \int_0^z dz' 2\delta_m(z', a) = 2d_m(a)^{2m} \int_0^z dz'(1/r)^{2m+1}. \quad (15.14.31)$$

The integrals appearing in (14.31) can be found in H. B. Dwight's book. See the references at the end of this chapter. There one finds, for the first few, the results

$$\int_0^z dz'(1/r)^3 = (1/a)^2(z/r), \quad (15.14.32)$$

$$\int_0^z dz'(1/r)^5 = (1/a)^4[(z/r) - (1/3)(z/r)^3], \quad (15.14.33)$$

$$\int_0^z dz'(1/r)^7 = (1/a)^6[(z/r) - (2/3)(z/r)^3 + (1/5)(z/r)^5]. \quad (15.14.34)$$

Note that, consistent with (14.27), they are all *odd* functions of  $z$ . We remark that the sequence of relations beginning with (14.32) can be found by a recursion relation involving simple differentiation. See Exercise 14.1. Simple calculation shows that there is the limiting property

$$\lim_{z \rightarrow +\infty} (z/r) = 1. \quad (15.14.35)$$

Employing (14.32) and (14.33) in (14.31) gives for the first two  $\operatorname{sgn}_m(z, a)$  functions the results

$$\operatorname{sgn}_1(z, a) = 2d_1(a)^2 \int_0^z dz'(1/r)^3 = 2d_1(z/r) = z/(z^2 + a^2)^{1/2} \quad (15.14.36)$$

and

$$\begin{aligned} \operatorname{sgn}_2(z, a) &= 2d_2(a)^4 \int_0^z dz'(1/r)^5 = 2d_2[(z/r) - (1/3)(z/r)^3] \\ &= (3/2)[z/(z^2 + a^2)^{1/2} - (1/3)z^3/(z^2 + a^2)^{3/2}]. \end{aligned} \quad (15.14.37)$$

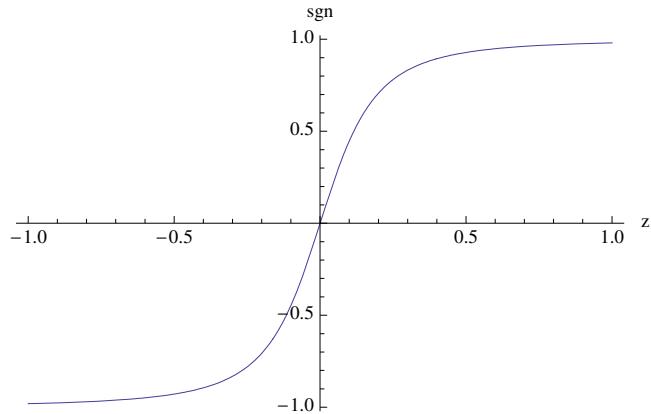


Figure 15.14.5: The approximating signum function (14.36) when  $a = .2$ .

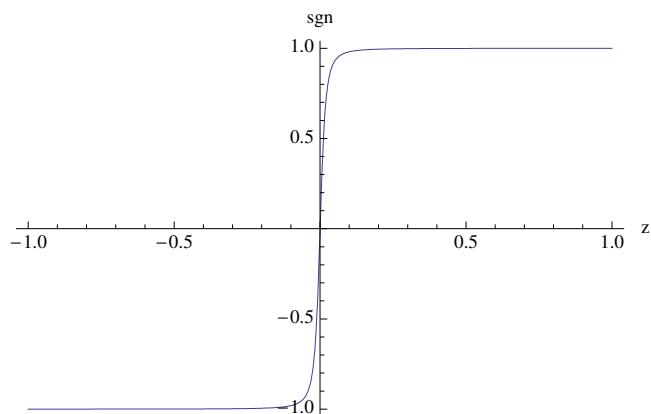


Figure 15.14.6: The approximating signum function (14.36) when  $a = .02$ .

As a sanity check, the limits (14.28) can be verified explicitly for the cases  $m = 1$  and  $m = 2$  by using (4.35) in (4.36) and (4.37). Figures 14.5 and 14.6 above display  $\text{sgn}_1(z, a)$  for two values of  $a$ . Inspection shows that the properties (14.26) through (14.28) are evident.

The next question to be addressed is how rapidly these limits are approached as  $z \rightarrow \pm\infty$ . Since  $\text{sgn}_m(z, a)$  is an odd function, see (14.27), it suffices to examine the case  $z > 0$ . For  $\text{sgn}_1(z, a)$  we see from (14.36) that (for  $z > 0$ ) there is the expansion

$$\begin{aligned}\text{sgn}_1(z, a) &= z/(z^2 + a^2)^{1/2} = z/[z(1 + a^2/z^2)^{1/2}] = (1 + a^2/z^2)^{-1/2} \\ &= 1 - (1/2)(a/z)^2 + O[(a/z)^4] \text{ as } z \rightarrow +\infty.\end{aligned}\quad (15.14.38)$$

What about the functions  $\text{sgn}_m(z, a)$  for  $m > 1$ ? We know that the  $\delta_m(z, a)$  become evermore sharply peaked as  $m$  increases, and therefore we expect the  $\text{sgn}_m(z, a)$  to take on their limiting values evermore rapidly as  $m$  increases. To explore this possibility, make the large  $z$  Ansätze

$$\text{sgn}_m(z, a) = 1 - \gamma_m(a/z)^{2n_m} + \dots \text{ as } z \rightarrow +\infty. \quad (15.14.39)$$

where  $\gamma_m$  and  $n_m$  are to be determined. From this hypothesis it follows that

$$\partial_z \text{sgn}_m(z, a) = \gamma_m 2n_m (1/a)(a/z)^{2n_m+1} + \dots \quad (15.14.40)$$

But, from (14.25) and (14.30), we find that

$$\begin{aligned}\partial_z \text{sgn}_m(z, a) &= 2\delta_m(z, a) = 2d_m(a)^{2m}(1/r)^{2m+1} \\ &= 2d_m(1/a)(a/z)^{2m+1} + O[(a/z)^{2m+3}].\end{aligned}\quad (15.14.41)$$

Comparison of (14.40) and (14.41) yields the relations

$$n_m = m \quad (15.14.42)$$

and

$$\gamma_m = d_m(1/m). \quad (15.14.43)$$

Thus, we have the result

$$\text{sgn}_m(z, a) = 1 - d_m(1/m)(a/z)^{2m} + O[(a/z)^{2m+2}] \text{ as } z \rightarrow +\infty. \quad (15.14.44)$$

Note that (14.44) agrees with (14.38) in the case  $m = 1$ , as it should. It can also be verified directly that (14.44) is correct in the case  $m = 2$ . See Exercise 14.2.

### 15.14.3 Approximating Bump Functions

The *true* (also called *hard-edge*) bump function is defined by the rule,

$$\begin{aligned}\text{bump}(z, L) &= 1 \text{ for } z \in (0, L), \\ \text{bump}(z, L) &= 0 \text{ for } z \notin [0, L], \\ \text{bump}(0, L) &= \text{bump}(L, L) = 1/2.\end{aligned}\quad (15.14.45)$$

Note that, by definition, for the true bump function there is the result

$$\int_{-\infty}^{\infty} dz \text{bump}(z, L) = L. \quad (15.14.46)$$

The true bump function can evidently also be defined in terms of the delta function by the relation

$$\text{bump}(z, L) = \int_0^L dz' \delta(z - z'). \quad (15.14.47)$$

From this integral representation we immediately see that

$$\begin{aligned} \int_{-\infty}^{\infty} dz \text{bump}(z, L) &= \int_{-\infty}^{\infty} dz \int_0^L dz' \delta(z - z') = \int_0^L dz' \int_{-\infty}^{\infty} dz \delta(z - z') \\ &= \int_0^L dz' = L, \end{aligned} \quad (15.14.48)$$

in agreement with (14.46).

Let us manipulate the integral representation (14.47). Since the delta function is even, we may also write

$$\text{bump}(z, L) = \int_0^L dz' \delta(z' - z). \quad (15.14.49)$$

Next change variables of integration by writing

$$w = z' - z \quad (15.14.50)$$

and further manipulate so that (14.49) becomes

$$\begin{aligned} \text{bump}(z, L) &= \int_{-z}^{L-z} dw \delta(w) = \int_{-z}^0 dw \delta(w) + \int_0^{L-z} dw \delta(w) \\ &= - \int_0^{-z} dw \delta(w) + \int_0^{L-z} dw \delta(w) \\ &= -(1/2)\text{sgn}(-z) + (1/2)\text{sgn}(L - z) \\ &= (1/2)\text{sgn}(z) - (1/2)\text{sgn}(z - L). \end{aligned} \quad (15.14.51)$$

[Here we have used the oddness property (14.21).] We have learned that the true bump function is the difference of two true signum functions.

With this background in mind, we are prepared to define the *approximating (soft-edge)* bump function  $\text{bump}_m(z, a, L)$ . In analogy to (14.51), we make the rule

$$\text{bump}_m(z, a, L) = (1/2)\text{sgn}_m(z, a) - (1/2)\text{sgn}_m(z - L, a). \quad (15.14.52)$$

Alternatively, in analogy to (14.49),  $\text{bump}_m(z, a, L)$  can also be defined in terms of an approximating delta function by the rule

$$\text{bump}_m(z, a, L) = \int_0^L dz' \delta_m(z - z', a). \quad (15.14.53)$$

In fact, the definitions (14.52) and (14.53) are equivalent, as can be seen by following steps that are analogous to those made in passing between (14.49) and (14.51).

It can be verified that the approximating bump function  $\text{bump}_m(z, a, L)$  has the properties

$$\text{bump}_m(z, a, L) \simeq 1 \text{ for } z \in (0, L), \quad (15.14.54)$$

$$\text{bump}_m(z, a, L) \simeq 0 \text{ for } z \text{ elsewhere}, \quad (15.14.55)$$

$$\lim_{a \rightarrow 0} \text{bump}_m(z, a, L) = \text{bump}(z, L). \quad (15.14.56)$$

At this point we should remark that we have been using the term *bump function* in a somewhat different way from that often employed in mathematics. In mathematics a bump function is generally a smooth ( $C^\infty$ ) function with exact value 1 over some region and exact value 0 slightly outside this region. By contrast, in (14.54) and (14.55), we require only that this be approximately true of  $\text{bump}_m(z, a, L)$ , and  $\text{bump}(z, L)$  is not  $C^\infty$ .

There are three other properties of  $\text{bump}_m(z, a, L)$  that are of interest. The first is that it is *even* about the point  $z = L/2$ ,

$$\text{bump}_m(L/2 + u, a, L) = \text{bump}_m(L/2 - u, a, L). \quad (15.14.57)$$

This result follows from (14.27) and the representation (14.52). See Exercise 14.3. The second is that, like (14.46), there is the integral relation

$$\int_{-\infty}^{\infty} dz \text{bump}_m(z, a, L) = L. \quad (15.14.58)$$

Using the same line of reasoning that connected (14.47) and (14.48), this result follows from the property (14.12) and the representation (14.53).

The third property concerns the falloff to the value zero as  $z$  becomes ever more distant from the interval  $[0, L]$ . For sufficiently small  $a$  an approximating bump function looks like a “box” with faces at  $z = 0$  and  $z = L$ . Let us consider the behavior of an approximating bump function beyond the face at  $z = L$ . Namely, let us write  $z = L + \zeta$  and examine the behavior of the approximating bump function as  $\zeta \rightarrow +\infty$ . [The behavior before the face  $z = 0$  then follows from the symmetry relation (14.57).] For the approximating bump function we will use the representation (14.52), and for the approximating signum functions we will use the expansion (14.44). So doing yields the expansion

$$\text{bump}_m(z, a, L) \simeq (1/2)(d_m/m)\{-[a/(L + \zeta)]^{2m} + (a/\zeta)^{2m}\} \text{ as } \zeta \rightarrow +\infty. \quad (15.14.59)$$

We suppose that  $L > a$  so that for  $a < \zeta < L$  the dominant term in (14.59) is that with the factor  $(a/\zeta)^{2m}$ . Therefore we expect the asymptotic behavior

$$\text{bump}_m(z, a, L) \simeq (1/2)(d_m/m)(a/\zeta)^{2m} \text{ for } \zeta \in (a, L). \quad (15.14.60)$$

For even larger values of  $\zeta$  there is the expansion

$$a/(L + \zeta) = (a/\zeta)(1 + L/\zeta)^{-1} \simeq (a/\zeta)(1 - L/\zeta), \quad (15.14.61)$$

from which it follows that, for sufficiently large  $\zeta$ , the function  $\text{bump}_m(z, a, L)$  has the even more rapid fall off

$$\text{bump}_m(z, a, L) \simeq (1/2)(d_m/m)(a/\zeta)^{2m}(L/\zeta) \text{ for } \zeta > L. \quad (15.14.62)$$

We close this subsection with a graphic portrayal of  $\text{bump}_1(z, a, L)$  for which, using (14.36) and (14.37), there is the result

$$\begin{aligned} \text{bump}_1(z, a, L) &= (1/2)\text{sgn}_1(z, a) - (1/2)\text{sgn}_1(z - L, a) \\ &= (1/2)(z)[z^2 + a^2]^{-1/2} - (1/2)(z - L)[(z - L)^2 + a^2]^{-1/2}. \end{aligned} \quad (15.14.63)$$

Figures 14.7 and 14.8 below illustrate, for two values of  $a$ , the properties (14.54) through (14.57) and the properties (14.60) and (14.62).

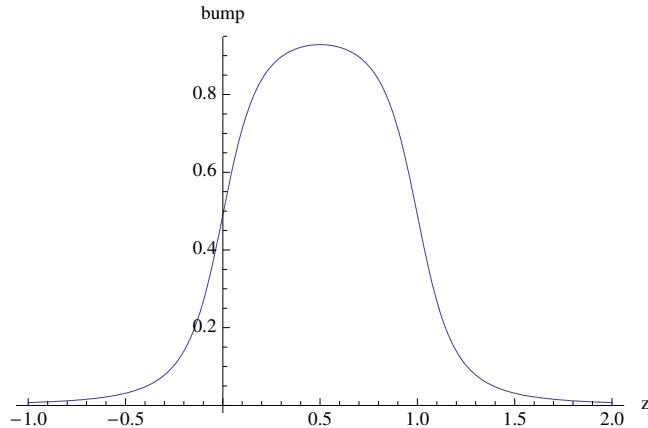


Figure 15.14.7: The approximating bump function (14.63) when  $a = .2$  and  $L = 1$ .

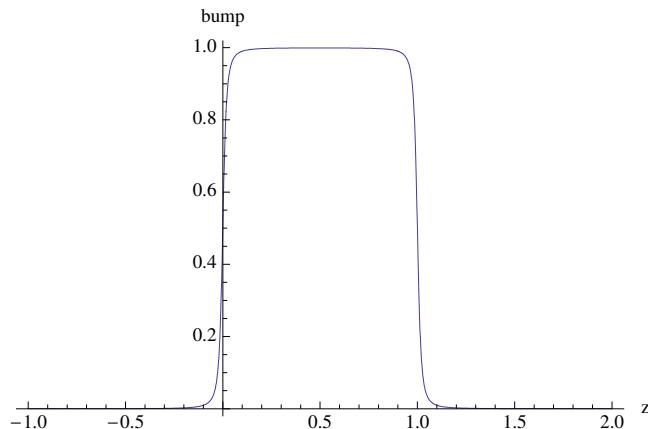


Figure 15.14.8: The approximating bump function (14.63) when  $a = .02$  and  $L = 1$ .

## Exercises

**15.14.1.** Verify (14.32) by applying  $\partial_z$  to both sides. Verify that applying  $\partial_a$  to both sides of (14.32) produces (14.33), etc.

**15.14.2.** Verify directly that (14.44) is correct in the case  $m = 2$  by making a large  $z$  expansion of the right side of (14.37).

**15.14.3.** Your task in this exercise is to verify (14.57). Using (14.27) and the representation (14.52), show that there is the result

$$\begin{aligned} \text{bump}_m(L/2 + u, a, L) &= (1/2)\text{sgn}_m(L/2 + u, a) - (1/2)\text{sgn}_m(L/2 + u - L, a) \\ &= (1/2)\text{sgn}_m(L/2 + u, a) - (1/2)\text{sgn}_m(-L/2 + u, a) \\ &= (1/2)\text{sgn}_m(L/2 + u, a) + (1/2)\text{sgn}_m(L/2 - u, a). \end{aligned} \quad (15.14.64)$$

Similarly, find a result for  $\text{bump}_m(L/2 - u, a, L)$ . Show that the two results you have found are equal.

**15.14.4.** In  $\text{bump}_1(z, a, L)$ , as given explicitly by (4.63), make the substitution  $z = L + \zeta$  and find its expansion in  $\zeta$  for large positive  $\zeta$ . Show that your results agree with (14.60) and (14.62).

## 15.15 Relation between Potentials Produced by Analogous Single- and Double-Layer Ring Sources

### 15.15.1 Double-Layer Ring Distributions

Defining integral relation for potential due to related double-layer ring source

In Section 16, for any value of  $m \geq 1$ , we defined potentials  $\psi^r$  due to single-layer ring source distributions. We will now define related potentials  $\psi_m^{dlr}(\mathbf{r})$ , due to *double-layer ring* source distributions, by writing

$$\psi_m^{dlr}(\mathbf{r}) = [1/(4\pi)] \int_S d\mathbf{S}' \cdot (\mathbf{r} - \mathbf{r}') \tau_m(\mathbf{r}') / ||\mathbf{r} - \mathbf{r}'||^3 \quad (15.15.1)$$

where

$$\tau_m(\mathbf{r}') = (4\pi)B_0a^2\delta(z')\sin m\phi' \quad (15.15.2)$$

and  $S$  is the surface of a cylinder of radius  $a$ . See (9.16). Note that  $\tau_m$  has units  $BL$  in accord with (9.19). Also note that  $\sigma_m$  as given by (16.2) and  $\tau_m$  as given by (17.2) are identical save for the extra factor of  $a$  in (17.2) needed for  $\tau_m$ , and hence  $\psi_m^{dlr}$ , to have the proper units.

For the further ingredients in the integral we employ cylindrical coordinates and write

$$\mathbf{r} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z; \quad (15.15.3)$$

$$\mathbf{r}' = x'\mathbf{e}_x + y'\mathbf{e}_y + z'\mathbf{e}_z; \quad (15.15.4)$$

$$x = \rho \cos \phi, \quad x' = a \cos \phi'; \quad (15.15.5)$$

$$y = \rho \sin \phi, \quad y' = a \sin \phi'; \quad (15.15.6)$$

$$\mathbf{r}' = a(\cos \phi' \mathbf{e}_x + \sin \phi' \mathbf{e}_y) + z' \mathbf{e}_z; \quad (15.15.7)$$

$$\mathbf{r} - \mathbf{r}' = (x - a \cos \phi') \mathbf{e}_x + (y - a \sin \phi') \mathbf{e}_y + (z - z') \mathbf{e}_z; \quad (15.15.8)$$

$$\begin{aligned} \|\mathbf{r} - \mathbf{r}'\|^2|_{z'=0} &= (x - x')^2 + (y - y')^2 + z^2 \\ &= x^2 + y^2 - 2(xx' + yy') + (x')^2 + (y')^2 + z^2 \\ &= \rho^2 - 2a\rho(\cos \phi \cos \phi' + \sin \phi \sin \phi') + a^2 + z^2 \\ &= \rho^2 - 2a\rho \cos(\phi - \phi') + a^2 + z^2; \end{aligned} \quad (15.15.9)$$

$$d\mathbf{S}' = a(d\phi' dz')(\cos \phi' \mathbf{e}_x + \sin \phi' \mathbf{e}_y). \quad (15.15.10)$$

Then we have

$$\begin{aligned} d\mathbf{S}' \cdot (\mathbf{r} - \mathbf{r}') &= a(d\phi' dz')(x \cos \phi' - a \cos^2 \phi' + y \sin \phi' - a \sin^2 \phi' \\ &= a(d\phi' dz')(x \cos \phi' + y \sin \phi' - a) \\ &= a(d\phi' dz')[\rho \cos(\phi - \phi') - a] \end{aligned} \quad (15.15.11)$$

since

$$x \cos \phi' + y \sin \phi' = \rho(\cos \phi \cos \phi' + \sin \phi \sin \phi') = \rho \cos(\phi - \phi'). \quad (15.15.12)$$

Putting everything together gives the net result

$$\begin{aligned} \psi_m^{dlr}(\mathbf{r}) &= [1/(4\pi)] \int_S d\mathbf{S}' \cdot (\mathbf{r} - \mathbf{r}') \tau_m(\mathbf{r}') / \|\mathbf{r} - \mathbf{r}'\|^3 = \\ &B_0 a^3 \int_{-\pi}^{\pi} d\phi' \int_{-\infty}^{\infty} dz' [\rho \cos(\phi - \phi') - a] \sin(m\phi') \delta(z') / \|\mathbf{r} - \mathbf{r}'\|^3 = \\ &B_0 a^3 \int_{-\pi}^{\pi} d\phi' [\rho \cos(\phi - \phi') - a] \sin(m\phi') / \|\mathbf{r} - \mathbf{r}'\|^3|_{z'=0} = \\ &B_0 a^3 \int_{-\pi}^{\pi} d\phi' [\rho \cos(\phi - \phi') - a] \sin m\phi' [\rho^2 - 2a\rho \cos(\phi - \phi') + a^2 + z^2]^{-3/2}. \end{aligned} \quad (15.15.13)$$

### Standard form integral relation

What is inviting now is manipulation of the integral on the far right side of (17.13). This manipulation will be similar to that at the end of Subsection 16.1 with the goal of again achieving a standard form.

Let us begin. Since the cosine function is even, we may first rewrite (17.13) in the form

$$\psi_m^{dlr}(\mathbf{r}) = B_0 a^3 \int_{-\pi}^{\pi} d\phi' [\rho \cos(\phi' - \phi) - a] \sin m\phi' [\rho^2 - 2a\rho \cos(\phi' - \phi) + a^2 + z^2]^{-3/2}. \quad (15.15.14)$$

Next make the steps (16.11) through (16.14), as done before, so that (17.14) becomes

$$\begin{aligned} \psi_m^{dlr}(\mathbf{r}) &= B_0 a^3 \int_{-\pi}^{\pi} d\theta \\ &[\rho \cos(\theta) - a] [\sin(m\theta) \cos(m\phi) + \cos(m\theta) \sin(m\phi)] [\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-3/2}. \end{aligned} \quad (15.15.15)$$

Evidently (17.15) can be written as the sum of two terms. Doing so gives the result

$$\begin{aligned} \psi_m^{dlr}(\mathbf{r}) &= \\ &B_0 a^3 \cos m\phi \int_{-\pi}^{\pi} d\theta [\rho \cos(\theta) - a] \sin(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-3/2} \\ &+ B_0 a^3 \sin m\phi \int_{-\pi}^{\pi} d\theta [\rho \cos(\theta) - a] \cos(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-3/2}. \end{aligned} \quad (15.15.16)$$

Observe that the integrand for the first integral on the right side of (17.16) is an *odd* function of  $\theta$ , and therefore this integral vanishes. We are left with the standard-form result

$$\begin{aligned} \psi_m^{dlr}(\mathbf{r}) &= \\ &B_0 a^3 \sin m\phi \int_{-\pi}^{\pi} d\theta [\rho \cos(\theta) - a] \cos(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-3/2}, \end{aligned} \quad (15.15.17)$$

which isolates the  $\phi$  dependence of  $\psi_m^{dlr}$ .

#### 15.15.2 Explicit Formula Relating Potentials

The stage has been set to find a very useful connection: Look at (16.17). Suppose we compute the quantity  $\{a^3 \partial_a [(1/a^2) \psi^r(\rho, \phi, z; m, s, a)]\}$ . Do so in steps: From (16.17) we have

$$\begin{aligned} [(1/a^2) \psi^r(\rho, \phi, z; m, s, a)] &= \\ &B_0 \sin m\phi \int_{-\pi}^{\pi} d\theta \cos(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-1/2}. \end{aligned} \quad (15.15.18)$$

It follows that

$$\begin{aligned} \partial_a[(1/a^2)\psi^r(\rho, \phi, z; m, s, a)] &= \\ B_0 \sin m\phi \int_{-\pi}^{\pi} d\theta \cos(m\theta) \partial_a[\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-1/2} &= \\ B_0 \sin m\phi \int_{-\pi}^{\pi} d\theta [\rho \cos(\theta) - a] \cos(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-3/2}. \end{aligned} \quad (15.15.19)$$

Finally, we find that

$$\begin{aligned} a^3 \partial_a[(1/a^2)\psi^r(\rho, \phi, z; m, s, a)] &= \\ B_0 a^3 \sin m\phi \int_{-\pi}^{\pi} d\theta [\rho \cos(\theta) - a] \cos(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-3/2} &= \\ = \psi_m^{dlr}(\mathbf{r}). \end{aligned} \quad (15.15.20)$$

[Recall (17.17).] That is, we have the key result

$$\psi_m^{dlr}(\mathbf{r}) = a^3 \partial_a[(1/a^2)\psi^r(\rho, \phi, z; m, s, a)]. \quad (15.15.21)$$

We have been able to find the *double-layer ring* result  $\psi_m^{dlr}(\mathbf{r})$ , which according to (17.13) appears to be a reasonably complicated integral, in terms of an easy derivative involving the known *single-layer ring* result  $\psi^r$ . That these two results should be related . . .

Suppose we write out (17.17) in more detail to make the definition

$$\begin{aligned} \psi_m^{dlr}(\mathbf{r}, a) &= \\ B_0 a^3 \sin m\phi \int_{-\pi}^{\pi} d\theta [\rho \cos(\theta) - a] \cos(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-3/2}, \end{aligned} \quad (15.15.22)$$

and write out (16.17) in more detail to make the definition

$$\psi_m^r(\mathbf{r}, a) = B_0 a^2 \sin m\phi \int_{-\pi}^{\pi} d\theta \cos(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + z^2]^{-1/2}. \quad (15.15.23)$$

Then, employing this notation, (17.21) takes the form

$$\psi_m^{dlr}(\mathbf{r}, a) = a^3 \partial_a[(1/a^2)\psi_m^r(\mathbf{r}, a)]. \quad (15.15.24)$$

### 15.15.3 Explicit Formula Relating On-Axis Gradients

### 15.15.4 Explicit Formula for Potential due to Double-Layer Ring Source

## Exercises

**15.15.1.** Look at the standard form result for  $\psi^r$  given by (16.17). Verify that it can also be written in the form

$$\psi^r(\rho, \phi, z; m, s, a) = 2B_0(\sin m\phi)(a^2)[1/(2a\rho)]^{1/2} \int_0^\pi d\theta \cos(m\theta)[p - \cos(\theta)]^{-1/2} \quad (15.15.25)$$

where

$$p = (\rho^2 + z^2 + a^2)/(2\rho a). \quad (15.15.26)$$

The integral appearing in (17.30) is known to have the value

$$\int_0^\pi d\theta \cos(m\theta)[p - \cos(\theta)]^{-1/2} = Q_{m-1/2}(p) \quad (15.15.27)$$

where  $Q_\beta$  denotes a Legendre function of the *second kind*. [They are called *fractional* Legendre functions of the second kind if  $\beta$  is not an integer, as is the case in (17.32).] Therefore, putting everything together, we have the final result

$$\psi^r(\rho, \phi, z; m, s, a) = 2B_0(\sin m\phi)(a^2)[1/(2a\rho)]^{1/2}Q_{m-1/2}(p). \quad (15.15.28)$$

A fair amount is known about second-kind Legendre functions, and consequently (17.33) is potentially a useful result. One application is a study of the analytic properties of  $\psi^r$ . It can be shown that  $Q_\beta(p)$  is analytic in  $p$  save for branch points at  $p = \pm 1$  if  $\beta$  is integer, and also at  $p = -\infty$  if  $\beta$  is non-integer/fractional.

In particular, and for our purposes, we will explore the analytic properties of  $Q_{m-1/2}(p)$ . Evidently, as it stands in (17.32),  $Q_{m-1/2}(p)$  is analytic in  $p$  as long as  $[p - \cos(\theta)]$  does not vanish during the course of integration. Verify that in the course of integration  $[p - \cos(\theta)]$  will lie on the straight line connecting  $p - 1$  and  $p + 1$ . So there is surely analyticity in  $p$  as long as either  $p$  is complex or is real with  $p > 1$ . In particular, it can be shown that there is analyticity in the  $p$  plane cut along the real interval/line segment  $p \in [-\infty, 1]$ .

We next note that, according to (17.31),  $p$  is real as long as  $\rho, z$ , and  $a$  are real, which we expect them to be. Therefore, for our purposes, we are interested in knowing what the restrictions on  $\rho, z$ , and  $a$  are so that  $p \geq 1$ . Verify that

$$(a - \rho)^2 = a^2 + \rho^2 - 2\rho a \quad (15.15.29)$$

from which it follows that

$$\begin{aligned} p &= (\rho^2 + a^2 + z^2)/(2\rho a) = [2\rho a + z^2 + (a - \rho)^2]/(2\rho a) = \\ &1 + z^2/(2\rho a) + (a - \rho)^2/(2\rho a). \end{aligned} \quad (15.15.30)$$

Show that  $p > 1$  unless  $z = 0$  and  $\rho = a$ , in which case  $p = 1$  and  $\mathbf{r}$  is on the ring. We expect singular behavior on the ring because that is where the sources are.<sup>13</sup> Therefore, as far as the  $Q_{m-1/2}(p)$  term is involved,  $\psi^r$  is analytic *off* the ring.

What can be said about the  $[1/(2a\rho)]^{1/2}$  term in (17.33), which suggests that  $\psi^r$  may be singular for  $\mathbf{r}$  on the  $z$  axis? Show from (16.17) that there is actually no problem on the  $z$  axis, as is physically reasonable since then  $\mathbf{r}$  is far away from any source. Consequently, the only (real) singularities for  $\psi^r$  are on the ring.<sup>14</sup>

## 15.16 Potentials Produced by Single-Layer and Double-Layer Ring Sources Uniformly Distributed on a Cylindrical Surface

### 15.16.1 Use of Single-Layer Ring Source

### 15.16.2 Use of Double-Layer Ring Source

## 15.17 Closing Remarks

### 15.17.1 Caveat about Significance of Integrated Multipoles

Suppose the relations (3.65) and (3.66) are used to compute the integrals of the transverse field components  $B_x(x, y, z)$  and  $B_y(x, y, z)$  over the range  $z = -\infty$  to  $z = +\infty$ . We observe that for  $n \geq 0$  there are the relations

$$\int_{-\infty}^{\infty} dz C_0^{[n+2]}(z) = C_0^{[n+1]}(z)|_{-\infty}^{\infty}, \quad (15.17.1)$$

$$\int_{-\infty}^{\infty} dz C_{m,\alpha}^{[n+1]}(z) = C_{m,\alpha}^{[n]}(z)|_{-\infty}^{\infty}. \quad (15.17.2)$$

Also we know that the  $C_0^{[n+1]}(z)$  and the  $C_{m,\alpha}^{[n]}(z)$  vanish at  $z = \pm\infty$ . Consequently, all of the terms in the sums (3.65) and (3.66) integrate to zero save for those that involve  $C_0^{[1]}(z)$ .

---

<sup>13</sup>As stated earlier, it can be shown that  $Q_{m-1/2}(p)$  has a branch point at  $p = 1$ . This branch point is logarithmic in nature in that it is infinitely sheeted.

<sup>14</sup>Note that, for  $\rho$  in the complex neighborhood of zero, according to (17.31)  $p$  can take on values in the complex neighborhood of  $-\infty$ . Apparently the singularity of  $Q_{m-1/2}(p)$  in this neighborhood compensates for that in  $(1/\rho)^{1/2}$  so the net result is that for  $\psi^r$  there is analyticity in  $\rho$  in the complex neighborhood of zero.

and the  $C_{m,\alpha}^{[0]}(z)$ . We these observations in mind, we find the results

$$\begin{aligned} \int_{-\infty}^{\infty} dz B_x(x, y, z) &= \sum_{m=0}^{\infty} (m+1)\rho^m \cos(m\phi) \int_{-\infty}^{\infty} dz C_{m+1,c}^{[0]}(z) \\ &\quad + \sum_{m=0}^{\infty} (m+1)\rho^m \sin(m\phi) \int_{-\infty}^{\infty} dz C_{m+1,s}^{[0]}(z) \\ &= \sum_{m'=1}^{\infty} m' \rho^{m'-1} \cos[(m'-1)\phi] \int_{-\infty}^{\infty} dz C_{m',c}^{[0]}(z) \\ &\quad + \sum_{m'=1}^{\infty} m' \rho^{m'-1} \sin[(m'-1)\phi] \int_{-\infty}^{\infty} dz C_{m',s}^{[0]}(z), \end{aligned} \tag{15.17.3}$$

$$\begin{aligned} \int_{-\infty}^{\infty} dz B_y(x, y, z) &= \sum_{m=0}^{\infty} (m+1)\rho^m \cos(m\phi) \int_{-\infty}^{\infty} dz C_{m+1,s}^{[0]}(z) \\ &\quad - \sum_{m=0}^{\infty} (m+1)\rho^m \sin(m\phi) \int_{-\infty}^{\infty} dz C_{m+1,c}^{[0]}(z) \\ &= \sum_{m'=1}^{\infty} m' \rho^{m'-1} \cos[(m'-1)\phi] \int_{-\infty}^{\infty} dz C_{m',s}^{[0]}(z) \\ &\quad - \sum_{m'=1}^{\infty} m' \rho^{m'-1} \sin[(m'-1)\phi] \int_{-\infty}^{\infty} dz C_{m',c}^{[0]}(z). \end{aligned} \tag{15.17.4}$$

Note that (11.3) and (11.4) are consistent with (3.77), (3.78), (3.80), and (3.81).

What are we to conclude from these results? The “multipole” content of a magnet is often specified, in effect, in terms of the *integrated multipole* quantities  $\int_{-\infty}^{\infty} dz C_{m',\alpha}^{[0]}(z)$  for  $m' = 1, 2, \dots$ . This is because magnet measurements are frequently made using spinning coils whose length is such that they extend beyond the ends of the magnets to include the fringe-field regions. (See Appendix K.) Hence, the use of such coils measures  $\int_{-\infty}^{\infty} dz B_x(x, y, z)$  and  $\int_{-\infty}^{\infty} dz B_y(x, y, z)$  which, according to (11.3) and (11.4), is equivalent to measuring the integrated multipoles. Moreover, the size of the integrated multipoles is often taken as a figure of merit for any given magnet.

Is this reasonable? We know that some terms of the form  $\exp(: f_3 :)\exp(: f_4 :)\dots$  in the transfer map can have deleterious effects on the dynamic aperture. Recall, for example, Section 1.2.3 which illustrated the effect of the term  $\exp(: q^3 :)$  in the simplest nonlinear case. We also know that the generators  $f_3, f_4, \dots$  arise from  $H_3, H_4, \dots$  terms in the Hamiltonian. Finally, we know that nonzero on-axis gradients of the form  $C_{3,\alpha}^{[0]}(z), C_{4,\alpha}^{[0]}(z), \dots$  produce nonzero terms of the form  $H_3, H_4, \dots$  in the Hamiltonian.<sup>15</sup> Therefore, if the integrated

<sup>15</sup>However, nonzero on-axis gradients of the form  $C_{3,\alpha}^{[0]}(z), C_{4,\alpha}^{[0]}(z), \dots$  are not the only source of  $H_3, H_4, \dots$  terms in the Hamiltonian. Such terms also arise, for example, from the expansion of the square root in (1.4) and occur even if  $A_x = A_y = 0$ .

multipoles are large for  $m' = 3, 4, \dots$ , we expect that nonlinear terms in the map will be important and the dynamic aperture is likely be small. Consequently, a good rule of thumb would appear to be that the integrated  $m' = 3, 4, \dots$  multipole terms should be small to minimize possibly deleterious nonlinear terms in the transfer map.

But, while minimizing the integrated multipoles would seem to be a possible way of minimizing the nonlinear terms in the transfer map, so doing is *not necessarily sufficient*. Consider, for example, the transfer map for a composite system consisting of two identical back-to back sextuples save that they are oppositely powered. All integrated multipoles for such a system would be exactly zero. However, the transfer map for this system could still have large nonlinear terms including those with  $f_3 \neq 0$ .<sup>16</sup> Observe also that fringe-field terms contribute to  $H_3, H_4, \dots$ , and therefore can produce nonlinear terms in the transfer map.<sup>17</sup> However, as noted earlier, all of the terms in the sums (3.65) and (3.66) integrate to zero save for those that involve  $C_0^{[1]}(z)$  and the  $C_{m,\alpha}^{[0]}(z)$ . Consequently all fringe-field terms integrate to zero, and their presence is therefore undetectable solely from an examination of the values of the integrated multipoles. Note also that  $m = 0$  (solenoid) terms make no contribution to the integrated multipoles. However, it can be shown that solenoid fringe-fields make nonlinear contributions to the transfer map.<sup>18</sup>

We conclude and emphasize that what is needed for a realistic calculation of transfer maps are the functions  $C_0^{[1]}(z)$  and the  $C_{m,\alpha}^{[0]}(z)$  *themselves*, and *not* just their integrals.

### 15.17.2 Need for On-Axis Gradients and the Use of Surface Data

From the work of the previous sections, we have learned that the dynamics of a charged particle passing through a region of space occupied by a magnetic field described by the scalar potential (3.33), or the azimuthal-free vector potential  $\mathbf{A}$  given by (4.21) through (4.26), or the symmetric Coulomb gauge vector potential  $\hat{\mathbf{A}}$  given by (5.89) through (5.94), or their vertical-free and possibly further adjusted variants as illustrated for the normal dipole, are completely determined by a knowledge of the on-axis gradient functions  $C_0^{[1]}(z)$  and  $C_{m,\alpha}^{[0]}(z)$  and their derivatives. In Chapter 16 we will treat cases for which the on-axis gradients can be computed analytically. In Chapters 17 through 21 we will describe several general methods for computing the on-axis gradients and their derivatives numerically based on the use of numerical field data on a *surface*. The surfaces employed will be those of cylinders with circular, elliptical, or rectangular cross sections. These methods are *smoothing*. That is, they have the virtue of being relatively insensitive to errors in the input data. Consequently,

---

<sup>16</sup>There are at least two other instances of this apparently malevolent principle: The first involves superconducting dipoles. They are often equipped with multipole-corrector coil packages at each end. Even if these coils are powered so that the composite system (dipole plus correctors) has net integrated multipole values of zero for the first few  $m'$  values (with  $m' \geq 3$ ), so doing does not guarantee that the net transfer map is free of  $f_n$  terms for the first few values of  $n \geq 3$ . The second concerns room-temperature quadrupoles. Sometimes they are hand-fitted during manufacture with end shims so that the net integrated  $m' = 6$  multipole (which, according to Subsection 3.5, is allowed) is in fact zero. So doing does not guarantee that the net transfer map is free of  $f_6$  terms. For a further discussion of correction methods, see Section 12.11.

<sup>17</sup>For example, dipole fringe-field effects in the hard-edge limit produce an  $f_4$  some of whose terms are infinite.

<sup>18</sup>For example, solenoid fringe-field effects in the hard-edge limit also produce an  $f_4$  some of whose terms are infinite.

they are ideally suited for numerical use.

### 15.17.3 Limitations Imposed by Symmetry and Hamilton and Maxwell

In the introduction to this chapter we noted that there are possible limitations on what transfer maps can be achieved. The first limitation is that the transfer map must be symplectic. The second arises from the fact that, in many instances, the electric and magnetic fields within beam-line elements must arise from fields that satisfy the source-free Maxwell equations. These limitations, combined with symmetry assumptions, may place restrictions upon what can actually be achieved. For example there is a remarkable theorem, due to Scherzer, which states that any imaging system having cylindrical symmetry must have negative spherical aberration. Consequently it is impossible to design, using only electric and magnetic elements with cylindrical symmetry, an electron microscope that is free of spherical aberration. As a practical consequence, the resolution of such microscopes is limited to a few Angstroms. To achieve zero spherical aberration it is necessary to break cylindrical symmetry with the careful use of nonlinear elements such as sextuples or octupoles. This is now done in the highest resolution electron microscopes with the result that it is now possible to achieve resolution at the atomic and subatomic level. For a discussion of Scherzer's theorem, and the possible correction of spherical aberration, see the reference at the end of the bibliography for this chapter.

## Exercises

**15.17.1.** Verify that the integrated transverse fields satisfy the transverse Laplace equation,

$$\nabla_{\perp}^2 \int_{-\infty}^{\infty} dz B_x(x, y, z) = \nabla_{\perp}^2 \int_{-\infty}^{\infty} dz B_y(x, y, z) = 0. \quad (15.17.5)$$

Hint: See Exercise 3.4.



# Bibliography

## The Poincaré-Coulomb Gauge

- [1] W. Brittin, W. Smythe, and W. Wyss, “Poincaré gauge in electrodynamics”, *Am. J. Phys.* **50**, p.693 (1982).
- [2] F. Cornish, “The Poincaré and related gauges in electromagnetic theory”, *Am. J. Phys.* **52**, p.460 (1984).
- [3] J. D. Jackson, “From Lorenz to Coulomb and other explicit gauge transformations”, LBNL-50079 (2002) and *Am. J. Phys.* **70**, 917 (2002). Alternatively, see the Web site <http://arxiv.org/ftp/physics/papers/0204/0204034.pdf>.

## Bessel, Legendre, and Other Special Functions

- [4] G.N. Watson, *A Treatise on the Theory of Bessel Functions*, Cambridge (1922).
- [5] E.T. Whittaker and G.N. Watson, *A Course of Modern Analysis*, Cambridge (1952).
- [6] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover (1972). Also available on the Web by Googling “abramowitz and stegun 1972”.
- [7] 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/>.
- [8] H. Bateman, A. Erdelyi, W. Magnus, F. Oberhettinger, F.G. Tricomi, *Higher Transcendental Functions*, 3 Volumes, McGraw-Hill (1953). Chapter 7 in Volume 2 treats Bessel functions.
- [9] H. Bateman, *Tables of Integral Transforms*, Volumes I, II, and III, McGraw-Hill (1954).
- [10] Chester Snow, *Hypergeometric and Legendre Functions with Applications to Integral Equations of Potential Theory*, Second Edition, U.S. Dept. of Commerce, National Bureau of Standards (now NIST), Applied Math. Series 19, U.S. Government Printing Office, 1952.
- [11] D. D. Šiljac, *Nonlinear Systems*, pp. 384-393, John Wiley (1969).

## Integral Tables

- [12] H. B. Dwight, *Tables of Integrals and Other Numerical Data*, Fourth Edition, Macmillan (1961).

## Electromagnetic Fields

- [13] W. Panofsky and M. Phillips, *Classical Electricity and Magnetism*, Second Edition, Addison-Wesley (1962) and Dover (2005).
- [14] J. D. Jackson, *Classical Electrodynamics*, John Wiley (1999).
- [15] G. Arfken, *Mathematical Methods for Physicists*, Second Edition, Academic Press (1970).
- [16] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, Vols. 1 and 2, McGraw-Hill (1953).
- [17] S. Russenschuck, *Field Computation for Accelerator Magnets: Analytical and Numerical Methods for Electromagnetic Design and Optimization*, Wiley (2010).

## Magnetic Optics and Electron Microscopes

- [18] A. Dragt, “Numerical third-order transfer map for solenoid”, *Nuclear Instruments and Methods in Physics Research A***298**, p. 441 (1990).
- [19] M. Bassetti and C. Biscari, “Analytic Formulae for Magnetic Multipoles”, *Particle Accelerators*, **52**, 221-250 (1996).
- [20] M. Bassetti and C. Biscari, “Cylinder Model of Multipoles”, Section 2.2.2 of *Handbook of Accelerator Physics and Engineering*, A. Chao and M. Tigner, Eds., World Scientific (2002).
- [21] P. W. Hawkes and E. Kasper, *Principles of Electron Optics*, Vols. 1 through 3, Academic Press (1996).
- [22] A. B. El-Kareh and J. C. J. El-Kareh, *Electron Beams, Lenses, and Optics*, Vols. 1 and 2, Academic Press (1970).
- [23] P. A. Sturrock, “The Aberrations of Magnetic Electron Lenses due to Asymmetries”, *Philosophical Transactions of the Royal Society of London, Series A, Mathematical and Physical Sciences*, Vol. 243, No. 868, pp. 387-429 (1951).
- [24] P. W. Hawkes, Edit., *Advances in Imaging and Electron Physics*, Academic Press, Elsevier Inc. This continuing journal merges the long-running journals *Advances in Electronics and Electron Physics* and *Advances in Optical and Electron Microscopy*.

## Scherzer’s Theorem

- [25] A. Dragt and E. Forest, "Lie Algebraic Theory of Charged-Particle Optics and Electron Microscopes", *Advances in Electronics and Electron Physics*, Volume 67, pp. 65-117, P. W. Hawkes, Edit. (1986). See the reference above to this journal.



# Chapter 16

## Realistic Transfer Maps for Straight Iron-Free Beam-Line Elements

Chapter 15 described cylindrical harmonic expansions for straight elements, and showed how the scalar potential, magnetic field, and various vector potentials can be described in terms of on-axis gradients. In most cases these on-axis gradients must be computed numerically. How this may be done in general for straight elements is described in Chapters 17 through 21. However, in some iron-free cases the on-axis gradients can be computed analytically, including their fringe-field behavior. This chapter treats several of these cases. We remark that although our discussion is limited to magnetic beam-line elements, electrostatic beam-line elements can be treated in an analogous way. In principle, the fringe field of any individual beam-line element at either end of the element has *infinite* extent. (This is particularly true of iron-free elements, but is also a consideration even in the case of some iron-dominated elements.) However in practice in many instances we may wish to regard a beam line as a collection of separated/isolated elements. To do this it is necessary to make an approximation in which leading and trailing end fields are “terminated” in some way. The crucial problem is how to relate canonical coordinates in the absence of a magnetic field with canonical coordinates in the presence of a magnetic field. The first part of this chapter is devoted to describing how this problem may be treated in general for straight beam-line elements. The remaining part of the chapter treats various specific straight iron-free beam-line elements.

### 16.1 Terminating End Fields

#### 16.1.1 Preliminary Observations

We begin with some preliminary observations. In Cartesian coordinates the Hamiltonian describing charged-particle motion with  $z$  as the independent variable is given by the relation

$$K = -[(p_t^{\text{can}})^2/c^2 - m^2c^2 - (p_x^{\text{can}} - qA_x)^2 - (p_y^{\text{can}} - qA_y)^2]^{1/2} - qA_z. \quad (16.1.1)$$

Here we have assumed that the electric scalar potential  $\psi$  vanishes and  $\mathbf{A}$  is static so that there is no electric field. Also, we have used the notation  $p_x^{\text{can}}$ ,  $p_y^{\text{can}}$ , and  $p_t^{\text{can}}$  to indicate

that it is the components of the *canonical* momenta that are involved in a Hamiltonian description of motion. See (1.6.16).

According to Hamilton's equations of motion, the change of a coordinate, say  $x(z)$ , with  $z$  is given by

$$\begin{aligned} dx/dz &= \partial K / \partial p_x^{\text{can}} \\ &= (p_x^{\text{can}} - qA_x) / [(p_t^{\text{can}})^2/c^2 - m^2c^2 - (p_x^{\text{can}} - qA_x)^2 - (p_y^{\text{can}} - qA_y)^2]^{1/2} \\ &= (p_x^{\text{can}} - qA_x) / [-(K + qA_z)]. \end{aligned} \quad (16.1.2)$$

Let us verify that this result agrees with what we already know. Recall that

$$K = -p_z^{\text{can}}. \quad (16.1.3)$$

See (1.6.6). It follows that (1.2) can be rewritten in the form

$$dx/dz = (p_x^{\text{can}} - qA_x) / (p_z^{\text{can}} - qA_z). \quad (16.1.4)$$

According to (1.5.27) through (1.5.30) there is the relation

$$\mathbf{p}^{\text{can}} - q\mathbf{A} = \mathbf{p}^{\text{mech}} \quad (16.1.5)$$

where  $\mathbf{p}^{\text{mech}}$  is the *mechanical* momentum given by

$$\mathbf{p}^{\text{mech}} = \gamma m \mathbf{v}. \quad (16.1.6)$$

Consequently, (1.4) can be rewritten in the form

$$dx/dz = p_x^{\text{mech}} / p_z^{\text{mech}} = \gamma m v_x / (\gamma m v_z) = v_x / v_z = \frac{dx/dt}{dz/dt}. \quad (16.1.7)$$

Evidently, the far left and far right sides of (1.7) agree. It is also easy to see that results analogous to those just found also hold for  $y(z)$ .

To complete the story we need to examine also the equation of motion for  $t(z)$ . In this case application of the standard Hamiltonian rules gives the result

$$\begin{aligned} dt/dz &= \partial K / \partial p_t^{\text{can}} \\ &= (-p_t^{\text{can}}/c^2) / [(p_t^{\text{can}})^2/c^2 - m^2c^2 - (p_x^{\text{can}} - qA_x)^2 - (p_y^{\text{can}} - qA_y)^2]^{1/2} \\ &= (-p_t^{\text{can}}/c^2) / [-(K + qA_z)]. \end{aligned} \quad (16.1.8)$$

Now use of (1.3), (1.5), (1.6), and (1.6.17) yields the relation

$$dt/dz = (-p_t^{\text{can}}/c^2) / p_z^{\text{mech}} = \gamma m / (\gamma m v_z) = \frac{1}{dz/dt} \quad (16.1.9)$$

so that the far left and far right sides of (1.9) also agree.

### 16.1.2 Matching Conditions

We now consider what matching conditions should be imposed upon entry and exit of fringe-field regions. To proceed further it is useful to introduce some notation. Let  $z^{\text{en}}$  denote the  $z$  value where a transition is to be made from a region where the magnetic field is *taken* to vanish to the beginning of the leading fringe-field region. That is, any charged particle in question *enters* the leading fringe-field region when  $z = z^{\text{en}}$ . Similarly, let  $z^{\text{ex}}$  denote the  $z$  value where a transition is to be made from the end of a trailing fringe-field region to a region where the magnetic field is again taken to vanish. That is, any charged particle in question *exits* the trailing fringe-field region when  $z = z^{\text{ex}}$ . Our task is to find matching relations at  $z^{\text{en}}$  and  $z^{\text{ex}}$ .

#### 16.1.2.1 Entering a Leading Fringe-Field Region

Let us begin with a consideration of the transition between a field-free region and a leading fringe-field region. Let  $K^{\text{ben}}$  be the Hamiltonian *before entry* into the fringe-field region, and let  $K^{\text{aen}}$  be the Hamiltonian *after entry* into the fringe-field region. Then, since the magnetic field and its associated vector potential are assumed to vanish before entry, we have the relation

$$K^{\text{ben}} = -[(p_t^{\text{canben}})^2/c^2 - m^2c^2 - (p_x^{\text{canben}})^2 - (p_y^{\text{canben}})^2]^{1/2}. \quad (16.1.10)$$

And, since the magnetic field (and therefore also the vector potential) is nonzero after entry, we have the relation

$$K^{\text{aen}} = -[(p_t^{\text{canaen}})^2/c^2 - m^2c^2 - (p_x^{\text{canaen}} - qA_x)^2 - (p_y^{\text{canaen}} - qA_y)^2]^{1/2} - qA_z. \quad (16.1.11)$$

Here we have added the suffixes *ben* and *aen* to the phase-space coordinates to denote their values *before entry* and *after entry*. Our task is to relate these phase-space coordinates.

As a first step, we naturally require that the coordinates be continuous at  $z^{\text{en}}$ ,

$$x^{\text{aen}} = x^{\text{ben}}, \quad (16.1.12)$$

$$y^{\text{aen}} = y^{\text{ben}}, \quad (16.1.13)$$

$$t^{\text{aen}} = t^{\text{ben}}, \quad (16.1.14)$$

when  $z = z^{\text{en}}$ . The next step is to specify what is to be done with the momenta.

One possibility is to require that the slopes/“velocities”  $dx/dz$ ,  $dy/dz$ , and  $dt/dz$  be continuous at  $z^{\text{en}}$ . Let us work out the consequences of such a requirement. Before entry we have the result

$$\begin{aligned} dx/dz &= \partial K^{\text{ben}} / \partial p_x^{\text{canben}} = \\ p_x^{\text{canben}} &/ [(p_t^{\text{canben}})^2/c^2 - m^2c^2 - (p_x^{\text{canben}})^2 - (p_y^{\text{canben}})^2]^{1/2}, \end{aligned} \quad (16.1.15)$$

and after entry there is the result

$$\begin{aligned} dx/dz &= \partial K^{\text{aen}} / \partial p_x^{\text{canaen}} = \\ (p_x^{\text{canaen}} - qA_x) &/ [(p_t^{\text{canaen}})^2/c^2 - m^2c^2 - (p_x^{\text{canaen}} - qA_x)^2 - (p_y^{\text{canaen}} - qA_y)^2]^{1/2}. \end{aligned} \quad (16.1.16)$$

See (1.2). An analogous result holds for  $dy/dz$ . Finally, for  $dt/dz$  there is the before entry result

$$dt/dz = \partial K^{\text{ben}} / \partial p_t^{\text{canben}} = \\ (-p_t^{\text{canben}}/c^2)/[(p_t^{\text{canben}})^2/c^2 - m^2c^2 - (p_x^{\text{canben}})^2 - (p_y^{\text{canben}})^2]^{1/2}, \quad (16.1.17)$$

and the after entry result

$$dt/dz = \partial K^{\text{aen}} / \partial p_t^{\text{canaen}} = \\ (-p_t^{\text{canaen}}/c^2)/[(p_t^{\text{canaen}})^2/c^2 - m^2c^2 - (p_x^{\text{canaen}} - qA_x)^2 - (p_y^{\text{canaen}} - qA_y)^2]^{1/2}. \quad (16.1.18)$$

See (1.8). Now equate the far right sides of (1.15) and (1.16), the far right sides of their  $dy/dz$  counterparts, and the far right sides of (1.17) and (1.18). So doing yields the transition matching relations

$$p_x^{\text{canaen}} - qA_x = p_x^{\text{canben}}, \quad (16.1.19)$$

$$p_y^{\text{canaen}} - qA_y = p_y^{\text{canben}}, \quad (16.1.20)$$

$$p_t^{\text{canaen}} = p_t^{\text{canben}}. \quad (16.1.21)$$

In view of (1.5) the relations (1.19) and (1.20) can also be written in the form

$$p_x^{\text{mechaen}} = p_x^{\text{mechben}}, \quad (16.1.22)$$

$$p_y^{\text{mechaen}} = p_y^{\text{mechben}}. \quad (16.1.23)$$

Moreover, under our assumption that the electrical potential  $\psi = 0$ , (1.21) and (1.6.17) yield the relation

$$\mathbf{p}^{\text{mechaen}} \cdot \mathbf{p}^{\text{mechaen}} = \mathbf{p}^{\text{mechben}} \cdot \mathbf{p}^{\text{mechben}}. \quad (16.1.24)$$

This relation, when combined with (1.22) and (1.23), yields the further result

$$p_z^{\text{mechaen}} = p_z^{\text{mechben}}. \quad (16.1.25)$$

We conclude that imposition of the requirement that the slopes/“velocities” be continuous entails that the mechanical momenta be continuous.

The relation (1.21) is satisfactory because we would hope that the energy would not change upon entry into the leading fringe-field region. Again recall (1.6.17). However, we also desire that the phase-space transition matching relations be a symplectic transformation. Calculation shows that the transformation given by (1.12) through (1.14) and (1.19) through (1.21) is *not* symplectic. Compute the Poisson bracket of the left sides of (1.19) and (1.20) to find the result

$$[p_x^{\text{canaen}} - qA_x, p_y^{\text{canaen}} - qA_y] = [p_x^{\text{canaen}}, -qA_y] + [-qA_x, p_y^{\text{canaen}}] \\ = q\{\partial A_y/\partial x^{\text{aen}} - \partial A_x/\partial y^{\text{aen}}\} = qB_z. \quad (16.1.26)$$

[Recall (1.7.40).] While hopefully small, generally  $B_z(x, y, z^{\text{en}})$  differs from zero at the beginning of the leading fringe-field region. On the other hand, the Poisson bracket of

the right sides of (1.19) and (1.20) must vanish since  $p_x^{\text{canben}}$  and  $p_y^{\text{canben}}$  are supposed to be canonical momenta. Therefore the phase-space transformation given by (1.12) through (1.14) and (1.19) through (1.21) is generally *not* symplectic. Review, at this point, Exercise 6.4.11.

We expect that neglect of the magnetic field in the region  $z < z^{\text{en}}$  will lead to some error in trajectories. However, we do not want this error to violate the symplectic condition. The simplest way to maintain the symplectic condition is to retain the relations (1.12) through (1.14) and replace the relations (1.19) through (1.21) by the relations

$$p_x^{\text{canaen}} = p_x^{\text{canben}}, \quad (16.1.27)$$

$$p_y^{\text{canaen}} = p_y^{\text{canben}}, \quad (16.1.28)$$

$$p_t^{\text{canaen}} = p_t^{\text{canben}}. \quad (16.1.29)$$

In this case the transition matching relations (1.12) through (1.14) and (1.27) through (1.29) amount to the identity map  $\mathcal{I}$ , and the symplectic condition is trivially satisfied. Now, however, the error in trajectories manifests itself in that the slopes/“velocities”  $dx/dz$ ,  $dy/dz$ , and  $dt/dz$  may be expected to be discontinuous at  $z^{\text{en}}$ . Inspection of (1.15) and (1.16), their  $dy/dz$  counterparts, and (1.17) and (1.18) shows that, in lowest approximation, these discontinuities are proportional to components of  $\mathbf{A}(x, y, z^{\text{en}})$ . Indeed, again in view of (1.3) and (1.5), the transition relations (1.27) and (1.28) can be written in the form

$$\Delta p_x^{\text{mech}} = p_x^{\text{mechaen}} - p_x^{\text{mechben}} = qA_x(x, y, z^{\text{en}}), \quad (16.1.30)$$

$$\Delta p_y^{\text{mech}} = p_y^{\text{mechaen}} - p_y^{\text{mechben}} = qA_y(x, y, z^{\text{en}}). \quad (16.1.31)$$

Also, in view of (1.29), (1.24) continues to hold. Therefore, upon combining (1.29) through (1.31), we see that  $p_z^{\text{mechaen}}$  is given by the relation

$$\begin{aligned} p_z^{\text{mechaen}} &= [(p_z^{\text{mechben}})^2 + (p_x^{\text{mechben}})^2 - (p_x^{\text{mechaen}})^2 + (p_y^{\text{mechben}})^2 - (p_y^{\text{mechaen}})^2]^{1/2} \\ &= [(p_z^{\text{mechben}})^2 - (\Delta p_x^{\text{mech}})(\Sigma p_x^{\text{mech}}) - (\Delta p_y^{\text{mech}})(\Sigma p_y^{\text{mech}})]^{1/2} \\ &= [(p_z^{\text{mechben}})^2 - qA_x(x, y, z^{\text{en}})(\Sigma p_x^{\text{mech}}) - qA_y(x, y, z^{\text{en}})(\Sigma p_y^{\text{mech}})]^{1/2} \end{aligned} \quad (16.1.32)$$

where

$$\Sigma p_x^{\text{mech}} = p_x^{\text{mechaen}} + p_x^{\text{mechben}} = 2p_x^{\text{mechben}} + qA_x(x, y, z^{\text{en}}), \quad (16.1.33)$$

$$\Sigma p_y^{\text{mech}} = p_y^{\text{mechaen}} + p_y^{\text{mechben}} = 2p_y^{\text{mechben}} + qA_y(x, y, z^{\text{en}}). \quad (16.1.34)$$

We conclude that imposition of continuity in the canonical momenta as expressed by (1.27) through (1.29) entails a discontinuity in the mechanical momenta, and this discontinuity depends on the size of  $\mathbf{A}(x, y, z^{\text{en}})$ . It is therefore desirable to work in a gauge where  $\mathbf{A}(x, y, z^{\text{en}})$  is as *small* as possible.<sup>1</sup> Subsequently, we will explore the use of the minimum vector potential of Subsection 15.2.5 for this purpose.

---

<sup>1</sup>To see that the choice of gauge can affect the size of nonlinear terms, look at Exercise 15.5.7. There it is observed, as an example, that for a normal quadrupole the vector potentials in the azimuthal-free and symmetric Coulomb gauges are different in the fringe-field regions. In particular, the nonlinearities are smaller in the symmetric Coulomb gauge.

### 16.1.2.2 Exiting a Trailing Fringe-Field Region

The transition between a trailing fringe-field region and a subsequent field-free region may be considered in an analogous way. We again require continuity in the coordinates and canonical momenta. As described earlier, let  $z^{\text{ex}}$  denote the  $z$  value where a transition is to be made from the end of a trailing fringe-field region to a region where the magnetic field is again taken to vanish. That is, any charged particle in question *exits* the trailing fringe-field region when  $z = z^{\text{ex}}$ . We will also add the suffixes *aex* and *bex* to the phase-space coordinates to denote their values *after* and *before exit*. In terms of this notation we impose the matching conditions

$$x^{\text{aex}} = x^{\text{bex}}, \quad (16.1.35)$$

$$y^{\text{aex}} = y^{\text{bex}}, \quad (16.1.36)$$

$$t^{\text{aex}} = t^{\text{bex}}, \quad (16.1.37)$$

$$p_x^{\text{canaaex}} = p_x^{\text{canbex}}, \quad (16.1.38)$$

$$p_y^{\text{canaaex}} = p_y^{\text{canbex}}, \quad (16.1.39)$$

$$p_t^{\text{canaaex}} = p_t^{\text{canbex}} \quad (16.1.40)$$

when  $z = z^{\text{ex}}$ . So so doing entails discontinuities in the mechanical momenta given by the relations

$$\Delta p_x^{\text{mech}} = p_x^{\text{mechaex}} - p_x^{\text{mechbex}} = qA_x(x, y, z^{\text{ex}}), \quad (16.1.41)$$

$$\Delta p_y^{\text{mech}} = p_y^{\text{mechaex}} - p_y^{\text{mechbex}} = qA_y(x, y, z^{\text{ex}}), \quad (16.1.42)$$

$$\begin{aligned} p_z^{\text{mechaex}} &= [(p_z^{\text{mechbex}})^2 + (p_x^{\text{mechbex}})^2 - (p_x^{\text{mechaex}})^2 + (p_y^{\text{mechbex}})^2 - (p_y^{\text{mechaex}})^2]^{1/2} \\ &= [(p_z^{\text{mechbex}})^2 - (\Delta p_x^{\text{mech}})(\Sigma p_x^{\text{mech}}) - (\Delta p_y^{\text{mech}})(\Sigma p_y^{\text{mech}})]^{1/2} \\ &= [(p_z^{\text{mechbex}})^2 - qA_x(x, y, z^{\text{ex}})(\Sigma p_x^{\text{mech}}) - qA_y(x, y, z^{\text{ex}})(\Sigma p_y^{\text{mech}})]^{1/2} \end{aligned} \quad (16.1.43)$$

where

$$\Sigma p_x^{\text{mech}} = p_x^{\text{mechaex}} + p_x^{\text{mechbex}} = 2p_x^{\text{mechbex}} + qA_x(x, y, z^{\text{ex}}), \quad (16.1.44)$$

$$\Sigma p_y^{\text{mech}} = p_y^{\text{mechaex}} + p_y^{\text{mechbex}} = 2p_y^{\text{mechbex}} + qA_y(x, y, z^{\text{ex}}). \quad (16.1.45)$$

That is, imposition of continuity in the canonical momenta as expressed by (1.38) through (1.40) again entails discontinuities in the associated mechanical momenta. It is therefore also desirable to work in a gauge where  $\mathbf{A}(x, y, z^{\text{ex}})$  is as small as possible.

### 16.1.2.3 Modified Hamiltonian, Vector Potential, Magnetic Field, and Current

One way to view the symplectic matching relations (1.12) through (1.14), (1.27) through (1.29), and (1.35) through (1.40) is to replace the Hamiltonian (1.1) by a modified Hamiltonian  $K^{\text{mod}}$  given by

$$K^{\text{mod}} = -[(p_t^{\text{can}})^2/c^2 - m^2c^2 - (p_x^{\text{can}} - qA_x^{\text{mod}})^2 - (p_y^{\text{can}} - qA_y^{\text{mod}})^2]^{1/2} - qA_z^{\text{mod}} \quad (16.1.46)$$

where

$$\mathbf{A}^{\text{mod}}(x, y, z) = \theta(z - z^{\text{en}})\theta(z^{\text{ex}} - z)\mathbf{A}(x, y, z). \quad (16.1.47)$$

That is, the vector potential is taken to vanish for  $z < z^{\text{en}}$ , turns on at  $z = z^{\text{en}}$ , and again turns off for  $z > z^{\text{ex}}$ . A little thought shows that integrating the equations of motion associated with this modified Hamiltonian automatically produces the matching relations (1.12) through (1.14), (1.27) through (1.29), and (1.35) through (1.40).

What is the modified magnetic field  $\mathbf{B}^{\text{mod}}$  associated with this modified vector potential? Evaluation of

$$\mathbf{B}^{\text{mod}} = \nabla \times \mathbf{A}^{\text{mod}} \quad (16.1.48)$$

gives the relations

$$\begin{aligned} B_x^{\text{mod}}(x, y, z) &= \partial_y A_z^{\text{mod}} - \partial_z A_y^{\text{mod}} \\ &= \theta(z - z^{\text{en}})\theta(z^{\text{ex}} - z)B_x(x, y, z) \\ &\quad - [\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]A_y(x, y, z), \end{aligned} \quad (16.1.49)$$

$$\begin{aligned} B_y^{\text{mod}}(x, y, z) &= \partial_z A_x^{\text{mod}} - \partial_x A_z^{\text{mod}} \\ &= \theta(z - z^{\text{en}})\theta(z^{\text{ex}} - z)B_y(x, y, z) \\ &\quad + [\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]A_x(x, y, z), \end{aligned} \quad (16.1.50)$$

$$\begin{aligned} B_z^{\text{mod}}(x, y, z) &= \partial_x A_y^{\text{mod}} - \partial_y A_x^{\text{mod}} \\ &= \theta(z - z^{\text{en}})\theta(z^{\text{ex}} - z)B_z(x, y, z). \end{aligned} \quad (16.1.51)$$

By the construction (1.48), the modified magnetic field is divergence free,

$$\nabla \cdot \mathbf{B}^{\text{mod}} = 0, \quad (16.1.52)$$

as required. [Note that making the simple Ansatz  $\mathbf{B}^{\text{mod}} = \theta(z - z^{\text{en}})\theta(z^{\text{ex}} - z)\mathbf{B}$  violates the requirement (1.52). It is this Ansatz that would arise naturally if one were integrating the non-canonical Lorentz-force equations given in Exercise 1.6.16.]

What current  $\mathbf{j}^{\text{mod}}$  produces this modified magnetic field? It is specified by employing  $\mathbf{B}^{\text{mod}}$  as given by (1.49) through (1.51) in the relation

$$\mu_0 \mathbf{j}^{\text{mod}} = \nabla \times \mathbf{B}^{\text{mod}}. \quad (16.1.53)$$

Doing so directly leads to considerable algebra, which can be bypassed with the use of suitable vector identities. Proceed as follows: Combining (1.48) and (1.53) gives the relation

$$\mu_0 \mathbf{j}^{\text{mod}} = \nabla \times (\nabla \times \mathbf{A}^{\text{mod}}) = \nabla(\nabla \cdot \mathbf{A}^{\text{mod}}) - \nabla^2 \mathbf{A}^{\text{mod}}. \quad (16.1.54)$$

Let us work on the first term on the right side of (1.54). From the definition (1.47) there is the result

$$\nabla \cdot \mathbf{A}^{\text{mod}} = \mathbf{A} \cdot \nabla[\theta(z - z^{\text{en}})\theta(z^{\text{ex}} - z)] + \theta(z - z^{\text{en}})\theta(z^{\text{ex}} - z)\nabla \cdot \mathbf{A}. \quad (16.1.55)$$

Evaluation of the first term on the right in (1.55) gives the result

$$\begin{aligned}\mathbf{A} \cdot \nabla[\theta(z - z^{\text{en}})\theta(z^{\text{ex}} - z)] &= A_z \partial_z[\theta(z - z^{\text{en}})\theta(z^{\text{ex}} - z)] \\ &= [\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]A_z.\end{aligned}\quad (16.1.56)$$

We also assume that  $\mathbf{A}$  is in a Coulomb gauge,  $\nabla \cdot \mathbf{A} = 0$ , so that (1.55) becomes

$$\nabla \cdot \mathbf{A}^{\text{mod}} = [\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]A_z,\quad (16.1.57)$$

and therefore

$$\begin{aligned}\nabla(\nabla \cdot \mathbf{A}^{\text{mod}}) &= e_x[\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]\partial_x A_z \\ &\quad + e_y[\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]\partial_y A_z \\ &\quad + e_z[\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]\partial_z A_z \\ &\quad + e_z[\delta'(z - z^{\text{en}}) + \delta'(z^{\text{ex}} - z)]A_z.\end{aligned}\quad (16.1.58)$$

Next we turn to working out  $-\nabla^2 \mathbf{A}^{\text{mod}}$ , the second term on the right side of (1.54). For the  $x$  component we have the intermediate result

$$\begin{aligned}-\nabla^2 A_x^{\text{mod}} &= -\nabla^2[\theta(z - z^{\text{en}})\theta(z^{\text{ex}} - z)A_x] \\ &= -\theta(z - z^{\text{en}})\theta(z^{\text{ex}} - z)(\partial_x^2 + \partial_y^2)A_x - \partial_z^2[\theta(z - z^{\text{en}})\theta(z^{\text{ex}} - z)A_x].\end{aligned}\quad (16.1.59)$$

By the product rule there is the relation

$$\begin{aligned}\partial_z[\theta(z - z^{\text{en}})\theta(z^{\text{ex}} - z)A_x] &= [\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]A_x \\ &\quad + \theta(z - z^{\text{en}})\theta(z^{\text{ex}} - z)\partial_z A_x,\end{aligned}\quad (16.1.60)$$

from which it follows that

$$\begin{aligned}\partial_z^2[\theta(z - z^{\text{en}})\theta(z^{\text{ex}} - z)A_x] &= [\delta'(z - z^{\text{en}}) + \delta'(z^{\text{ex}} - z)]A_x \\ &\quad + 2[\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]\partial_z A_x \\ &\quad + \theta(z - z^{\text{en}})\theta(z^{\text{ex}} - z)\partial_z^2 A_x.\end{aligned}\quad (16.1.61)$$

Combining (1.59) and (1.61) yields the next intermediate result

$$-\nabla^2 A_x^{\text{mod}} = -2[\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]\partial_z A_x - [\delta'(z - z^{\text{en}}) + \delta'(z^{\text{ex}} - z)]A_x,\quad (16.1.62)$$

Here we have used the fact that  $\mathbf{A}$  is harmonic. See (15.5.4). Similarly, there are the analogous next intermediate results

$$-\nabla^2 A_y^{\text{mod}} = -2[\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]\partial_z A_y - [\delta'(z - z^{\text{en}}) + \delta'(z^{\text{ex}} - z)]A_y,\quad (16.1.63)$$

$$-\nabla^2 A_z^{\text{mod}} = -2[\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]\partial_z A_z - [\delta'(z - z^{\text{en}}) + \delta'(z^{\text{ex}} - z)]A_z.\quad (16.1.64)$$

We are now able to combine the two terms on the right side of (1.54), using (1.58) and (1.62) through (1.64), to yield the desired final results

$$\begin{aligned}\mu_0 j_x^{\text{mod}} &= [\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)][\partial_x A_z - 2\partial_z A_x] \\ &\quad - [\delta'(z - z^{\text{en}}) + \delta'(z^{\text{ex}} - z)]A_x,\end{aligned}\tag{16.1.65}$$

$$\begin{aligned}\mu_0 j_y^{\text{mod}} &= [\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)][\partial_y A_z - 2\partial_z A_y] \\ &\quad - [\delta'(z - z^{\text{en}}) + \delta'(z^{\text{ex}} - z)]A_y,\end{aligned}\tag{16.1.66}$$

$$\mu_0 j_z^{\text{mod}} = -[\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]\partial_z A_z.\tag{16.1.67}$$

Evidently requiring the vector potential to vanish for  $z < z^{\text{en}}$ , turn on at  $z = z^{\text{en}}$ , and again turn off for  $z > z^{\text{ex}}$  is equivalent to introducing sheet (corresponding to the  $\delta$  functions) and double-sheet (corresponding to the  $\delta'$  functions) currents at  $z = z^{\text{en}}$  and  $z = z^{\text{ex}}$ . And the strengths of these currents are proportional to the values of  $\mathbf{A}$  and its first derivatives at  $z = z^{\text{en}}$  and  $z = z^{\text{ex}}$ .

### 16.1.3 Changing Gauge

It may be useful to change gauges at various points during the course of integrating a trajectory and computing an associated transfer map. For example, to minimize end-field termination effects, it is desirable to change to minimum vector potentials at  $z = z^{\text{en}}$  and  $z = z^{\text{ex}}$ . Suppose the gauge is to be *changed* at the point  $z = z^c$ . Let  $x^b$ ,  $y^b$ , and  $t^b$  denote coordinate functions *before* the change, and let  $x^a$ ,  $y^a$ , and  $t^a$  denote coordinate functions *after* the change. Also, let  $\mathbf{A}^b(x^b, y^b; z)$  and  $\mathbf{A}^a(x^a, y^a; z)$  be the vector potentials before ( $z < z^c$ ) and after ( $z > z^c$ ) the change point  $z^c$ . Finally, let  $p_x^{\text{canb}}$ ,  $p_y^{\text{canb}}$ ,  $p_t^{\text{canb}}$  be the canonical momentum functions before the change, and let  $p_x^{\text{cana}}$ ,  $p_y^{\text{cana}}$ ,  $p_t^{\text{cana}}$  be the canonical momentum functions after the change. In terms of these quantities, the before and after Hamiltonians  $K^b$  and  $K^a$  are given by the relations

$$K^b = -[(p_t^{\text{canb}})^2/c^2 - m^2 c^2 - (p_x^{\text{canb}} - qA_x^b)^2 - (p_y^{\text{canb}} - qA_y^b)^2]^{1/2} - qA_z^b \text{ for } z < z^c,\tag{16.1.68}$$

$$K^a = -[(p_t^{\text{cana}})^2/c^2 - m^2 c^2 - (p_x^{\text{cana}} - qA_x^a)^2 - (p_y^{\text{cana}} - qA_y^a)^2]^{1/2} - qA_z^a \text{ for } z > z^c.\tag{16.1.69}$$

What should be the matching relations between the phase-space quantities before and after? Since the choice of gauge should have no physical effect, there is the immediate requirement that the coordinate functions be continuous:

$$\begin{aligned}x^a(z) &= x^b(z) \text{ when } z = z^c, \\ y^a(z) &= y^b(z) \text{ when } z = z^c, \\ t^a(z) &= t^b(z) \text{ when } z = z^c.\end{aligned}\tag{16.1.70}$$

For the same reason, we require that the velocities, and hence the mechanical momenta, be continuous. From (1.5) and (1.6) we see that this requirement is equivalent to the relations

$$\mathbf{p}^{\text{cana}} - q\mathbf{A}^a = \mathbf{p}^{\text{canb}} - q\mathbf{A}^b \text{ when } z = z^c.\tag{16.1.71}$$

In terms of components, (1.71) yields the matching relations

$$\begin{aligned} p_x^{\text{cana}} &= p_x^{\text{canb}} + q(A_x^a - A_x^b) \text{ when } z = z^c, \\ p_y^{\text{cana}} &= p_y^{\text{canb}} + q(A_y^a - A_y^b) \text{ when } z = z^c. \end{aligned} \quad (16.1.72)$$

Finally, the total energy cannot change under a gauge transformation and therefore, since we have assumed that the electric scalar potential  $\psi$  vanishes, there is the matching relation

$$p_t^{\text{cana}} = p_t^{\text{canb}} \text{ when } z = z^c. \quad (16.1.73)$$

We note that this relation also follows from (1.6.17).

We assume there is some common overlap region where both  $\mathbf{A}^b$  and  $\mathbf{A}^a$  are defined. Since they both give rise to the same magnetic field, there is the relation

$$\nabla \times (\mathbf{A}^a - \mathbf{A}^b) = 0. \quad (16.1.74)$$

It follows that there is a function  $\chi$  such that

$$\mathbf{A}^a - \mathbf{A}^b = \nabla \chi. \quad (16.1.75)$$

Consequently, the relations (1.72) can be rewritten in the form

$$\begin{aligned} p_x^{\text{cana}} &= p_x^{\text{canb}} + q(\partial/\partial x)\chi \text{ when } z = z^c, \\ p_y^{\text{cana}} &= p_y^{\text{canb}} + q(\partial/\partial y)\chi \text{ when } z = z^c. \end{aligned} \quad (16.1.76)$$

There is one last step. Let  $\mathcal{T}^c$  be the symplectic *transformation* map defined by the relation

$$\mathcal{T}^c = \exp(q : \chi :). \quad (16.1.77)$$

With aid of this map it is easily verified that the relations (1.70), (1.72), and (1.73) can be rewritten in the form

$$\begin{aligned} x^a(z) &= \exp(q : \chi :)x^b(z) \text{ with } z = z^c, \\ y^a(z) &= \exp(q : \chi :)y^b(z) \text{ with } z = z^c, \\ t^a(z) &= \exp(q : \chi :)t^b(z) \text{ with } z = z^c; \end{aligned} \quad (16.1.78)$$

$$\begin{aligned} p_x^{\text{cana}}(z) &= \exp(q : \chi :)p_x^{\text{canb}}(z) \text{ with } z = z^c, \\ p_y^{\text{cana}}(z) &= \exp(q : \chi :)p_y^{\text{canb}}(z) \text{ with } z = z^c, \\ p_t^{\text{cana}}(z) &= \exp(q : \chi :)p_t^{\text{canb}}(z) \text{ with } z = z^c. \end{aligned} \quad (16.1.79)$$

We have determined that a change in gauge amounts to making a symplectic transformation. Review Exercises 6.2.8 and 6.5.3.

### 16.1.4 Application to Fringe-Field Termination

GENMAP equations may be integrated in any convenient gauge. Change from the Poincaré-Coulomb gauge to the convenient gauge upon entry, and change back from the convenient gauge to the Poincaré-Coulomb gauge upon exit. Concatenate the maps associated with these gauge transformations with the map produced by GENMAP to yield the net map for the “terminated” beam-line element. (Subsequently we will find that, when the design orbit is straight, these gauge-transformation maps are near identity maps.) Note that, in addition to minimizing termination error, this procedure produces a map that does not depend on the choice of convenient gauge. Put another way we have, in effect, specified/*fixed* the effective gauge in a way that minimizes termination error.

Use of the Poincaré-Coulomb gauge minimizes velocity discontinuities associated with fringe-field termination, and correspondingly, fictitious sheet currents. Moreover, in the case where the design orbit is a straight line (taken to be the  $z$  axis), terminating fringe fields in the Poincaré-Coulomb gauge does not affect the design orbit itself since then the vector potential vanishes at the termination point on the  $z$  axis. See (15.2.73). Thus, the design orbit is like the ray along the optical axis in geometrical light-ray optics.

In summary, the desire to maintain the symplectic condition and honor the Maxwell equations in a way that minimizes discontinuities leads to the prescription for fringe-field termination just described when the design orbit is straight. It is an approximation, but it is the best one can hope to do if one is to maintain the fiction of separate elements.

## Exercises

**16.1.1.** Show that integrating the equations of motion associated with the modified Hamiltonian given by (1.46) and (1.47) automatically produces the matching relations (1.12) through (1.14), (1.27) through (1.29), and (1.35) through (1.40).

## 16.2 Solenoids

The remainder of this chapter is devoted to the treatment of various specific straight beam-line elements for which the on-axis gradients can be found analytically. We begin with the case of solenoids.<sup>2</sup>

---

<sup>2</sup>The term *solenoid* for a long cylindrical winding was coined in 1823 by André-Marie Ampère (1775-1836) based on the French word *solénoïd* (based in turn on the Greek word  $\sigma\omega\lambda\eta\nu\omega\varepsilon\iota\eta\varsigma$ ) which means “like a pipe”. Indeed, the magnetic field lines inside a solenoid do resemble the flow lines of water in a pipe. For his 1827 magnum opus ‘Memoir on the Mathematical Theory of Electrodynamic Phenomena Deduced Solely from Experiment’ he was awarded the title “the Newton of electricity”. Ampère’s name is one of the 72 names inscribed on the Eiffel Tower. Tragically, Ampère’s father did not live to see his son’s great achievements because, while Ampère was a teenager, in 1793 his father was guillotined during the French revolution.

### 16.2.1 Preliminaries

A solenoid is a straight beam-line element whose field is described by a cylindrical harmonic expansion that contains (ideally) only an  $m = 0$  term. Figure 2.1 illustrates a Cartesian coordinate system for the treatment of a solenoid. We recall from Section 15.3.3 that in this case the magnetic scalar potential  $\psi$  has the expansion

$$\begin{aligned}\psi(x, y, z) &= \psi_0(x, y, z) = \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! \ell!} C_0^{[2\ell]}(z) \rho^{2\ell} \\ &= C_0^{[0]}(z) - (1/4)(x^2 + y^2) C_0^{[2]}(z) + \dots\end{aligned}\quad (16.2.1)$$

with

$$\rho^2 = x^2 + y^2. \quad (16.2.2)$$

See (15.3.53) and (15.5.5). Correspondingly, the associated magnetic field has the expansion

$$B_x = \partial_x \psi_0 = -(1/2)x C_0^{[2]}(z) + \dots, \quad (16.2.3)$$

$$B_y = \partial_y \psi_0 = -(1/2)y C_0^{[2]}(z) + \dots, \quad (16.2.4)$$

$$\begin{aligned}B_z &= \partial_z \psi_0 = \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! \ell!} C_0^{[2\ell+1]}(z) \rho^{2\ell} \\ &= C_0^{[1]}(z) - (1/4)(x^2 + y^2) C_0^{[3]}(z) + \dots.\end{aligned}\quad (16.2.5)$$

In particular, there is the result

$$\mathbf{B}(0, 0, z) = C_0^{[1]}(z) \mathbf{e}_z. \quad (16.2.6)$$

Also, according to Section 15.5.1, there is a suitable associated vector potential  $\hat{\mathbf{A}}^0$  (in the symmetric Coulomb gauge which, in the case of a solenoid, is also the Poincaré-Coulomb gauge) given by the relation

$$\hat{A}_x^0 = -yU, \quad (16.2.7)$$

$$\hat{A}_y^0 = xU, \quad (16.2.8)$$

$$\hat{A}_z^0 = 0, \quad (16.2.9)$$

where  $U$  is defined to be

$$V(\rho, z) = (1/2) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! (\ell + 1)!} C_0^{[2\ell+1]}(z) \rho^{2\ell}. \quad (16.2.10)$$

Correspondingly, the vector potential will have an expansion in  $x$  and  $y$  of the form

$$\hat{A}_x^0 = -yU = -y(1/2)[C_0^{[1]}(z) - (1/8)C_0^{[3]}(z)(x^2 + y^2) + \dots], \quad (16.2.11)$$

$$\hat{A}_y^0 = xU = x(1/2)[C_0^{[1]}(z) - (1/8)C_0^{[3]}(z)(x^2 + y^2) + \dots], \quad (16.2.12)$$

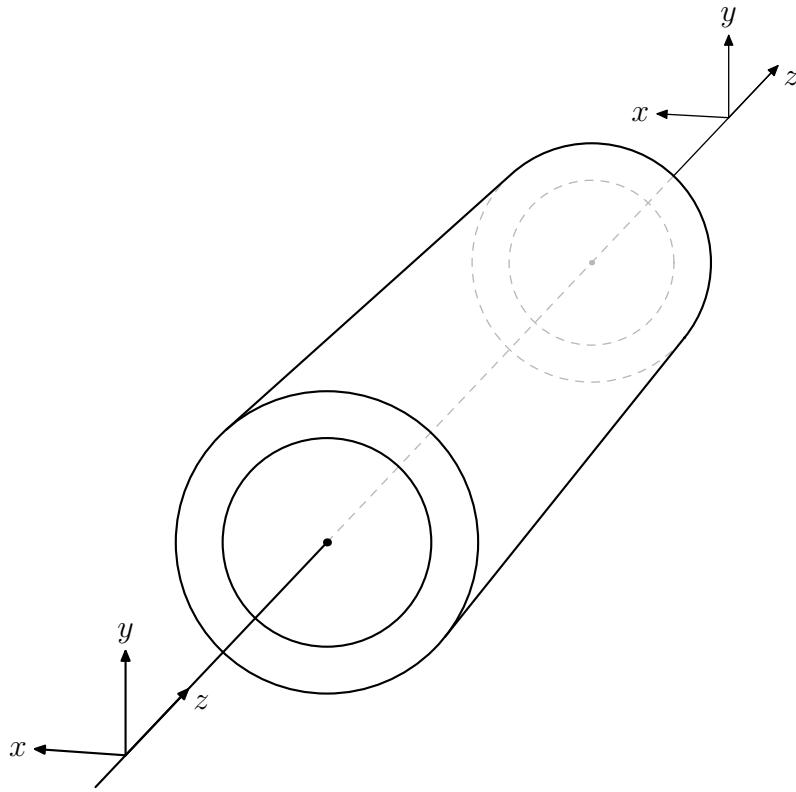


Figure 16.2.1: Coordinate system for a solenoid.

$$\hat{\mathbf{A}}_z^0 = 0, \quad (16.2.13)$$

which can be written in the vector form

$$\hat{\mathbf{A}}^0(\mathbf{r}) = -(1/2)\mathbf{r} \times \mathbf{B}(0, 0, z) + \text{higher order terms}. \quad (16.2.14)$$

Note the resemblance between (2.14) and (15.2.61). This resemblance should not surprise us because we know from Section 15.7.1 that  $\hat{\mathbf{A}}^0$  is in the Poincaré-Coulomb gauge with respect to any origin on the  $z$  axis.

From (2.3) through (2.5) and (2.7) through (2.13) we see that both  $\mathbf{B}$  and  $\hat{\mathbf{A}}^0$  are completely specified in terms of a single “master” function  $C_0^{[1]}(z)$  and its derivatives. Moreover, according to (2.6), the function  $C_0^{[1]}(z)$  is given in terms of the longitudinal on-axis field by the relation

$$C_0^{[1]}(z) = B_z(0, 0, z). \quad (16.2.15)$$

We observe that for a long uniform solenoid the on-axis field  $B_z(0, 0, z)$  will be nearly constant in the body of the solenoid, and therefore the quantities  $C_0^{[n]}(z)$  will be small in this region for  $n > 1$ . However, these derivatives may be large in fringe-field regions. We will next see what can be said more specifically about the master function  $C_0^{[1]}(z)$  in various cases.

## 16.2.2 Simple Air-Core Solenoid

### 16.2.2.1 General Properties of the On-Axis Field

For simplicity, consider initially the case of a simple air-core solenoid consisting of a single-layer circular cylindrical uniform winding of length  $L$  and radius  $\rho = a$ , and powered so that the interior field (in the infinite-length limit) is  $B$ . For such a solenoid it can be shown that the on-axis field is given by the relation

$$B_z(0, 0, z) = B \{z/[z^2 + a^2]^{1/2} - (z - L)/[(z - L)^2 + a^2]^{1/2}\}/2 \quad (16.2.16)$$

where the cylinder axis is the  $z$  axis and the winding extends from  $z = 0$  to  $z = L$ . The on-axis fields of more general air-core solenoids can be found from (2.16) by superposition. See Subsection 2.4. Here we assume that the effect of a solenoidal winding is well approximated by a uniform current sheet (or a collection of uniform current sheets) in the  $e_\phi$  direction. In actuality, a single layer solenoidal winding is a helix. In assuming a uniform current sheet, we ignore helical effects. For a discussion of helical effects, see the book by W. Smythe cited in the references at the end of this chapter.

Suppose we define a soft-edge “bump” function  $\text{bump}(z, a, L)$  by the rule

$$\text{bump}(z, a, L) = \{z/[z^2 + a^2]^{1/2} - (z - L)/[(z - L)^2 + a^2]^{1/2}\}/2 \quad (16.2.17)$$

so that (2.16) can be written in the form

$$B_z(0, 0, z) = B \text{ bump}(z, a, L). \quad (16.2.18)$$

Then there is also the result

$$C_0^{[1]}(z) = B_z(0, 0, z) = B \text{ bump}(z, a, L). \quad (16.2.19)$$

It can be verified that the soft-edge bump function has the properties

$$\text{bump}(z, a, L) \simeq 1 \text{ for } z \in [0, L], \quad (16.2.20)$$

$$\text{bump}(z, a, L) \simeq 0 \text{ elsewhere,} \quad (16.2.21)$$

$$\text{bump}(L/2 + w, a, L) = \text{bump}(L/2 - w, a, L), \quad (16.2.22)$$

It can also be shown that

$$\int_{-\infty}^{\infty} \text{bump}(z, a, L) dz = L. \quad (16.2.23)$$

See Exercise 2.3. In particular, from the results above, it follows that for a simple air-core solenoid there is the relation

$$\int_{-\infty}^{\infty} C_0^{[1]}(z) dz = \int_{-\infty}^{\infty} B_z(0, 0, z) dz = BL. \quad (16.2.24)$$

At this point two remarks are in order: The first is that we have been using the term *bump function* in a slightly different way from that often employed in mathematics. In mathematics a bump function is generally a smooth ( $C^\infty$ ) function with exact value 1 over

some region and exact value 0 slightly outside this region. By contrast, in (2.20) and (2.21), we require only that this be approximately true. The second is that the relation (2.24) also holds in the case of a *thick* air-core solenoid. See Subsection 2.4 and Exercise 2.9.

We also observe that, according to (2.17), the soft-edge bump function can be written in the form

$$\text{bump}(z, a, L) = [\text{sgn}(z, a) - \text{sgn}(z - L, a)]/2 \quad (16.2.25)$$

where  $\text{sgn}(z, a)$  is an approximating “signum” function given by the relation

$$\text{sgn}(z, a) = z/[z^2 + a^2]^{1/2}. \quad (16.2.26)$$

Figures 2.2 and 2.3 illustrate the behavior of the approximating signum function for two different values of  $a$ . Evidently the approximating signum function becomes the true signum function in the limit that  $a$  goes to zero,

$$\lim_{a \rightarrow 0} \text{sgn}(z, a) = \text{sgn}(z). \quad (16.2.27)$$

Recall that the true signum function has the definition

$$\begin{aligned} \text{sgn}(z) &= 1 \text{ if } z > 0, \\ \text{sgn}(z) &= 0 \text{ if } z = 0, \\ \text{sgn}(z) &= -1 \text{ if } z < 0. \end{aligned} \quad (16.2.28)$$

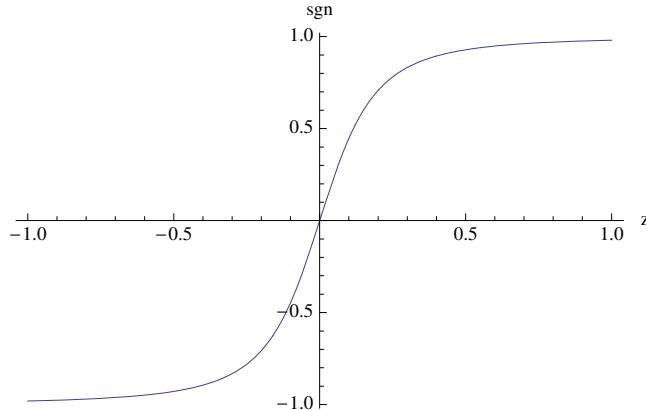


Figure 16.2.2: The approximating signum function (2.26) when  $a = .2$ .

In this same limit the soft-edge bump function becomes the hard-edge bump function,

$$\lim_{a \rightarrow 0} \text{bump}(z, a, L) = \text{bump}(z, L). \quad (16.2.29)$$

The hard-edge bump function has the properties,

$$\begin{aligned} \text{bump}(z, L) &= 1 \text{ for } z \in (0, L), \\ \text{bump}(0, L) &= \text{bump}(L, L) = 1/2, \\ \text{bump}(z, L) &= 0 \text{ elsewhere.} \end{aligned} \quad (16.2.30)$$

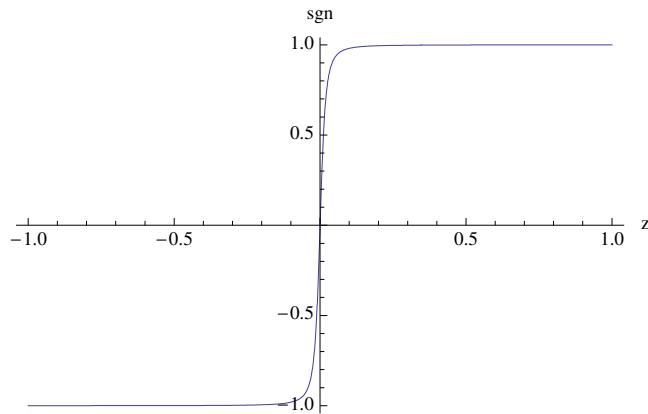


Figure 16.2.3: The approximating signum function (2.26) when  $a = .02$ .

Figures 2.4 and 2.5 illustrate the bump-function properties (2.20) through (2.22) for a fixed value of  $L$  and two different values of  $a$ . As expected, the simple air-core solenoid soft-edge bump function approaches a hard-edge bump function in the limit  $a \rightarrow 0$ . We also see that the quantity  $a$  plays the role of a *characteristic length* that controls the rate of fall off. The fringe-field region is large if  $a$  is large, and vanishes as  $a$  goes to zero. Finally, we note that  $\text{sgn}(z, a)$  is *analytic* as a function of  $z$  save for branch points at  $z = \pm ia$ . Correspondingly,  $\text{bump}(z, a, L)$  and hence also all the  $C_0^{[n]}(z)$  are analytic in  $z$  save for branch points at  $z = \pm ia$  and  $z = L \pm ia$ . Therefore approximating  $C_0^{[1]}(z)$  by a series of straight-line segments, as is sometimes done in the literature, violates its fundamental analytic properties.

Note also that use of the term *fringe field* to describe what is going on here can be a bit misleading. It is true that the field does extend beyond/outside the solenoid boundaries  $z = 0$  and  $z = L$ , but it is also affected/diminished *inside* the boundaries, particularly noticeably in the vicinity of the boundaries. Finally note that (2.24) holds for all  $a$  and does not depend on  $a$ . Thus, so to speak, whatever on-axis field “disappears” from inside the boundaries of the solenoid due to fringing behavior is in fact found outside the boundaries.

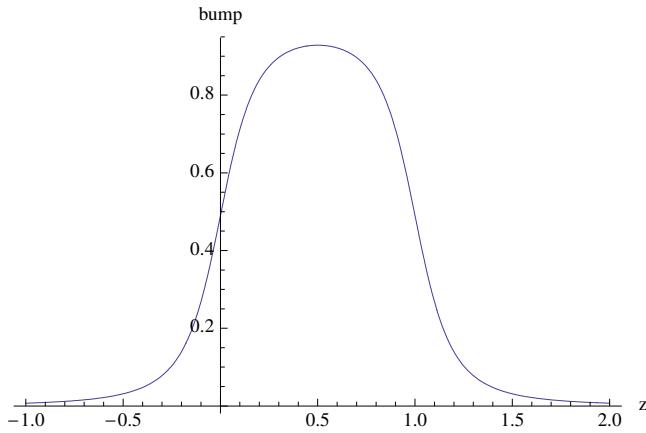


Figure 16.2.4: The soft-edge bump function (2.17) when  $a = .2$  and  $L = 1$ .

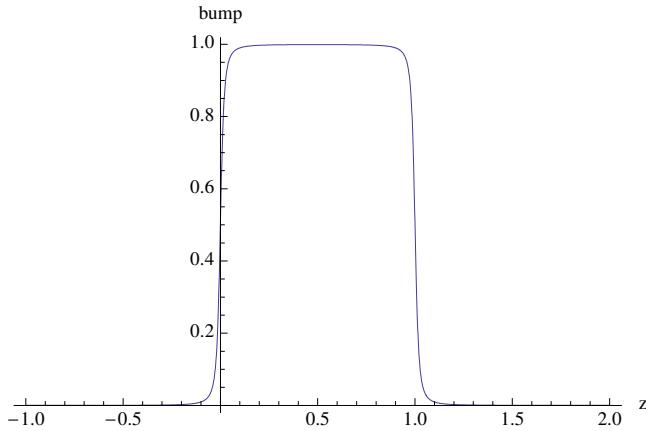


Figure 16.2.5: The soft-edge bump function (2.17) when  $a = .02$  and  $L = 1$ .

### 16.2.2.2 Asymptotic Behavior of the On-Axis Field

Let us examine the behaviors of the approximating signum function and soft-edge bump function in more detail. Evidently the approximating signum function is an odd function of  $z$ ,

$$\operatorname{sgn}(-z, a) = -\operatorname{sgn}(z, a), \quad (16.2.31)$$

and, correspondingly, satisfies the relation

$$\operatorname{sgn}(0, a) = 0. \quad (16.2.32)$$

It is also easy to verify that

$$\operatorname{sgn}(\pm a, a) = \pm 1/\sqrt{2} = \pm .707\cdots. \quad (16.2.33)$$

Finally, It can be verified from (2.26) that, when  $(a/z)^2 < 1$ , there are the asymptotic behaviors

$$\operatorname{sgn}(z, a) = 1 - (1/2)(a/z)^2 + (3/8)(a/z)^4 - (15/48)(a/z)^6 + \cdots \text{ as } z \rightarrow +\infty, \quad (16.2.34)$$

$$\operatorname{sgn}(z, a) = -1 + (1/2)(a/z)^2 - (3/8)(a/z)^4 + (15/48)(a/z)^6 - \dots \text{ as } z \rightarrow -\infty. \quad (16.2.35)$$

The behavior of the soft-edge bump function is a bit more complicated but, according to (2.25), follows from that of the approximating signum function. We begin with two simple observations. From (2.25) and (2.32) we see that there are the “end” ( $z = 0$  and  $z = L$ ) values

$$\operatorname{bump}(0, a, L) = \operatorname{bump}(L, a, L) = (1/2)L/(L^2 + a^2)^{1/2} \rightarrow (1/2) \text{ as } (a/L) \rightarrow 0, \quad (16.2.36)$$

and the “center” ( $z = L/2$ ) value

$$\operatorname{bump}(L/2, a, L) = L/(L^2 + 4a^2)^{1/2} \rightarrow 1 \text{ as } (a/L) \rightarrow 0. \quad (16.2.37)$$

We will next study the soft-edge bump function’s *near* leading/entering end behavior, its behavior when  $z \in (-L, 0)$ . [Its near trailing/exiting end behavior, its behavior when  $z \in (L, 2L)$ , then follows by symmetry.] For this purpose, according to (2.25), we need to know the behavior of both  $\operatorname{sgn}(z, a)$  and  $\operatorname{sgn}(z - L, a)$  when  $z < 0$ . The behavior of  $\operatorname{sgn}(z, a)$  for  $z < 0$  is given by (2.35). For  $\operatorname{sgn}(z - L, a)$  we find (when  $z \approx 0$ ) the expansion

$$\begin{aligned} \operatorname{sgn}(z - L, a) &= (z - L)/[(z - L)^2 + a^2]^{1/2} = -[1 + a^2/(z - L)^2]^{1/2} \\ &= -1 + (1/2)[a/(z - L)]^2 - (3/8)[a/(z - L)]^4 + \dots \\ &= -1 + (1/2)(a/L)^2 + (a^2 z/L^3) + \dots \end{aligned} \quad (16.2.38)$$

(Here we assume that both  $z/L$  and  $a/L$  are small.) Consequently, for the range  $z \approx 0$  but  $z < -a$  so that (2.35) holds, we conclude from (2.25) there is the expansion

$$\begin{aligned} \operatorname{bump}(z, a, L) &= [\operatorname{sgn}(z, a) - \operatorname{sgn}(z - L, a)]/2 \\ &= -1/2 + (1/4)(a/z)^2 - (3/16)(a/z)^4 + \dots \\ &\quad + 1/2 - (1/4)(a/L)^2 - (1/2)(a^2 z/L^3) - \dots \\ &= +(1/4)(a/z)^2 - (1/4)(a/L)^2 - (1/2)(a/L)^2(z/L) - \dots . \end{aligned} \quad (16.2.39)$$

We see that the  $(a/z)^2$  term dominates for small  $z$  and therefore  $\operatorname{bump}(z, a, L)$  decreases as  $1/z^2$  as  $z$  becomes more negative. Upon reflection, this result should be expected. Close by the end of a solenoid the external field looks like a *monopole* field; and the field of a monopole falls off with distance as the inverse square.

Complete asymptotic behavior does not set in until  $z < -L$ . In that case, using (2.35), we see that  $\operatorname{sgn}(z - L, a)$  has the expansion

$$\begin{aligned} \operatorname{sgn}(z - L, a) &= -1 + (1/2)[a/(z - L)]^2 - (3/8)[a/(z - L)]^4 + \dots \\ &= -1 + (1/2)(a/z)^2 + (La^2/z^3) \\ &\quad + (3/8)(4a^2 L^2 - a^4)/z^4 + \dots \text{ when } z < -L. \end{aligned} \quad (16.2.40)$$

We now find from (2.25) the result

$$\begin{aligned} \operatorname{bump}(z, a, L) &= [\operatorname{sgn}(z, a) - \operatorname{sgn}(z - L, a)]/2 \\ &= -1/2 + (1/4)(a/z)^2 - (3/16)(a/z)^4 + \dots \\ &\quad + 1/2 - (1/4)(a/z)^2 - (1/2)(La^2/z^3) + \dots \\ &= -(1/2)La^2/z^3 + O(1/z^4) \text{ when } z \rightarrow -\infty. \end{aligned} \quad (16.2.41)$$

Correspondingly, the on-axis gradient  $C_0^{[1]}(z)$  and on-axis field  $B_z(0, 0, z)$  fall off for very large ( $z < -L$ ) distances as

$$C_0^{[1]}(z) = B_z(0, 0, z) = -(1/2)BLa^2/z^3 + \dots \text{ when } z \rightarrow -\infty. \quad (16.2.42)$$

Analogous fall off occurs when  $z > 2L$ . This result is also to be expected. From far enough away, the end fields of a solenoid of length  $L$  look like those of two monopoles of opposite signs a distance  $L$  apart, and therefore combine to appear as the field of a *dipole* once one is more than a distance  $L$  away. Finally, at large distances, the field of a dipole falls off as  $1/|z|^3$ . Thus, for a simple air-core solenoid, there is a  $1/|z|^3$  fall off for the fringe field. Subsequently, in Subsection 21.1.3, we will see that the fringe field can fall off *exponentially* in the case of an iron-dominated solenoid.

We close this subsection with the admonition that, although the asymptotic expansions we have examined are illuminating, there is no substitute for computing  $B_z(0, 0, z)$  exactly using (2.17) and (2.18).

### 16.2.2.3 Properties of the Vector Potential

The computation of orbits in and transfer maps for solenoids, using a Hamiltonian formulation, requires the use of a vector potential. We will employ the vector potential given by (2.11) through (2.13). Evidently, depending on the order to which we wish to work, we need the functions  $C_0^{[1]}(z)$ ,  $C_0^{[3]}(z)$ ,  $C_0^{[5]}(z), \dots$ . To get a feel for what is involved, let us examine, for example, the function  $C_0^{[3]}(z)$ . From (2.15) and (2.19) we see that

$$C_0^{[3]}(z) = (\partial/\partial z)^2 B_z(0, 0, z) = B (\partial/\partial z)^2 \text{bump}(z, a, L) = B \text{bump}''(z, a, L). \quad (16.2.43)$$

Figures (2.6) and (2.7) illustrate the function  $\text{bump}''$  for a fixed value of  $L$  and two different values of  $a$ . Evidently the function  $\text{bump}''$  becomes quite singular at the ends of the solenoid in the limit  $a \rightarrow 0$ . Indeed it approaches the function  $\delta'(z)$  at the leading end, and the function  $-\delta'(z - L)$  at the trailing end. Moreover, it falls off quite rapidly beyond the fringe-field regions. From (2.39) and (2.41) we conclude that there is the near-by asymptotic behavior

$$\text{bump}''(z, a, L) = (3/2)a^2/z^4 + \dots \text{ as } z \rightarrow -\infty, \quad (16.2.44)$$

and the far asymptotic behavior

$$\text{bump}''(z, a, L) = -6La^2/z^5 + \dots \text{ as } z \rightarrow -\infty, \quad (16.2.45)$$

The still higher derivatives of the bump function, needed to compute the  $C_0^{[n]}(z)$  for still larger values of  $n$ , are even more singular in the limit  $a \rightarrow 0$ , and fall off ever more rapidly as  $z \rightarrow -\infty$ . Analogous results hold for  $z > L$  and  $z > 2L$ .

### 16.2.3 Opposing Simple Solenoid Doublet

We have seen that the on-axis field of a single simple solenoid has the *far* asymptotic behavior (2.42). A sequence of solenoids, all having the same “sign”, will have the same far

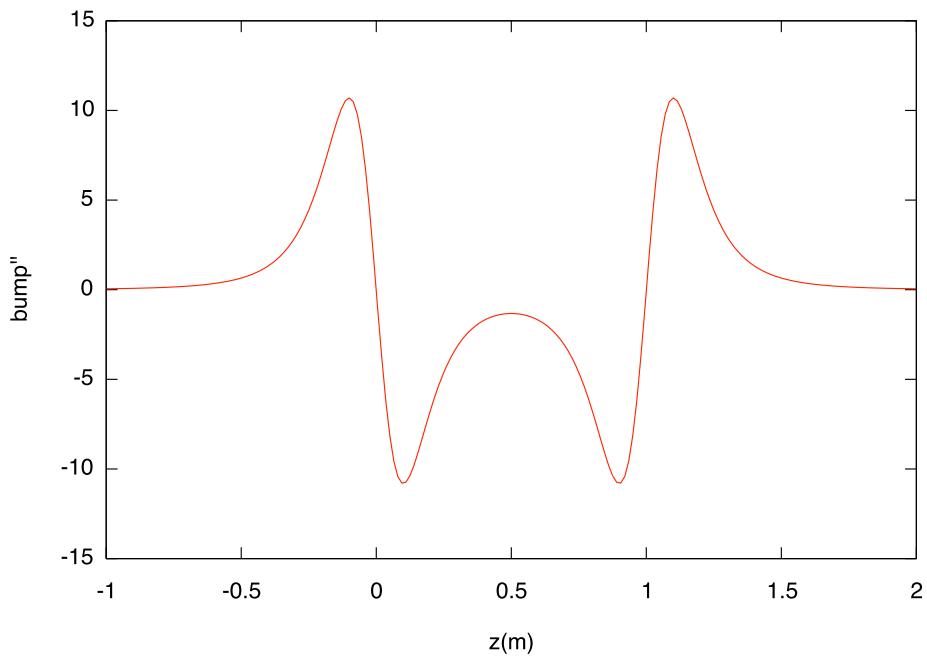


Figure 16.2.6: The function  $\text{bump}''$  when  $a = .2$  and  $L = 1$ .

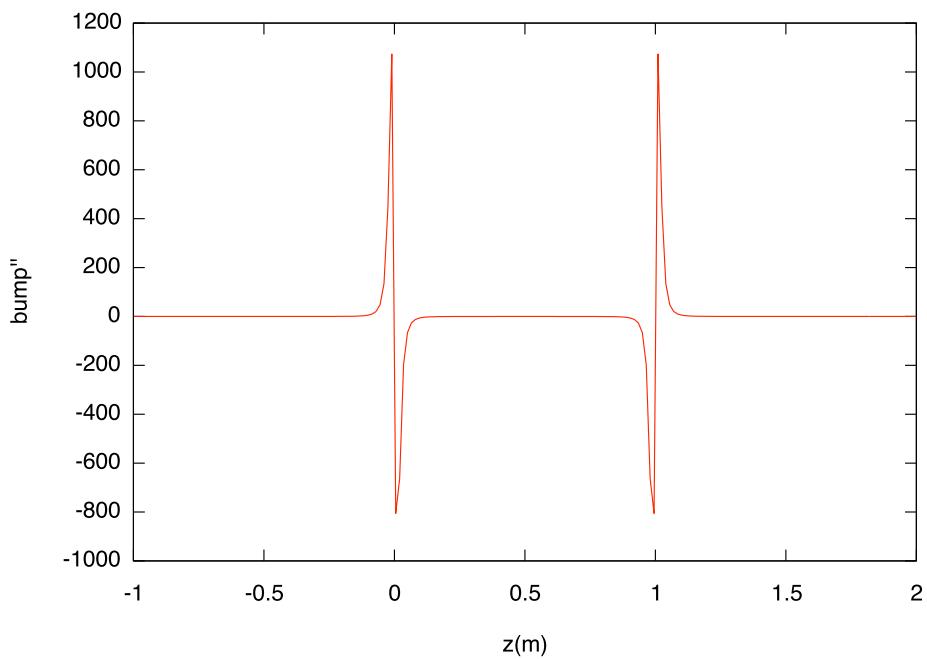


Figure 16.2.7: The function  $\text{bump}''$  when  $a = .02$  and  $L = 1$ .

fall-off behavior. In this subsection we will study the far fall-off behavior for what we call an *opposing solenoid doublet*. By this term we mean a pair of solenoids, each of length  $L$ , separated by a space  $D$ , and having opposite strengths. By superposition, the on-axis field for such a pair of solenoids is given by the relation

$$\begin{aligned} B_z^{\text{osd}}(0, 0, z) &= B \text{ bump}(z, a, L) - B \text{ bump}(z - L - D, a, L) \\ &= B\{z/[z^2 + a^2]^{1/2} - (z - L)/[(z - L)^2 + a^2]^{1/2}\}/2 \\ &\quad - B\{(z - L - D)/[(z - L - D)^2 + a^2]^{1/2} - (z - 2L - D)/[(z - 2L - D)^2 + a^2]^{1/2}\}/2. \\ &= B[\text{sgn}(z, a) - \text{sgn}(z - L, a) - \text{sgn}(z - L - D, a) + \text{sgn}(z - 2L - D, a)]. \end{aligned} \tag{16.2.46}$$

Computation using expansions such as (2.42) shows that for this opposing solenoid doublet there is the far fall-off behavior

$$C_0^{[1]}(z) = B_z^{\text{osd}}(0, 0, z) = 3Ba^2L(L + D)/z^4 + O(1/z^5) \text{ when } z \rightarrow -\infty. \tag{16.2.47}$$

Comparison with (2.42) shows that this  $1/z^4$  far fall-off behavior is one order higher in  $1/z$  than that for a single solenoid. This result is to be expected because the end of each solenoid looks like a monopole, at a far distance the four ends of the two solenoids in the opposing solenoid doublet look like an in-line quadrupole, and the field of a quadrupole falls off as  $1/z^4$ .

### 16.2.4 More Complicated Air-Core Solenoids

The fields for more complicated air-core solenoids can be found from those of simple single-layer air-core solenoids by superposition. Consider, for example, the on-axis field of an air-core solenoid that has a multi-layer winding with inner radius  $a_1$  and outer radius  $a_2$ . We will call such a solenoid a *thick* solenoid. We observe that there is the integral relation

$$\int_{a_1}^{a_2} da \{1/[z^2 + a^2]^{1/2}\} = \log \left( \{[z^2 + a_2^2]^{1/2} + a_2\} / \{[z^2 + a_1^2]^{1/2} + a_1\} \right). \tag{16.2.48}$$

Correspondingly, the on-axis field of such a thick solenoid is given by the relation

$$\begin{aligned} B_z(0, 0, z) &= (B/2)[1/(a_2 - a_1)] \left[ z \log \left( \{[z^2 + a_2^2]^{1/2} + a_2\} / \{[z^2 + a_1^2]^{1/2} + a_1\} \right) \right. \\ &\quad \left. - (z - L) \log \left( \{[(z - L)^2 + a_2^2]^{1/2} + a_2\} / \{[(z - L)^2 + a_1^2]^{1/2} + a_1\} \right) \right]. \end{aligned} \tag{16.2.49}$$

Here again the winding extends from  $z = 0$  to  $z = L$  and the interior field (in the infinite length limit) is  $B$ .

Evidently, from this result and by superposition, analytic on-axis results can be obtained for any combination of concentric coils of various lengths, thicknesses, locations, and powerings. Note, because we have assumed cylindrical symmetry in all cases, only the  $m = 0$  terms are present in the expansion (15.3.33) so that (2.1) continues to hold.

### 16.2.5 Computation of Transfer Map

In this subsection we will compute the transfer map for a solenoid (or a collection of solenoids) including fringe-field effects. To do so we begin with the Hamiltonian (1.1) and employ the vector potential given by (2.7) through (2.10). We then introduce dimensionless scaled deviation variables and the associated scaled deviation-variable Hamiltonian. Finally, we expand the scaled deviation-variable Hamiltonian in a Taylor series, and employ this Taylor series to compute the transfer map.

#### 16.2.5.1 Dimensionless Scaled Deviation Variables and Scaled Deviation-Variable Hamiltonian

A solenoid is an example of a straight beam-line element for which the design orbit may be taken to be the  $z$  axis (a straight line) traversed with constant velocity. According to the results of Section 13.1.5, the scaled deviation variable Hamiltonian  $H(X, Y, \tau, P_x, P_y, P_\tau; z)$  for any such element is given by

$$\begin{aligned} H(X, Y, \tau, P_x, P_y, P_\tau; z) = & \\ - (1/\ell) \{ [1 - (2P_\tau/\beta_0) + P_\tau^2 - (P_x - A_x^s)^2 - (P_y - A_y^s)^2]^{1/2} + (P_\tau/\beta_0) - (1/\beta_0^2) \}. & \end{aligned} \quad (16.2.50)$$

Here the dimensionless scaled deviation variables  $(X, Y, \tau, P_x, P_y, P_\tau)$  are defined in terms of the original variables  $(x, y, t, p_x, p_y, p_t)$  by the relations (13.1.21) through (13.1.26), and the *scaled* vector potential  $\mathbf{A}^s$  is defined in terms of the original vector potential  $\hat{\mathbf{A}}^0$  by the relation

$$\mathbf{A}^s(X, Y, z) = (q/p^0) \hat{\mathbf{A}}^0(\ell X, \ell Y, z). \quad (16.2.51)$$

We have also used (2.9).

The relations (2.7) through (2.10) and (2.51) can be used to find  $\mathbf{A}^s$  for solenoids. To do so it is convenient to rewrite (2.10) in the form

$$V(\rho, z) = (1/2) \sum_{n=0}^{\infty} (-1)^n \frac{1}{2^{2n} n! (n+1)!} B^{[2n]}(z) \rho^{2n}. \quad (16.2.52)$$

Here we have introduced, in accord with (2.15), the notation

$$B^{[0]}(z) = C_0^{[1]}(z) = B_z(0, 0, z), \quad (16.2.53)$$

$$B^{[2n]}(z) = (\partial/\partial z)^{2n} B^{[0]}(z) = (\partial/\partial z)^{2n} C_0^{[1]}(z) = C_0^{[2n+1]}(z). \quad (16.2.54)$$

Combining these relations yields the results

$$A_x^s = -Y U^s(X, Y, z) \quad (16.2.55)$$

$$A_y^s = X U^s(X, Y, z), \quad (16.2.56)$$

$$A_z^s = 0, \quad (16.2.57)$$

where

$$U^s(X, Y, z) = \ell(q/p^0)V(\ell X, \ell Y, z) = (1/2) \sum_{n=0}^{\infty} (-1)^n \frac{1}{2^{2n} n! (n+1)!} b^{[2n]}(z) (X^2 + Y^2)^n. \quad (16.2.58)$$

Here we have introduced the notation

$$b^{[2n]}(z) = (q/p^0)\ell^{2n+1}B^{[2n]}(z), \quad (16.2.59)$$

and observe, in view of (1.5.81) and (2.54), that the quantities  $b^{[2n]}(z)$  are dimensionless. Also, from (2.50), we conclude that the scaled deviation variable Hamiltonian  $H$  has dimensions of 1/length.

### 16.2.5.2 Symmetry of Scaled Deviation-Variable Hamiltonian

The Hamiltonian  $H$  given by (2.50) has a symmetry that is worth noting. Define two two-dimensional vectors  $\mathbf{Q}$  and  $\mathbf{P}$  by the rules

$$\mathbf{Q} = X\mathbf{e}_x + Y\mathbf{e}_y, \quad (16.2.60)$$

$$\mathbf{P} = P_x\mathbf{e}_x + P_y\mathbf{e}_y. \quad (16.2.61)$$

Also make the definitions

$$Q^2 = \mathbf{Q} \cdot \mathbf{Q} = X^2 + Y^2, \quad (16.2.62)$$

$$P^2 = \mathbf{P} \cdot \mathbf{P} = P_x^2 + P_y^2, \quad (16.2.63)$$

$$J_z = (\mathbf{Q} \times \mathbf{P}) \cdot \mathbf{e}_z = XP_y - YP_x. \quad (16.2.64)$$

With the aid of these definitions we may write

$$\begin{aligned} (P_x - A_x^s)^2 + (P_y - A_y^s)^2 &= (P_x)^2 + (P_y)^2 + (A_x^s)^2 + (A_y^s)^2 - 2P_x A_x^s - 2P_y A_y^s \\ &= \mathbf{P} \cdot \mathbf{P} + \mathbf{A}^s \cdot \mathbf{A}^s - 2\mathbf{P} \cdot \mathbf{A}^s. \end{aligned} \quad (16.2.65)$$

We also observe with the aid of (2.55) and (2.56) that there are the relations

$$\mathbf{P} \cdot \mathbf{A}^s = (XP_y - YP_x)U^s = J_z U^s, \quad (16.2.66)$$

$$\mathbf{Q} \cdot \mathbf{A}^s = 0, \quad (16.2.67)$$

and

$$\mathbf{A}^s \cdot \mathbf{A}^s = Q^2(U^s)^2. \quad (16.2.68)$$

The relation (2.67) is a consequence of our decision to employ the Poincaré-Coulomb gauge. Also note that, according to (2.58),  $U^s$  depends only on  $Q^2$  and  $z$ . With the aid of these relations we may also write

$$(P_x - A_x^s)^2 + (P_y - A_y^s)^2 = P^2 + Q^2(U^s)^2 - 2J_z U^s. \quad (16.2.69)$$

It is easily verified that  $J_z$  has the properties

$$: J_z : X = [J_z, X] = [(XP_y - YP_x), X] = Y, \quad (16.2.70)$$

$$: J_z : Y = -X, \quad (16.2.71)$$

$$: J_z : P_x = P_y, \quad (16.2.72)$$

$$: J_z : P_y = -P_x, \quad (16.2.73)$$

$$: J_z : \tau =: J_z : P_\tau = 0. \quad (16.2.74)$$

Consequently, as the notation is meant to suggest, the Lie operator  $: J_z :$  is the generator of rotations about the  $z$  axis. It follows that there are the relations

$$: J_z : Q^2 =: J_z : U^s =: J_z : P^2 =: J_z : (\mathbf{P} \cdot \mathbf{Q}) = 0. \quad (16.2.75)$$

We remark that the last relation in (2.75) is consistent with the identity

$$(\mathbf{P} \cdot \mathbf{Q})^2 = Q^2 P^2 - J_z^2. \quad (16.2.76)$$

We also see from (2.55) and (2.56) that there are the relations

$$: J_z : A_x^s == [J_z, -YU^s] = -[J_z, Y]U^s = XU^s = A_y^s \quad (16.2.77)$$

and

$$: J_z : A_y^s == [J_z, XU^s] = [J_z, X]U^s = YU^s = -A_x^s. \quad (16.2.78)$$

It follows that

$$: J_z : (\mathbf{A}^s \cdot \mathbf{A}^s) =: J_z : (\mathbf{Q} \cdot \mathbf{A}^s) =: J_z : (\mathbf{P} \cdot \mathbf{A}^s) = 0. \quad (16.2.79)$$

We note that these last results can be viewed as a consequence of (2.68), (2.67), and (2.66). But they can also be viewed as consequence of the relations (2.70) through (2.73) and (2.77) and (2.78).

Based on the work so far  $H$  as given by (2.50) can be rewritten in either of the forms

$$\begin{aligned} H(X, Y, \tau, P_x, P_y, P_\tau; z) &= \\ &- (1/\ell) \{ [1 - (2P_\tau/\beta_0) + P_\tau^2 - \mathbf{P} \cdot \mathbf{P} - \mathbf{A}^s \cdot \mathbf{A}^s + 2\mathbf{P} \cdot \mathbf{A}^s]^{1/2} + (P_\tau/\beta_0) - (1/\beta_0^2) \}. \end{aligned} \quad (16.2.80)$$

and

$$\begin{aligned} H(X, Y, \tau, P_x, P_y, P_\tau; z) &= \\ &- (1/\ell) \{ [1 - (2P_\tau/\beta_0) + P_\tau^2 - P^2 - Q^2(U^s)^2 + 2J_z U^s]^{1/2} + (P_\tau/\beta_0) - (1/\beta_0^2) \}. \end{aligned} \quad (16.2.81)$$

From either of these forms it is evident, using (2.75) and (2.79), that

$$: J_z : H = [J_z, H] = 0. \quad (16.2.82)$$

That is,  $H$  is invariant under rotations about the  $z$  axis and, conversely,  $J_z$  is an integral of motion. (We also say that  $H$  and  $J_z$  are in involution.) Note that this invariance stems from the fact that we are dealing with the the magnetic field case described by an  $m = 0$  scalar potential  $\psi$ . This scalar potential has rotational symmetry about the  $z$  axis, and the gauge for the associated vector potential has been judiciously chosen to maintain this symmetry. See Exercise 2.10.

### 16.2.5.3 Properties of Transfer Map and Factorization of Linear Part

Because of the  $J_z U^s$  term in (2.81) [or the  $\mathbf{P} \cdot \mathbf{A}^s$  term in (2.80)] the map  $\mathcal{M}$  generated by  $H$  produces rotations about the  $z$  axis as well as other effects. Consequently  $\mathcal{M}$  does not preserve the  $X, P_x$  and  $Y, P_y$  planes. When performing fitting operations it is easier to understand what is happening when motion in the  $X, P_x$  and  $Y, P_y$  planes is uncoupled. Uncoupling can be accomplished by a trick. Define a Hamiltonian with no rotational parts, call it  $H^{\text{nonrot}}$ , by removing the  $J_z U^s$  term in (2.81):

$$\begin{aligned} H^{\text{nonrot}}(X, Y, \tau, P_x, P_y, P_\tau; z) = \\ - (1/\ell) \{ [1 - (2P_\tau/\beta_0) + P_\tau^2 - P^2 - Q^2(U^s)^2]^{1/2} + (P_\tau/\beta_0) - (1/\beta_0^2) \}. \end{aligned} \quad (16.2.83)$$

The map generated by  $H^{\text{nonrot}}$ , call it  $\mathcal{M}^{\text{nonrot}}$ , will preserve the  $X, P_x$  and  $Y, P_y$  planes. See Exercise 2.17. To proceed, first carry out the desired fitting operation using  $\mathcal{M}^{\text{nonrot}}$  in place of  $\mathcal{M}$ . After a fit has been achieved using  $\mathcal{M}^{\text{nonrot}}$  in place of  $\mathcal{M}$ , continue on using the associated full  $\mathcal{M}$  in subsequent calculations.

In general the maps  $\mathcal{M}$  and  $\mathcal{M}^{\text{nonrot}}$  do not commute. However, as shown in Exercise 2.15, the matrices  $M$  and  $M^{\text{nonrot}}$  associated with their linear parts do commute,

$$M^{\text{nonrot}} M = M M^{\text{nonrot}}. \quad (16.2.84)$$

In this case it is possible to define a matrix  $M^{\text{rot}}$  by the rule

$$M^{\text{rot}} = (M^{\text{nonrot}})^{-1} M = M (M^{\text{nonrot}})^{-1} \quad (16.2.85)$$

so that

$$M = M^{\text{rot}} M^{\text{nonrot}} = M^{\text{nonrot}} M^{\text{rot}}. \quad (16.2.86)$$

As the notation suggests,  $M^{\text{rot}}$  describes rotations about the  $z$  axis.

For example, for the matrix  $M^{\text{body}}$  displayed in Exhibit 2.11 in Subsection 2.6.2, Exhibits 2.1 and 2.2 below display the factors  $M^{\text{nonrot}}$  and  $M^{\text{rot}}$ .

Exhibit 16.2.1: The matrix  $M^{\text{nonrot}}$  factor of the  $M^{\text{body}}$  displayed in Exhibit 2.11

$$\begin{array}{ccccccc} 9.94943E-01 & 9.98265E-01 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 \\ -1.01064E-02 & 9.94943E-01 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 \\ 0.00000E+00 & 0.00000E+00 & 9.94943E-01 & 9.98265E-01 & 0.00000E+00 & 0.00000E+00 \\ 0.00000E+00 & 0.00000E+00 & -1.01064E-02 & 9.94943E-01 & 0.00000E+00 & 0.00000E+00 \\ 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 1.00000E+00 & 4.11143E-01 \\ 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 1.00000E+00 \end{array}$$

Exhibit 16.2.2: The matrix  $M^{\text{rot}}$  factor of the  $M^{\text{body}}$  displayed in Exhibit 2.11

$$\begin{array}{ccccccc} 9.94963E-01 & 0.00000E+00 & 1.00240E-01 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 \\ 0.00000E+00 & 9.94963E-01 & 0.00000E+00 & 1.00240E-01 & 0.00000E+00 & 0.00000E+00 \\ -1.00240E-01 & 0.00000E+00 & 9.94963E-01 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 \\ 0.00000E+00 & -1.00240E-01 & 0.00000E+00 & 9.94963E-01 & 0.00000E+00 & 0.00000E+00 \\ 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 1.00000E+00 & 0.00000E+00 \\ 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 1.00000E+00 \end{array}$$

As can be seen, the effect of  $M^{\text{nonrot}}$  is to produce (equal) *focussing* in both planes. (Note also that  $M^{\text{nonrot}}$  does not introduce any coupling between planes.) And  $M^{\text{rot}}$  produces a rotation about the  $z$  axis by an angle  $\theta_{\text{rot}}$ . In the case that the entrance and exit planes are *well outside* the fringe-field regions [which, as can be seen from Figure 2.5 and (2.42), is not quite true for this example],  $\theta_{\text{rot}}$  is given (in radians) by the relation

$$\theta_{\text{rot}} = [1/(2 \text{ brho})] \int_{-\infty}^{+\infty} B_z(0, 0, z) dz = BL/(2 \text{ brho}). \quad (16.2.87)$$

Here brho is the magnetic rigidity. See Exercise 1.5.9. Note that the result (2.87) does not depend on  $a$  in the case of a simple solenoid or on  $a_1$  and  $a_2$  in the case of a thick solenoid.

#### 16.2.5.4 Expansion of Scaled Deviation-Variable Hamiltonian

To find the transfer map  $\mathcal{M}$  about the design orbit it is necessary to express the scaled deviation-variable Hamiltonian  $H$  as a sum of homogeneous polynomials,

$$H = \sum_{m=0}^{\infty} H_m. \quad (16.2.88)$$

Doing so for the Hamiltonian (2.50) gives, for the first few terms, the results

$$H_0 = 1/(\beta_0^2 \gamma_0^2 \ell), \quad (16.2.89)$$

$$H_1 = 0, \quad (16.2.90)$$

$$\begin{aligned} H_2 &= [1/(2\ell)](P_x^2 + P_y^2) - [b^{[0]}/(2\ell)](XP_y - YP_x) \\ &\quad + [(b^{[0]})^2/(8\ell)](X^2 + Y^2) + [1/(2\beta_0^2 \gamma_0^2 \ell)]P_\tau^2, \end{aligned} \quad (16.2.91)$$

$$\begin{aligned} H_3 &= [1/(2\beta_0 \ell)]P_\tau(P_x^2 + P_y^2) - [b^{[0]}/(2\beta_0 \ell)]P_\tau(XP_y - YP_x) \\ &\quad + [(b^{[0]})^2/(8\beta_0 \ell)]P_\tau(X^2 + Y^2) + [1/(\beta_0^3 \gamma_0^2 \ell)]P_\tau^3, \end{aligned} \quad (16.2.92)$$

$$\begin{aligned} H_4 &= (1/8\ell)(P_x^4 + 2P_x^2 P_y^2 + P_y^4) - [b^{[0]}/(4\ell)](P_x^2 + P_y^2)(XP_y - YP_x) \\ &\quad + [(b^{[0]})^2/(16\ell)](X^2 P_x^2 + Y^2 P_y^2) + [3(b^{[0]})^2/(16\ell)](X^2 P_y^2 + Y^2 P_x^2) \\ &\quad - [(b^{[0]})^2/(4\ell)](XP_x YP_y) + \{[b^{[2]} - (b^{[0]})^3]/(16\ell)\}(X^2 + Y^2)(XP_y - YP_x) \\ &\quad + \{[(b^{[0]})^4 - 4b^{[0]} b^{[2]}]/(128\ell)\}(X^4 + 2X^2 Y^2 + Y^4) - [(3 - \beta_0^3)/(4\beta_0^2 \ell)]P_\tau^2(P_x^2 + P_y^2) \\ &\quad - [b^{[0]}(3 - \beta_0^3)/(4\beta_0^2 \ell)]P_\tau^2(XP_y - YP_x) + [(b^{[0]})^2(3 - \beta_0^3)/(16\beta_0^2 \ell)]P_\tau^2(X^2 + Y^2) \\ &\quad + [(5 - \beta_0)^2]/(8\beta_0^4 \gamma_0^2 \ell)]P_\tau^4. \end{aligned} \quad (16.2.93)$$

These are the terms required to compute  $\mathcal{M}$  through third order. Note that  $H_1$  vanishes as it should. If it did not, the phase-space path obtained by setting all deviation variables to zero (the design orbit) would not be a solution of the equations of motion. We also remark that the constant piece  $H_0$  is irrelevant to the actual motion, and does not enter into the

calculation of  $\mathcal{M}$ . It is presented only as an aid for those who wish to check the expansion (2.88) through (2.93).

In view of the symmetry (2.82), it is also instructive to have an expansion of  $H$  in terms of the variables  $P_\tau$ ,  $P^2$ ,  $J_z$ , and  $Q^2$ . In terms of these variables there are the results

$$H_2 = [1/(2\ell)]P^2 - [b^{[0]}/(2\ell)]J_z + [(b^{[0]})^2/(8\ell)]Q^2 + [1/(2\beta_0^2\gamma_0^2\ell)]P_\tau^2, \quad (16.2.94)$$

$$H_3 = (1/\beta_0)P_\tau H_2, \quad (16.2.95)$$

$$\begin{aligned} H_4 = & [1/(8\ell)](P^2)^2 - [b^{[0]}/4\ell]P^2 J_z + [b^{[0]}/(8\ell)]J_z^2 + [-b^{[0]}/(8\ell) + 3(b^{[0]})^2/(16\ell)]P^2 Q^2 \\ & + \{[b^{[2]} - (b^{[0]})^3]/(16\ell)\}Q^2 J_z + \{[(b^{[0]})^4 - 4b^{[0]}b^{[2]}]/(128\ell)\}(Q^2)^2 \\ & + [(3 - \beta_0^2)/(4\beta_0^2\ell)]P_\tau^2 P^2 - [(3 - \beta_0^2)b^{[0]}/(4\beta_0^2\ell)]P_\tau^2 J_z \\ & + [(3 - \beta_0^2)(b^{[0]})^2/(16\beta_0^2\ell)]P_\tau^2 Q^2 + [(5 - \beta_0)^2]/(8\beta_0^4\gamma_0^2\ell)]P_\tau^4. \end{aligned} \quad (16.2.96)$$

Note that, because  $H_3$  as given by (2.95) is proportional to  $P_\tau$ , all second-order aberrations for any solenoid transfer map are purely chromatic.

We close this subsection with the remark that the result given by (2.94) through (2.96) holds for any solenoid. For simplicity, in subsequent sections we will apply them to the case of a simple air-core solenoid, in which case (2.19) holds. But they are also applicable to more complicated air-core solenoids as described in Subsection 2.4 as well as solenoids containing iron.

## 16.2.6 Solenoidal Fringe-Field Effects: Attempts to Hard-Edge Model Them

### 16.2.6.1 Convergence and Divergence

Suppose we wish to make a *simple* model of fringe-field effects. The hope would be to find a model whose fields are not too different from those that can be attained by feasible magnet construction and for which analytic calculations can be made using simple approximations and without too much difficulty, thereby bypassing the need for detailed numerical calculation involving a detailed knowledge of the functions  $b^{[n]}(z)$ . One idea for doing so is to consider a model in which the bump function in (2.18) and (2.19) is replaced by a bump function having the properties (2.30). This so-called *hard-edge* model, for which the on-axis field begins and ends abruptly, has only limited utility. Here are several objections to this approach:

- Real solenoids, and in particular multi-layer solenoids as described in Subsection 2.4, have extended fringe fields. From (2.49) one sees that the on-axis field involves both the inner radius  $a_1$  and the outer radius  $a_2$ . For a realistic/*thick* multi-layer solenoid  $a_2$  is relatively large. Correspondingly, the fringe fields falls off only slowly. Therefore beginning and terminating the on-axis field abruptly is a poor approximation for real solenoids.
- Suppose we restrict our attention to single-layer solenoids as described in Subsection 2.2. In this case (which we have called the *simple* solenoid case), as examination

of Figures 2.4 and 2.5 illustrates, it might be useful to attempt a hard-edge model. That is, we might attempt to compute the transfer map  $\mathcal{M}$  when  $a = 0$  because then, according to (2.29), the soft-edge bump function becomes the hard-edge bump function. In this case however, as described in Subsection 2.2.3, the on-axis gradient  $C_0^{[3]}$  involves  $bump''(z, a, L)$  which takes on the appearance of  $\delta'(z)$  and  $-\delta'(z - L)$  in the hard-edge limit. See Figure (2.7). We note that the appearance of the  $\delta'$  functions is a consequence of the representation (2.1) which itself is a consequence of the Maxwell equations for  $\mathbf{B}$ . We also note that  $H_4$  as given by (2.96) involves  $b^{[2]}(z)$  which in turn, according to (2.43) and (2.59), can involve the pesky  $\delta'$  functions. Therefore the differential equation (10.5.61) for  $f_4$  is ill defined in the hard-edge case. One might hope to deal with this complication by making calculations for  $a \neq 0$ , including all fringe-field effects, and then taking the limit  $a \rightarrow 0$ . When this is done it can be shown that some of the the third- and higher-order aberrations (described by the  $f_n$  with  $n \geq 4$ ) of the transfer map  $\mathcal{M}$  for a solenoid become *infinite* in the hard-edge ( $a \rightarrow 0$ ) limit! Thus, the hard-edge limit is unphysical for a solenoid if these aberrations are important.<sup>3</sup> Correspondingly, third-order solenoid aberrations can be reduced by making the fringe-field regions large. It also helps to make the solenoid weak since aberrations are proportional to  $B$ . If this is done, the solenoid must also be made long (to compensate for the small  $B$ ) in order to maintain the desired paraxial properties.

- Eventually we will discover that, for a simple solenoid, all second-order aberrations due to fringe fields *vanish* in the hard-edge limit, and all third-order aberrations due to fringe fields also vanish in the hard-edge limit save for three which *diverge* as  $1/a$ . (See Exercises 2.22 and 2.23.) There are *no* second and third-order aberrations due to fringe fields that take on finite nonzero limiting values in the hard-edge limit.
- Finally we must acknowledge the obvious but irritating fact that the aperture of the simple solenoid, which must contain the beam, shrinks to zero as  $a \rightarrow 0$ .

To illustrate some of these points, let us examine the transfer map for the specific simple air-core solenoid we have been discussing. To do so we will employ the Lie-algebraic charged particle beam transport code MaryLie. Among the beam-line elements it treats is the simple air-core solenoid. Exhibits 2.3 through 2.5 below show (through third order) the transfer map  $\mathcal{M}$  for the three cases  $a = 0.2$ ,  $a = .02$ , and  $a = .002$ . (Here we use the indexing scheme of Table 39.2.1.) In all cases the solenoid has length  $L = 1$ , and the entry and exit planes are taken to be at  $z = z^{\text{en}} = -1$  and  $z = z^{\text{ex}} = 2$ , respectively. All lengths are in meters, and we have used the terminology of Subsection 1.2. See Figures 2.4 and 2.5. The

<sup>3</sup>There is confusion/error on this point in the literature. Some authors give aberration results through third order and in the hard-edge limit for many common beam-line elements, but give results for simple solenoids only through second order. And their accompanying discussion can be read to imply that no difficulty is expected in extending the simple solenoid results through third order. Other authors propose formulas for the third-order aberrations of a simple solenoid in the hard-edge limit, and these formulas are independent of  $a$  and thus yield *finite* results when  $a = 0$ . Yet other authors correctly recognize that attempting to make the fringe-field region very small, say by adding extra coils at solenoid ends, leads to some very large third-order aberrations.

quantity  $B$  has the value  $B = 1$  Tesla and the magnetic rigidity (brho) is that for 800 MeV protons. Finally, the scale length is taken to be  $\ell = 1$  meter. Numerical integration of the differential equations (described in Section 10.5.2) to compute  $\mathcal{M}$  was carried out employing the Adams10 routine described in Appendix B.8. Both these differential equations and the Adams10 routine are incorporated into MaryLie. The number of integrations steps was 5000 for the cases  $a = 0.2$  and  $a = .02$ , and 10,000 for the case  $a = .002$ . Results are accurate to at least 10 significant figures.

Evidently the matrices for the three cases are not very different. Moreover, the  $f_3$  Lie generators (which describe second-order aberrations), those with indices 28 through 83, are comparable for the three cases. Both these behaviors are consistent with some sort of convergence occurring for the matrix and  $f_3$  entries as  $a \rightarrow 0$ . See Exercise 2.16. Examination of the equations of motion for these quantities when  $H_2$  and  $H_3$  are given by (2.91) and (2.92), see (10.5.32) and (10.5.60), shows that convergence is to be expected. However, the matter is delicate because the function  $b^{[0]}(z)$  that appears in  $H_2$  and hence also in  $H_3$  [see (2.94) and (2.95)] is discontinuous at  $z = 0$  and  $z = L$  in the limit  $a \rightarrow 0$ . (See also Exercise 2.2.) Therefore the assumptions of Theorems 1.3.1 and 1.3.2 are violated in the limit  $a \rightarrow 0$ . (Note that in this context  $a$  is a parameter.)

**Exhibit 16.2.3: Transfer map for the case  $a = 0.2$**

matrix for map is :

9.83483E-01	2.96811E+00	1.00099E-01	3.02094E-01	0.00000E+00	0.00000E+00
-7.58348E-03	9.83483E-01	-7.71845E-04	1.00099E-01	0.00000E+00	0.00000E+00
-1.00099E-01	-3.02094E-01	9.83483E-01	2.96811E+00	0.00000E+00	0.00000E+00
7.71845E-04	-1.00099E-01	-7.58348E-03	9.83483E-01	0.00000E+00	0.00000E+00
0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	1.00000E+00	1.23343E+00
0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	1.00000E+00

nonzero elements in generating polynomial are :

```

f( 33)=f( 20 00 01 )=-4.56819202124767E-03
f( 38)=f( 11 00 01 )= 1.16503508346131E-04
f( 45)=f( 10 01 01 )= 0.12049092663579
f( 53)=f( 02 00 01 )= -1.7728464321596
f( 57)=f( 01 10 01 )=-0.12049092663579
f( 67)=f( 00 20 01 )=-4.56819202124767E-03
f( 70)=f( 00 11 01 )= 1.16503508346112E-04
f( 76)=f( 00 02 01 )= -1.7728464321596
f( 83)=f( 00 00 03 )=-0.73260547246490
f( 84)=f( 40 00 00 )=-3.74253654402362E-03
f( 85)=f( 31 00 00 )= 2.24430345842169E-02
f( 87)=f( 30 01 00 )= 8.95982012441282E-04
f( 90)=f( 22 00 00 )=-5.26850707023760E-02
f( 91)=f( 21 10 00 )=-8.95982012441342E-04
f( 92)=f( 21 01 00 )=-2.48332996215296E-03
f( 95)=f( 20 20 00 )=-7.48507308804724E-03
f( 96)=f( 20 11 00 )= 2.24430345842169E-02
f( 99)=f( 20 02 00 )=-2.26497144701727E-02

```

```

f(104)=f( 20 00 02 )=-6.25873997450046E-03
f(105)=f( 13 00 00 )= 5.13684894940268E-02
f(106)=f( 12 10 00 )= 2.48332996215340E-03
f(107)=f( 12 01 00 )= 2.86233975277233E-02
f(110)=f( 11 20 00 )= 2.24430345842170E-02
f(111)=f( 11 11 00 )=-6.00707124644067E-02
f(114)=f( 11 02 00 )= 5.13684894940265E-02
f(119)=f( 11 00 02 )= 3.65870158733148E-04
f(121)=f( 10 21 00 )= 8.95982012441354E-04
f(124)=f( 10 12 00 )=-2.48332996215361E-03
f(130)=f( 10 03 00 )= 2.86233975277249E-02
f(135)=f( 10 01 02 )= 0.16398429796274
f(140)=f( 04 00 00 )=-0.39062543855982
f(141)=f( 03 10 00 )=-2.86233975277245E-02
f(145)=f( 02 20 00 )=-2.26497144701727E-02
f(146)=f( 02 11 00 )= 5.13684894940265E-02
f(149)=f( 02 02 00 )=-0.78125087711963
f(154)=f( 02 00 02 )= -2.4021742345697
f(155)=f( 01 30 00 )=-8.95982012441327E-04
f(156)=f( 01 21 00 )= 2.48332996215335E-03
f(159)=f( 01 12 00 )=-2.86233975277243E-02
f(164)=f( 01 10 02 )=-0.16398429796274
f(175)=f( 00 40 00 )=-3.74253654402362E-03
f(176)=f( 00 31 00 )= 2.24430345842170E-02
f(179)=f( 00 22 00 )=-5.26850707023762E-02
f(184)=f( 00 20 02 )=-6.25873997450046E-03
f(185)=f( 00 13 00 )= 5.13684894940276E-02
f(190)=f( 00 11 02 )= 3.65870158733149E-04
f(195)=f( 00 04 00 )=-0.39062543855982
f(200)=f( 00 02 02 )= -2.4021742345697
f(209)=f( 00 00 04 )=-0.93366303371890

```

Exhibit 16.2.4: Transfer map for the case a=0.02

matrix for map is :

9.79608E-01	2.96235E+00	1.00691E-01	3.04490E-01	0.00000E+00	0.00000E+00
-1.00980E-02	9.79608E-01	-1.03794E-03	1.00691E-01	0.00000E+00	0.00000E+00
-1.00691E-01	-3.04490E-01	9.79608E-01	2.96235E+00	0.00000E+00	0.00000E+00
1.03794E-03	-1.00691E-01	-1.00980E-02	9.79608E-01	0.00000E+00	0.00000E+00
0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	1.00000E+00	1.23343E+00
0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	1.00000E+00

nonzero elements in generating polynomial are :

```

f( 33)=f( 20 00 01 )=-6.10129602475707E-03
f( 38)=f( 11 00 01 )= 2.06122460889397E-04
f( 45)=f( 10 01 01 )= 0.12167479425840
f( 53)=f( 02 00 01 )= -1.7698863189223
f( 57)=f( 01 10 01 )=-0.12167479425840
f( 67)=f( 00 20 01 )=-6.10129602475707E-03
f( 70)=f( 00 11 01 )= 2.06122460889378E-04

```

```

f( 76)=f( 00 02 01 )= -1.7698863189223
f( 83)=f( 00 00 03 )=-0.73260547246490
f( 84)=f( 40 00 00 )=-3.82799704943991E-02
f( 85)=f( 31 00 00 )= 0.22920563229564
f( 87)=f( 30 01 00 )= 7.71671926080641E-04
f( 90)=f( 22 00 00 )=-0.56748370148935
f( 91)=f( 21 10 00 )=-7.71671926080904E-04
f( 92)=f( 21 01 00 )=-2.05451760035706E-03
f( 95)=f( 20 20 00 )=-7.65599409887983E-02
f( 96)=f( 20 11 00 )= 0.22920563229564
f( 99)=f( 20 02 00 )=-0.19593987204417
f(104)=f( 20 00 02 )=-8.37702532799553E-03
f(105)=f( 13 00 00 )= 0.66447262222389
f(106)=f( 12 10 00 )= 2.05451760035958E-03
f(107)=f( 12 01 00 )= 2.70686638900989E-02
f(110)=f( 11 20 00 )= 0.22920563229564
f(111)=f( 11 11 00 )=-0.74308765889035
f(114)=f( 11 02 00 )= 0.66447262222389
f(119)=f( 11 00 02 )= 6.47127530769082E-04
f(121)=f( 10 21 00 )= 7.71671926081047E-04
f(124)=f( 10 12 00 )=-2.05451760035932E-03
f(130)=f( 10 03 00 )= 2.70686638901053E-02
f(135)=f( 10 01 02 )= 0.16559550393811
f(140)=f( 04 00 00 )=-0.67707437134121
f(141)=f( 03 10 00 )=-2.70686638901064E-02
f(145)=f( 02 20 00 )=-0.19593987204417
f(146)=f( 02 11 00 )= 0.66447262222389
f(149)=f( 02 02 00 )= -1.3541487426824
f(154)=f( 02 00 02 )= -2.3947144417881
f(155)=f( 01 30 00 )=-7.71671926080822E-04
f(156)=f( 01 21 00 )= 2.05451760035805E-03
f(159)=f( 01 12 00 )=-2.70686638901048E-02
f(164)=f( 01 10 02 )=-0.16559550393811
f(175)=f( 00 40 00 )=-3.82799704943991E-02
f(176)=f( 00 31 00 )= 0.22920563229564
f(179)=f( 00 22 00 )=-0.56748370148935
f(184)=f( 00 20 02 )=-8.37702532799553E-03
f(185)=f( 00 13 00 )= 0.66447262222389
f(190)=f( 00 11 02 )= 6.47127530769100E-04
f(195)=f( 00 04 00 )=-0.67707437134121
f(200)=f( 00 02 02 )= -2.3947144417881
f(209)=f( 00 00 04 )=-0.93366303371890

```

Exhibit 16.2.5: Transfer map for the case  $a=0.002$

matrix for map is :

nonzero elements in generating polynomial are :

```

f( 33)=f( 20 00 01 )=-6.27795556809269E-03
f( 38)=f( 11 00 01 )= 2.16268873229983E-04
f( 45)=f( 10 01 01 )= 0.12168683913260
f( 53)=f( 02 00 01 )= -1.7696108701855
f( 57)=f( 01 10 01 )=-0.12168683913260
f( 67)=f( 00 20 01 )=-6.27795556809270E-03
f( 70)=f( 00 11 01 )= 2.16268873229994E-04
f( 76)=f( 00 02 01 )= -1.7696108701855
f( 83)=f( 00 00 03 )=-0.73260547246506
f( 84)=f( 40 00 00 )=-0.38235532452970
f( 85)=f( 31 00 00 )= 2.2888185879143
f( 87)=f( 30 01 00 )= 8.00590462204597E-04
f( 90)=f( 22 00 00 )= -5.7102701093391
f( 91)=f( 21 10 00 )=-8.00590462205045E-04
f( 92)=f( 21 01 00 )=-2.13388593660145E-03
f( 95)=f( 20 20 00 )=-0.76471064905939
f( 96)=f( 20 11 00 )= 2.2888185879143
f( 99)=f( 20 02 00 )= -1.9103970318647
f(104)=f( 20 00 02 )=-8.62088196355748E-03
f(105)=f( 13 00 00 )= 6.8352404267418
f(106)=f( 12 10 00 )= 2.13388593640862E-03
f(107)=f( 12 01 00 )= 2.71197052999494E-02
f(110)=f( 11 20 00 )= 2.2888185879143
f(111)=f( 11 11 00 )= -7.5997461549496
f(114)=f( 11 02 00 )= 6.8352404267445
f(119)=f( 11 00 02 )= 6.78947912878182E-04
f(121)=f( 10 21 00 )= 8.00590462204979E-04
f(124)=f( 10 12 00 )=-2.13388593639684E-03
f(130)=f( 10 03 00 )= 2.71197052987057E-02
f(135)=f( 10 01 02 )= 0.16561189662672
f(140)=f( 04 00 00 )= -3.5930682684087
f(141)=f( 03 10 00 )=-2.71197052987419E-02
f(145)=f( 02 20 00 )= -1.9103970318647
f(146)=f( 02 11 00 )= 6.8352404267445
f(149)=f( 02 02 00 )= -7.1861365368258
f(154)=f( 02 00 02 )= -2.3940220818360
f(155)=f( 01 30 00 )=-8.00590462202088E-04
f(156)=f( 01 21 00 )= 2.13388593658432E-03
f(159)=f( 01 12 00 )=-2.71197052999558E-02
f(164)=f( 01 10 02 )=-0.16561189662672
f(175)=f( 00 40 00 )=-0.38235532452970
f(176)=f( 00 31 00 )= 2.2888185879143
f(179)=f( 00 22 00 )= -5.7102701093391
f(184)=f( 00 20 02 )=-8.62088196355748E-03
f(185)=f( 00 13 00 )= 6.8352404267418
f(190)=f( 00 11 02 )= 6.78947912878175E-04
f(195)=f( 00 04 00 )= -3.5930682684087
f(200)=f( 00 02 02 )= -2.3940220818360
f(209)=f( 00 00 04 )=-0.93366303371922

```

What can be said about the  $f_4$  Lie generators (which describe third-order aberrations), those with indices 84 through 209? Some of them are quite different for the three values of  $a$ . For example, the values of  $f(84)$ , which are the coefficients of  $X^4$  in the  $f_4$  Lie generators for the three cases, are quite different. Examination shows that these values are  $f(84) = -3.74253654402362E - 03$ ,  $f(84) = -3.82799704943991E - 02$ , and  $f(84) = -0.38235532452970$  for the cases  $a = 0.2$ ,  $a = .02$ , and  $a = .002$ , respectively. This behavior suggests the coefficient of  $X^4$  is *diverging* (in magnitude) to  $\infty$  as  $a \rightarrow 0$ . The same is true of some of the other  $f_4$  entries. Indeed, it can be illustrated numerically and demonstrated analytically that the divergent  $f_4$  entries behave as  $1/a$  as  $a \rightarrow 0$  so that, for example, the product  $af(84)$  approaches a *constant* as  $a \rightarrow 0$ . Table 2.1 below illustrates this divergence/behavior for the case of  $f(84)$ .

Table 16.2.1: Numerical behavior of  $f(84)$  for small values of  $a$ .

$a$	$f(84)$	$af(84)$
.2	-3.7425E-3	-7.4850E-4
.02	-3.8279E-2	-7.6558E-4
.002	-3.8235E-1	-7.6470E-4

We have made a preliminary study of the  $a \rightarrow 0$  behavior of the transfer map  $\mathcal{M}$  for a simple solenoid. In the rest of this subsection we will examine the matter in greater detail.

### 16.2.6.2 Behavior of Linear Part

#### 16.2.6.2.1 Factorization into Three $a$ Dependent Maps/Matrices

Something more can be said about  $M$ , the matrix for the *linear* part of  $\mathcal{M}$ , if one attempts to form hard-edge limits/approximations. Let  $\mathcal{M}_{-1 \rightarrow 2}$  be the transfer map between the planes  $z = -1$  and  $z = 2$ , respectively. It is the map displayed in Exhibits 2.3 through 2.5 for three different values of  $a$ . Also, employing analogous notation, consider the maps  $\mathcal{M}_{-1 \rightarrow 0}$ ,  $\mathcal{M}_{0 \rightarrow 1}$ , and  $\mathcal{M}_{1 \rightarrow 2}$ . Then we have the relation

$$\mathcal{M}_{-1 \rightarrow 2} = \mathcal{M}_{-1 \rightarrow 0} \mathcal{M}_{0 \rightarrow 1} \mathcal{M}_{1 \rightarrow 2}. \quad (16.2.97)$$

Next, let  $\mathcal{D}$  be the map for a drift of length 1 meter. Employ this map to define implicitly two other maps  $\mathcal{M}^{\text{lff}}$  and  $\mathcal{M}^{\text{tff}}$  by writing

$$\mathcal{M}_{-1 \rightarrow 0} = \mathcal{D} \mathcal{M}^{\text{lff}} \quad (16.2.98)$$

and

$$\mathcal{M}_{1 \rightarrow 2} = \mathcal{M}^{\text{tff}} \mathcal{D}. \quad (16.2.99)$$

Then, particularly when  $a$  is small, we may view  $\mathcal{M}^{\text{lff}}$  and  $\mathcal{M}^{\text{tff}}$  as *leading* and *trailing fringe-field* maps. Of course, (2.98) and (2.99) can be solved to give the explicit definitions

$$\mathcal{M}^{\text{lff}} = \mathcal{D}^{-1} \mathcal{M}_{-1 \rightarrow 0} \quad (16.2.100)$$

$$\mathcal{M}^{\text{tf}} = \mathcal{M}_{1 \rightarrow 2} \mathcal{D}^{-1}. \quad (16.2.101)$$

Also, make the definition

$$\mathcal{M}^{\text{body}} = \mathcal{M}_{0 \rightarrow 1}. \quad (16.2.102)$$

Then we have the factorization

$$\mathcal{M}_{-1 \rightarrow 2} = \mathcal{D} \mathcal{M}^{\text{lf}} \mathcal{M}^{\text{body}} \mathcal{M}^{\text{tf}} \mathcal{D}. \quad (16.2.103)$$

Note that all these maps are symplectic. Note also that, as illustrated in Figure 2.5, when  $a$  is sufficiently small, the map  $\mathcal{M}_{-1 \rightarrow 2}$  essentially describes transport through a 1 meter drift followed by transport through a 1 meter solenoid followed by transport through a final 1 meter drift. We therefore expect, when  $a$  is sufficiently small, that  $\mathcal{M}^{\text{lf}}$  and  $\mathcal{M}^{\text{tf}}$  will describe leading and trailing fringe-field effects *outside* the solenoid, and  $\mathcal{M}^{\text{body}}$  will describe all effects occurring *within* the solenoid itself.

Finally, in view of (2.103), we make the definition

$$\mathcal{M}_{\text{solenoid}} = \mathcal{M}^{\text{lf}} \mathcal{M}^{\text{body}} \mathcal{M}^{\text{tf}}. \quad (16.2.104)$$

Note that  $\mathcal{M}_{\text{solenoid}}$  has been factored into *three*  $a$  dependent maps. Correspondingly for the associated linear parts there will be the relation

$$M_{\text{solenoid}} = M^{\text{tf}} M^{\text{body}} M^{\text{lf}}. \quad (16.2.105)$$

The matrix  $M_{\text{solenoid}}$  has also been factorized into three  $a$  dependent matrices.

From the previous discussion we expect that some of the third- and higher-order aberration parts of the maps  $\mathcal{M}^{\text{lf}}$ ,  $\mathcal{M}^{\text{tf}}$ , and  $\mathcal{M}^{\text{body}}$  may diverge as  $a \rightarrow 0$ . But for now let us examine the linear/matrix parts of the maps  $\mathcal{M}^{\text{lf}}$ ,  $\mathcal{M}^{\text{body}}$ , and  $\mathcal{M}^{\text{tf}}$  in the hard-edge limit  $a \rightarrow 0$ . We begin with the maps  $\mathcal{M}^{\text{lf}}$  and  $\mathcal{M}^{\text{tf}}$ . Exhibits 2.6 through 2.9 show the matrices associated with these maps for the values  $a = 0.2$  and  $a = 0.02$ .

Exhibit 16.2.6: The matrix  $M^{\text{lf}}$  for the case  $a = 0.2$

$$\begin{array}{ccccccc} 9.99944E-01 & 3.31737E-06 & 8.72573E-03 & 2.89481E-08 & 0.00000E+00 & 0.00000E+00 \\ -2.11420E-04 & 9.99980E-01 & -1.84490E-06 & 8.72604E-03 & 0.00000E+00 & 0.00000E+00 \\ -8.72573E-03 & -2.89481E-08 & 9.99944E-01 & 3.31737E-06 & 0.00000E+00 & 0.00000E+00 \\ 1.84490E-06 & -8.72604E-03 & -2.11420E-04 & 9.99980E-01 & 0.00000E+00 & 0.00000E+00 \\ 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 1.00000E+00 & 0.00000E+00 \\ 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 1.00000E+00 \end{array}$$

Exhibit 16.2.7: The matrix  $M^{\text{lf}}$  for the case  $a = 0.02$

$$\begin{array}{ccccccc} 9.99999E-01 & 4.78327E-09 & 1.00901E-03 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 \\ -2.24996E-05 & 1.00000E+00 & -2.27024E-08 & 1.00901E-03 & 0.00000E+00 & 0.00000E+00 \\ -1.00901E-03 & 0.00000E+00 & 9.99999E-01 & 4.78327E-09 & 0.00000E+00 & 0.00000E+00 \\ 2.27024E-08 & -1.00901E-03 & -2.24996E-05 & 1.00000E+00 & 0.00000E+00 & 0.00000E+00 \\ 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 1.00000E+00 & 0.00000E+00 \\ 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 1.00000E+00 \end{array}$$

Exhibit 16.2.8: The matrix  $M^{\text{tf}}$  for the case  $a = 0.2$

```
9.99980E-01 3.31737E-06 8.72604E-03 2.89481E-08 0.00000E+00 0.00000E+00
-2.11420E-04 9.99944E-01 -1.84490E-06 8.72573E-03 0.00000E+00 0.00000E+00
-8.72604E-03 -2.89481E-08 9.99980E-01 3.31737E-06 0.00000E+00 0.00000E+00
1.84490E-06 -8.72573E-03 -2.11420E-04 9.99944E-01 0.00000E+00 0.00000E+00
0.00000E+00 0.00000E+00 0.00000E+00 0.00000E+00 1.00000E+00 0.00000E+00
0.00000E+00 0.00000E+00 0.00000E+00 0.00000E+00 0.00000E+00 1.00000E+00
```

Exhibit 16.2.9: The matrix  $M^{\text{tf}}$  for the case  $a = 0.02$

1.00000E+00	4.78328E-09	1.00901E-03	0.00000E+00	0.00000E+00	0.00000E+00
-2.24996E-05	9.99999E-01	-2.27024E-08	1.00901E-03	0.00000E+00	0.00000E+00
-1.00901E-03	0.00000E+00	1.00000E+00	4.78328E-09	0.00000E+00	0.00000E+00
2.27024E-08	-1.00901E-03	-2.24996E-05	9.99999E-01	0.00000E+00	0.00000E+00
0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	1.00000E+00	0.00000E+00
0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	1.00000E+00

Upon comparing Exhibits 2.6 and 2.7 we see that there appears to be the limiting behavior

$$\lim_{a \rightarrow 0} M^{\text{tf}} = I. \quad (16.2.106)$$

And, upon comparing Exhibits 2.8 and 2.9, we see that there appears to be the limiting behavior

$$\lim_{a \rightarrow 0} M^{\text{tf}} = I. \quad (16.2.107)$$

Thus it appears that, for a solenoid in the hard-edge limit, there are *no* effects on the *linear* part of the transfer map due to fringe fields outside the solenoid. These results can be proved analytically using the  $H_2$  given by (2.91) to compute  $M_{-1 \rightarrow 0}$  and  $M_{1 \rightarrow 2}$  since in the hard-edge limit  $b^{[0]}(z)$  vanishes for  $z < 0$  and  $z > L = 1$ , and therefore the resulting  $M$  for such computations will simply be that for a 1 meter drift.

What can be said about the linear part of  $\mathcal{M}^{\text{body}}$ ? Exhibits 2.10 and 2.11 show the matrices associated with these maps for the values  $a = 0.2$  and  $a = 0.02$ .

Exhibit 16.2.10: The matrix  $M^{\text{body}}$  for the case  $a = 0.2$

9.92886E-01	9.95137E-01	8.35777E-02	8.37672E-02	0.00000E+00	0.00000E+00
-7.17625E-03	9.92886E-01	-6.04072E-04	8.35777E-02	0.00000E+00	0.00000E+00
-8.35777E-02	-8.37672E-02	9.92886E-01	9.95137E-01	0.00000E+00	0.00000E+00
6.04072E-04	-8.35777E-02	-7.17625E-03	9.92886E-01	0.00000E+00	0.00000E+00
0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	1.00000E+00	4.11143E-01
0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	1.00000E+00

Exhibit 16.2.11: The matrix  $M^{\text{body}}$  for the case  $a = 0.02$

9.89931E-01	9.93237E-01	9.97335E-02	1.00067E-01	0.00000E+00	0.00000E+00
-1.00555E-02	9.89931E-01	-1.01307E-03	9.97335E-02	0.00000E+00	0.00000E+00
-9.97335E-02	-1.00067E-01	9.89931E-01	9.93237E-01	0.00000E+00	0.00000E+00
1.01307E-03	-9.97335E-02	-1.00555E-02	9.89931E-01	0.00000E+00	0.00000E+00
0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	1.00000E+00	4.11143E-01
0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	1.00000E+00

Comparison of the matrices in Exhibits 2.10 and 2.11 shows that some sort of limit also appears to be approached by  $M^{\text{body}}$  as  $a \rightarrow 0$ . But what is this limit? Exhibit 2.12 shows  $M^{\text{uniform}}$ , the matrix computed (numerically) using the Hamiltonian  $H_2$  given by (2.91) with  $b^{[0]}(z)$  having a *constant* value. (This matrix can be computed analytically as well as numerically. The analytic result is that quoted in Section 13.4. See Exercise 2.16.)

Examination of the matrices in the Exhibits 2.10 through 2.12 shows that there appears to be the limiting behavior

$$\lim_{a \rightarrow 0} M^{\text{body}} = M^{\text{uniform}}. \quad (16.2.108)$$

This result can be proved analytically using the  $H_2$  given by (2.91) to compute  $M_{0 \rightarrow 1}$  since in the hard-edge limit  $b^{[0]}(z)$  is constant for  $z$  in the *open* interval  $z \in (0, 1)$ .

Exhibit 16.2.12: The matrix  $M^{\text{uniform}}$  for the case of a uniform field

$$\begin{array}{ccccccc} 9.89543E-01 & 9.93019E-01 & 1.01722E-01 & 1.02079E-01 & 0.00000E+00 & 0.00000E+00 \\ -1.04201E-02 & 9.89543E-01 & -1.07116E-03 & 1.01722E-01 & 0.00000E+00 & 0.00000E+00 \\ -1.01722E-01 & -1.02079E-01 & 9.89543E-01 & 9.93019E-01 & 0.00000E+00 & 0.00000E+00 \\ 1.07116E-03 & -1.01722E-01 & -1.04201E-02 & 9.89543E-01 & 0.00000E+00 & 0.00000E+00 \\ 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 1.00000E+00 & 4.11143E-01 \\ 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 0.00000E+00 & 1.00000E+00 \end{array}$$

From (2.106) through (2.108) we conclude that in the hard-edge limit there are *no* fringe-field contributions to the *linear* part of the transfer map for a solenoid. That is, there is the limiting result

$$\lim_{a \rightarrow 0} M_{\text{solenoid}} = M^{\text{uniform}}. \quad (16.2.109)$$

(Curiously, this is the same result as that found by neglecting fringe fields entirely!) Some other authors have also reached the same conclusion by other methods. However we hasten to emphasize, as we have already seen, that in the hard-edge limit there are disastrous fringe-field effects for some third- and higher-order aberrations. It is therefore highly desirable, in the case of a solenoid, to treat fringe-field effects with care (which must be done numerically) using realistic profiles  $b^{[n]}(z)$ .

Yet other authors provide formulas for matrices  $M_{\text{fringe}}$  and  $M_{\text{longitudinal}}$  which are meant to play roles analogous to  $M^{\text{lff}}$ ,  $M^{\text{tf}}$ , and  $M^{\text{body}}$ . These matrices differ from those given by the linear parts of (2.100) through (2.102); and their limiting values differ from those given by (2.106) through (2.108). They are also *not* symplectic. However, when  $a = 0$ , their product does give the symplectic result

$$M_{\text{fringe}} M_{\text{longitudinal}} M_{\text{fringe}}^{-1} = M^{\text{uniform}}. \quad (16.2.110)$$

Based on the results of these authors one might be tempted to conclude that, at least in some way, fringe fields even in the hard-edge limit do play some role in determining the linear part of the transfer map for a solenoid. [Note however that their net effect cancels out because of the result (2.110).] The explanation for this confusing circumstance is that these authors employ in essence mechanical rather than canonical momenta in their calculations. So doing is expected to yield nonsymplectic results in the presence of magnetic fields. Recall Exercise 1.7.5. But these nonsymplectic results can ultimately cancel, as in (2.110), when a full/complete calculation is made providing the magnetic field vanishes before entry into and after exit from the solenoid. See Exercise 6.4.11.

### 16.2.6.2.2 Factorization Involving Only Two $a$ Dependent Maps/Matrices

There is another variation on the theme we have been exploring. Suppose instead of (2.103) we attempt an Ansatz of the form

$$\mathcal{M}_{-1 \rightarrow 2} = \mathcal{D} \mathcal{M}^{\text{LFF}} \mathcal{M}^{\text{uniform}} \mathcal{M}^{\text{TFF}} \mathcal{D}. \quad (16.2.111)$$

Here  $\mathcal{M}^{\text{uniform}}$  is the map computed for a length of 1 meter using the Hamiltonian  $H$  given by (2.88) through (2.93) with  $b^{[0]}$  having a *constant* value and correspondingly  $b^{[n]} = 0$  for  $n > 0$ . (For future reference, we will call this Hamiltonian  $H^{\text{uniform}}$ .) If the Ansatz (2.111) is successful, the maps  $\mathcal{M}^{\text{LFF}}$  and  $\mathcal{M}^{\text{TFF}}$  will describe *both* the effects arising from the depletion of the field within the body of the solenoid and the effects of the fields that extend beyond the ends of the solenoid. In view of (2.111) we make the definition

$$\mathcal{M}_{\text{solenoid}} = \mathcal{M}^{\text{LFF}} \mathcal{M}^{\text{uniform}} \mathcal{M}^{\text{TFF}}. \quad (16.2.112)$$

Assuming success of the Ansatz (2.111), we see that  $\mathcal{M}_{\text{solenoid}}$  has been factorized in a way that involves only *two*  $a$  dependent maps, namely  $\mathcal{M}^{\text{LFF}}$  and  $\mathcal{M}^{\text{TFF}}$ , and one  $a$  independent map, namely  $\mathcal{M}^{\text{uniform}}$ . Correspondingly for the associated linear parts there will be the relation

$$M_{\text{solenoid}} = M^{\text{TFF}} M^{\text{uniform}} M^{\text{LFF}}. \quad (16.2.113)$$

The matrix  $M_{\text{solenoid}}$  has been factorized in a way that involves only two  $a$  dependent matrices and one  $a$  independent matrix

Let us pause momentarily at this point to compare the factorizations (2.104) and (2.112). The map  $\mathcal{M}_{\text{solenoid}}$  is the same in both. [Correspondingly, the matrices  $M_{\text{solenoid}}$  given by (2.105) and (2.113) are the same.] But (2.104) may be viewed as a kind of *local* factorization in that it treats *separately* effects that occur before, within, and after the body of the solenoid. By contrast (2.112) may be viewed as a *nonlocal/lumped* factorization in that  $\mathcal{M}^{\text{LFF}}$  describes effects that occur both before and after entry into the body of the solenoid and  $\mathcal{M}^{\text{TFF}}$  describes effects that occur both within the body of the solenoid and after exit from the body of the solenoid. No attempt is made to describe separately what occurs only within the body of the solenoid itself.<sup>4</sup>

To continue, how can we determine the maps  $\mathcal{M}^{\text{LFF}}$  and  $\mathcal{M}^{\text{TFF}}$ ? Let  $\mathcal{H}$  be the map of a uniform “half” solenoid, the map computed for a uniform solenoid with a length of 1/2 meter using the Hamiltonian  $H$  given by (2.88) through (2.93) with  $b^{[0]}$  having a constant value and correspondingly  $b^{[n]} = 0$  for  $n > 0$ . Then evidently there will be the relation

$$\mathcal{M}^{\text{uniform}} = \mathcal{H} \mathcal{H} \quad (16.2.114)$$

and the Ansatz (2.111) becomes

$$\mathcal{M}_{-1 \rightarrow 2} = \mathcal{D} \mathcal{M}^{\text{LFF}} \mathcal{H} \mathcal{H} \mathcal{M}^{\text{TFF}} \mathcal{D}. \quad (16.2.115)$$

---

<sup>4</sup>Strictly speaking, the factorization (2.104) is not completely local since, according to the definition (2.98),  $\mathcal{M}^{\text{lf}}$  lumps together at the end of the leading drift all the fringe-field effects that have accumulated prior to the body of the solenoid; and, according to (2.99),  $\mathcal{M}^{\text{tf}}$  lumps together at the beginning of the trailing drift all the fringe-field effects that will accumulate after the body of the solenoid.



Exhibit 16.2.14: The matrix  $M^{\text{LFF}}$  for the case  $a = 0.02$

Exhibit 16.2.15: The matrix  $M^{\text{LFF}}$  for the case  $a = 0.002$

1.00000E+00	-1.10603E-08	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
1.64409E-05	1.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
0.00000E+00	0.00000E+00	1.00000E+00	-1.10603E-08	0.00000E+00	0.00000E+00
0.00000E+00	0.00000E+00	1.64409E-05	1.00000E+00	0.00000E+00	0.00000E+00
0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	1.00000E+00	0.00000E+00
0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	1.00000E+00

Exhibit 16.2.16: The matrix  $M^{\text{TFF}}$  for the case  $a = 0.2$

Exhibit 16.2.17: The matrix  $M^{\text{TFF}}$  for the case  $a = 0.02$

Exhibit 16.2.18: The matrix  $M^{\text{TFF}}$  for the case  $a = 0.002$

$1.00000E+00$	$-1.10596E-08$	$0.00000E+00$	$0.00000E+00$	$0.00000E+00$	$0.00000E+00$
$1.64409E-05$	$1.00000E+00$	$0.00000E+00$	$0.00000E+00$	$0.00000E+00$	$0.00000E+00$
$0.00000E+00$	$0.00000E+00$	$1.00000E+00$	$-1.10596E-08$	$0.00000E+00$	$0.00000E+00$
$0.00000E+00$	$0.00000E+00$	$1.64409E-05$	$1.00000E+00$	$0.00000E+00$	$0.00000E+00$
$0.00000E+00$	$0.00000E+00$	$0.00000E+00$	$0.00000E+00$	$1.00000E+00$	$0.00000E+00$
$0.00000E+00$	$0.00000E+00$	$0.00000E+00$	$0.00000E+00$	$0.00000E+00$	$1.00000E+00$

From Exhibits 2.13 through 2.15 we infer that there is the limiting behavior

$$\lim_{a \rightarrow 0} M^{\text{LFF}} = I. \quad (16.2.121)$$

And from Exhibits 2.16 through 2.18 we infer that there is the limiting behavior

$$\lim_{a \rightarrow 0} M^{\text{TFF}} = I. \quad (16.2.122)$$

These limiting behaviors can also be verified analytically, and are to be expected. Again we conclude that in the hard-edge limit there are *no* fringe-field contributions to the *linear* part of the transfer map for a solenoid. That is, (2.109) again holds.

There is some irony here. We have seen that the matrices for the linear parts of  $\mathcal{M}^{\text{LFF}}$  and  $\mathcal{M}^{\text{TFF}}$  have the benign limiting behavior (2.121) and (2.122). We also know from (2.112) that the only  $a$  dependence in  $\mathcal{M}_{\text{solenoid}}$  arises from that in  $\mathcal{M}^{\text{LFF}}$  and  $\mathcal{M}^{\text{TFF}}$ , and we have seen that  $\mathcal{M}_{\text{solenoid}}$  has some divergent third-order aberrations as  $a \rightarrow 0$ . We conclude that while the limiting behavior of the linear parts of  $\mathcal{M}^{\text{LFF}}$  and  $\mathcal{M}^{\text{TFF}}$  is benign, that of some of the nonlinear parts of  $\mathcal{M}^{\text{LFF}}$  and  $\mathcal{M}^{\text{TFF}}$  is pathological.

What can be said about the linear parts of  $\mathcal{M}^{\text{LFF}}$  and  $\mathcal{M}^{\text{TFF}}$  when  $a$  is small but *nonzero*? From Exhibits 2.13 through 2.15 we see that, for small  $a$ , the effect of the leading fringe field is to produce *identical defocussing* in both planes. And from Exhibits 2.16 through 2.18 we see that the same is true for the effect of the trailing fringe field.<sup>5</sup> The modest and small  $a$  effect of fringe fields is to *decrease* the focussing effect of a solenoid compared to that predicted by  $M^{\text{uniform}}$ . Recall (2.113).

Moreover, as illustrated in Table 2.2 below, for small  $a$  the defocussing strength behaves linearly in  $a$ . That is, for example, the quantity  $M_{21}^{\text{LFF}} = M_{43}^{\text{LFF}}$  is proportional to  $a$  when  $a$  is sufficiently small so that the product  $(1/a)M_{21}^{\text{LFF}}$  approaches a *constant* as  $a \rightarrow 0$ .<sup>6</sup> Identical results hold for  $M_{21}^{\text{TFF}} = M_{43}^{\text{TFF}}$ .

Finally, we remark that these results are consistent with those obtained by some other authors using other methods.

---

<sup>5</sup>In passing we also note that, for a given value of  $a$  and within the announced numerical accuracy, the matrices  $M^{\text{LFF}}$  and  $M^{\text{TFF}}$  differ only by permutations of various *diagonal* entries. This result is a consequence of *reversal symmetry*. That is, for a given value of  $a$ , the matrices  $M^{\text{LFF}}$  and  $M^{\text{TFF}}$  are *reverses* of each other. See Chapter 36.

<sup>6</sup>For modest values of  $a$  such as  $a = 0.2$ , and still larger values of  $a$ , there are some effects on  $M_{21}^{\text{LFF}}$  that arise from the approximation we have made that there are no leading external fringe-field effects before  $z^{\text{en}} = -1$ . Similarly there are some effects on  $M_{21}^{\text{TFF}}$  that arise from the approximation we have made that there are no trailing external fringe-field effects after  $z^{\text{ex}} = 2$ . These effects disappear as  $a \rightarrow 0$  because then external fringe fields become more and more confined to the vicinities of the entrance and exit of the solenoid. As an indication of the size of these effects, suppose the leading external fringe field region is taken to begin at a larger negative value of  $z$ , for example  $z = z^{\text{en}} = -10$ . (Of course, in this case we must also employ for  $\mathcal{D}$  the map of a 10 meter drift.) Then, for  $a = 0.2$ , there are the results  $M_{21}^{\text{LFF}} = 1.42982\text{E-}03$  and  $(1/a)M_{21}^{\text{LFF}} = 7.14910\text{E-}03$ , which are to be compared with those in the first line of Table 2.2. Evidently in this case the effects of the approximations we have made are small.

Table 16.2.2: Numerical behavior of  $M_{21}^{\text{LFF}}$  for small values of  $a$ .

$a$	$M_{21}^{\text{LFF}}$	$(1/a)M_{21}^{\text{LFF}}$
.2	1.42997E-03	7.14985E-03
.02	1.62696E-04	8.13480E-03
.002	1.64409E-05	8.22045E-03

### 16.2.6.3 Behavior of Nonlinear Part

We have concluded that, while the limiting behavior of the linear parts of  $\mathcal{M}^{\text{LFF}}$  and  $\mathcal{M}^{\text{TFF}}$  is benign, that of some of the nonlinear parts of  $\mathcal{M}^{\text{LFF}}$  and  $\mathcal{M}^{\text{TFF}}$  is pathological. In this subsection we will examine in more detail the behavior of the nonlinear part of  $\mathcal{M}^{\text{LFF}}$  in the limit  $a \rightarrow 0$ . For brevity, we will not present the behavior of  $\mathcal{M}^{\text{TFF}}$ . But, as expected, it is found to be analogous to that of  $\mathcal{M}^{\text{LFF}}$ .

Exhibits 2.19 through 2.21 display (through third order) the maps  $\mathcal{M}^{\text{LFF}}$  given by (2.119) for the cases  $a = 0.2$ ,  $a = 0.02$ , and  $a = 0.002$ , respectively. For these exhibits the rotational parts of the solenoidal maps have *not* been removed.

Exhibit 16.2.19: The map  $\mathcal{M}^{\text{LFF}}$  for the case  $a = 0.2$

matrix for map is :

```

9.99702E-01 -7.47109E-05 -9.30614E-06 6.95479E-10 0.00000E+00 0.00000E+00
1.42984E-03 1.00030E+00 -1.33102E-08 -9.31170E-06 0.00000E+00 0.00000E+00
9.30614E-06 -6.95479E-10 9.99702E-01 -7.47109E-05 0.00000E+00 0.00000E+00
1.33102E-08 9.31170E-06 1.42984E-03 1.00030E+00 0.00000E+00 0.00000E+00
0.00000E+00 0.00000E+00 0.00000E+00 0.00000E+00 1.00000E+00 0.00000E+00
0.00000E+00 0.00000E+00 0.00000E+00 0.00000E+00 0.00000E+00 1.00000E+00

```

nonzero elements in generating polynomial are :

```

f( 33)=f( 20 00 01 )= 8.48173754616857E-04
f( 38)=f( 11 00 01 )= 7.08661171957406E-04
f( 45)=f( 10 01 01 )=-1.10582080947857E-05
f( 53)=f( 02 00 01 )= 1.33079847471684E-04
f( 57)=f( 01 10 01 )= 1.10582080948066E-05
f( 67)=f( 00 20 01 )= 8.48173754616855E-04
f( 70)=f( 00 11 01 )= 7.08661171957389E-04
f( 76)=f( 00 02 01 )= 1.33079847476181E-04
f( 84)=f( 40 00 00 )=-1.86910737357663E-03
f( 85)=f( 31 00 00 )=-2.23275870608366E-03
f( 87)=f( 30 01 00 )=-1.51855928877267E-04
f( 90)=f( 22 00 00 )=-3.06010303211954E-04
f( 91)=f( 21 10 00 )= 1.51855928877281E-04
f( 92)=f( 21 01 00 )= 2.37154993794211E-02
f( 95)=f( 20 20 00 )=-3.73821474715327E-03
f( 96)=f( 20 11 00 )=-2.23275870608358E-03

```

```

f( 99)=f( 20 02 00 )= 8.51691896034609E-04
f(104)=f( 20 00 02 )= 1.15254677042892E-03
f(105)=f( 13 00 00 )= 1.26460888238766E-04
f(106)=f( 12 10 00 )=-2.37154993794227E-02
f(107)=f( 12 01 00 )=-9.17588458782363E-04
f(110)=f( 11 20 00 )=-2.23275870608357E-03
f(111)=f( 11 11 00 )=-2.31540439851645E-03
f(114)=f( 11 02 00 )= 1.26460888456775E-04
f(119)=f( 11 00 02 )= 1.38437434049872E-03
f(121)=f( 10 21 00 )=-1.51855928877260E-04
f(124)=f( 10 12 00 )= 2.37154993794226E-02
f(130)=f( 10 03 00 )=-9.17588458807038E-04
f(135)=f( 10 01 02 )=-1.50498675709195E-05
f(140)=f( 04 00 00 )= 4.36074169958001E-05
f(141)=f( 03 10 00 )= 9.17588458803895E-04
f(145)=f( 02 20 00 )= 8.51691896035069E-04
f(146)=f( 02 11 00 )= 1.26460888454029E-04
f(149)=f( 02 02 00 )= 8.72148317848653E-05
f(154)=f( 02 00 02 )= 3.39093545585142E-04
f(155)=f( 01 30 00 )= 1.51855928877314E-04
f(156)=f( 01 21 00 )=-2.37154993794229E-02
f(159)=f( 01 12 00 )= 9.17588458801365E-04
f(164)=f( 01 10 02 )= 1.50498675709225E-05
f(175)=f( 00 40 00 )=-1.86910737357663E-03
f(176)=f( 00 31 00 )=-2.23275870608367E-03
f(179)=f( 00 22 00 )=-3.06010303212333E-04
f(184)=f( 00 20 02 )= 1.15254677042892E-03
f(185)=f( 00 13 00 )= 1.26460888246612E-04
f(190)=f( 00 11 02 )= 1.38437434049872E-03
f(195)=f( 00 04 00 )= 4.36074169358758E-05
f(200)=f( 00 02 02 )= 3.39093545579732E-04

```

Exhibit 16.2.20: The map  $\mathcal{M}^{\text{LFF}}$  for the case  $a = 0.02$

matrix for map is :

9.99992E-01	-1.22277E-06	-9.20989E-08	0.00000E+00	0.00000E+00	0.00000E+00
1.62716E-04	1.00001E+00	0.00000E+00	-9.21003E-08	0.00000E+00	0.00000E+00
9.20989E-08	0.00000E+00	9.99992E-01	-1.22277E-06	0.00000E+00	0.00000E+00
0.00000E+00	9.21003E-08	1.62716E-04	1.00001E+00	0.00000E+00	0.00000E+00
0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	1.00000E+00	0.00000E+00
0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	1.00000E+00

nonzero elements in generating polynomial are :

```

f( 33)=f( 20 00 01 )= 9.66314417891064E-05
f( 38)=f( 11 00 01 )= 1.85289937516709E-05
f( 45)=f( 10 01 01 )=-1.09406551700164E-07
f( 53)=f( 02 00 01 )= 2.17830287646548E-06
f( 57)=f( 01 10 01 )= 1.09406551727920E-07
f( 67)=f( 00 20 01 )= 9.66314417891068E-05
f( 70)=f( 00 11 01 )= 1.85289937516412E-05

```

```

f( 76)=f( 00 02 01 )= 2.17830287385645E-06
f( 84)=f( 40 00 00 )=-1.93149067689109E-02
f( 85)=f( 31 00 00 )=-2.61905549343168E-03
f( 87)=f( 30 01 00 )=-1.86391130390008E-05
f( 90)=f( 22 00 00 )=-9.05018682344674E-05
f( 91)=f( 21 10 00 )= 1.86391130388983E-05
f( 92)=f( 21 01 00 )= 2.55870298344304E-02
f( 95)=f( 20 20 00 )=-3.86298135378219E-02
f( 96)=f( 20 11 00 )=-2.61905549343095E-03
f( 99)=f( 20 02 00 )= 7.83183807162944E-05
f(104)=f( 20 00 02 )= 1.31485684215657E-04
f(105)=f( 13 00 00 )=-7.58534145784010E-07
f(106)=f( 12 10 00 )=-2.55870298344239E-02
f(107)=f( 12 01 00 )=-1.03168131481354E-05
f(110)=f( 11 20 00 )=-2.61905549343097E-03
f(111)=f( 11 11 00 )=-3.37640497875350E-04
f(114)=f( 11 02 00 )=-7.58534755148649E-07
f(119)=f( 11 00 02 )= 3.62090101634249E-05
f(121)=f( 10 21 00 )=-1.86391130388087E-05
f(124)=f( 10 12 00 )= 2.55870298344245E-02
f(130)=f( 10 03 00 )=-1.03168130791351E-05
f(135)=f( 10 01 02 )=-1.48898803105739E-07
f(140)=f( 04 00 00 )= 5.20628154886127E-07
f(141)=f( 03 10 00 )= 1.03168130669920E-05
f(145)=f( 02 20 00 )= 7.83183807188067E-05
f(146)=f( 02 11 00 )=-7.58534727284429E-07
f(149)=f( 02 02 00 )= 1.04126206200428E-06
f(154)=f( 02 00 02 )= 5.55068286516206E-06
f(155)=f( 01 30 00 )= 1.86391130389137E-05
f(156)=f( 01 21 00 )=-2.55870298344221E-02
f(159)=f( 01 12 00 )= 1.03168129842943E-05
f(164)=f( 01 10 02 )= 1.48898803052262E-07
f(175)=f( 00 40 00 )=-1.93149067689109E-02
f(176)=f( 00 31 00 )=-2.61905549343156E-03
f(179)=f( 00 22 00 )=-9.05018682329818E-05
f(184)=f( 00 20 02 )= 1.31485684215654E-04
f(185)=f( 00 13 00 )=-7.58534226352400E-07
f(190)=f( 00 11 02 )= 3.62090101634544E-05
f(195)=f( 00 04 00 )= 5.20628151791380E-07
f(200)=f( 00 02 02 )= 5.55068286537210E-06

```

Exhibit 16.2.21: The map  $\mathcal{M}^{\text{LFF}}$  for the case  $a = 0.002$

matrix for map is :

1.00000E+00	-1.27806E-08	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
1.64617E-05	1.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00
0.00000E+00	0.00000E+00	1.00000E+00	-1.27806E-08	0.00000E+00	0.00000E+00
0.00000E+00	0.00000E+00	1.64617E-05	1.00000E+00	0.00000E+00	0.00000E+00
0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	1.00000E+00	0.00000E+00
0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	1.00000E+00

nonzero elements in generating polynomial are :

```

f( 33)=f( 20 00 01 )= 9.77739063782491E-06
f( 38)=f( 11 00 01 )= 2.99993492913329E-07
f( 45)=f( 10 01 01 )= 1.11383166578882E-10
f( 53)=f( 02 00 01 )= 2.27936177710220E-08
f( 57)=f( 01 10 01 )=-1.11383180456670E-10
f( 67)=f( 00 20 01 )= 9.77739063782491E-06
f( 70)=f( 00 11 01 )= 2.99993492886962E-07
f( 76)=f( 00 02 01 )= 2.27936205465795E-08
f( 84)=f( 40 00 00 )=-0.19316037246934
f( 85)=f( 31 00 00 )=-2.62331874601927E-03
f( 87)=f( 30 01 00 )=-1.89599356992097E-06
f( 90)=f( 22 00 00 )=-9.71312429641274E-06
f( 91)=f( 21 10 00 )= 1.89599356725453E-06
f( 92)=f( 21 01 00 )= 2.56091096783283E-02
f( 95)=f( 20 20 00 )=-0.38632074493868
f( 96)=f( 20 11 00 )=-2.62331874601768E-03
f( 99)=f( 20 02 00 )= 7.73683945196578E-06
f(104)=f( 20 00 02 )= 1.33064336534576E-05
f(105)=f( 13 00 00 )=-3.17816277210040E-08
f(106)=f( 12 10 00 )=-2.56091096782642E-02
f(107)=f( 12 01 00 )=-1.02864023718979E-07
f(110)=f( 11 20 00 )=-2.62331874601767E-03
f(111)=f( 11 11 00 )=-3.48999279807779E-05
f(114)=f( 11 02 00 )=-3.17767246773918E-08
f(119)=f( 11 00 02 )= 5.86324364974507E-07
f(121)=f( 10 21 00 )=-1.89599356821861E-06
f(124)=f( 10 12 00 )= 2.56091096783162E-02
f(130)=f( 10 03 00 )=-1.02863765470695E-07
f(135)=f( 10 01 02 )= 1.51604383154817E-10
f(140)=f( 04 00 00 )= 5.33338799513228E-09
f(141)=f( 03 10 00 )= 1.02862986930269E-07
f(145)=f( 02 20 00 )= 7.73683948127914E-06
f(146)=f( 02 11 00 )=-3.17772725634192E-08
f(149)=f( 02 02 00 )= 1.06253442294646E-08
f(154)=f( 02 00 02 )= 5.80574061863604E-08
f(155)=f( 01 30 00 )= 1.89599356946647E-06
f(156)=f( 01 21 00 )=-2.56091096783008E-02
f(159)=f( 01 12 00 )= 1.02862771318019E-07
f(164)=f( 01 10 02 )=-1.51604352741914E-10
f(175)=f( 00 40 00 )=-0.19316037246934
f(176)=f( 00 31 00 )=-2.62331874601742E-03
f(179)=f( 00 22 00 )=-9.71312426696364E-06
f(184)=f( 00 20 02 )= 1.33064336534594E-05
f(185)=f( 00 13 00 )=-3.17829202628548E-08
f(190)=f( 00 11 02 )= 5.86324364877742E-07
f(195)=f( 00 04 00 )= 5.34906990923290E-09
f(200)=f( 00 02 02 )= 5.80574133293362E-08

```

Examination of these exhibits shows that, as expected, (2.121) continues to hold even when the rotational parts of  $M^{LFF}$  have not been removed. We reiterate that, in the

hard-edge limit and when canonical coordinates are employed, there are *no* fringe-field contributions to the *linear* part of the transfer map for a solenoid.

What can be said about the  $\exp(: f_3 :)$  content of  $\mathcal{M}^{\text{LFF}}$ ? Examination of the  $f_3$  contents of these same exhibits, the generators with indices 28 through 83, shows that numerically there is the limiting behavior

$$\lim_{a \rightarrow 0} f_3^{\text{LFF}} = 0. \quad (16.2.123)$$

The same result can be obtained analytically. Consequently, in the hard-edge limit and when canonical coordinates are employed, there are also *no* fringe-field contributions to the *quadratic* part of the transfer map for a solenoid. All second-order aberrations associated with  $\mathcal{M}^{\text{LFF}}$  vanish in the hard-edge limit.

What can be said about the  $\exp(: f_4 :)$  content of  $\mathcal{M}^{\text{LFF}}$ ? Examination of the generators with indices 84 through 209 shows that numerically some of the  $f_4^{\text{LFF}}$  generators *grow/diverge* in magnitude as  $a \rightarrow 0$ . Indeed, it can be illustrated numerically and demonstrated analytically that the divergent  $f_4^{\text{LFF}}$  entries behave as  $1/a$  as  $a \rightarrow 0$ . For example, Table 2.3 below illustrates this divergence for the case of  $f^{\text{LFF}}(84)$ .

Table 16.2.3: Numerical behavior of  $f^{\text{LFF}}(84)$  for small values of  $a$ .

$a$	$f^{\text{LFF}}(84)$	$a f^{\text{LFF}}(84)$
.2	-1.8691E-3	-3.7382E-4
.02	-1.9315E-2	-3.8630E-4
.002	-1.9316E-1	-3.8632E-4

At this point here are two other remarks to be made. First, there are also divergent  $f_n^{\text{LFF}}$  generators for  $n > 4$ . Second, as already observed, the only  $a$  dependence in (2.112) is in the maps  $\mathcal{M}^{\text{LFF}}$  and  $\mathcal{M}^{\text{TFF}}$ . We conclude that *all* divergent aberrations for a simple solenoid in the hard-edge limit arise from divergencies in the fringe-field maps.

Upon thinking in more detail, there is more that can be inferred about the  $a$  dependence of the  $f_4^{\text{LFF}}$  generators. We have already seen that  $H$  and  $J_z$  are in involution. Recall (2.82). Indeed, upon examination we see that all the terms in the expansion of  $H$  given in (2.94) through (2.96) are separately in involution with  $J_z$ . Therefore, since  $f_4$  is constructed from the ingredients of  $H$  using only Lie algebraic operations, we expect that  $f_4^{\text{LFF}}$  and  $J_z$  will be in involution. Below we list the various possible *static* ( $\tau$  independent)  $f_4$  polynomials, call them  $I_j$ , that are in *involution* with  $J_z$  (and hence *invariant* under the action of rotations generated by  $: J_z :$ ) and display their monomial content.

$$I_1 = (P^2)^2 = (P_x^2 + P_y^2)^2 = P_x^4 + 2P_x^2P_y^2 + P_y^4, \quad (16.2.124)$$

$$\begin{aligned} I_2 &= P^2 J_z = (P_x^2 + P_y^2)(XP_y - YP_x) \\ &= XP_x^2P_y - P_x^3Y + XP_y^3 - P_xYP_y^2, \end{aligned} \quad (16.2.125)$$

$$I_3 = J_z^2 = (XP_y - YP_x)^2 = X^2P_y^2 - 2XP_xYP_y + P_x^2Y^2, \quad (16.2.126)$$

$$I_4 = P^2 Q^2 = (P_x^2 + P_y^2)(X^2 + Y^2) = X^2 P_x^2 + X^2 P_y^2 + P_x^2 Y^2 + Y^2 P_y^2, \quad (16.2.127)$$

$$\begin{aligned} I_5 &= Q^2 J_z = (X^2 + Y^2)(XP_y - YP_x) \\ &= X^3 P_y - X^2 P_x Y + XY^2 P_y - P_x Y^3, \end{aligned} \quad (16.2.128)$$

$$I_6 = (Q^2)^2 = (X^2 + Y^2)^2 = X^4 + 2X^2 Y^2 + Y^4, \quad (16.2.129)$$

$$I_7 = P_\tau^2 P^2 = P_x^2 P_\tau^2 + P_y^2 P_\tau^2, \quad (16.2.130)$$

$$I_8 = P_\tau^2 J_z = XP_y P_\tau^2 - P_x Y P_\tau^2, \quad (16.2.131)$$

$$I_9 = P_\tau^2 Q^2 = X^2 P_\tau^2 + Y^2 P_\tau^2, \quad (16.2.132)$$

$$I_{10} = P_\tau^4. \quad (16.2.133)$$

Note that the monomials  $X^2 P_y^2$  and  $P_x^2 Y^2$  appear in both the invariants  $I_3$  and  $I_4$ . All other monomials appear at most once in the invariants  $I_j$ .

According to the reasoning of the previous paragraph, we expect that  $f_4^{\text{LFF}}$  can be expressed/expanded (in a unique way) in terms of the  $I_j$  for any value of  $a$ . This is indeed the case. The relations (2.134) through (2.136) below display this expansion for the values  $a = 0.2$ ,  $a = 0.02$ , and  $a = 0.002$ , respectively.

$$\begin{aligned} f_4^{\text{LFF}}|_{a=0.2} &= (4.36074169958001E - 05)I_1 - (9.17588458782363E - 04)I_2 \\ &+ (1.15770219925823E - 03)I_3 - (3.06010303211954E - 04)I_4 \\ &- (1.51855928877267E - 04)I_5 - (1.86910737357663E - 03)I_6 \\ &+ (3.39093545585142E - 04)I_7 - (1.50498675709195E - 05)I_8 \\ &+ (1.15254677042892E - 03)I_9 + (0)I_{10}; \end{aligned} \quad (16.2.134)$$

$$\begin{aligned} f_4^{\text{LFF}}|_{a=0.02} &= (5.20628154886127E - 07)I_1 - (1.03168131481354E - 05)I_2 \\ &+ (1.68820248937675E - 04)I_3 - (9.05018682344674E - 05)I_4 \\ &- (1.86391130390008E - 05)I_5 - (1.93149067689109E - 02)I_6 \\ &+ (5.55068286516206E - 06)I_7 - (1.48898803105739E - 07)I_8 \\ &+ (1.31485684215657E - 04)I_9 + (0)I_{10}; \end{aligned} \quad (16.2.135)$$

$$\begin{aligned} f_4^{\text{LFF}}|_{a=0.002} &= (5.33338799513228E - 09)I_1 - (1.02864023718979E - 07)I_2 \\ &+ (1.74499639903890E - 05)I_3 - (9.71312429641274E - 06)I_4 \\ &- (1.89599356992097E - 06)I_5 - (0.19316037246934)I_6 \\ &+ (5.80574061863604E - 08)I_7 + (1.51604383154817E - 10)I_8 \\ &+ (1.33064336534576E - 05)I_9 + (0)I_{10}. \end{aligned} \quad (16.2.136)$$

Inspection of (2.134) through (2.136) shows that, save for  $I_6$ , the coefficients of all the  $I_j$  vanish as  $a \rightarrow 0$ . See Exercise 2.21. By contrast, the aberrations associated with the ingredients of  $I_6$  grow as  $1/a$  as  $a \rightarrow 0$ . Recall Table 2.3 and see Exercise 2.22. Thus the behavior of  $f_4^{\text{LFF}}$  is *simple* in the hard-edge limit in that most (all but three) of its entries vanish in this limit. But it is also *pathological* in that the entries for the generators  $X^4$ ,  $X^2Y^2$ , and  $Y^4$ , those that occur in  $I_6$ , diverge in this limit.<sup>7</sup> Finally we remark that, while we have illustrated these results numerically, they can also be proven analytically. The pathological behavior arises from the appearance of the the  $\delta'$  functions that occur in the  $a \rightarrow 0$  limit, and must occur in any hard-edge model.

### 16.2.7 Consequences of Terminating Solenoidal End Fields

Suppose we wish to find the transfer map for a solenoid with the approximation that the leading fringe-field region begins at the “entry” point  $z = z^{\text{en}}$  and the trailing fringe-field region ends at the “exit” point  $z = z^{\text{ex}}$ . That is, we make the approximation that the vector potential is to be set to zero for  $z < z^{\text{en}}$  and  $z > z^{\text{ex}}$ . Since  $\hat{\mathbf{A}}^0$ , the vector potential we will employ, is in the Poincaré-Coulomb gauge with respect to any origin on the  $z$  axis, we may use (following the methods of Section 1) this vector potential to terminate end fields both before entry of the leading fringe field and after exit of the trailing fringe field. That is, there is no need to make gauge transformations at these points because the vector potential is already in the minimum gauge. In this subsection we will study the consequences of terminating end fields using the Poincaré-Coulomb gauge.

Let us begin by finding the associated discontinuities in the mechanical momenta as given by (1.30), (1.31), (1.41), and (1.42). For the vector potential we use (2.7) through (2.10). So doing using (1.30) and (1.31) gives, upon entry, the results

$$\begin{aligned}\Delta p_x^{\text{mech}} &= qA_x(x, y, z^{\text{en}}) = -qyU(\rho, z^{\text{en}}) \\ &= -qy(1/2)[C_0^{[1]}(z^{\text{en}}) - (1/8)C_0^{[3]}(z^{\text{en}})(x^2 + y^2) + \dots] \\ &= -qy(1/2)[B_z(0, 0, z^{\text{en}}) - (1/8)B_z''(0, 0, z^{\text{en}})(x^2 + y^2) + \dots] \\ &= -qy(B/2)[\text{bump}(z^{\text{en}}, a, L) - (1/8)\text{bump}''(z^{\text{en}}, a, L)(x^2 + y^2) + \dots],\end{aligned}\tag{16.2.137}$$

$$\begin{aligned}\Delta p_y^{\text{mech}} &= qA_y(x, y, z^{\text{en}}) = qxU(\rho, z^{\text{en}}) \\ &= qx(1/2)[C_0^{[1]}(z^{\text{en}}) - (1/8)C_0^{[3]}(z^{\text{en}})(x^2 + y^2) + \dots] \\ &= qx(1/2)[B_z(0, 0, z^{\text{en}}) - (1/8)B_z''(0, 0, z^{\text{en}})(x^2 + y^2) + \dots] \\ &= qx(B/2)[\text{bump}(z^{\text{en}}, a, L) - (1/8)\text{bump}''(z^{\text{en}}, a, L)(x^2 + y^2) + \dots].\end{aligned}\tag{16.2.138}$$

---

<sup>7</sup>We hasten to add that the aberrations associated with  $\mathcal{M}^{\text{LFF}}$  and  $\mathcal{M}^{\text{TFF}}$  are not the only aberrations for  $\mathcal{M}_{\text{solenoid}}$ . According to (2.112) there will also be aberrations associated with  $\mathcal{M}^{\text{uniform}}$ . But they are always finite and, by definition,  $a$  independent. They are also relatively easy to compute because  $\mathcal{M}^{\text{uniform}}$  arises from the  $z$  independent Hamiltonian  $H^{\text{uniform}}$ .

Here, in writing the last lines of (2.137) and (2.138), we have assumed the field profile of Subsection 2.2. Similarly, upon exit, we find from (1.41), and (1.42) the discontinuity results

$$\begin{aligned}\Delta p_x^{\text{mech}} &= qA_x(x, y, z^{\text{ex}}) = -qyU(\rho, z^{\text{ex}}) \\ &= -qy(1/2)[C_0^{[1]}(z^{\text{ex}}) - (1/8)C_0^{[3]}(z^{\text{ex}})(x^2 + y^2) + \dots] \\ &= -qy(1/2)[B_z(0, 0, z^{\text{ex}}) - (1/8)B_z''(0, 0, z^{\text{ex}})(x^2 + y^2) + \dots] \\ &= -qy(B/2)[\text{bump}(z^{\text{ex}}, a, L) - (1/8)\text{bump}''(z^{\text{ex}}, a, L)(x^2 + y^2) + \dots],\end{aligned}\tag{16.2.139}$$

$$\begin{aligned}\Delta p_y^{\text{mech}} &= qA_y(x, y, z^{\text{ex}}) = qxU(\rho, z^{\text{ex}}) \\ &= qx(1/2)[C_0^{[1]}(z^{\text{ex}}) - (1/8)C_0^{[3]}(z^{\text{ex}})(x^2 + y^2) + \dots] \\ &= qx(1/2)[B_z(0, 0, z^{\text{ex}}) - (1/8)B_z''(0, 0, z^{\text{ex}})(x^2 + y^2) + \dots] \\ &= qx(B/2)[\text{bump}(z^{\text{ex}}, a, L) - (1/8)\text{bump}''(z^{\text{ex}}, a, L)(x^2 + y^2) + \dots].\end{aligned}\tag{16.2.140}$$

We see that in all cases the discontinuities are proportional to  $B_z(0, 0, z)$  and its derivatives at  $z = z^{\text{en}}$  or  $z = z^{\text{ex}}$ . See Figures 2.4 through 2.7 for examples of how these functions behave in the case of a simple air-code solenoid. Moreover, the discontinuities also vanish as the spatial deviations from the  $z$  axis (the design orbit) become small.

We close this subsection by finding, at entry and exit, the surface currents implied by our termination procedure/approximation. Since  $\hat{A}_z^0 = 0$  in the Poincaré-Coulomb gauge for any solenoid or collection of solenoids, the relations (1.65) through (1.67) take the form

$$\begin{aligned}\mu_0 j_x^{\text{mod}} &= -2[\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]\partial_z \hat{A}_x^0 \\ &\quad - [\delta'(z - z^{\text{en}}) + \delta'(z^{\text{ex}} - z)]\hat{A}_x^0,\end{aligned}\tag{16.2.141}$$

$$\begin{aligned}\mu_0 j_y^{\text{mod}} &= -2[\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]\partial_z \hat{A}_y^0 \\ &\quad - [\delta'(z - z^{\text{en}}) + \delta'(z^{\text{ex}} - z)]\hat{A}_y^0,\end{aligned}\tag{16.2.142}$$

$$\mu_0 j_z^{\text{mod}} = 0.\tag{16.2.143}$$

Let us evaluate (2.141) and (2.142) using the explicit form for  $\hat{\mathbf{A}}^0$  given by (2.7) through (2.9). Doing so gives the intermediate results

$$\begin{aligned}\mu_0 j_x^{\text{mod}} &= 2[\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]y\partial_z U \\ &\quad + [\delta'(z - z^{\text{en}}) + \delta'(z^{\text{ex}} - z)]yU,\end{aligned}\tag{16.2.144}$$

$$\begin{aligned}\mu_0 j_y^{\text{mod}} &= -2[\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]x\partial_z U \\ &\quad - [\delta'(z - z^{\text{en}}) + \delta'(z^{\text{ex}} - z)]xU.\end{aligned}\tag{16.2.145}$$

At this point it is convenient to employ cylindrical components for  $\mathbf{j}^{\text{mod}}$  using the relations

$$j_\rho^{\text{mod}} = \cos \phi j_x^{\text{mod}} + \sin \phi j_y^{\text{mod}}, \quad (16.2.146)$$

$$j_\phi^{\text{mod}} = -\sin \phi j_x^{\text{mod}} + \cos \phi j_y^{\text{mod}}. \quad (16.2.147)$$

Recall (15.2.22) and (15.2.23). Implementing these substitutions gives the results

$$\mu_0 j_\rho^{\text{mod}} = 0, \quad (16.2.148)$$

$$\begin{aligned} \mu_0 j_\phi^{\text{mod}} &= -2[\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]\rho \partial_z U \\ &\quad - [\delta'(z - z^{\text{en}}) + \delta'(z^{\text{ex}} - z)]\rho U. \end{aligned} \quad (16.2.149)$$

Here we have used the relations

$$y \cos \phi - x \sin \phi = 0, \quad (16.2.150)$$

$$y \sin \phi + x \cos \phi = \rho. \quad (16.2.151)$$

We see that  $\mathbf{j}^{\text{mod}}$ , the current that is required to cancel the residual solenoidal fringe field, has only a  $\phi$  component. This is to be expected since the current that produces the solenoidal field itself has only a  $\phi$  component. The last step is to use the expansion (2.10) for  $U$ . With the aid of this expansion we find the final result

$$\begin{aligned} \mu_0 j_\phi^{\text{mod}} &= -2[\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]\rho \partial_z U \\ &\quad - [\delta'(z - z^{\text{en}}) + \delta'(z^{\text{ex}} - z)]\rho U \\ &= [\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]\rho [C_0^{[2]}(z) - (1/8)C_0^{[4]}(z)(x^2 + y^2) + \dots] \\ &\quad + (1/2)[\delta'(z - z^{\text{en}}) + \delta'(z^{\text{ex}} - z)]\rho [C_0^{[1]}(z) - (1/8)C_0^{[3]}(z)(x^2 + y^2) + \dots] \\ &= [\delta(z - z^{\text{en}}) - \delta(z^{\text{ex}} - z)]\rho [B'_z(0, 0, z) - (1/8)B'''_z(0, 0, z)(x^2 + y^2) + \dots] \\ &\quad + (1/2)[\delta'(z - z^{\text{en}}) + \delta'(z^{\text{ex}} - z)]\rho [B_z(0, 0, z) - (1/8)B''_z(0, 0, z)(x^2 + y^2) + \dots]. \end{aligned} \quad (16.2.152)$$

Like the discontinuities in the mechanical momenta,  $\mathbf{j}^{\text{mod}}$  is also proportional to  $B_z(0, 0, z)$  and its derivatives at  $z = z^{\text{en}}$  or  $z = z^{\text{ex}}$ , and also vanishes as the spatial deviations from the  $z$  axis (the design orbit) become small.

## Exercises

**16.2.1.** This exercise treats properties of the simple solenoid. Begin by assuming that  $B_z(0, 0, z)$  as given by (2.16) describes the on-axis field of a simple air-core solenoid. Verify that there is the result

$$B = \mu_0 I N / L \quad (16.2.153)$$

where  $I$  is the current in the coil,  $N$  is the number of turns in the single-layer winding, and  $L$  is the length of the coil. Hint: Use (2.24) and Ampère's law. It is of historical interest to note that the name *solenoid* was coined by Ampère.

The next step is to verify, using Biot-Savart, that  $B_z(0, 0, z)$  as given by (2.16) does indeed describe the on-axis field of a simple air-core solenoid.

**16.2.2.** Show that for the on-axis field of a simple solenoid as given by (2.16) there are, for the fields at the midpoint/center and either end, the limiting behaviors

$$\lim_{L \rightarrow \infty} B_z(0, 0, L/2) = B, \quad (16.2.154)$$

$$\lim_{L \rightarrow \infty} B_z(0, 0, 0) = \lim_{L \rightarrow \infty} B_z(0, 0, L) = B/2. \quad (16.2.155)$$

Verify that the same is true for any bump function model that is constructed from approximating signum functions.

**16.2.3.** The purpose of this exercise is to verify the relations (2.20) through (2.23). Show that the approximating signum function (2.26) has the properties

$$\text{sgn}(-z, a) = -\text{sgn}(z, a), \quad (16.2.156)$$

$$\lim_{z \rightarrow \pm\infty} \text{sgn}(z, a) = \pm 1. \quad (16.2.157)$$

Verify the limiting behavior (2.27). Sketch  $\text{sgn}(z, a)$ ,  $-\text{sgn}(z - L, a)$ , and  $\text{bump}(z, a, L)$  as given by (2.25) to verify the relations (2.20) through (2.22).

What remains is to prove the relation (2.23). Begin by writing

$$\int_{-\infty}^{\infty} \text{bump}(z, a, L) dz = \lim_{w \rightarrow \infty} \int_{-w}^w \text{bump}(z, a, L) dz. \quad (16.2.158)$$

Next verify from the representation (2.25) that

$$\int_{-w}^w \text{bump}(z, a, L) dz = (1/2) \int_{-w}^w \text{sgn}(z, a) dz - (1/2) \int_{-w}^w \text{sgn}(z - L, a) dz. \quad (16.2.159)$$

Show that the first integral on the right side of (2.159) vanishes because of (2.156). Show that by making the change of variables  $x = z - L$  the second integral on the right side of (2.159) becomes

$$\begin{aligned} -(1/2) \int_{-w}^w \text{sgn}(z - L, a) dz &= -(1/2) \int_{-w-L}^{w-L} \text{sgn}(x, a) dx \\ &= -(1/2) \int_{-w-L}^{w+L} \text{sgn}(x, a) dx + (1/2) \int_{w-L}^{w+L} \text{sgn}(x, a) dx. \end{aligned} \quad (16.2.160)$$

Verify that the first integral in the second line of (2.160) vanishes, again because of (2.156). It follows that there is the result

$$\int_{-w}^w \text{bump}(z, a, L) dz = (1/2) \int_{w-L}^{w+L} \text{sgn}(x, a) dx. \quad (16.2.161)$$

Show from (2.157) that there is the result

$$\lim_{w \rightarrow \infty} (1/2) \int_{w-L}^{w+L} \text{sgn}(x, a) dx = (1/2)2L = L. \quad (16.2.162)$$

Put all your intermediate results together to obtain the final result

$$\int_{-\infty}^{\infty} \text{bump}(z, a, L) dz = L, \quad (16.2.163)$$

as desired. Note that the proof of this result has depended only on the representation (2.25) and properties (2.156) and (2.157), which are required properties of any approximating signum function.

**16.2.4.** Verify the fall-off relations (2.34) and (2.35) and the limiting behaviors (2.36) and (2.37).

**16.2.5.** Verify the expansion (2.38) and the near leading end fall-off behavior (2.39).

**16.2.6.** Verify the expansion (2.40) and the far fall-off behavior given by (2.41) and (2.42).

**16.2.7.** Verify the near-by and far asymptotic behaviors (2.44) and (2.45).

**16.2.8.** Verify the far fall-off behavior (2.47).

**16.2.9.** Verify the relation (2.49) for the on-axis field of a thick solenoid. Show that (2.24), (2.154), and (2.155) continue to hold. Show that (2.153) also holds providing that  $N$  is now the total of number of turns in the whole multilayer winding. Finally, use the notation  $B_z(0, 0, z; a)$  and  $B_z(0, 0, z; a_1, a_2)$  to denote the right sides of (2.16) and (2.49), respectively. Show that there is the limiting relation

$$\lim_{a_2 \rightarrow a} B_z(0, 0, z; a, a_2) = B_z(0, 0, z; a). \quad (16.2.164)$$

Hint: Use the representation (2.48).

**16.2.10.** Observe that (2.7) through (2.9) are the same as (15.5.7) through (15.5.9). Verify that using (15.2.22), (15.2.23), and (15.5.7) through (15.5.9) gives the results (15.5.37) through (15.5.39). That is,  $\hat{\mathbf{A}}^0$  and hence  $\mathbf{A}^s$  has only a  $\phi$  component. Verify that this component manifests rotational symmetry about the  $z$  axis by having no  $\phi$  dependence.

**16.2.11.** Verify that the  $b^{[2n]}(z)$  as given by (2.59) are dimensionless.

**16.2.12.** In (2.64) the quantity  $J_z$  is defined in terms of canonical variables. What happens if mechanical momenta are employed instead? Define *scaled* mechanical momenta  $P_x^{\text{mech}}$  and  $P_y^{\text{mech}}$  by writing

$$P_x^{\text{mech}} = P_x - A_x^s, \quad (16.2.165)$$

$$P_y^{\text{mech}} = P_y - A_y^s; \quad (16.2.166)$$

from which it follows that there are the relations

$$P_x = P_x^{\text{mech}} + A_x^s, \quad (16.2.167)$$

$$P_y = P_y^{\text{mech}} + A_y^s. \quad (16.2.168)$$

[Recall the unscaled results (1.5.30).] Show that the expression for  $J_z$  in terms of mechanical momenta is given by the relation

$$J_z = X P_y^{\text{mech}} - Y P_x^{\text{mech}} + X A_y^s - Y A_x^s. \quad (16.2.169)$$

Verify, using (2.55), (2.56), and (2.58), that there is the result

$$\begin{aligned} X A_y^s - Y A_x^s &= (X^2 + Y^2) U^s(X, Y, z) \\ &= (1/2) \sum_{n=0}^{\infty} (-1)^n \frac{1}{2^{2n} n! (n+1)!} b^{[2n]}(z) (X^2 + Y^2)^{n+1}. \end{aligned} \quad (16.2.170)$$

Show from (15.5.38) and (2.52) that there is the result

$$\begin{aligned} \rho \hat{A}_\phi^0 &= \rho^2 U(\rho, z) = (1/2) \sum_{n=0}^{\infty} (-1)^n \frac{1}{2^{2n} n! (n+1)!} C_0^{[2n+1]}(z) \rho^{2n+2} \\ &= (1/2) \sum_{n=0}^{\infty} (-1)^n \frac{1}{2^{2n} n! (n+1)!} B^{[2n]}(z) \rho^{2n+2}. \end{aligned} \quad (16.2.171)$$

Verify from (13.1.21) and (13.1.22) that

$$\rho^{2n+2} = \ell^{2n+2} (X^2 + Y^2)^{n+1}. \quad (16.2.172)$$

Therefore (2.171) can also be written in the form

$$\rho \hat{A}_\phi^0 = (1/2) \sum_{n=0}^{\infty} (-1)^n \frac{1}{2^{2n} n! (n+1)!} B^{[2n]}(z) \ell^{2n+2} (X^2 + Y^2)^{n+1}. \quad (16.2.173)$$

Verify from (2.59) that

$$B^{[2n]}(z) \ell^{2n+2} = (\ell p^0 / q) b^{[2n]}(z) \quad (16.2.174)$$

so that

$$\begin{aligned} \rho \hat{A}_\phi^0 &= (\ell p^0 / q) (1/2) \sum_{n=0}^{\infty} (-1)^n \frac{1}{2^{2n} n! (n+1)!} b^{[2n]}(z) (X^2 + Y^2)^{n+1} \\ &= (\ell p^0 / q) (X^2 + Y^2) U^s(X, Y, z) = (\ell p^0 / q) (X A_y^s - Y A_x^s). \end{aligned} \quad (16.2.175)$$

Make the definition

$$J_z^{\text{mech}} = X P_y^{\text{mech}} - Y P_x^{\text{mech}}. \quad (16.2.176)$$

Consequently, verify that (2.169) can be rewritten in the form

$$J_z = J_z^{\text{mech}} + [q/(\ell p^0)] \rho \hat{A}_\phi^0. \quad (16.2.177)$$

According to (2.82)  $J_z$  is conserved. Also, far outside a solenoid,  $\hat{A}_\phi^0$  and hence  $\rho \hat{A}_\phi^0$  vanish. Therefore, (2.177) shows that  $J_z^{\text{mech}}$  must have different values inside and outside a solenoid and what their difference must be. Show that if a particle enters a solenoid from a zero field region with some initial value of  $J_z$  and ultimately exits into a second zero field region, then its final value of  $J_z$  must be the same as its initial value.

**16.2.13.** Verify the identity (2.76). Show that (2.75) and (2.79) are a consequence of (2.70) through (2.73) and (2.77) and (2.78).

**16.2.14.** Compare the  $f_4$  contents of the maps in Exhibits 2.3, 2.4, and 2.5. Which  $f_4$  entries appear to be diverging (in magnitude) to  $\infty$  as  $a \rightarrow 0$ ? Why, for a given value of  $a$ , are there the relations  $f(84) = f(175)$  and  $f(95) = 2f(84)$ ? Can you find other relations of this kind? Hint: Show that the Lie generators  $f_n$  for a solenoid map must satisfy

$$: J_z : f_n = 0. \quad (16.2.178)$$

Show that the same is true for a drift map and therefore for a composite of drift and solenoid maps.

**16.2.15.** The purpose of this exercise is to verify the factorization of the linear part as described in Subsection 2.5.3. Begin by writing the transfer map for a solenoid in the general form

$$\mathcal{M} = \mathcal{R} \exp(: f_3 :) \exp(: f_4 :) \cdots. \quad (16.2.179)$$

Then  $\mathcal{R}$  will be determined by  $H_2$  as given by (2.94). Verify that  $H_2$  can be written in the form

$$H_2 = H_2^{\text{nonrot}} + H_2^{\text{rot}} \quad (16.2.180)$$

where

$$H_2^{\text{nonrot}} = [1/(2\ell)]P^2 + \{[b^{[0]}(z)]^2/(8\ell)\}Q^2 + [1/(2\beta_0^2\gamma_0^2\ell)]P_\tau^2 \quad (16.2.181)$$

and

$$H_2^{\text{rot}} = -[b^{[0]}(z)/(2\ell)]J_z. \quad (16.2.182)$$

Let  $\mathcal{R}$ ,  $\mathcal{R}^{\text{nonrot}}$ , and  $\mathcal{R}^{\text{rot}}$  be the maps generated by  $H_2$ ,  $H_2^{\text{nonrot}}$ , and  $H_2^{\text{rot}}$ , respectively. Verify that  $: H_2^{\text{nonrot}} :$  and  $: H_2^{\text{rot}} :$  commute and consequently prove, using the results of Exercise 10.2.2, that there are the relations

$$\mathcal{R} = \mathcal{R}^{\text{nonrot}} \mathcal{R}^{\text{rot}} = \mathcal{R}^{\text{rot}} \mathcal{R}^{\text{nonrot}}. \quad (16.2.183)$$

Correspondingly, there are the associated matrix relations

$$R = R^{\text{rot}} R^{\text{nonrot}} = R^{\text{nonrot}} R^{\text{rot}}. \quad (16.2.184)$$

**16.2.16.** Through terms of second order, the transfer map for a solenoid can be written in the general form

$$\mathcal{M} = \mathcal{R} \exp(: f_3 :). \quad (16.2.185)$$

The goal of this exercise is to compute for a simple solenoid, in the hard-edge limit  $a \rightarrow 0$ , both the matrix  $R$  associated with  $\mathcal{R}$  and the Lie generator  $f_3$ . For a further discussion of motion in a uniform magnetic field, see Exercise 32.2.7.

Examination of  $H_2$  and  $H_3$  as given by (2.94) and (2.95) shows that their  $z$  dependence is given entirely in terms of  $b^{[0]}(z)$ . This function is bounded for all  $a$ , and in the hard-edge limit takes on a *constant* value in the open interval  $z \in (0, L)$ . See Figures 2.4 and 2.5. We recall that here  $z$  plays the role of the independent variable, and therefore in the hard-edge limit  $H_2$  and  $H_3$  do not depend on the independent variable. Consequently show that, in

the hard-edge limit and through terms of second order, the transfer map for a solenoid can be written in the form

$$\mathcal{M} = \exp(-L : H_2 + H_3 :). \quad (16.2.186)$$

Recall (7.4.1). Next verify that  $:H_2:$  and  $:H_3:$  commute under the assumption that  $b^{[0]}(z)$  is constant in the open interval  $z \in (0, L)$ . It follows that (2.186) can be rewritten in the form

$$\mathcal{M} = \exp(-L : H_2 :) \exp(-L : H_3 :). \quad (16.2.187)$$

Comparison of (2.185) and (2.187) yields the results

$$\mathcal{R} = \exp(-L : H_2 :) \quad (16.2.188)$$

and

$$f_3 = -LH_3. \quad (16.2.189)$$

What remains is to find the matrix  $R$  associated with  $\mathcal{R}$ . Review Exercise 2.15. The discussion there holds for all (well behaved) functions  $b^{[0]}(z)$  and therefore also holds when  $b^{[0]}(z)$  is constant in the open interval  $z \in (0, L)$ . In the case that  $b^{[0]}(z)$  is constant in the open interval  $z \in (0, L)$  there are the results

$$\mathcal{R}^{\text{nonrot}} = \exp(-L : H_2^{\text{nonrot}} :) \quad (16.2.190)$$

and

$$\mathcal{R}^{\text{rot}} = \exp(-L : H_2^{\text{rot}} :). \quad (16.2.191)$$

At this point we pause to observe that  $f_3$  as given by (2.95) and (2.189) can also be decomposed into two parts whose associated Lie operators commute. We may write

$$f_3 = f_3^{\text{nonrot}} + f_3^{\text{rot}} \quad (16.2.192)$$

where

$$f_3^{\text{nonrot}} = -L(1/\beta_0)P_\tau H_2^{\text{nonrot}} \quad (16.2.193)$$

and

$$f_3^{\text{rot}} = -L(1/\beta_0)P_x H_2^{\text{rot}}. \quad (16.2.194)$$

To continue, your next task is to compute  $R^{\text{nonrot}}$ , the matrix associated with  $\mathcal{R}^{\text{nonrot}}$ . Begin by verifying that  $H_2^{\text{nonrot}}$  as given by (2.181) consists of three pieces associated with the  $x, P_x$ ;  $y, P_y$ ; and  $\tau, P_\tau$  planes, and that the Lie operators associated with different pieces all commute. Correspondingly  $R^{\text{nonrot}}$  has nonzero entries consisting only of  $2 \times 2$  matrices centered on the diagonal. What are these matrices? Consider the Lie transformation

$$\exp\{[-L/(2\ell)]P_x^2 - [L(b^{[0]})^2/(8\ell)]X^2 : \} \quad (16.2.195)$$

that is associated with the  $x, P_x$  part of  $\mathcal{R}^{\text{nonrot}}$  when  $b^{[0]}(z)$  is constant in the open interval  $z \in (0, L)$ . Let  $R_{XP_x}$  be the  $2 \times 2$  matrix that describes the action of this Lie transformation on the  $X, P_x$  plane. Use the formalism and results of Section 8.7.2 to make the identifications

$$b = L/\ell, \quad (16.2.196)$$

$$a = 0, \quad (16.2.197)$$

$$c = [L/(4\ell)](b^{[0]})^2, \quad (16.2.198)$$

from which it follows that

$$\Delta = -[L/(2\ell)]^2(b^{[0]})^2 \quad (16.2.199)$$

and

$$\Delta^{1/2} = i[L/(2\ell)](b^{[0]}). \quad (16.2.200)$$

Show using (8.7.35) that there is the result

$$R_{XP_x} = \begin{pmatrix} \cosh(\Delta^{1/2}) & b[\sinh(\Delta^{1/2})]/\Delta^{1/2} \\ -c[\sinh(\Delta^{1/2})]/\Delta^{1/2} & \cosh(\Delta^{1/2}) \end{pmatrix}. \quad (16.2.201)$$

Also verify that there are the relations

$$b/\Delta^{1/2} = (L/\ell)(-i)(2\ell/L)(1/b^{[0]}) = -i(2/b^{[0]}) \quad (16.2.202)$$

and

$$c/\Delta^{1/2} = [L/(4\ell)](b^{[0]})^2(-i)(2\ell/L)(1/b^{[0]}) = -i(b^{[0]}/2). \quad (16.2.203)$$

Introduce the notation

$$k = (b^{[0]}/2) \quad (16.2.204)$$

and

$$\psi = [L/(2\ell)](b^{[0]}) = k(L/\ell) \quad (16.2.205)$$

so that (2.200), (2.202), and (2.203) take the forms

$$\Delta^{1/2} = i\psi, \quad (16.2.206)$$

$$b/\Delta^{1/2} = -i/k, \quad (16.2.207)$$

and

$$c/\Delta^{1/2} = -ik. \quad (16.2.208)$$

Verify, using this notation, that (2.201) can be written in the final form

$$R_{XP_x} = \begin{pmatrix} \cos(\psi) & (1/k)\sin(\psi) \\ -k\sin(\psi) & \cos(\psi) \end{pmatrix}. \quad (16.2.209)$$

With regard of the action of  $\mathcal{R}^{\text{nonrot}}$  on the  $Y, P_y$  plane, verify from symmetry considerations that

$$R_{YP_y} = R_{XP_x}. \quad (16.2.210)$$

Finally show that the effect of the Lie transformation  $\exp\{[-L/(2\beta_0^2\gamma_0^2\ell)]P_\tau^2\}$  on the  $\tau, P_\tau$  plane is given by the relations

$$\exp\{[-L/(2\beta_0^2\gamma_0^2\ell)]P_\tau^2\}\tau = \tau + [L/(\beta_0^2\gamma_0^2\ell)]P_\tau, \quad (16.2.211)$$

$$\exp\{[-L/(2\beta_0^2\gamma_0^2\ell)]P_\tau^2\}P_\tau = P_\tau. \quad (16.2.212)$$

Consequently there is the corresponding matrix

$$R_{\tau P_\tau} = \begin{pmatrix} 1 & \eta \\ 0 & 1 \end{pmatrix} \quad (16.2.213)$$

where

$$\eta = L/(\beta_0^2 \gamma_0^2 \ell). \quad (16.2.214)$$

You have shown that

$$R^{\text{nonrot}} = \begin{pmatrix} C & (1/k)S & 0 & 0 & 0 & 0 \\ -kS & C & 0 & 0 & 0 & 0 \\ 0 & 0 & C & (1/k)S & 0 & 0 \\ 0 & 0 & -kS & C & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \eta \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (16.2.215)$$

where we have used the short-hand notation

$$C = \cos(\psi) \quad (16.2.216)$$

and

$$S = \sin(\psi). \quad (16.2.217)$$

Note that the entries in  $R^{\text{nonrot}}$  are *even* functions of  $k$  and hence  $R^{\text{nonrot}}$  is *invariant* under the replacement  $b^{[0]} \rightarrow -b^{[0]}$ . This symmetry is also evident from the form of  $H_2^{\text{nonrot}}$  as given in (2.167).

Now turn to the calculation of  $R^{\text{rot}}$ , the matrix associated with  $\mathcal{R}^{\text{rot}}$ . Verify that  $\mathcal{R}^{\text{rot}}$  can be written on the form

$$\mathcal{R}^{\text{rot}} = \exp(-L : H_2^{\text{rot}} :) = \exp(\psi : J_z :). \quad (16.2.218)$$

Use the results (2.70) and (2.71) to verify the relation

$$\begin{aligned} \mathcal{R}^{\text{rot}} X &= \exp(\psi : J_z :) X = \\ &X + \psi : J_z : X + \psi^2 (1/2!) : J_z :^2 X + \psi^3 (1/3!) : J_z :^3 X + \cdots = \\ &X + \psi Y - \psi^2 (1/2!) X - \psi^3 (1/3!) Y + \cdots = \\ &X[1 - \psi^2 (1/2!) + \cdots] + Y[\psi - \psi^3 (1/3!) + \cdots] = \\ &X \cos(\psi) + Y \sin(\psi). \end{aligned} \quad (16.2.219)$$

In a similar way verify that there is the relation

$$\mathcal{R}^{\text{rot}} Y = -X \sin(\psi) + Y \cos(\psi). \quad (16.2.220)$$

Use the results (2.72) and (2.73) to find the analogous relations

$$\mathcal{R}^{\text{rot}} P_x = P_x \cos(\psi) + P_y \sin(\psi) \quad (16.2.221)$$

and

$$\mathcal{R}^{\text{rot}} P_y = -P_x \sin(\psi) + P_y \cos(\psi). \quad (16.2.222)$$

Finally observe from (2.74) that  $\mathcal{R}^{\text{rot}}$  leaves  $\tau$  and  $P_\tau$  in peace. You have shown that

$$R^{\text{rot}} = \begin{pmatrix} C & 0 & S & 0 & 0 & 0 \\ 0 & C & 0 & S & 0 & 0 \\ -S & 0 & C & 0 & 0 & 0 \\ 0 & -S & 0 & C & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (16.2.223)$$

Note that, according to (2.218), making the replacement  $b^{[0]} \rightarrow -b^{[0]}$  entails the replacement  $R^{\text{rot}} \rightarrow (R^{\text{rot}})^{-1}$ .

We are almost done. According to (2.184) there will be the matrix relation

$$R = R^{\text{rot}} R^{\text{nonrot}} = R^{\text{nonrot}} R^{\text{rot}} = M^{\text{uniform}}. \quad (16.2.224)$$

Verify that

$$R = M^{\text{uniform}} = \begin{pmatrix} C^2 & (1/k)SC & SC & (1/k)S^2 & 0 & 0 \\ -kSC & C^2 & -kS^2 & SC & 0 & 0 \\ -SC & -(1/k)S^2 & C^2 & (1/k)SC & 0 & 0 \\ kS^2 & -SC & -kSC & C^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \eta \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (16.2.225)$$

Show that (2.225) agrees with (13.4.1) when use is made of (2.59) and (1.5.81).

Verify that, if we wish, we may write

$$R = (R^{\text{rot}})^{1/2} R^{\text{nonrot}} (R^{\text{rot}})^{1/2} \quad (16.2.226)$$

where

$$(R^{\text{rot}})^{1/2} = \exp[-(L/2) : H_2^{\text{rot}} :] = \exp[(\psi/2) : J_z :]. \quad (16.2.227)$$

However, there does not seem to be much point in doing so except, perhaps, to exhibit the reversibility properties of  $R$ . See Section 36.2.

**16.2.17.** Review Section 2.5.3. This exercise explores properties of  $H^{\text{nonrot}}$  and  $\mathcal{M}^{\text{nonrot}}$ . We will employ the notation

$$\zeta = (X, P_x, Y, P_y, \tau, P_\tau). \quad (16.2.228)$$

Begin by considering *initial* conditions with the property

$$\zeta_1^{\text{in}} = X^{\text{in}} = 0, \quad (16.2.229)$$

$$\zeta_2^{\text{in}} = P_x^{\text{in}} = 0, \quad (16.2.230)$$

$$\zeta_3^{\text{in}} = Y^{\text{in}} = \text{anything}, \quad (16.2.231)$$

$$\zeta_4^{\text{in}} = P_y^{\text{in}} = \text{anything}, \quad (16.2.232)$$

$$\zeta_5^{\text{in}} = \tau^{\text{in}} = \text{anything}, \quad (16.2.233)$$

$$\zeta_6^{\text{in}} = P_{\tau}^{\text{in}} = \text{anything}. \quad (16.2.234)$$

These are the initial conditions for motion that is, at least initially, in the vertical plane. Show that the resulting subsequent motion, when governed by  $H^{\text{nonrot}}$ , remains in the vertical plane. In particular, show that then the associated *final* conditions under the action of  $\mathcal{M}^{\text{nonrot}}$  have the property

$$X^{\text{fin}} = 0, \quad (16.2.235)$$

$$P_x^{\text{fin}} = 0. \quad (16.2.236)$$

Next show that

$$: J_z : \mathcal{M}^{\text{nonrot}} = \mathcal{M}^{\text{nonrot}} : J_z :, \quad (16.2.237)$$

and consequently

$$\mathcal{R}^{\text{rot}} \mathcal{M}^{\text{nonrot}} = \mathcal{M}^{\text{nonrot}} \mathcal{R}^{\text{rot}}. \quad (16.2.238)$$

Suppose we summarize the results (2.229) through (2.236) by writing

$$\zeta^{\text{fin}} = \mathcal{M}^{\text{nonrot}} \zeta^{\text{in}}. \quad (16.2.239)$$

Use (2.238) and (2.239) to show that

$$\mathcal{M}^{\text{nonrot}} \mathcal{R}^{\text{rot}} \zeta^{\text{in}} = \mathcal{R}^{\text{rot}} \zeta^{\text{fin}}. \quad (16.2.240)$$

You have shown that  $\mathcal{M}^{\text{nonrot}}$  preserves *all* planes obtained by rotating, by any angle  $\psi$ , the vertical plane about the  $z$  axis. This includes, of course, the horizontal plane.

**16.2.18.** Review the discussion of Section 15.11.1. Using (2.3) through (2.6), show that for a solenoid the transverse components of its magnetic field have the integral property

$$\int_{-\infty}^{\infty} dz B_x(x, y, z) = \int_{-\infty}^{\infty} dz B_y(x, y, z) = 0. \quad (16.2.241)$$

**16.2.19.** Review Exercise 2.1. Consider a “one-turn” solenoid consisting of a single circular loop of radius  $a$  lying in the  $z = 0$  plane, centered on the origin, and carrying a current  $I$ . Show that in this case

$$B_z^{\text{one turn}}(0, 0, z) = \mu_0 I \delta(z, a) \quad (16.2.242)$$

where  $\delta(z, a)$  is an approximating delta function defined by

$$\delta(z, a) = (a^2/2)/(z^2 + a^2)^{3/2}. \quad (16.2.243)$$

Recall the discussion of Approximating Delta Functions in Subsection 15.11.6 and the relation (15.11.39). From this discussion one sees that

$$\int_{-\infty}^{\infty} dz \delta(z, a) = 1. \quad (16.2.244)$$

Show that there is the relation

$$\text{bump}(z, a, L) = \int_0^L dz' \delta(z - z', a). \quad (16.2.245)$$

Show that (2.23) follows from (2.244) and (2.245).

**16.2.20.** Verify that (2.134) through (2.136) reproduce the  $f_4$  content of Exhibits 2.19 through 2.21.

**16.2.21.** The purpose of this exercise is to study how the coefficients of the  $I_i$ , save for  $I_6$ , in the relations (2.134) through (2.136) tend to zero as  $a \rightarrow 0$ . Let us write relations of this kind in the general form

$$f_4^{\text{LFF}}(a) = \sum_{i=1}^{10} c_i^{\text{LFF}}(a) I_i \quad (16.2.246)$$

where we have indicated that  $f_4^{\text{LFF}}$  is  $a$  dependent. Verify that for the case of  $c_1^{\text{LFF}}(a)$  one may make the Table 2.4 below, and conclude that  $c_1^{\text{LFF}}(a)$  vanishes as  $a^2$  when  $a$  goes to zero. Verify that the same is true for  $c_2^{\text{LFF}}(a)$ ,  $c_7^{\text{LFF}}(a)$ , and  $c_8^{\text{LFF}}(a)$ . By contrast, verify that  $c_3^{\text{LFF}}(a)$ ,  $c_4^{\text{LFF}}(a)$ ,  $c_5^{\text{LFF}}(a)$ , and  $c_9^{\text{LFF}}(a)$  vanish as  $a^1$  when  $a$  goes to zero. Evidently  $c_{10}^{\text{LFF}}(a)$  vanishes for all values of  $a$ .

Table 16.2.4: Numerical behavior of  $c_1^{\text{LFF}}(a)$  for small values of  $a$ .

$a$	$c_1^{\text{LFF}}$	$(1/a^2)c_1^{\text{LFF}}$
.2	4.3607E-5	1.09019E-3
.02	5.2063E-7	1.30158E-3
.002	5.3334E-9	1.33335E-3

**16.2.22.** By making a suitable table, illustrate that  $c_6^{\text{LFF}}(a)$  diverges as  $1/a$  when  $a$  goes to zero.

**16.2.23.** Show that  $H_4$  as given by (2.96) can be written in the form

$$H_4 = \sum_{i=1}^{10} d_i(z) I_i, \quad (16.2.247)$$

and verify that it is  $d_5$  and  $d_6$  that contain the pesky  $\delta'$  functions. In particular, verify that  $d_5$  contains  $b^{[2]}$  (which involves  $\delta'$  functions) and  $d_6$  contains the product  $b^{[0]}b^{[2]}$ . Reason that the latter is more singular than the former because  $b^{[0]}$  is discontinuous in the hard-edge limit. Show that there are the results

$$\text{sgn}'(z, a) = \partial_z \text{sgn}(z, a) = 2\delta(z, a) \quad (16.2.248)$$

and

$$\begin{aligned} \int_{-\infty}^{\infty} dz \text{sgn}(z, a) \delta'(z, a) &= [\text{sgn}(z, a) \delta(z, a)]|_{z=-\infty}^{z=+\infty} - \int_{-\infty}^{\infty} dz \text{sgn}'(z, a) \delta(z, a) \\ &= 0 - 2 \int_{-\infty}^{\infty} dz \delta(z, a) \delta(z, a) = -2 \int_{-\infty}^{\infty} dz \delta^2(z, a) = -(3\pi/16)(1/a). \end{aligned} \quad (16.2.249)$$

Review and, if you have not already done so, perform Exercise 2.22. In agreement with the results of this exercise, make the small  $a$  Ansatz

$$c_6^{\text{LFF}}(a) \simeq e^{\text{LFF}}/a \quad (16.2.250)$$

where  $e^{\text{LFF}}$  is a coefficient to be determined. Find a formula for  $e^{\text{LFF}}$  in terms of beam parameters and the parameters for a simple solenoid.

**16.2.24.** Exercise on the vector spherical harmonic decomposition of the vector potential for the magnetic field of a simple solenoid.

## 16.3 Two Common Iron-Free Dipoles

### 16.3.1 Preliminaries

A *straight* dipole is a straight beam-line element whose field is described by a cylindrical harmonic expansion that contains primarily an  $m = 1$  term.<sup>8</sup> We recall from Section 15.3.3 that for the  $m = 1$  case the magnetic scalar potential  $\psi$  has the expansion

$$\psi(x, y, z) = \psi_{1,s}(x, y, z) = y[C_{1,s}^{[0]}(z) - (1/8)(x^2 + y^2)C_{1,s}^{[2]}(z) + \dots]. \quad (16.3.1)$$

See (15.3.57). [Here we have retained only the *normal*  $m = 1$  term, but a skew ( $\psi_{1,c}$ ) term is also possible. See (15.3.40) and Exercise 15.4.1.] Correspondingly, the associated magnetic field has the expansion

$$B_x = \partial_x \psi_{1,s} = -(1/4)xyC_{1,s}^{[2]}(z) + \dots, \quad (16.3.2)$$

$$B_y = \partial_y \psi_{1,s} = C_{1,s}^{[0]}(z) - (1/8)(x^2 + 3y^2)C_{1,s}^{[2]}(z) + \dots, \quad (16.3.3)$$

$$B_z = \partial_z \psi_{1,s} = y[C_{1,s}^{[1]}(z) - (1/8)(x^2 + y^2)C_{1,s}^{[3]}(z) + \dots]. \quad (16.3.4)$$

From (3.2) through (3.4) we see that  $\mathbf{B}$  is completely specified in terms of a single “master” function  $C_{1,s}^{[0]}(z)$  and its derivatives. Moreover, according to (3.3), the on-axis field has only a  $B_y$  component, and it is given by the relation

$$B_y(0, 0, z) = C_{1,s}^{[0]}(z). \quad (16.3.5)$$

See Exercise 1.5.7.

There is also a suitable associated vector potential  $\hat{\mathbf{A}}^{1,s}$  given (in symmetric Coulomb gauge) by the relations

$$\hat{A}_x^{1,s} = (1/4)(x^2 - y^2)C_{1,s}^{[1]}(z) - (1/48)(x^4 - y^4)C_{1,s}^{[3]}(z) + \dots, \quad (16.3.6)$$

$$\hat{A}_y^{1,s} = (1/2)xyC_{1,s}^{[1]}(z) - (1/24)(x^3y + xy^3)C_{1,s}^{[3]}(z) + \dots, \quad (16.3.7)$$

---

<sup>8</sup>In practice dipoles are often *bent* because the design orbit in a dipole is *curved*. In this case a cylindrical harmonic expansion is of limited use. See Subsection 3.7.

$$\hat{A}_z^{1,s} = -x C_{1,s}^{[0]}(z) + (1/8)(x^3 + xy^2) C_{1,s}^{[2]}(z) - (1/192)(x^5 + 2x^3y^2 + xy^4) C_{1,s}^{[4]}(z) + \dots \quad (16.3.8)$$

Recall (15.5.97) through (15.5.99). We observe that  $\hat{\mathbf{A}}^{1,s}$ , like  $\mathbf{B}$ , is also completely specified in terms of  $C_{1,s}^{[0]}(z)$  and its derivatives. Vector potentials in other gauges are also possible, in particular the azimuthal-free gauge, and they too are completely specified in terms of  $C_{1,s}^{[0]}(z)$  and its derivatives. Recall Sections 15.4 through 15.7.

Eventually we will want to examine in some mathematical detail what can be said about the master function  $C_{1,s}^{[0]}(z)$  in various cases. But first we will describe two commonly used current windings.

### 16.3.2 Current Windings for two Common Air-Core Dipoles

So far, unlike the case of a solenoid where a current distribution was specified, no prescription has been given of current windings/distributions that could be fabricated to produce the field of a reasonably “pure dipole”. The aim of this subsection is to describe two commonly used (or contemplated) thin-shell windings on a straight circular cylinder of radius  $a$  and length  $L$  such that the magnetic field produced within the bore has *primarily* an  $m = 1$  component.

Then, in the next two subsections, we will consider windings such that *only* the  $m = 1$  on-axis gradient is nonzero for the field produced by such windings. We will call such a winding an *ideal* air-core dipole. However it is not the case, even for an ideal air-core dipole, that the field is that of a *perfect* dipole, a field, say, only in the  $\mathbf{e}_y$  direction. According to (15.2.61) through (15.2.64) there are additional components in the fringe-field regions at the ends of the dipole where  $C_{1,s}^{[0]}(z)$  is changing.

As already stated, we will consider air-core dipoles that consist of a thin-shell winding placed on a circular cylinder of radius  $\rho = a$  and length  $L$ . We will also arrange the coordinate system so that the winding begins at  $z = 0$  and ends at  $z = L$ .

There are two commonly used or considered approaches as to what the nature of this winding should be. One approach is to arrange to have most of the winding running on straight lines parallel to the cylinder axis to form what are called *saddle coils*. See Figures 3.1 and 3.2. Moreover, the spacing between successive straight lines is arranged so that the cross-sectional current density for the straight-line portion of the winding has (effectively on average) a  $\cos(\theta)$  distribution. [Here suppose the coordinate system shown in Figure 2.1 is also employed in part *c* of Figure 3.1 below. Then, following customary nomenclature,  $\theta$  is the angle  $\phi$  defined by (15.2.12) through (15.2.14). From now on we will refer to a  $\cos(\phi)$  distribution.] This can be accomplished by placing the winding in grooves machined into the underlying cylinder or by placing appropriately sized variable-width spacers (not depicted) between successive straight wires.

It can be shown that for a  $\cos(\phi)$  current distribution for the straight-line portion of the winding, and assuming the coils are long so that  $L$  is large, the field in the vicinity of  $z = L/2$  is that of a reasonably pure dipole. See Exercise 3.5. That is, we already know from symmetry considerations, see Section 15.3.5, that only the  $C_{m,s}^{[0]}(z)$  for odd  $m$  can be nonzero. And for a  $\cos(\phi)$  current distribution for the straight-line portion of the winding it can be shown that  $C_{1,s}^{[0]}(z)$  is substantial, and the remaining  $C_{m,s}^{[0]}(z)$  are small, in the vicinity of  $z = L/2$ .



Figure 16.3.1: Coils and cylinders: Part *c* of this figure shows coils draped like saddles, above and below, over a circular cylinder. Apart from the coil ends, most of the winding runs along straight lines parallel to the cylinder axis.

What can be said about the field *away* from the central region  $z \simeq L/2$ ? There the  $C_{3,s}^{[0]}(z)$ ,  $C_{5,s}^{[0]}(z)$ ,  $\dots$  can be substantial due to the currents flowing around the cylinder at the coil ends. Thus, an air-core dipole made with saddle  $\cos(\phi)$  coils is not ideal as defined above. Moreover, in general there will be the integral results

$$\int_{-\infty}^{\infty} dz C_{m,s}^{[0]}(z) \neq 0 \text{ for } m \text{ odd.} \quad (16.3.9)$$

In particular the undesired  $m = 3, m = 5, \dots$  integrated multipole strengths will in general be nonzero.

But it is in principle possible to drive various undesired *integrated* multipole strengths to zero by placing and appropriately powering suitable multipole corrector windings at or near the ends of the main coil. Of course, when this is done, the net  $C_{3,s}^{[0]}(z)$ ,  $C_{5,s}^{[0]}(z)$ ,  $\dots$  remain nonzero. Only their integrals vanish.

A second approach, variously called “canted  $\cos(\theta)$ ”, “double helix”, or “tilted solenoid”, is based on a simple configuration where a conductor is wound around a cylinder as two oppositely *tilted* solenoids. Figure 3.3 illustrates such a winding.<sup>9</sup> See the Ph.D. thesis of *Brouwer* cited at the end of this chapter for an extensive treatment of canted  $\cos(\theta)$  dipoles.

It can be shown that two tilted solenoids when powered as shown produce what is primarily a dipole field. Each layer produces a combination of vertical and solenoidal fields. If the currents are directed as shown, the vertical components *add* to produce a primarily dipole field and the solenoidal components *cancel* (save for end effects which are modest and also integrate to zero).<sup>10</sup> (And if the current in one of the layers is directed as shown and the other is reversed, the vertical components essentially cancel and the solenoidal components

<sup>9</sup>The use of the term “ $\cos(\theta)$ ” in this context may seem slightly confusing since each tilted solenoid is wound *uniformly* with no variable width spacers between turns. However, as study of Figure 3.3 suggests, it can be shown that the net effect of the two tilted layers is to produce, in the overlap region and in the  $z$  direction, a  $\cos(\phi)$  current distribution.

<sup>10</sup>Here we assume the layers have infinitesimal thickness so that they both have the same radii. If not,

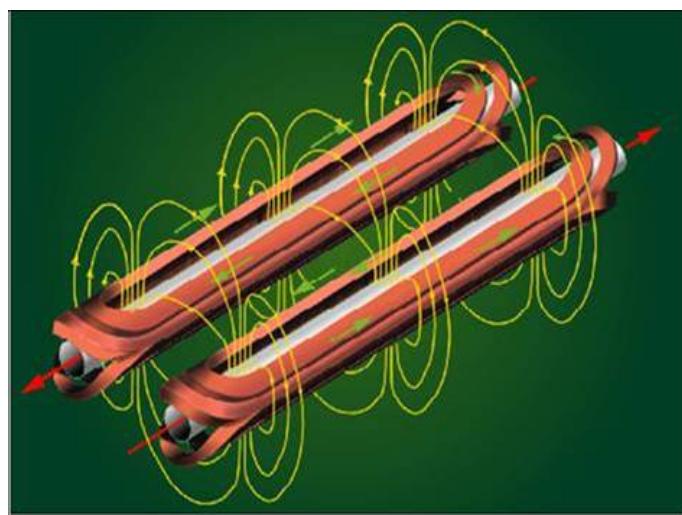


Figure 16.3.2: Artist's illustration of coils for the two-in-one 15 meter long dipoles of the CERN Large Hadron Collider. Note that the fringe fields near the dipole ends are *not* portrayed.

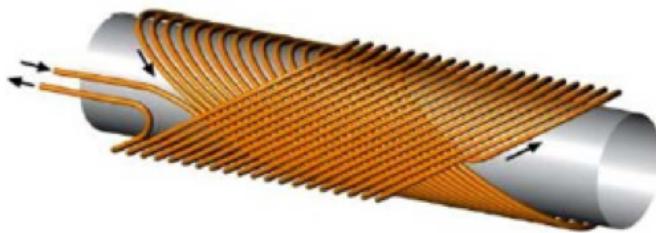


Figure 16.3.3: A winding composed of two oppositely tilted solenoids to form a canted  $\cos(\theta)$  dipole.

add to produce a primarily solenoidal field.) Figure 3.4 shows, in Tesla, the solenoidal field component  $B_z(0, 0, z) = C_0^{[1]}(z)$  for the individual layers as well as the net  $C_0^{[1]}(z)$  for a two-layer canted  $\cos(\theta)$  dipole. Note that, due to the advertised cancellation, the net solenoidal field is small save for end effects. [And at the ends the solenoidal field oscillates (takes both signs) so that it essentially integrates to zero.] The solenoidal field at the center location  $z = L/2$  is 0.003 Tesla and arises from the layers having slightly different radii.

With regard to the dipole and higher multipole fields, and again from symmetry considerations, only the  $C_{m,s}^{[0]}(z)$  for odd  $m$  can be nonzero. And, if  $L$  is large, it can be shown that  $C_{1,s}^{[0]}(z)$  is substantial, and the remaining  $C_{m,s}^{[0]}(z)$  are small, in the vicinity of  $z = L/2$ . Figure 3.5 shows, for example,  $C_{1,s}^{[0]}(z)$ ; and Figure 3.6 shows  $C_{3,s}^{[0]}(z)$  and  $C_{5,s}^{[0]}(z)$ . But note that the  $C_{3,s}^{[0]}(z)$ ,  $C_{5,s}^{[0]}(z)$ , ... are *not* zero for all  $z$ , and therefore this air-core dipole is also not ideal.

Nevertheless, this air-core dipole does have a remarkable property: It can be shown that  $\int_{-\infty}^{\infty} dz C_{1,s}^{[0]}(z)$  is substantial, and for all the remaining  $C_{m,s}^{[0]}(z)$  there is the relation

$$\int_{-\infty}^{\infty} dz C_{m,s}^{[0]}(z) = 0. \quad (16.3.10)$$

That is, the *integrated* strengths of all undesired multipoles vanish for this air-core winding!<sup>11</sup> Again see Figure 3.6.<sup>11</sup>

---

cancellation is not perfect so that there is a small residual solenoidal field even apart from end effects. In any event the effect of solenoidal fields can presumably be compensated if desired by the addition of other windings or the use of skew quadrupoles. For example, the  $x, y$  coupling effect of the strong solenoidal fields associated with some detectors in storage rings/colliders is routinely compensated by the use of skew quadrupoles.

<sup>11</sup>Windings (for air-core magnets) which have the property that all integrated multipoles vanish, save for some desired multipole, are sometimes called *Lambertson* windings. See the references at the end of this chapter.

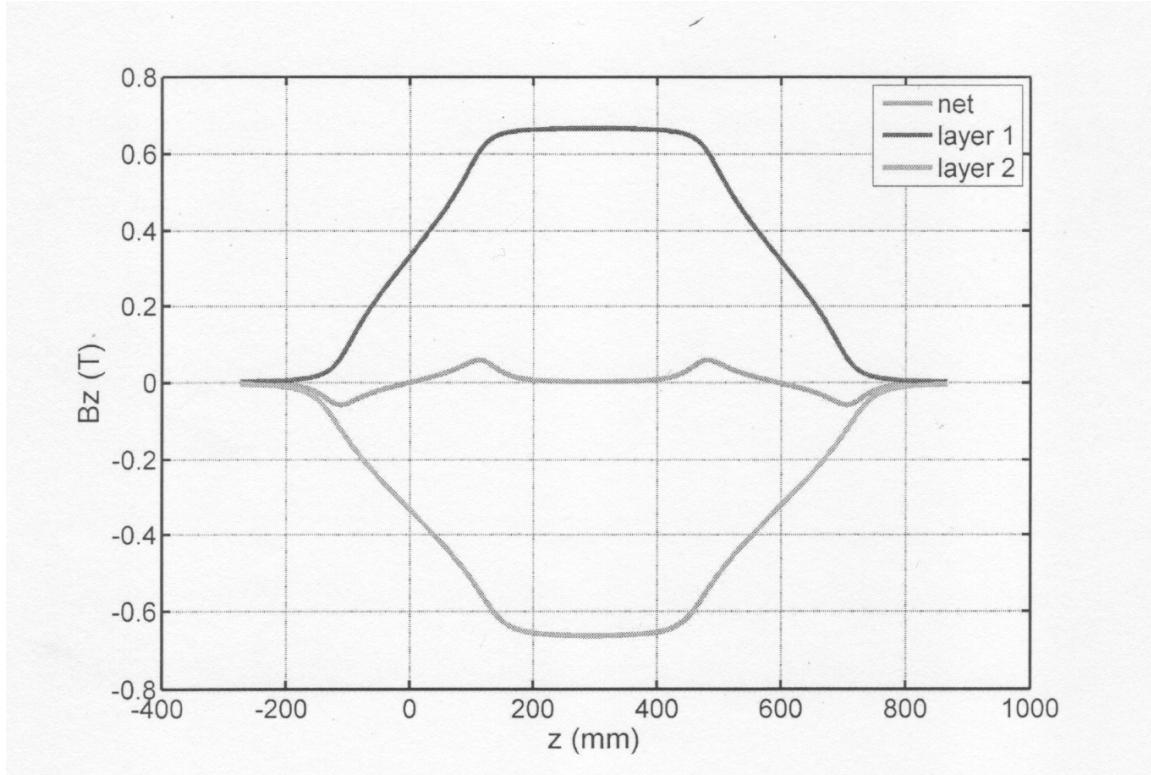


Figure 16.3.4: The individual and net  $C_0^{[1]}(z)$  for a canted  $\cos(\theta)$  dipole.

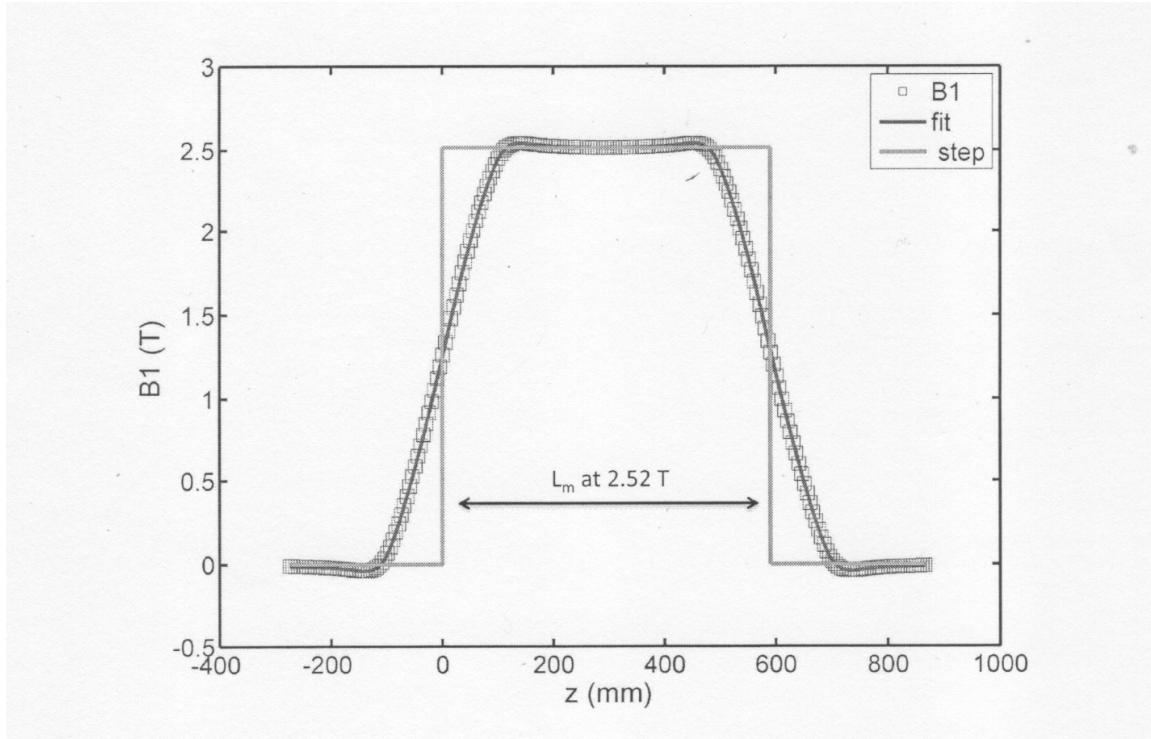


Figure 16.3.5: The on-axis gradient  $C_{1,s}^{[0]}(z)$  in Tesla for a canted  $\cos(\theta)$  dipole. Also shown is a hard-edge bump function approximation.

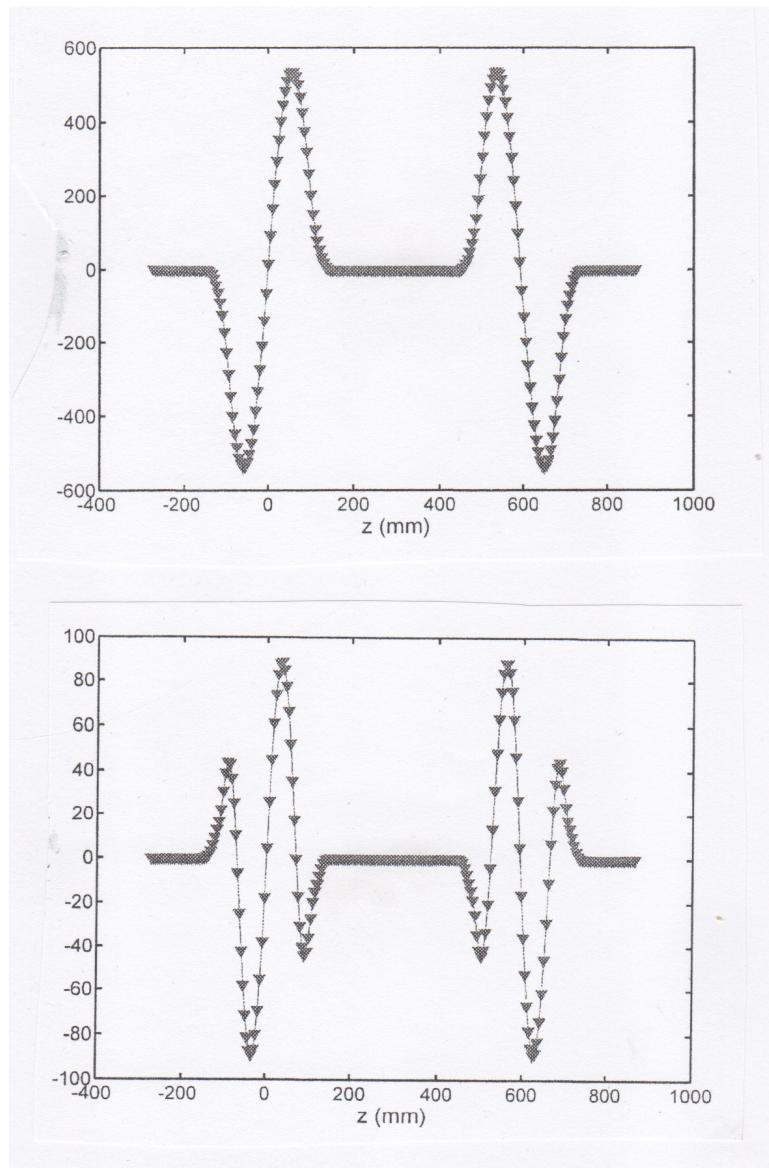


Figure 16.3.6: The on-axis gradient  $C_{3,s}^{[0]}(z)$  (above) and on-axis gradient  $C_{5,s}^{[0]}(z)$  (below) in dimensionless units for a canted  $\cos(\theta)$  dipole. They are small for  $z \simeq L/2$ , but do not vanish everywhere. Nevertheless their integrated strengths do vanish.

## 16.4 Bassetti-Biscari Windings for Pure Multipoles

We have seen two examples of windings that fail to produce ideal air-core dipoles. Is there a winding that succeeds? Yes, there are several. This section will describe windings due to Bassetti and Biscari that produce pure multipoles, are easy to visualize, and for which it is possible to compute the  $C_{m,s}^{[n]}(z)$  *analytically*. See their work cited in the references at the end of this chapter. In the next section we will describe stream-function windings that also produce pure multipoles.

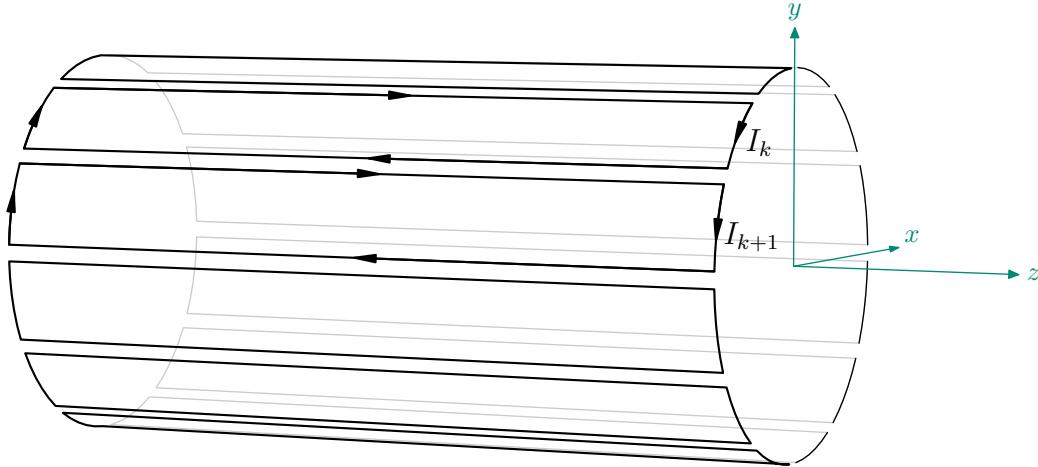


Figure 16.4.1: A net of  $n$  coils draped over a cylinder. The  $k^{\text{th}}$  coil carries a current  $I_k$ .

### 16.4.1 Winding Geometry

Figure 4.1 above shows a net of  $n$  separate single-turn coils placed on a circular cylinder. And Figure 4.2 below shows a top view of the right end of the cylinder and illustrates the sign convention for describing the current flow in the right (*positive z*) end of each coil and the currents as they flow along the long sides of each coil. A dot denotes the tip of an arrow as it comes up from below the plane of the figure, and a cross denotes the feather of an arrow as it goes down below the plane of the figure. In this figure  $n = 12$  coils are displayed with the circular arc of each  $I_k$  current subtending an angle of  $\Delta = 2\pi/n = 30^\circ$ . We also define angles  $\phi_k$  by the rule

$$\phi_k = k\Delta. \quad (16.4.1)$$

Thus, at each angular location  $\phi_k$  there is an upward and downward (increasing  $z$  and decreasing  $z$ ) current pair so that, for example, there is a *net* current  $(I_1 - I_n)$  along the side of the cylinder in the  $z$  direction at the location  $\phi_0 = 0$ .

### 16.4.2 Specification of Currents for a Pure Dipole

In order to achieve an effective  $\cos\phi$  current distribution along the *length* of the cylinder, the currents in adjacent coils are required to be related by the rules

$$I_2 - I_1 = \hat{I} \cos \phi_1, \quad (16.4.2)$$

$$\begin{aligned} I_3 - I_2 &= \hat{I} \cos \phi_2, \\ I_{k+1} - I_k &= \hat{I} \cos \phi_k, \\ \dots &\dots, \\ I_n - I_{n-1} &= \hat{I} \cos \phi_{n-1}, \\ I_1 - I_n &= \hat{I} \cos \phi_0 = \hat{I}, \end{aligned} \quad (16.4.3)$$

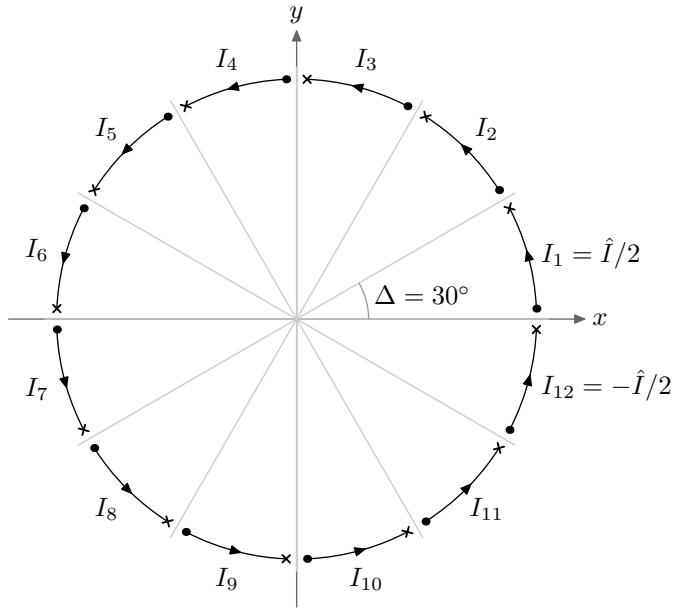


Figure 16.4.2: Top view of the right ends of the coils shown in Figure 3.7. The  $z$  axis comes out of the plane of the paper.

where  $\hat{I}$  is some amount of current yet to be determined. As a sanity check, we observe that the sum of the left sides of (3.12) through (3.13) vanishes. And for the right sides computation shows that the sum also vanishes, as desired,

$$\begin{aligned} \hat{I}[1 + \cos \phi_1 + \cos \phi_2 + \cdots + \cos \phi_{n-1}] &= \hat{I}\Re\left\{\sum_{k=0}^{n-1} \exp(ik\Delta)\right\} \\ &= \hat{I}\Re\{[1 - \exp(in\Delta)]/[1 - \exp(i\Delta)]\} \\ &= \hat{I}\Re\{[1 - \exp(2\pi i)]/[1 - \exp(i2\pi/n)]\} = 0. \end{aligned} \quad (16.4.4)$$

Observe that the left sides of (3.12) through (3.13) only involve current *differences*, and therefore the currents themselves still need to be determined. This can be done by specifying that

$$I_1 = \hat{I}/2. \quad (16.4.5)$$

Note that this specification, when employed with the last equation (3.13), produces the pleasant result

$$I_n = -\hat{I}/2. \quad (16.4.6)$$

It can be shown that (3.12) through (3.13) and (3.15), when taken together, yield the relations

$$I_k = (\hat{I}/2)[\sin(\Delta/2)]^{-1} \sin \psi_k \quad (16.4.7)$$

where

$$\psi_k = (k - 1/2)\Delta. \quad (16.4.8)$$

See Exercise \*. Note that  $\psi_k$  is the angular location of the *midpoint* of the arc that carries the current  $I_k$ . For example,  $\psi_1 = \Delta/2$ .

### 16.4.3 Description of Resulting On-Axis Gradient and On-Axis Field

Consider cases for which  $n$  has a value of the form  $n = 4\ell$  where  $\ell$  is an integer. Then, in the *continuum* large  $\ell$  limit (and with  $\hat{I}$  adjusted accordingly), it can be shown that the collection of coils of the kind shown schematically in Figure 3.7 and more precisely in Figure 3.8 (for the case  $\ell = 3$ ), with the  $I_k$  described by (3.17) and (3.18), produces an ideal dipole field. That is,  $C_{1,s}^{[0]}(z)$  is substantial, and the remaining  $C_{m,s}^{[0]}(z)$  all vanish. See Exercise \*.

Moreover, in this case  $C_{1,s}^{[0]}(z)$ , and the  $C_{1,s}^{[n]}(z)$ , can be computed analytically. If the winding is on a cylinder of radius  $a$  that begins at  $z = 0$  and extends to  $z = L$ , then the on-axis gradient for such an air-core dipole, as is the case for the on-axis gradient for a simple air-core solenoid, can be described in terms of a soft-edge bump function which, for our *ideal air-core dipole*, we will call  $\text{bump}_{\text{iacd}}(z, a, L)$ . That is, the on-axis gradient  $C_{1,s}^{[0]}(z)$  can be written in the form

$$C_{1,s}^{[0]}(z) = B_0 \text{bump}_{\text{iacd}}(z, a, L) \quad (16.4.9)$$

where  $B_0$  is again the dipole strength at  $z = L/2$  in the infinite length limit.

Like the soft-edge bump function for a simple air-core solenoid, the soft-edge bump function for this ideal air-core dipole can be written in terms of an associated approximating signum function  $\text{sgn}_{\text{iacd}}(z, a)$  in the form

$$\text{bump}_{\text{iacd}}(z, a, L) = [\text{sgn}_{\text{iacd}}(z, a) - \text{sgn}_{\text{iacd}}(z - L, a)]/2. \quad (16.4.10)$$

It can be shown that for this ideal air-core dipole the approximating signum function is given by the relation

$$\text{sgn}_{\text{iacd}}(z, a) = z(z^2 + 2a^2)/(z^2 + a^2)^{3/2}. \quad (16.4.11)$$

See Exercise 6.1. Figures 3.9 and 3.10 illustrate the behavior of this approximating signum function for two different values of  $a$ . Evidently the approximating signum function (3.21) becomes the true signum function in the limit  $a \rightarrow 0$ .

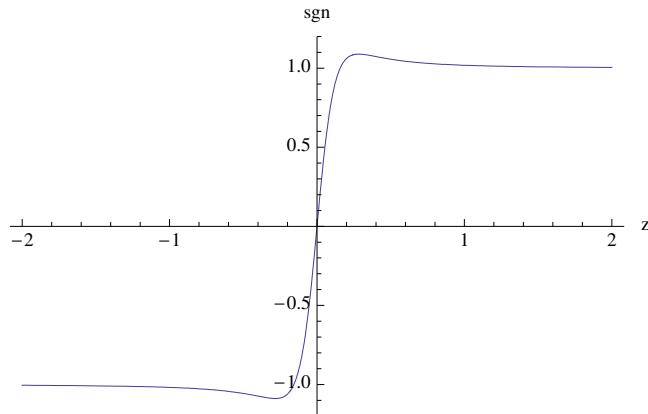


Figure 16.4.3: The approximating signum function (3.21) when  $a = .2$ .

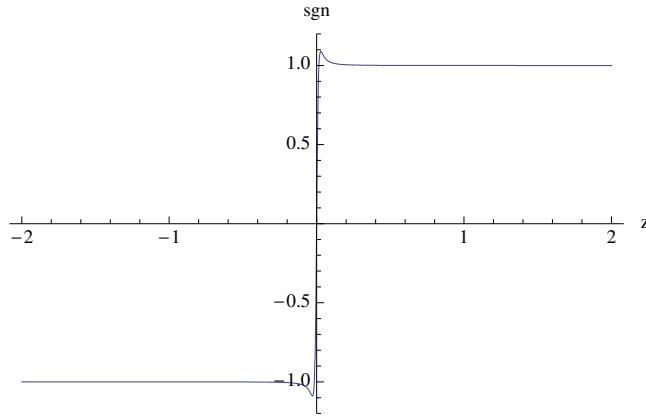


Figure 16.4.4: The approximating signum function (3.21) when  $a = .02$ .

It follows from (3.20) and (3.21) that, like the soft-edge bump functions for a simple air-core solenoid, the soft-edge bump function for this ideal air-core dipole satisfies the relations (2.20) through (2.23). Recall Exercise 2.3. We see from (2.23) and (3.19) that for this ideal air-core dipole there is the relation

$$\int_{-\infty}^{\infty} C_{1,s}^{[0]}(z) dz = B_0 L. \quad (16.4.12)$$

Figures 3.11 and 3.12 illustrate the properties (2.20) through (2.22) for a fixed value of  $L$  and two different values of the radius  $a$ . Evidently the ideal air-core dipole soft-edge bump function given by (3.20) and (3.21) becomes a hard-edge bump function in the limit  $a \rightarrow 0$ . The radius  $a$  plays the role of a characteristic length that controls the rate of fall off. The fringe-field region is large if  $a$  is large, and vanishes as  $a$  goes to zero.

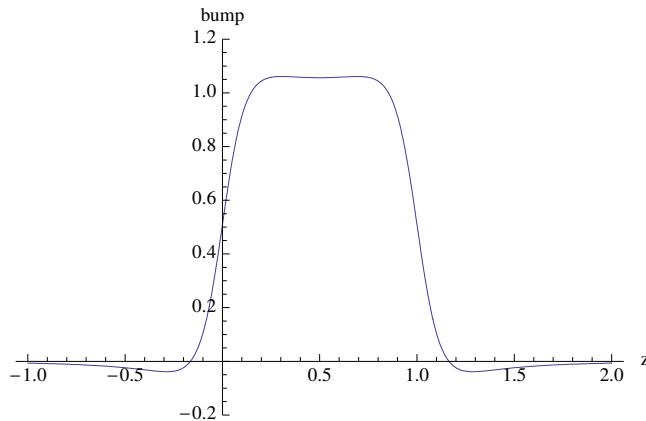


Figure 16.4.5: The soft-edge bump function given by (3.20) and (3.21) when  $a = 0.2$  and  $L = 1$ .

From (3.21) and (3.20) we find the asymptotic behaviors

$$\text{sgn}_{\text{iacd}}(z, a) = 1 + (1/2)a^2/z^2 + O(1/z^4) \text{ as } z \rightarrow +\infty, \quad (16.4.13)$$

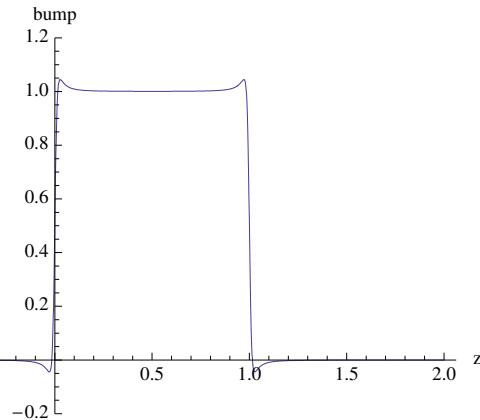


Figure 16.4.6: The soft-edge bump function given by (3.20) and (3.21) when  $a = 0.02$  and  $L = 1$ .

$$\text{bump}_{\text{iacd}}(z, a, L) = -(1/2)L a^2 / |z|^3 + O(1/|z|^4) \text{ as } |z| \rightarrow \infty. \quad (16.4.14)$$

Consequently  $C_{1,s}^{[0]}(z)$  falls off for large distances as

$$C_{1,s}^{[0]}(z) = -(1/2)B_0 L a^2 / |z|^3 + O(1/|z|^4) \text{ as } |z| \rightarrow \infty. \quad (16.4.15)$$

We see that the fall off goes as  $1/|z|^3$ , just as it does for the simple air-core solenoid. Compare (2.42) and (3.25). Note, however, that the approximating signum functions (2.26) and (3.21) are different. Compare, for example, Figures 2.2 and 3.9. Also, compare (2.34) with (3.24). Correspondingly, the bump functions for the model of Subsection 3.2 and this idealized air-core dipole are different, and the relations (2.42) and (3.25) differ in sign. Finally we observe, for example from Figures 3.11 and 3.12, that for this ideal air-core dipole the bump function  $\text{bump}_{\text{iacd}}(z, a, L)$ , while positive in the center of the dipole, can be *negative* outside the dipole. This change in sign *cannot* be modeled by an *Enge* function profile.<sup>12</sup>

#### 16.4.4 (Place Holder) Mathematical Model for a Pure Normal Dipole Based on both Single Layer and Double Layer Monopole Distributions

#### 16.4.5 Specification of Currents for a Pure Quadrupole

##### Preliminaries

A quadrupole is a straight beam-line element whose field is described by a cylindrical harmonic expansion that contains primarily an  $m = 2$  term. We recall from Section 15.3.3 that in this case the magnetic scalar potential  $\psi$  has the expansion

$$\psi(x, y, z) = \psi_{2,s}(x, y, z) = 2xy[C_{2,s}^{[0]}(z) - (1/24)(x^2 + y^2)C_{2,s}^{[2]}(z) + \dots]. \quad (16.4.16)$$

---

<sup>12</sup>Note that for a canted  $\cos(\theta)$  dipole, see Figure 3.5, the field can also be negative outside the dipole.

(Here we have retained only the *normal* term, but a skew term is also possible.) See (15.3.67). Correspondingly, the associated magnetic field has the expansion

$$B_x = \partial_x \psi_{2,s} = 2y C_{2,s}^{[0]}(z) - (1/12)(3x^2y + y^3)C_{2,s}^{[2]}(z) + \dots, \quad (16.4.17)$$

$$B_y = \partial_y \psi_{2,s} = 2x C_{2,s}^{[0]}(z) - (1/12)(x^3 + 3xy^2)C_{2,s}^{[2]}(z) + \dots, \quad (16.4.18)$$

$$B_z = \partial_z \psi_{2,s} = 2xy[C_{2,s}^{[1]}(z) - (1/24)(x^2 + y^2)C_{2,s}^{[3]}(z) + \dots]. \quad (16.4.19)$$

From (5.2) through (5.4) we see that  $\mathbf{B}$  is completely specified in terms of a single “master” function  $C_{2,s}^{[0]}(z)$  and its derivatives. Moreover, according to (5.2) and (5.3), the on-axis field is characterized by a quadrupole strength  $Q$  given by the relation

$$Q(0, 0, z) = 2C_{2,s}^{[0]}(z). \quad (16.4.20)$$

See Exercise 1.5.9. We will next examine what can be said about the master function  $C_{2,s}^{[0]}(z)$  in various cases.

### Current Winding for a particular Ideal Air-Core Quadrupole

An ideal air-core quadrupole consists of a thin-shell winding having a circular cross section of radius  $\rho = a$  and length  $L$ . The current in the winding ideally has a  $\cos(2\phi)$  dependence, which results in a “pure” quadrupole field within the bore. That is, *only* the  $m = 2$  on-axis gradient is nonzero for the field produced by such a winding. However, it is not the case that the field is that of a *perfect* quadrupole. According to (15.2.65) through (15.2.68) there are additional components in the fringe-field regions at the ends of the quadrupole where  $C_{2,s}^{[0]}(z)$  is changing.

#### 16.4.6 Description of Resulting On-Axis Gradient and On-Axis Field

If the winding begins at  $z = 0$  and extends to  $z = L$ , then the on-axis field gradient for such an *ideal air-core quadrupole* can again be described in terms of a soft-edge bump function which we will call  $\text{bump}_{\text{iacq}}(z, a, L)$ . That is, the on-axis field gradient  $C_{2,s}^{[0]}(z)$  can again be written in the form

$$C_{2,s}^{[0]}(z) = (Q_0/2)\text{bump}_{\text{iacq}}(z, a, L) \quad (16.4.21)$$

where  $Q_0$  is again the quadrupole strength in the infinite length limit.

Like the previous examples of soft-edge bump functions, the soft-edge bump function for an ideal air-core quadrupole can be written in terms of an associated approximating signum function  $\text{sgn}_{\text{iacq}}$  in the form

$$\text{bump}_{\text{iacq}}(z, a, L) = [\text{sgn}_{\text{iacq}}(z, a) - \text{sgn}_{\text{iacq}}(z - L, a)]/2. \quad (16.4.22)$$

It can be shown that for an ideal air-core quadrupole the approximating signum function  $\text{sgn}_{\text{iacq}}(z, a)$  is given by the relation

$$\begin{aligned} \text{sgn}_{\text{iacq}}(z, a) &= [z^5 + (5/2)z^3a^2 + (9/4)za^4]/(z^2 + a^2)^{5/2} \\ &= z[z^4 + (5/2)z^2a^2 + (9/4)a^4]/(z^2 + a^2)^{5/2}. \end{aligned} \quad (16.4.23)$$

See Exercise 6.1. Figures 5.7 and 5.8 illustrate the behavior of this approximating signum function for two different values of  $a$ . Evidently the approximating signum function (5.29) becomes the true signum function in the limit  $a \rightarrow 0$ .

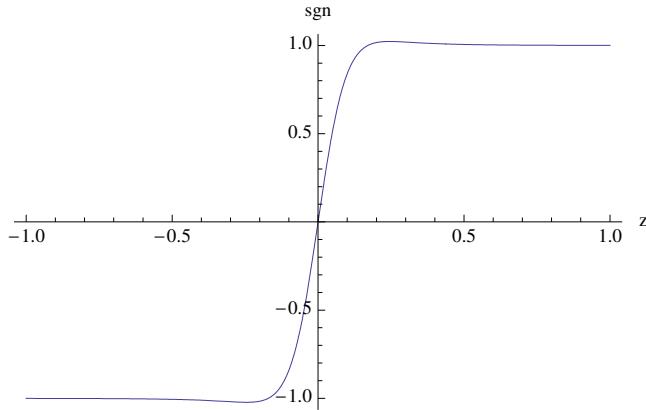


Figure 16.4.7: The approximating signum function (5.29) when  $a = .2$ .

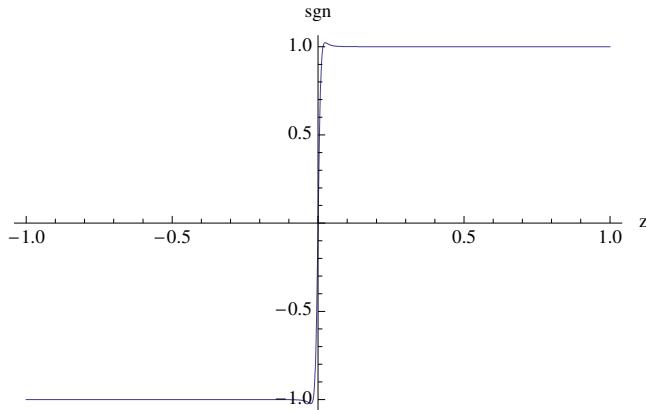


Figure 16.4.8: The approximating signum function (5.29) when  $a = .02$ .

It follows from (5.28) and (5.29) that, like the previous soft-edge bump functions, the soft-edge bump function for an idealized air-core quadrupole satisfies the relations (1.15) through (1.18). Recall Exercise 1.2. Also, the relation (5.23) continues to hold.

Figures 5.9 and 5.10 illustrate the properties (1.15) through (1.17) for a fixed value of  $L$  and two different values of the radius  $a$ . Evidently the ideal air-core quadrupole soft-edge bump function becomes a hard-edge bump function in the limit  $a \rightarrow 0$ . The radius  $a$  plays the role of a characteristic length that controls the rate of fall off. The fringe-field region is large if  $a$  is large, and vanishes as  $a$  goes to zero. From (5.29) and (5.28) we find the asymptotic behaviors

$$\text{sgn}_{\text{iacq}}(z, a) = 1 + (3/8)(a/z)^4 + O(1/z^6) \text{ as } z \rightarrow \infty, \quad (16.4.24)$$

$$\text{bump}_{\text{iacq}}(z, a, L) = -(3/4)L a^4 / |z|^5 + O(1/|z|^6) \text{ as } |z| \rightarrow \infty. \quad (16.4.25)$$

Consequently  $C_{2,s}^{[0]}(z)$  falls off for large distances as

$$C_{2,s}^{[0]}(z) = -(3/4)(Q/2)L a^4 / |z|^5 + O(1/|z|^6) \text{ as } |z| \rightarrow \infty. \quad (16.4.26)$$

We see that the fall off goes as  $1/|z|^5$ , which is the same rate as that for the cylindrical quadrupole consisting of a uniform distribution of pure quadrupole rings as described in Subsection 5.2. Note, however, that the sign on the right side of (5.31) is negative just as it is for the right side of (3.31). Therefore, the bump function for an idealized air-core quadrupole also cannot be modeled by an Enge function profile.

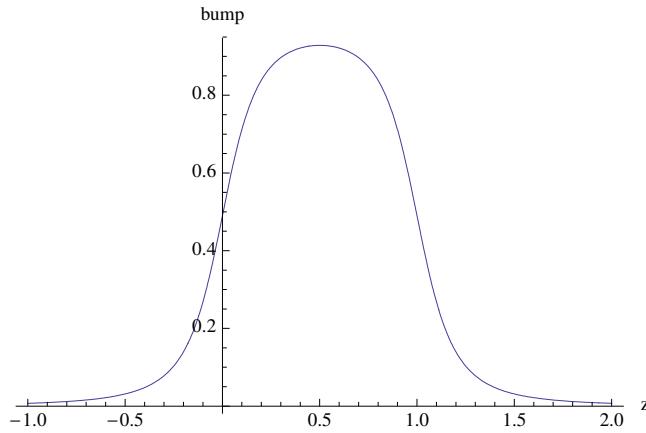


Figure 16.4.9: The soft-edge bump function given by (5.28) and (5.29) when  $a = .2$  and  $L = 1$ .

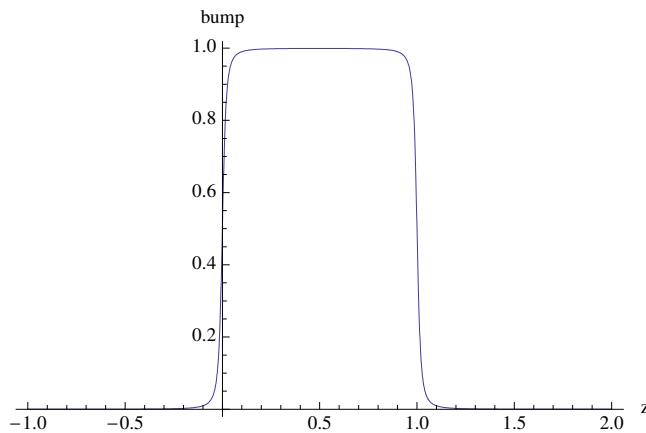


Figure 16.4.10: The soft-edge bump function given by (5.28) and (5.29) when  $a = .02$  and  $L = 1$ .

### 16.4.7 (Place Holder) Mathematical Model for a Pure Normal Quadrupole Based on both Single Layer and Double Layer Monopole Distributions

### 16.4.8 Sextupoles and Beyond

A sextupole is a beam-line element whose field is described by a cylindrical harmonic expansion that contains primarily an  $m = 3$  term. The simplest way to produce a sextupole field, as the name again suggests, is to properly locate and assign strengths to six monopoles. In the case of a sextupole the sextet of monopoles can be taken to be three doublets rotated successively by 60 degrees. We already know from the work of Section 15.8.2 that a monopole doublet produces an  $m = 3$  term. See (15.8.33). And (15.8.34) evaluated with  $m = 3$  describes how this term falls off for large  $|z|$ . We conclude that the on-axis gradient for a monopole sextet falls off as  $1/|z|^7$  for large  $|z|$ . In analogy with the case of dipoles and quadrupoles, we expect that the on-axis gradients for a line of monopole sextets and an idealized air-core sextupole will also fall off as  $1/|z|^7$  for large  $|z|$ .

## 16.5 Stream-Function Windings for Pure Multipoles

The work of Bassetti and Biscari employed what we may call *rectangular* windings. Recall Figure 4.1. As we have described, their use has the very pleasing feature that the on-axis gradients and their derivatives they produce, the  $C_{m,s}^{[n]}(z)$ , are known in explicit analytic form. In this section we will describe the use of stream-function windings. They are more general than Bassetti-Biscari windings, but have the complication that the on-axis gradients and their derivatives they produce, the  $C_{m,s}^{[n]}(z)$ , are generally available only (but reliably) in numerical form.

Consider a circular cylinder of radius  $a$ . Suppose there is a specified (and time-independent) current  $\mathbf{J}$  flowing in a winding on the *surface* of the cylinder. Assume also that the electric charge density  $\rho_e$  is time independent. Then a static magnetic field  $\mathbf{B}$  will be produced in the interior of the cylinder and, since the current is confined to the surface, the interior of the cylinder will be source free. Correspondingly, as described in Section 15.3, the interior field can be derived from a *scalar potential* expressed in terms of *on-axis gradients*. The aim of this section is to obtain the on-axis gradients in terms of  $\mathbf{J}$ .

### 16.5.1 General Current Ansatz

Since it is assumed that both the current and charge density are time independent, it follows from charge conservation that the current  $\mathbf{J}$  must be divergence free,

$$\nabla \cdot \mathbf{J} = -\partial_t \rho_e = 0. \quad (16.5.1)$$

For our purposes we therefore need a way to specify a *divergence free* current  $\mathbf{J}$  on the surface of a cylinder of radius  $a$ .

Consider the vector field  $\mathbf{F}(\rho, \phi, z)$  defined by

$$\mathbf{F}(\rho, \phi, z) = \mathbf{e}_\rho \delta(\rho - a) V(\phi, z) \quad (16.5.2)$$

where  $V$  is a function to be determined. Make the Ansatz

$$\mathbf{J} = \nabla \times \mathbf{F} \quad (16.5.3)$$

to find the results

$$J_\rho = 0, \quad (16.5.4)$$

$$J_\phi = \delta(\rho - a)\partial_z V, \quad (16.5.5)$$

$$J_z = \delta(\rho - a)(-1/\rho)\partial_\phi V = \delta(\rho - a)(-1/a)\partial_\phi V. \quad (16.5.6)$$

Evidently  $\mathbf{J}$  has support only on the cylinder and is tangential to the cylinder. Let us compute  $\nabla \cdot \mathbf{J}$ . We find

$$\nabla \cdot \mathbf{J} = (1/\rho)\partial_\phi J_\phi + \partial_z J_z = [(1/\rho)\delta(\rho - a)][\partial_\phi \partial_z - \partial_z \partial_\phi]V = 0, \quad (16.5.7)$$

as desired and expected from the Ansatz (5.3) because the divergence of the curl of any sufficiently differentiable vector field is zero.

### 16.5.2 Canonical Stream Function

Let us write  $\mathbf{J}$  in the form

$$\mathbf{J} = \delta(\rho - a)\mathbf{j} \quad (16.5.8)$$

where

$$j_\rho = 0, \quad (16.5.9)$$

$$j_\phi = \partial_z V, \quad (16.5.10)$$

$$j_z = (-1/a)\partial_\phi V. \quad (16.5.11)$$

Next introduce quantities  $Q$ ,  $P$ , and  $H$  by the rules

$$Q = z/a \Leftrightarrow z = Qa, \quad (16.5.12)$$

$$P = \phi, \quad (16.5.13)$$

and

$$H(Q, P) = -(1/a)V(P, Qa) = -(1/a)V(\phi, z). \quad (16.5.14)$$

Note that  $Q$  and  $P$  are dimensionless. We seek parameterized curves  $Q(\lambda), P(\lambda)$  in the  $Q, P$  plane that satisfy the differential equations

$$\dot{Q} = \partial H / \partial P \quad (16.5.15)$$

and

$$\dot{P} = -\partial H / \partial Q \quad (16.5.16)$$

where a dot denotes  $(d/d\lambda)$ .

Let us see if these Ansätze are sane. From (5.15), (5.14), (5.13), and (5.11) we find

$$\dot{Q} = \partial H / \partial P = -(1/a)\partial_\phi V = j_z. \quad (16.5.17)$$

From (5.16), (5.14), (5.12), and (5.10) we find

$$\dot{P} = -\partial H / \partial Q = \partial_z V = j_\phi. \quad (16.5.18)$$

We see that the flow lines in the  $Q, P$  plane generated by  $H$  are along the current  $\mathbf{j}$ . Put another way, the tangent vector to the curve  $Q(\lambda), P(\lambda)$  is  $\mathbf{j}$ .

Moreover, since  $H$  does not depend on  $\lambda$ , it is conserved. It follows that the level lines of  $H$  coincide with the flow lines, and hence are also along  $\mathbf{j}$ . We will call  $H(Q, P)$  the *canonical stream function*.

### Fourier Expansion for $V$

Since the current  $\mathbf{J}$  lies on a cylindrical surface, the function  $V$  must be periodic in  $\phi$  with period  $2\pi$ . It therefore has a Fourier expansion of the form

$$V(\phi, z) = f_0(z) + \sum_{m=1}^{\infty} [f_{ms}(z) \sin(m\phi) + f_{mc}(z) \cos(m\phi)]. \quad (16.5.19)$$

We will treat each of the terms in this expansion *separately*. We expect (and will show) that the currents described by the terms  $f_{ms}(z) \sin(m\phi)$  and  $f_{mc}(z) \cos(m\phi)$ , with  $m \geq 1$ , will produce normal and skew multipole magnetic fields, respectively. Since skew fields are simply related to normal fields by suitable rotations about the  $z$  axis, and normal fields due to their more frequent use are easier to visualize, we will devote our further attention when  $m \geq 1$  to only the *normal* case. That is, in effect, we will assume the  $f_{mc} = 0$ . [See the text surrounding (15.3.105) and (15.3.106).] According to (5.4) through (5.6), the  $f_0$  term produces a solenoidal current. We will treat this case, which may be viewed as the  $m = 0$  *skew* case, after treating the normal  $m \geq 1$  cases.

### Computation of $\mathbf{j}$ for the Normal Multipole Case

Begin by writing

$$\mathbf{j} = j_\rho \mathbf{e}_\rho + j_\phi \mathbf{e}_\phi + j_z \mathbf{e}_z. \quad (16.5.20)$$

Then, using (5.9) through (5.11) and (5.19), we find for  $\mathbf{j}$  in terms of cylindrical unit vectors the result

$$\mathbf{j} = \sin(m\phi) f'_{ms}(z) \mathbf{e}_\phi - (m/a) \cos(m\phi) f_{ms}(z) \mathbf{e}_z. \quad (16.5.21)$$

For future use it is also useful to introduce fixed Cartesian unit vectors. Recall that

$$\mathbf{e}_\phi = -\sin(\phi) \mathbf{e}_x + \cos(\phi) \mathbf{e}_y. \quad (16.5.22)$$

Therefore we may also write

$$\mathbf{j}(\mathbf{r}) = j_x \mathbf{e}_x + j_y \mathbf{e}_y + j_z \mathbf{e}_z \quad (16.5.23)$$

with

$$j_x = -\sin(\phi) \sin(m\phi) f'_{ms}(z), \quad (16.5.24)$$

$$j_y = \cos(\phi) \sin(m\phi) f'_{ms}(z), \quad (16.5.25)$$

$$j_z = -(m/a) \cos(m\phi) f_{ms}(z). \quad (16.5.26)$$

[Here is a possible source of confusion: In the Accelerator Physics literature the angle we have called  $\phi$  is often referred to as the angle *theta*. Observe that, according to (5.26),  $j_z \sim \cos(m\phi)$ . If the cylinder bearing the winding is long compared to its radius and if most of the winding is parallel to the  $z$  axis, then most of the current will be in the  $z$  direction save for the current at the ends of the cylinder.<sup>13</sup> And this  $j_z$  will produce a  $\mathbf{B}$  that is mostly in the  $\mathbf{e}_x$  and  $\mathbf{e}_y$  directions. See, for example, Figure 3.1 part *c* and Figure 3.2. For that reason coils for which  $V$  has the angular dependence  $\sin(m\phi)$  are often referred to in the Accelerator Physics literature as *cosine theta* windings.]

### Computation of $\mathbf{j}$ for the $m = 0$ (Solenoidal) Case

For the solenoid case we will continue to employ (5.9) through (5.11) with, following (5.19),

$$V(\phi, z) = f_0(z). \quad (16.5.27)$$

In this case

$$j_\rho = 0, \quad (16.5.28)$$

$$j_\phi = \partial_z V = f'_0(z), \quad (16.5.29)$$

$$j_z = (-1/a) \partial_\phi V = 0. \quad (16.5.30)$$

Correspondingly,  $\mathbf{J}$  will have *only* a  $\phi$  component and that component will be *independent* of  $\phi$ . Therefore we again find

$$\nabla \cdot \mathbf{J} = (1/\rho) \partial_\phi J_\phi + \partial_z J_z = 0, \quad (16.5.31)$$

as desired.

### Specification of Current Filaments/Windings

Eventually we must specify how to place current carrying wires on the surface of the cylinder to achieve a current distribution that approximates  $\mathbf{j}$ . The  $m = 0$  and  $m \neq 0$  cases require separate treatment.

### The $m = 0$ (Solenoidal) Case

It might appear to be straight forward to treat the  $m = 0$  (solenoidal) case. We simply make the Ansatz

$$j_\phi = g(z). \quad (16.5.32)$$

---

<sup>13</sup>Note that, according to (5.21), for most of the current to be in the  $z$  direction  $f'_{ms}(z)$  must be small for most of the winding save at the cylinder ends.

### Definition of Shape Functions

We will call  $g(z)$  the  $m = 0$  *shape function*. Similarly the functions  $f_{m\alpha}(z)$ , with  $\alpha = s$  or  $\alpha = c$ , are called the normal and skew multipole shape functions.

Note that, in terms of fixed Cartesian unit vectors,  $\mathbf{j}$  as given by (5.28) through (5.30), (5.32), and (5.22) has the components

$$j_x = -g(z) \sin(\phi), \quad (16.5.33)$$

$$j_y = g(z) \cos(\phi), \quad (16.5.34)$$

$$j_z = 0. \quad (16.5.35)$$

Thus for the  $m = 0$  case  $\mathbf{J}$  is a *sheet current* encircling (flowing about) the cylinder in the azimuthal direction, and its magnitude may depend on  $z$ . Alternatively, the winding consists of a collection of few-turn circular coils wound around the cylinder at various locations  $z$  and having net circulating currents proportional to  $g(z)$ .

However, these specifications fail to describe how current is supplied to the winding or collection of windings by current leads, and lead currents will generally not be purely azimuthal. Also, single-layer solenoidal windings are actually helical, so that their currents also have nonvanishing  $j_z$ . For a discussion of helical effects, see the book by W. Smythe cited in the references at the end of this chapter. Perhaps, to minimize helical effects, solenoids should be wound with two layers with  $j_z$  in the second layer being the negative of that in the first layer. Also, in that case both incoming and outgoing current leads would be at the same end of the solenoid, and by placing the current leads near each other the magnetic fields arising from the currents in the leads could be arranged to be self canceling in lowest approximation. We conclude that ideal solenoidal windings are not possible, and one can only hope to minimize non ideal effects in actual practice.

### Specification of Current Filaments/Windings for the $m \geq 1$ Cases

Even beyond the question of how to supply current by suitable current leads, the specification of windings for the general  $m \geq 1$  cases is more complicated. From (5.21) we see that, when  $m \geq 1$ ,  $\mathbf{j}$  has *both*  $\mathbf{e}_\phi$  and  $\mathbf{e}_z$  components. Therefore the specification of windings is more complicated than in the  $m = 0$  (solenoidal) case. One way to do so is to find level lines of  $H$  and then place suitably powered windings on these level lines. Since the exposition of this approach is fairly extensive, it has been placed in Appendix G.

#### 16.5.3 Plan of Attack

As stated earlier, the aim of this section is to obtain the on-axis gradients in terms of  $\mathbf{J}$ . More precisely, since  $\mathbf{J}$  is specified by the shape functions  $g(z)$  and  $f_{m\alpha}$  [see (5.33) through (5.35) and (5.24) through (5.26)], we wish to obtain the on-axis gradients in terms of the shape functions. We will show that one way to find on-axis gradients in terms of shape functions is as follows:

1. Express magnetic scalar potentials  $\psi$  in terms of on-axis gradients. For examples, for the  $m = 0$  case there is the result

$$\begin{aligned}\psi(x, y, z) &= \psi_0(x, y, z) = -\sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! \ell!} C_0^{[2\ell]}(z) \rho^{2\ell} \\ &= -C_0^{[0]}(z) + (1/4)(x^2 + y^2) C_0^{[2]}(z) + \dots \\ &= -C_0^{[0]}(z) + O[\rho^2].\end{aligned}\quad (16.5.36)$$

See the first line of (15.3.33). And for the  $m \geq 1$  normal cases there are the results

$$\begin{aligned}\psi(\rho, \phi, z) &= \psi_{m,s}(\rho, \phi, z) = -\sin(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,s}^{[2\ell]}(z) \rho^{2\ell+m} \\ &= -\rho^m \sin(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,s}^{[2\ell]}(z) \rho^{2\ell} \\ &= -\rho^m \sin(m\phi) C_{m,s}^{[0]}(z) + \rho^m \sin(m\phi) \rho^2 \{(1/4)[m!/(1+m)!]\} C_{m,s}^{[2]}(z) + \dots \\ &= -\rho^m \sin(m\phi) C_{m,s}^{[0]}(z) + O[\rho^{m+2}].\end{aligned}\quad (16.5.37)$$

See the last line of (15.3.33).

2. Express magnetic fields  $\mathbf{B}$  in terms of magnetic scalar potentials  $\psi$ , and hence in terms of on-axis gradients. For examples, for the  $m = 0$  case there is the result

$$\begin{aligned}B_z &= -\partial_z \psi_0 = \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! \ell!} C_0^{[2\ell+1]}(z) \rho^{2\ell} \\ &= C_0^{[1]}(z) - (1/4)(x^2 + y^2) C_0^{[3]}(z) + \dots \\ &= C_0^{[1]}(z) + O[\rho^2].\end{aligned}\quad (16.5.38)$$

And for the  $m \geq 1$  normal cases we will see that there are the results

$$B_y = -\partial_y \psi = m \rho^{m-1} \cos[(m-1)\phi] C_{m,s}^{[0]}(z) + O[\rho^{m+1}]. \quad (16.5.39)$$

3. Express magnetic fields  $\mathbf{B}$ , initially in terms of  $\mathbf{J}$  and ultimately in terms of the shape functions, using the law of Biot and Savart. Given a *static* and divergence free  $\mathbf{J}$ , Biot and Savart state that it produces a magnetic field  $\mathbf{B}$  specified by the relation

$$\mathbf{B}(\mathbf{r}) = [\mu_0/(4\pi)] \int_V d^3 \mathbf{r}' [\mathbf{J}(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}')] / ||\mathbf{r} - \mathbf{r}'||^3. \quad (16.5.40)$$

4. Equate the two results for  $\mathbf{B}$  and then perform additional manipulations to obtain on-axis gradients, initially in terms of  $\mathbf{J}$  and ultimately in terms of the shape functions  $g(z)$  or the  $f_{m\alpha}(z)$ .

Here is another variation on the plan of attack: First compute  $\mathbf{J}$  using (5.8), (5.33) through (5.35), and (5.24) through (5.26). Employ this  $\mathbf{J}$  in (5.40) to compute  $\mathbf{B}$ . From this  $\mathbf{B}$  determine  $\psi$ . Verify that this  $\psi$  is of the forms (5.36) or (5.37), and read off  $C_0^{[1]}(z)$  or  $C_{m,s}^{[0]}(z)$ .

### 16.5.4 Execution of Plan for the Solenoidal ( $m = 0$ ) Case

#### What May We Expect?

As a warmup exercise, we will first treat the easiest case,  $m = 0$ . A solenoid is a straight beam-line element whose magnetic field is related to a cylindrical harmonic expansion that contains (ideally) only an  $m = 0$  term. That is, its magnetic scalar potential  $\psi$  has the expansion (5.36). Correspondingly the associated magnetic field has the expansion given by (5.38) and the complementary relations

$$B_x = -\partial_x \psi_0 = -(1/2)x C_0^{[2]}(z) + \dots, \quad (16.5.41)$$

$$B_y = -\partial_y \psi_0 = -(1/2)y C_0^{[2]}(z) + \dots. \quad (16.5.42)$$

From (5.38), (5.41), and (5.42) we see that there are the *on-axis* results

$$B_x(0, 0, z) = 0, \quad (16.5.43)$$

$$B_y(0, 0, z) = 0, \quad (16.5.44)$$

$$B_z(0, 0, z) = C_0^{[1]}(z). \quad (16.5.45)$$

#### Ingredients for and Use of Biot-Savart Law

Following the plan of attack, the next step is to employ the Biot-Savart law (5.40). Below are the ingredients for the Biot-Savart integrand when cylindrical coordinates are employed: Begin with

$$\begin{aligned} \mathbf{r} - \mathbf{r}' &= (x - x')\mathbf{e}_x + (y - y')\mathbf{e}_y + (z - z')\mathbf{e}_z \\ &= [\rho \cos(\phi) - \rho' \cos(\phi')]\mathbf{e}_x + [\rho \sin(\phi) - \rho' \sin(\phi')]\mathbf{e}_y + (z - z')\mathbf{e}_z \\ &= [\rho \cos(\phi) - a \cos(\phi')]\mathbf{e}_x + [\rho \sin(\phi) - a \sin(\phi')]\mathbf{e}_y + (z - z')\mathbf{e}_z. \end{aligned} \quad (16.5.46)$$

Here, in anticipation of the effect of the delta function appearing in the definition of  $\mathbf{J}$ , we have set  $\rho' = a$ . To facilitate future calculations, define a displacement vector  $\Delta$  by writing

$$\Delta = \mathbf{r} - \mathbf{r}' = \Delta_x \mathbf{e}_x + \Delta_y \mathbf{e}_y + \Delta_z \mathbf{e}_z \quad (16.5.47)$$

so that

$$\Delta_x = [\rho \cos(\phi) - a \cos(\phi')], \quad (16.5.48)$$

$$\Delta_y = [\rho \sin(\phi) - a \sin(\phi')], \quad (16.5.49)$$

$$\Delta_z = (z - z'). \quad (16.5.50)$$

Define a *source* vector  ${}^0\mathbf{S}$  by writing

$${}^0\mathbf{S} = \mathbf{j}(\mathbf{r}') \times \Delta \quad (16.5.51)$$

with  $\mathbf{j}$  given by (5.33) through (5.35). So doing gives the results

$$\begin{aligned} {}^0S_x &= j_y\Delta_z - j_z\Delta_y \\ &= \cos(\phi')g(z')(z - z'), \end{aligned} \quad (16.5.52)$$

$$\begin{aligned} {}^0S_y &= j_z\Delta_x - j_x\Delta_z \\ &= \sin(\phi')g(z')(z - z'), \end{aligned} \quad (16.5.53)$$

$$\begin{aligned} {}^0S_z &= j_x\Delta_y - j_y\Delta_x \\ &= -\sin(\phi')g(z')[\rho\sin(\phi) - a\sin(\phi')] \\ &\quad - \cos(\phi')g(z')[\rho\cos(\phi) - a\cos(\phi')]. \end{aligned} \quad (16.5.54)$$

Here we have placed a leading superscript 0 on  $\mathbf{S}$  to emphasize that it arises from an  $m = 0$   $\mathbf{j}$ , and we will do the same for the  $\mathbf{B}$  it produces. Also,

$$\begin{aligned} \|\mathbf{r} - \mathbf{r}'\|^2 &= (x - x')^2 + (y - y')^2 + (z - z')^2 \\ &= x^2 + y^2 - 2(xx' + yy') + (x')^2 + (y')^2 + (z - z')^2 \\ &= \rho^2 - 2a\rho(\cos\phi\cos\phi' + \sin\phi\sin\phi') + a^2 + (z - z')^2 \\ &= \rho^2 - 2a\rho\cos(\phi - \phi') + a^2 + (z - z')^2. \end{aligned} \quad (16.5.55)$$

Finally,

$$d^3\mathbf{r}' = \rho'd\rho'd\phi'dz'. \quad (16.5.56)$$

### Net Result

In terms of these ingredients, and employing cylindrical coordinates, use of Biot-Savart gives the relation

$${}^0\mathbf{B} = [\mu_0/(4\pi)]a \int_{-\infty}^{+\infty} dz' \int_{-\pi}^{+\pi} d\phi' {}^0\mathbf{S}/[\rho^2 - 2a\rho\cos(\phi - \phi') + a^2 + (z - z')^2]^{3/2}. \quad (16.5.57)$$

Let us evaluate this relation for  $\rho = 0$ . Observe that

$$[\rho^2 - 2a\rho\cos(\phi - \phi') + a^2 + (z - z')^2]|_{\rho=0} = a^2 + (z - z')^2. \quad (16.5.58)$$

Consequently, we have the on-axis result

$${}^0\mathbf{B}(0, 0, z) = [\mu_0/(4\pi)]a \int_{-\infty}^{+\infty} dz' [a^2 + (z - z')^2]^{-3/2} \int_{-\pi}^{+\pi} d\phi' {}^0\mathbf{S}^0. \quad (16.5.59)$$

Here  ${}^0\mathbf{S}^0$  is the on-axis value of  ${}^0\mathbf{S}$ ,

$${}^0\mathbf{S}^0 = {}^0\mathbf{S}|_{\rho=0}. \quad (16.5.60)$$

From (5.52) through (5.54) we see that

$${}^0S_x^0 = \cos(\phi')g(z')(z - z'), \quad (16.5.61)$$

$${}^0 S_y^0 = \sin(\phi') g(z')(z - z'), \quad (16.5.62)$$

$${}^0 S_z^0 = ag(z'). \quad (16.5.63)$$

From (5.61) through (5.63) we see that in (5.59) there are the angular integral results

$$\int_{-\pi}^{+\pi} d\phi' {}^0 S_x^0 = \int_{-\pi}^{+\pi} d\phi' {}^0 S_y^0 = 0, \quad (16.5.64)$$

$$\int_{-\pi}^{+\pi} d\phi' {}^0 S_z^0 = 2\pi ag(z'). \quad (16.5.65)$$

It follows that

$$B_x(0, 0, z) = B_y(0, 0, z) = 0 \quad (16.5.66)$$

in accord with our expectations (5.43) and (5.44). And for  $B_z(0, 0, z)$  we find using (5.59) and (5.65) the result

$${}^0 B_z(0, 0, z) = \mu_0(a^2/2) \int_{-\infty}^{+\infty} dz' g(z')/[a^2 + (z - z')^2]^{3/2}. \quad (16.5.67)$$

Upon comparing (5.45) and (5.67) we see that

$$C_0^{[1]}(z) = \mu_0(a^2/2) \int_{-\infty}^{+\infty} dz' g(z')/[a^2 + (z - z')^2]^{3/2}. \quad (16.5.68)$$

We have found the master function  $C_0^{[1]}(z)$  in terms of the  $m = 0$  shape function  $g(z')$ , as desired.

### Definition of Kernels and Computation of Derivatives of On-Axis Gradients

We would also like to have formulas for the  $z$  derivatives of  $C_0^{[1]}(z)$ . Introduce functions  $\Lambda(1, 0; z, z')$  by the rules

$$\Lambda(1, 0; z, z') = [a^2 + (z - z')^2]^{-3/2} \quad (16.5.69)$$

and, for  $k \geq 1$ ,

$$\Lambda(k, 0; z, z') = \partial_z^{k-1} [a^2 + (z - z')^2]^{-3/2}. \quad (16.5.70)$$

With the definition (5.69) we see that (5.68) can be written in the form

$$C_0^{[1]}(z) = \mu_0(a^2/2) \int_{-\infty}^{+\infty} dz' \Lambda(1, 0; z, z') g(z'). \quad (16.5.71)$$

And, by differentiation under the integral sign, we see that the  $C_0^{[k]}(z)$  can be written in the form

$$C_0^{[k]}(z) = \mu_0(a^2/2) \int_{-\infty}^{+\infty} dz' \Lambda(k, 0; z, z') g(z'). \quad (16.5.72)$$

The functions  $\Lambda(k, 0; z, z')$  act as *integration kernels* that produce the  $C_0^{[k]}(z)$  in terms of the  $m = 0$  shape function  $g(z')$ . And, looking at the functional form of the  $\Lambda(k, m; z, z')$  as given by (5.70), it is clear that they can be found analytically using, if necessary, *Mathematica*. Thus the  $\Lambda(k, 0; z, z')$  are *known* functions. For example,

$$\Lambda(2, 0; z, z') = \partial_z [a^2 + (z - z')^2]^{-3/2} = -3(z - z')[a^2 + (z - z')^2]^{-5/2}. \quad (16.5.73)$$

### Sanity Check

As a check on our  $m = 0$  work, let us consider a simple case for the  $m = 0$  shape function. Suppose that  $g(z')$  has support only in the interval  $z' \in [0, L]$  and that in this interval it has the constant value  $g_0$ . That is, we make for the  $m = 0$  shape function the Ansatz

$$g(z') = g_0 \text{bump}(z', L). \quad (16.5.74)$$

For this choice of the  $m = 0$  shape function there is the result

$$\begin{aligned} C_0^{[1]}(z) &= {}^0B_z(0, 0, z) = \mu_0(a^2/2)g_0 \int_0^L dz' [a^2 + (z - z')^2]^{-3/2} \\ &= * \mu_0(a^2/2)g_0 \int_0^L dz' \delta_1(z, a) = B_0 \text{ bump}_1(z, a, L). \end{aligned} \quad (16.5.75)$$

Here  $B_0$  is the on-axis field at  $z = L/2$  in the infinite length limit, and is given by

$$B_0 = *g_0. \quad (16.5.76)$$

See \*. This result agrees with our earlier result \*. A bump-function Ansatz for the  $m = 0$  shape function  $g(z')$  produces an-axis gradient that is proportional to the approximating bump function  $\text{bump}_1$ .

### 16.5.5 Preliminary Work for the Normal $m \geq 1$ Cases

We now move on to the more difficult  $m \geq 1$  cases. We will now use (5.39) (still to be proved) and will again want to use the Biot-Savart law (5.40). For insertion into the Biot-Savart law we will now define a source vector  ${}^m\mathbf{S}$  by writing

$${}^m\mathbf{S} = \mathbf{j}(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}') = \mathbf{j}(\mathbf{r}') \times \Delta \quad (16.5.77)$$

where  $\mathbf{j}$  is given by (5.23) through (5.26). So doing, and again using (5.46) through (5.50) and (5.78), gives the result

$$\begin{aligned} {}^m S_x &= j_y \Delta_z - j_z \Delta_y = \\ &\{ \cos(\phi') \sin(m\phi') f'_{ms}(z')(z - z') \\ &+ (m/a) \cos(m\phi') f_{ms}(z')[\rho \sin(\phi) - a \sin(\phi')]\}, \end{aligned} \quad (16.5.78)$$

$$\begin{aligned} {}^m S_y &= j_z \Delta_x - j_x \Delta_z = \\ &\{ -(m/a) \cos(m\phi') f_{ms}(z')[\rho \cos(\phi) - a \cos(\phi')] \\ &+ \sin(\phi') \sin(m\phi') f'_{ms}(z')(z - z')\}, \end{aligned} \quad (16.5.79)$$

$$\begin{aligned} {}^m S_z &= j_x \Delta_y - j_y \Delta_x = \\ &\{ -\sin(\phi') \sin(m\phi') f'_{ms}(z')[\rho \sin(\phi) - a \sin(\phi')] \\ &- \cos(\phi') \sin(m\phi') f'_{ms}(z')[\rho \cos(\phi) - a \cos(\phi')]\}. \end{aligned} \quad (16.5.80)$$

Also, recall (5.55) and (5.56). Putting everything together gives for  $m \geq 1$  the net result

$${}^m\mathbf{B} = [\mu_0/(4\pi)]a \int_{-\infty}^{+\infty} dz' \int_{-\pi}^{+\pi} d\phi' {}^m\mathbf{S} / [\rho^2 - 2a\rho \cos(\phi - \phi') + a^2 + (z - z')^2]^{3/2}. \quad (16.5.81)$$

### 16.5.6 Normal Dipole ( $m = 1$ ) Case

Suppose  $m = 1$ , which is the simplest  $m \geq 1$  case. Then, if (5.37) holds, we have the relation

$$\psi = \psi_1 = -\rho \sin(\phi) C_{1,s}^{[0]}(z) + O[\rho^3] = -y C_{1,s}^{[0]}(z) + O[\rho^3]. \quad (16.5.82)$$

It follows from

$${}^1\mathbf{B} = -\nabla \psi_1 \quad (16.5.83)$$

that in this case we expect the *on-axis* results

$${}^1B_x(0, 0, z) = 0, \quad (16.5.84)$$

$${}^1B_y(0, 0, z) = C_{1,s}^{[0]}(z), \quad (16.5.85)$$

$${}^1B_z(0, 0, z) = 0. \quad (16.5.86)$$

We note that that (5.86) is the  $m = 1$  and  $\rho = 0$  case of (5.39). Let us see if the Biot-Savart  ${}^1\mathbf{B}$  given by (5.82) with  $m = 1$  satisfies (5.85) through (5.87). With reference to (5.55), there is the on-axis result (5.58). Therefore the term  $[\rho^2 - 2a\rho \cos(\phi - \phi') + a^2 + (z - z')^2]$  in (5.82) has no angular dependence when  $\rho = 0$ . Also, let  ${}^1\mathbf{S}^0$  denote  ${}^1\mathbf{S}$  evaluated at  $\rho = 0$  so that (when  $m = 1$ )

$$\begin{aligned} {}^1S_x^0 &= \\ &\{\cos(\phi') \sin(\phi') f'_{1s}(z')(z - z') \\ &+(1/a) \cos(\phi') f_{1s}(z')[-a \sin(\phi')]\}, \end{aligned} \quad (16.5.87)$$

$$\begin{aligned} {}^1S_y^0 &= \\ &\{-(1/a) \cos(\phi') f_{1s}(z')[-a \cos(\phi')] \\ &+\sin(\phi') \sin(\phi') f'_{1s}(z')(z - z')\}, \end{aligned} \quad (16.5.88)$$

$$\begin{aligned} {}^1S_z^0 &= \\ &\{-\sin(\phi') \sin(\phi') f'_{1s}(z')[-a \sin(\phi')] \\ &-\cos(\phi') \sin(\phi') f'_{1s}(z')[-a \cos(\phi')]\}. \end{aligned} \quad (16.5.89)$$

Taking into account (5.58) and (5.87) through (5.89) we can now write (5.81) evaluated at  $(0, 0, z)$  in the component forms

$${}^1B_x(0, 0, z) = [\mu_0/(4\pi)]a \int_{-\infty}^{+\infty} dz' \{1/[a^2 + (z - z')^2]^{3/2}\} \int_{-\pi}^{+\pi} d\phi' {}^1S_x^0, \quad (16.5.90)$$

$${}^1B_y(0, 0, z) = [\mu_0/(4\pi)]a \int_{-\infty}^{+\infty} dz' \{1/[a^2 + (z - z')^2]^{3/2}\} \int_{-\pi}^{+\pi} d\phi' {}^1S_y^0, \quad (16.5.91)$$

$${}^1B_z(0, 0, z) = [\mu_0/(4\pi)]a \int_{-\infty}^{+\infty} dz' \{1/[a^2 + (z - z')^2]^{3/2}\} \int_{-\pi}^{+\pi} d\phi' {}^1S_z^0. \quad (16.5.92)$$

Examination of (5.87) and (5.89) shows that both  ${}^1S_x^0$  and  ${}^1S_z^0$  are *odd* functions of  $\phi'$ . Therefore the angular integrals in (5.90) and (5.92) vanish. Consequently, as hoped, the expectations (5.84) and (5.86) are met.

What remains is the case of  ${}^1B_y$ , for which

$${}^1S_y^0 = [\cos^2(\phi')f_{1s}(z') + \sin^2(\phi')f'_{1s}(z')(z - z')], \quad (16.5.93)$$

so that

$$\begin{aligned} \int_{-\pi}^{+\pi} d\phi' {}^1S_y^0 &= \int_{-\pi}^{+\pi} d\phi' [\cos^2(\phi')f_{1s}(z') + \sin^2(\phi')f'_{1s}(z')(z - z')] \\ &= \pi[f_{1s}(z') + f'_{1s}(z')(z - z')]. \end{aligned} \quad (16.5.94)$$

It follows that

$${}^1B_y(0, 0, z) = (\mu_0/4)a \int_{-\infty}^{+\infty} dz' [f_{1s}(z') + f'_{1s}(z')(z - z')]/[a^2 + (z - z')^2]^{3/2}. \quad (16.5.95)$$

### Manipulation and Integration by Parts

This result can be simplified by manipulation and integration by parts so that only  $f_{1s}$  and *not*  $f'_{1s}$  appears. Observe that the integral in (5.95) can be written as the sum of two terms,

$$\begin{aligned} &\int_{-\infty}^{+\infty} dz' [f_{1s}(z') + f'_{1s}(z')(z - z')]/[a^2 + (z - z')^2]^{3/2} \\ &= \int_{-\infty}^{+\infty} dz' f_{1s}(z')/[a^2 + (z - z')^2]^{3/2} \\ &+ \int_{-\infty}^{+\infty} dz' f'_{1s}(z')(z - z')/[a^2 + (z - z')^2]^{3/2}. \end{aligned} \quad (16.5.96)$$

The first term can be manipulated to become

$$\begin{aligned} \text{first term} &= \int_{-\infty}^{+\infty} dz' f_{1s}(z')/[a^2 + (z - z')^2]^{3/2} \\ &= \int_{-\infty}^{+\infty} dz' f_{1s}(z')[a^2 + (z - z')^2]/[a^2 + (z - z')^2]^{5/2}. \end{aligned} \quad (16.5.97)$$

The second term,

$$\text{second term} = \int_{-\infty}^{+\infty} dz' f'_{1s}(z')(z - z')/[a^2 + (z - z')^2]^{3/2}, \quad (16.5.98)$$

can be integrated by parts using the template

$$\int u dv = uv - \int v du \quad (16.5.99)$$

with

$$u = (z - z')/[a^2 + (z - z')^2]^{3/2} \quad (16.5.100)$$

and

$$v = f_{1s}(z'). \quad (16.5.101)$$

Begin the process by writing

$$\begin{aligned} \text{second term} &= \int_{-\infty}^{+\infty} dz' f'_{1s}(z')(z - z')/[a^2 + (z - z')^2]^{3/2} \\ &= \{f_{1s}(z')(z - z')/[a^2 + (z - z')^2]^{3/2}\}|_{z'=-\infty}^{z'=+\infty} \\ &\quad - \int_{-\infty}^{+\infty} dz' f_{1s}(z') \partial_{z'} \{(z - z')/[a^2 + (z - z')^2]^{3/2}\} \\ &= - \int_{-\infty}^{+\infty} dz' f_{1s}(z') \partial_{z'} \{(z - z')/[a^2 + (z - z')^2]^{3/2}\}. \end{aligned} \quad (16.5.102)$$

Here we have assumed that

$$\{f_{1s}(z')(z - z')/[a^2 + (z - z')^2]^{3/2}\}|_{z'=-\infty}^{z'=+\infty} = 0, \quad (16.5.103)$$

which will be the case if  $f_{1s}(z')$  is bounded at  $z' = \pm\infty$ . Moreover, we also have

$$\begin{aligned} \partial_{z'} \{(z - z')/[a^2 + (z - z')^2]^{3/2}\} &= \\ -1/[a^2 + (z - z')^2]^{3/2} + 3(z - z')^2/[a^2 + (z - z')^2]^{5/2}. \end{aligned} \quad (16.5.104)$$

It follows, upon combining (5.102) and (5.104) and further manipulation, that

$$\begin{aligned} \text{second term} &= \int_{-\infty}^{+\infty} dz' f'_{1s}(z')(z - z')/[a^2 + (z - z')^2]^{3/2} \\ &= - \int_{-\infty}^{+\infty} dz' f_{1s}(z') \partial_{z'} \{(z - z')/[a^2 + (z - z')^2]^{3/2}\} \\ &= \int_{-\infty}^{+\infty} dz' f_{1s}(z') \{1/[a^2 + (z - z')^2]^{3/2} - 3(z - z')^2/[a^2 + (z - z')^2]^{5/2}\}. \end{aligned} \quad (16.5.105)$$

Evidently the second term has two parts. The first part can be written and manipulated to give the result

$$\begin{aligned} &\text{first part of second term} \\ &= \int_{-\infty}^{+\infty} dz' f_{1s}(z')/[a^2 + (z - z')^2]^{3/2} \\ &= \int_{-\infty}^{+\infty} dz' f_{1s}(z') [a^2 + (z - z')^2]/[a^2 + (z - z')^2]^{5/2}. \end{aligned} \quad (16.5.106)$$

The second part is given by the relation

$$\begin{aligned} &\text{second part of second term} \\ &= \int_{-\infty}^{+\infty} dz' f_{1s}(z') [-3(z - z')^2]/[a^2 + (z - z')^2]^{5/2}. \end{aligned} \quad (16.5.107)$$

We can now add these two parts together, thereby regaining the full second term, to find

$$\begin{aligned}
 & \text{sum of parts} = \text{full second term} \\
 &= \int_{-\infty}^{+\infty} dz' f_{1s}(z') [a^2 + (z - z')^2] / [a^2 + (z - z')^2]^{5/2} \\
 &+ \int_{-\infty}^{+\infty} dz' f_{1s}(z') [-3(z - z')^2] / [a^2 + (z - z')^2]^{5/2} \\
 &= \int_{-\infty}^{+\infty} dz' f_{1s}(z') [a^2 - 2(z - z')^2] / [a^2 + (z - z')^2]^{5/2}. \tag{16.5.108}
 \end{aligned}$$

Finally, add together the results for the first and second terms to find the net result

$$\begin{aligned}
 & \int_{-\infty}^{+\infty} dz' [f_{1s}(z') + f'_{1s}(z')(z - z')] / [a^2 + (z - z')^2]^{3/2} \\
 &= \text{first term} + \text{second term} \\
 &= \int_{-\infty}^{+\infty} dz' f_{1s}(z') [a^2 + (z - z')^2] / [a^2 + (z - z')^2]^{5/2} \\
 &+ \int_{-\infty}^{+\infty} dz' f_{1s}(z') [a^2 - 2(z - z')^2] / [a^2 + (z - z')^2]^{5/2} \\
 &= \int_{-\infty}^{+\infty} dz' f_{1s}(z') [2a^2 - (z - z')^2] / [a^2 + (z - z')^2]^{5/2}. \tag{16.5.109}
 \end{aligned}$$

The work of manipulation and integration by parts is complete.

### Final $m = 1$ Results

We are now ready to find  ${}^1B_y(0, 0, z)$ . Combining (5.95) and (5.109) gives the result

$${}^1B_y(0, 0, z) = (\mu_0/4)a \int_{-\infty}^{+\infty} dz' f_{1s}(z') \{[2a^2 - (z - z')^2] / [a^2 + (z - z')^2]^{5/2}\}, \tag{16.5.110}$$

which can be rewritten as

$${}^1B_y(0, 0, z) = -(\mu_0/4)a \int_{-\infty}^{+\infty} dz' f_{1s}(z') \{[-2a^2 + (z - z')^2] / [a^2 + (z - z')^2]^{5/2}\}. \tag{16.5.111}$$

It follows from (5.85) that there is the ultimate result

$$C_{1,s}^{[0]}(z) = -\mu_0(a/4) \int_{-\infty}^{+\infty} dz' f_{1s}(z') \{[-2a^2 + (z - z')^2] / [a^2 + (z - z')^2]^{5/2}\}. \tag{16.5.112}$$

We have found, for the  $m = 1$  (and normal) case, the on-axis gradient  $C_{1,s}^{[0]}$  in terms of the  $m = 1$  shape function  $f_{1s}$ .

### Application to a Simple Example $f_{1s}$ Shape Function

Suppose, as a simple example, that  $f_{1s}(z')$  has support only in the interval  $z' \in [0, L]$  and that in this interval it has the constant value  $f_{1s}^0$ . That is, we make the bump-function Ansatz

$$f_{1s}(z') = f_{1s}^0 \text{bump}(z', L). \quad (16.5.113)$$

In this case we find, using (5.112) and (5.113), that for this choice of the  $m = 1$  (and normal) shape function there is the result

$$C_{1,s}^{[0]}(z) = -f_{1s}^0 \mu_0(a/4) \int_0^L dz' \{[-2a^2 + (z - z')^2]/[a^2 + (z - z')^2]^{5/2}\}. \quad (16.5.114)$$

The integral in (5.114) can be evaluated exactly by making the change of variables  $\zeta = z - z'$  and the use of integral tables. See \*. Greater insight is gained by observing that the integrand in (5.114) can be manipulated to give the relation

$$\begin{aligned} & [-2a^2 + (z - z')^2]/[a^2 + (z - z')^2]^{5/2} \\ &= [a^2 + (z - z')^2 - 3a^2]/[a^2 + (z - z')^2]^{5/2} \\ &= *\delta_1(z, a) + *\delta_2(z, a). \end{aligned} \quad (16.5.115)$$

See \*. Therefore we may write

$$\begin{aligned} C_{1,s}^{[0]}(z) &= -f_{1s}^0 \mu_0(a/4) \int_0^L dz' \{[-2a^2 + (z - z')^2]/[a^2 + (z - z')^2]^{5/2}\} \\ &= -f_{1s}^0 \mu_0(a/4) \int_0^L dz' [*\delta_1(z, a) + *\delta_2(z, a)] \\ &= -f_{1s}^0 \mu_0(a/4) [*\text{bump}_1(z, a, L) + *\text{bump}_2(z, a, L)]. \end{aligned} \quad (16.5.116)$$

See \*. A bump-function Ansatz for the shape function  $f_{1s}(z')$  produces an-axis gradient that is proportional to a weighted sum of two approximating bump functions, namely  $\text{bump}_1$  and  $\text{bump}_2$ .

### 16.5.7 Continued Work on the General $m > 1$ Normal Cases

#### What May We Expect?

Encouraged by success in the dipole case  $m = 1$ , let us work on the case of general  $m$ . For a given  $m$  the scalar potential for a normal multipole is given in terms of on-axis gradients by the relation

$$\begin{aligned} \psi_m(\rho, \phi, z) &= -\sin(m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell} \ell! (\ell + m)!} C_{m,s}^{[2\ell]}(z) \rho^{2\ell+m} \\ &= -\rho^m \sin(m\phi) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{m!}{2^{2\ell} \ell! (\ell + m)!} C_{m,s}^{[2\ell]}(z) \rho^{2\ell} \\ &= -\rho^m \sin(m\phi) C_{m,s}^{[0]}(z) + \rho^m \sin(m\phi) \rho^2 \{(1/4)[m!/(1+m)!]\} C_{m,s}^{[2]}(z) + \dots \end{aligned} \quad (16.5.117)$$

Let us work out the polynomial content of the factors  $\rho^m \sin(m\phi)$  and  $\rho^m \sin(m\phi)\rho^2$  for the first few values of  $m$ . And, since  ${}^m\mathbf{B} = -\nabla\psi_m$  and we are dealing with the *normal* multipole case, let us also work out the polynomial content of  ${}^m B_y$  which will be  $\partial_y[\rho^m \sin(m\phi)]$  plus terms of degree  $(m + 1)$  and higher. To do so recall that

$$\rho^m \sin m\phi = \Im[(x + iy)^m] = [1/(2i)][(x + iy)^m - (x - iy)^m], \quad (16.5.118)$$

$$\rho^m \cos m\phi = \Re[(x + iy)^m] = (1/2)[(x + iy)^m + (x - iy)^m], \quad (16.5.119)$$

and

$$\rho^2 = x^2 + y^2. \quad (16.5.120)$$

See Section 15.3.

Using these relations gives the results listed below:

### The case $m = 1$ .

$$\rho \sin(\phi) = y, \quad (16.5.121)$$

$$\partial_y[\rho \sin(\phi)] = 1 = \Re[(x + iy)^0] = m\rho^{m-1} \cos[(m-1)\phi] \text{ with } m = 1; \quad (16.5.122)$$

### The case $m = 2$ .

$$\rho^2 \sin(2\phi) = \Im(x + iy)^2 = 2xy, \quad (16.5.123)$$

$$\partial_y[\rho^2 \sin(2\phi)] = 2x = 2\Re[(x + iy)] = m\rho^{m-1} \cos[(m-1)\phi] \text{ with } m = 2. \quad (16.5.124)$$

### The general case.

Indeed, we find from (5.116) and (5.117) that

$$\begin{aligned} \partial_y[\rho^m \sin(m\phi)] &= \partial_y\{\Im[(x + iy)^m]\} = \partial_y[1/(2i)][(x + iy)^m - (x - iy)^m] \\ &= [1/(2i)][im][(x + iy)^{m-1} + (x - iy)^{m-1}] = m\Re[(x + iy)^{m-1}] \\ &= m\rho^{m-1} \cos[(m-1)\phi]. \end{aligned} \quad (16.5.125)$$

Similarly, for future use, we find from (5.117) that

$$\begin{aligned} \partial_x[\rho^m \cos(m\phi)] &= \partial_x\{\Re[(x + iy)^m]\} = \partial_x[1/(2)][(x + iy)^m + (x - iy)^m] \\ &= [1/(2)][m][(x + iy)^{m-1} + (x - iy)^{m-1}] = m\Re[(x + iy)^{m-1}] \\ &= m\rho^{m-1} \cos[(m-1)\phi]. \end{aligned} \quad (16.5.126)$$

We are now ready to draw a conclusion. Since  ${}^m\mathbf{B} = -\nabla\psi_m$ , it follows that

$$\begin{aligned} {}^m B_y &= -\partial_y\psi^m \\ &= C_{m,s}^{[0]}(z)\partial_y[\rho^m \sin(m\phi)] \\ &\quad + \text{terms of degree } (m+1) \text{ and higher.} \\ &= C_{m,s}^{[0]}(z)m\rho^{m-1} \cos[(m-1)\phi] \\ &\quad + \text{terms of degree } (m+1) \text{ and higher,} \end{aligned} \quad (16.5.127)$$

in agreement with (5.40). Therefore, if we are able by some means to work out  ${}^m B_y$  through terms of degree  $(m - 1)$  in  $x$  and  $y$  and determine, as a sanity check, that the result is of the form

$$\begin{aligned} {}^m B_y &= \text{common factor} \times \partial_y [\rho^m \sin(m\phi)] + \text{terms of degree } (m+1) \text{ and higher} \\ &= \{\text{common factor} \times m\rho^{m-1} \cos[(m-1)\phi]\} + O[\rho^{m+1}], \end{aligned} \quad (16.5.128)$$

then that common factor will be  $C_{m,s}^{[0]}(z)$ . Moreover, according to (3.94), there is the result

$${}^m B_y = [\mu_0/(4\pi)]a \int_{-\infty}^{+\infty} dz' \int_{-\pi}^{+\pi} d\phi' {}^m S_y / [\rho^2 - 2a\rho \cos(\phi - \phi') + a^2 + (z - z')^2]^{3/2} \quad (16.5.129)$$

where, according to (3.90),

$$\begin{aligned} {}^m S_y &= \\ &\{-(m/a) \cos(m\phi') f(z') [\rho \cos(\phi) - a \cos(\phi')] \\ &+ \sin(\phi') \sin(m\phi') f'(z') (z - z')\} \\ &= [-(m/a) f(z') \rho] \cos(m\phi') \cos(\phi) \\ &+ [mf(z')] \cos(m\phi') \cos(\phi') \\ &+ [f'(z')(z - z')] \sin(\phi') \sin(m\phi'). \end{aligned} \quad (16.5.130)$$

### Computation of ${}^m B_y$

From now on we will devote our effort to the computation of  ${}^m B_y$ . It will turn out to be all that is needed. But the methods to be developed are also applicable to the computation of  ${}^m B_x$  and  ${}^m B_z$  if desired.

### Standard Form Integral Relation

As was done in Section 15.12, let us try to isolate the  $\phi$  dependence by a change of integration variable. Write

$$\phi' = (\phi' - \phi) + \phi = \theta + \phi = \phi + \theta \quad (16.5.131)$$

where

$$\theta = \phi' - \phi. \quad (16.5.132)$$

We will use (3.146) to make a change of integration variables so that

$$d\phi' = d\theta. \quad (16.5.133)$$

Begin by observing that so doing transforms  ${}^m S_y$  to become  ${}^m S_y^{tr}$ . Also, as will be seen,  ${}^m S_y^{tr}$  can be written as the sum

$${}^m S_y^{tr} = {}^m S_y^{tre} + {}^m S_y^{tro} \quad (16.5.134)$$

where  ${}^m S_y^{tre}$  and  ${}^m S_y^{tro}$  are *even* and *odd* functions of  $\theta$ , respectively. This decomposition will facilitate integration over  $\theta$ .

For the trigonometric functions there are the relations

$$\sin(A)\cos(B) = (1/2)\sin(A+B) + (1/2)\sin(A-B), \quad (16.5.135)$$

$$\cos(A)\sin(B) = (1/2)\sin(A+B) - (1/2)\sin(A-B), \quad (16.5.136)$$

$$\cos(A)\cos(B) = (1/2)\cos(A+B) + (1/2)\cos(A-B), \quad (16.5.137)$$

$$\sin(A)\sin(B) = (-1/2)\cos(A+B) + (1/2)\cos(A-B), \quad (16.5.138)$$

$$\sin(n\phi') = \sin(n\phi + n\theta) = \sin(n\phi)\cos(n\theta) + \cos(n\phi)\sin(n\theta), \quad (16.5.139)$$

$$\cos(n\phi') = \cos(n\phi + n\theta) = \cos(n\phi)\cos(n\theta) - \sin(n\phi)\sin(n\theta). \quad (16.5.140)$$

Employ these relations to work out  ${}^mS_y^{tr}$ .

### Computation of ${}^mS_y^{tr}$

For  ${}^mS_y^{tr}$  we see from (3.145) that there are three factors, and use of (3.146) and application of (3.150) through (3.155) to these factors gives the results

#### Coefficient of $[-(m/a)f(z')\rho]$

$$\begin{aligned} \cos(m\phi')\cos(\phi) &= \cos(\phi)\cos(m\phi') = \cos(\phi)\cos(m\phi + m\theta) \\ &= \cos(\phi)[\cos(m\phi)\cos(m\theta) - \sin(m\phi)\sin(m\theta)] \\ &= [\cos(\phi)\cos(m\phi)]\cos(m\theta) - [\cos(\phi)\sin(m\phi)]\sin(m\theta) \\ &= \{(1/2)\cos[(m+1)\phi] + (1/2)\cos[(m-1)\phi]\}\cos(m\theta) \\ &\quad - \{(1/2)\sin[(m+1)\phi] + (1/2)\sin[(m-1)\phi]\}\sin(m\theta), \end{aligned} \quad (16.5.141)$$

#### Coefficient of $[mf(z')]$

$$\begin{aligned} \cos(m\phi')\cos(\phi') &= (1/2)\cos[(m+1)\phi'] + (1/2)\cos[(m-1)\phi'] \\ &= (1/2)\cos[(m+1)(\phi+\theta)] + (1/2)\cos[(m-1)(\phi+\theta)] \\ &= (1/2)\{\cos[(m+1)\phi]\cos[(m+1)\theta] - \sin[(m+1)\phi]\sin[(m+1)\theta]\} \\ &\quad + (1/2)\{\cos[(m-1)\phi]\cos[(m-1)\theta] - \sin[(m-1)\phi]\sin[(m-1)\theta]\}, \end{aligned} \quad (16.5.142)$$

#### Coefficient of $[f'(z')(z-z')]$

$$\begin{aligned} \sin(\phi')\sin(m\phi') &= -(1/2)\cos[(m+1)\phi'] + (1/2)\cos[(m-1)\phi'] \\ &= -(1/2)\cos[(m+1)(\phi+\theta)] + (1/2)\cos[(m-1)(\phi+\theta)] \\ &= -(1/2)\{\cos[(m+1)\phi]\cos[(m+1)\theta] - \sin[(m+1)\phi]\sin[(m+1)\theta]\} \\ &\quad + (1/2)\{\cos[(m-1)\phi]\cos[(m-1)\theta] - \sin[(m-1)\phi]\sin[(m-1)\theta]\}. \end{aligned} \quad (16.5.143)$$

**Computation of  ${}^m S_y^{tre}$** 

In accord with our earlier statement about the importance of only even in  $\theta$  terms, we extract them from (3.156) through (3.158) to retain the results

**Even Coefficient of  $[-(m/a)f(z')\rho]$** 

$$\{(1/2)\cos[(m+1)\phi] + (1/2)\cos[(m-1)\phi]\}\cos(m\theta), \quad (16.5.144)$$

**Even Coefficient of  $[mf(z')]$** 

$$\begin{aligned} & (1/2)\{\cos[(m+1)\phi]\cos[(m+1)\theta]\} \\ & + (1/2)\{\cos[(m-1)\phi]\cos[(m-1)\theta]\}, \end{aligned} \quad (16.5.145)$$

**Even Coefficient of  $[f'(z')(z-z')]$** 

$$\begin{aligned} & = -(1/2)\{\cos[(m+1)\phi]\cos[(m+1)\theta]\} \\ & + (1/2)\{\cos[(m-1)\phi]\cos[(m-1)\theta]\}. \end{aligned} \quad (16.5.146)$$

Upon combining the individual results (3.159) through (3.161) we find the grand result

$$\begin{aligned} {}^m S_y^{tre} = & (1/2)\{[mf(z')] + [f'(z')(z-z')]\}\cos[(m-1)\phi]\cos[(m-1)\theta] \\ & - [(m/a)f(z')\rho]\{(1/2)\cos[(m+1)\phi] + (1/2)\cos[(m-1)\phi]\}\cos(m\theta) \\ & - (1/2)\{[-mf(z')] - [-f'(z')(z-z')]\}\cos[(m+1)\phi]\cos[(m+1)\theta]. \\ = & (1/2)\{mf(z') + f'(z')(z-z')\}\cos[(m-1)\phi]\cos[(m-1)\theta] \\ & - (m/a)f(z')\rho\{(1/2)\cos[(m+1)\phi] + (1/2)\cos[(m-1)\phi]\}\cos(m\theta) \\ & + (1/2)\{mf(z') - f'(z')(z-z')\}\cos[(m+1)\phi]\cos[(m+1)\theta]. \end{aligned} \quad (16.5.147)$$

**Evaluation of Integrals**

Upon making the change of variables specified by (3.146) through (3.148), the integral (3.144) can be brought to the forms

$$\begin{aligned} {}^m B_y = & [\mu_0/(4\pi)]a \int_{-\infty}^{+\infty} dz' \int_{-\pi}^{+\pi} d\theta {}^m S_y^{tr}/[\rho^2 - 2a\rho\cos(-\theta) + a^2 + (z-z')^2]^{3/2} \\ = & [\mu_0/(4\pi)]a \int_{-\infty}^{+\infty} dz' \int_{-\pi}^{+\pi} d\theta {}^m S_y^{tr}/[\rho^2 - 2a\rho\cos(\theta) + a^2 + (z-z')^2]^{3/2}. \end{aligned} \quad (16.5.148)$$

Here we have used the fact that  $\cos(\theta)$  is an *even* function of  $\theta$ . Also, the integration kernel  $[\rho^2 - 2a\rho\cos(\theta) + a^2 + (z - z')^2]^{3/2}$  is an even function of  $\theta$ . Therefore, only the *even* components of  ${}^mS_y^{tr}$  contribute to the integrals, as claimed, and we may write

$${}^mB_y = [\mu_0/(4\pi)]a \int_{-\infty}^{+\infty} dz' \int_{-\pi}^{+\pi} d\theta {}^mS_y^{tre}/[\rho^2 - 2a\rho\cos(\theta) + a^2 + (z - z')^2]^{3/2}. \quad (16.5.149)$$

Examination of (3.162) shows that  ${}^mS_y^{tre}$  consists of factors of the form  $\cos(n\theta)$  multiplied by functions of  $\rho$ ,  $\phi$ ,  $z$ , and  $z'$ . Therefore, as far as the  $\theta$  integral is concerned, what we must ultimately evaluate are integrals  $I_n$  of the form

$$I_n = \int_{-\pi}^{+\pi} d\theta \cos(n\theta)[\rho^2 - 2a\rho\cos(\theta) + a^2 + (z - z')^2]^{-3/2}. \quad (16.5.150)$$

In analogy to (15.13.6), write

$$[\rho^2 - 2a\rho\cos(\theta) + a^2 + (z - z')^2] = [a^2 + (z - z')^2][1 - 2wh + h^2] \quad (16.5.151)$$

where

$$h = \rho/[a^2 + (z - z')^2]^{1/2} \quad (16.5.152)$$

and

$$w = \{a/[a^2 + (z - z')^2]^{1/2}\} \cos \theta = \beta \cos \theta \quad (16.5.153)$$

with

$$\beta = a/[a^2 + (z - z')^2]^{1/2}. \quad (16.5.154)$$

It follows that

$$[\rho^2 - 2a\rho\cos(\theta) + a^2 + (z - z')^2]^{-3/2} = [a^2 + (z - z')^2]^{-3/2}[1 - 2wh + h^2]^{-3/2}. \quad (16.5.155)$$

Correspondingly, the integral (3.165) can be rewritten in the form

$$I_n = [a^2 + (z - z')^2]^{-3/2} \int_{-\pi}^{+\pi} d\theta \cos(n\theta)[1 - 2wh + h^2]^{-3/2}. \quad (16.5.156)$$

Observe that the integrand in (3.170) is similar to the integrand in (15.13.3). The only major difference is the exponent  $-3/2$  rather than  $-1/2$ . See (15.13.6).

In Section 15.13 the integral (15.13.3) was subsequently treated with the aid of the Legendre function generating function (15.13.9). Can a similar approach be considered when the exponent is  $-3/2$  rather than  $-1/2$ ? The answer is *yes*. Thanks to the doctoral thesis and some of the subsequent work of *Leopold Gegenbauer* (1849-1903), it is known that there are the general generating function results

$$[1 - 2wh + h^2]^{-\lambda} = \sum_{\ell=0}^{\infty} h^\ell P_\ell^\lambda(w) \quad (16.5.157)$$

where the functions  $P_\ell^\lambda(w)$  are *polynomials* in  $w$  variously called Gegenbauer polynomials or *ultraspherical* polynomials.<sup>14</sup> [When  $\lambda = (1/2)$  the case of the Legendre polynomial generating function is recovered, and one has  $P_\ell^{1/2}(w) = P_\ell(w)$  where the  $P_\ell(w)$  are the usual Legendre polynomials.] For our purposes we are interested in pursuing the use of (3.171) for the case  $\lambda = 3/2$ . In order to simplify subsequent calculations let us write

$$I_n = [a^2 + (z - z')^2]^{-3/2} K_n \quad (16.5.158)$$

so that

$$K_n = \int_{-\pi}^{+\pi} d\theta \cos(n\theta) [1 - 2wh + h^2]^{-3/2}. \quad (16.5.159)$$

Our task is to explore what can be accomplished by combining (3.173) with (3.171) evaluated at  $\lambda = 3/2$ .

Combining (3.171) and (3.173) gives the result

$$\begin{aligned} K_n &= \int_{-\pi}^{+\pi} d\theta \cos(n\theta) \sum_{\ell=0}^{\infty} h^\ell P_\ell^{3/2}(w) \\ &= \sum_{\ell=0}^{\infty} \left\{ \rho / [a^2 + (z - z')^2]^{1/2} \right\}^\ell \int_{-\pi}^{+\pi} d\theta \cos(n\theta) P_\ell^{3/2}[\beta \cos(\theta)]. \end{aligned} \quad (16.5.160)$$

To continue, we observe that the  $P_\ell^{3/2}(w)$  are polynomials of degree  $\ell$  and are even or odd depending on whether  $\ell$  is even or odd, respectively. Indeed, there are the general formulas

$$P_\ell^{3/2}(w) = \sum_{n'=0}^{\ell/2} *w^{2n'} \text{ for } \ell \text{ even} \quad (16.5.161)$$

and

$$P_\ell^{3/2}(w) = \sum_{n'=0}^{(\ell-1)/2} *w^{1+2n'} \text{ for } \ell \text{ odd.} \quad (16.5.162)$$

---

<sup>14</sup>Gegenbauer polynomials in turn are proportional to special cases of the *Jacobi* polynomials  $P_\ell^{\alpha\beta}(w)$  with  $\alpha = \beta = \lambda - 1/2$ . With regard to mathematical genealogy, Gegenbauer's undergraduate teachers included Joseph Petzval of geometrical optics fame (see Subsection 3.3 of Appendix X) and Ludwig Boltzmann. As a graduate student Gegenbauer's teachers included Weierstrass, Helmholtz, and Kronecker.

They give, for example for the first seven  $P_\ell^{3/2}(w)$ , the results listed below.<sup>15</sup>

$$P_0^{3/2}(w) = 1, \quad (16.5.163)$$

$$P_1^{3/2}(w) = 3w, \quad (16.5.164)$$

$$P_2^{3/2}(w) = -(3/2) + (15/2)w^2, \quad (16.5.165)$$

$$P_3^{3/2}(w) = -(15/2)w + (35/2)w^3, \quad (16.5.166)$$

$$P_4^{3/2}(w) = (15/8) - (105/4)w^2 + (315/8)w^4, \quad (16.5.167)$$

$$P_5^{3/2}(w) = (105/8)w - (315/4)w^3 + (693/8)w^5, \quad (16.5.168)$$

$$P_6^{3/2}(w) = -(35/16) + (945/16)w^2 - (3465/16)w^4 + (3003/16)w^6. \quad (16.5.169)$$

For future use it will also be convenient to write relations such as (3.179) through (3.185) in the form

$$P_\ell^{3/2}(w) = \sum_{k=0}^{\ell} \gamma_{\ell k} w^k \quad (16.5.170)$$

where, since the  $P_\ell^{3/2}$  are of degree  $\ell$  and have definite parity, for a given  $\ell$  many of the coefficients  $\gamma_{\ell k}$  vanish:

$$\gamma_{\ell k} = 0 \text{ when } k > \ell \text{ or when } (-1)^k \neq (-1)^\ell \text{ or both are true.} \quad (16.5.171)$$

For modest values of  $\ell$  the  $\gamma_{\ell k}$  can be read off from (3.179) through (3.185), and for larger values of  $\ell$  can be found using (3.177) and (3.178) or, more conveniently, *Mathematica*. For convenience for future use, we list the first few  $\gamma_{\ell \ell}$  below:

$$\begin{aligned} \gamma_{00} &= 1, \quad \gamma_{11} = 3, \quad \gamma_{22} = 15/2, \quad \gamma_{33} = 35/2, \quad \gamma_{44} = 315/8, \quad \gamma_{55} = 693/8, \\ \gamma_{66} &= 3003/16, \quad \gamma_{77} = 6435/16, \quad \gamma_{88} = 109395/128, \quad \gamma_{99} = 230945/128. \end{aligned} \quad (16.5.172)$$

Moreover, it can be shown that there is the formula

$$\gamma_{nn} = (2n+1)!!/n!, \quad (16.5.173)$$

which can be employed in lieu of *Mathematica*. For future convenience, we observe this relation can be rewritten in the form

$$\gamma_{nn} = (2m-1)!!/(m-1)! \text{ when } n = m-1. \quad (16.5.174)$$

Finally we remark that, for our present purposes, all we need know about Gegenbauer polynomials are four items: they appear in the generating function relation (3.171), they have degree  $\ell$  and parity  $(-1)^\ell$ , and (for  $\lambda = 3/2$ ) the values of the  $\gamma_{nn}$ .

---

<sup>15</sup>*Mathematica* can compute such results to any reasonable order. In *Mathematica*, and much of the mathematics literature, Gegenbauer polynomials are denoted by the symbol  $C_\ell^{(\lambda)}$  rather than  $P_\ell^\lambda$ . We do not because we have already used  $C_{m,\alpha}^{[0]}$  to denote an on-axis gradient. Also, we are in splendid company because our notation follows that of the orthogonal polynomial master *Gábor Szegő* (1895–1985).

The stage is now set to analyze the nature of  $K_n$  as given by (3.174). Define coefficients  $c_{n\ell}$  by the rule

$$c_{n\ell} = \int_{-\pi}^{+\pi} d\theta \cos(n\theta) P_\ell^{3/2}[\beta \cos(\theta)]. \quad (16.5.175)$$

With the aid of this definition we may rewrite  $K_n$  in the form

$$K_n = \sum_{\ell=0}^{\infty} \{\rho/[a^2 + (z - z')^2]^{1/2}\}^\ell c_{n\ell}. \quad (16.5.176)$$

We will eventually show that

$$\begin{aligned} c_{n\ell} &= 0 \text{ for } \ell < n; c_{nn} \neq 0 \text{ and depends only on } \beta; \text{ and } c_{n\ell} = 0 \text{ for } \ell = n+1, \\ &\text{in fact } c_{n\ell} = 0 \text{ if } n \text{ and } \ell \text{ have different parity.} \end{aligned} \quad (16.5.177)$$

Assuming this to be true, we may rewrite (3.192) in the form

$$K_n = \{\rho/[a^2 + (z - z')^2]^{1/2}\}^n c_{nn} + \sum_{\ell=n+2}^{\infty \text{ by } 2s} \{\rho/[a^2 + (z - z')^2]^{1/2}\}^\ell c_{n\ell}. \quad (16.5.178)$$

Here the text “ $\infty$  by  $2s$ ” over the summation sign means that  $\ell$  is supposed to range from  $n+2$  to  $\infty$  in increments of 2 in accord with (3.193). Observe that the first term on the right side of (3.194) is proportional to  $\rho^n$  and the remaining terms go as  $\rho^{n+2}$  and still higher powers that successively increment by factors of  $\rho^2$ . Thus we may write

$$K_n = \rho^n [a^2 + (z - z')^2]^{-n/2} c_{nn} + O[\rho^{n+2}]. \quad (16.5.179)$$

The  $n$  value determines the lowest power in  $\rho$  that can occur, and the successive higher powers that increment by factors of  $\rho^2$ . That  $K_n$  has such a structure is a good omen, because we observe from (3.125) that  $\psi_m$  has a similar structure.

### 16.5.8 Final Results for the General $m > 1$ Normal Cases

After these several preparatory steps, we are ready for the climax. Consider the  $\theta$  dependence of  ${}^m S_y^{tre}$ . Observe that the *smallest*  $n$  value (angular frequency for the quantity  $\theta$ ) in  ${}^m S_y^{tre}$  is

$$n = m - 1. \quad (16.5.180)$$

See (3.162). It comes from terms involving  $\cos[(m-1)\theta]$ . Define  ${}^m s_y$  by the rule

$${}^m s_y = (1/2)\{mf(z') + f'(z')(z - z')\} \cos[(m-1)\phi] \cos[(m-1)\theta]. \quad (16.5.181)$$

It is the term in  ${}^m S_y^{tre}$  with the *smallest* (in  $\theta$ ) angular frequency. All other terms in  ${}^m S_y^{tre}$  have higher angular frequencies (higher  $n$  values) which therefore produce results that are higher order in  $\rho$ . The terms involving  $\cos[(m+1)\theta]$  produce results that are two orders higher in  $\rho$ . Nominally the terms involving  $\cos(m\theta)$  produce results that are one order higher in  $\rho$ . However, examination of (3.162) shows that such terms are themselves proportional

to  $\rho$ , and therefore also produce results that are two orders higher in  $\rho$ . Let us take these observations into account. Look again at (3.164). We conclude that there must be the result

$${}^m B_y = [\mu_0/(4\pi)]a \int_{-\infty}^{+\infty} dz' \int_{-\pi}^{+\pi} d\theta {}^m s_y / [\rho^2 - 2a\rho \cos(\theta) + a^2 + (z - z')^2]^{3/2} + O[\rho^{m+1}] \quad (16.5.182)$$

since we have seen that the terms omitted from  ${}^m S_y^{tre}$  produce results that are of order  $(m+1)$  in  $\rho$ .

Assuming we have correctly assessed what source terms lead to lowest order in  $\rho$  results, as will become ever more apparent later, let us press on to work out the consequences of (3.198). Begin by moving the factors in  ${}^m s_y$  around if they are not being integrated over so that (3.198) is rewritten in the form

$$\begin{aligned} {}^m B_y = & \{ [\mu_0/(4\pi)]a(1/2)[\cos[(m-1)\phi] \\ & \int_{-\infty}^{+\infty} dz' [mf(z') + f'(z')(z - z')] \\ & \int_{-\pi}^{+\pi} d\theta \cos[(m-1)\theta] / [\rho^2 - 2a\rho \cos(\theta) + a^2 + (z - z')^2]^{3/2} \} \\ & + O[\rho^{m+1}]. \end{aligned} \quad (16.5.183)$$

See (3.197). Next work on the theta integration to write

$$\begin{aligned} & \int_{-\pi}^{+\pi} d\theta \cos[(m-1)\theta] / [\rho^2 - 2a\rho \cos(\theta) + a^2 + (z - z')^2]^{3/2} \\ & = [a^2 + (z - z')^2]^{-3/2} \int_{-\pi}^{+\pi} d\theta \cos[(m-1)\theta] [1 - 2wh + h^2]^{-3/2}. \end{aligned} \quad (16.5.184)$$

Here we have used (3.166) and have moved  $z$  and  $z'$  dependent factor outside the theta integration. Therefore at this stage we have the result

$$\begin{aligned} {}^m B_y = & \{ [\mu_0/(4\pi)]a(1/2)[\cos[(m-1)\phi] \\ & \int_{-\infty}^{+\infty} dz' [mf(z') + f'(z')(z - z')] [a^2 + (z - z')^2]^{-3/2} \\ & \int_{-\pi}^{+\pi} d\theta \cos[(m-1)\theta] [1 - 2wh + h^2]^{-3/2} \} \\ & + O[\rho^{m+1}]. \end{aligned} \quad (16.5.185)$$

Now use (3.173) through (3.176) and (3.191) through (3.195) to further process the theta integral so that it becomes

$$\begin{aligned} & \int_{-\pi}^{+\pi} d\theta \cos(n\theta) [1 - 2wh + h^2]^{-3/2} = \\ & \{\rho^n [a^2 + (z - z')^2]^{-n/2}\} c_{nn} + O[\rho^{n+2}] \end{aligned} \quad (16.5.186)$$

with  $n$  given by (3.196). When this result (3.202) is subsequently used in (3.201), the factor  $\rho^n$  that appears on the right side of (3.202) may be moved to the left of the  $z'$  integration so that (3.201) becomes

$$\begin{aligned} {}^m B_y = & \{ [\mu_0/(4\pi)]a(1/2)[\rho^{m-1} \cos[(m-1)\phi] \\ & \int_{-\infty}^{+\infty} dz' [mf(z') + f'(z')(z - z')] [a^2 + (z - z')^2]^{-(n/2+3/2)} c_{nn}\} \\ & + O[\rho^{m+1}]. \end{aligned} \quad (16.5.187)$$

We will learn shortly that  $c_{nn}$  is a function only of  $\beta$  with  $\beta$  given by (5.160), and hence only a function of  $z$ ,  $z'$ , and  $a$ . Therefore, looking at (3.203), we see that we have isolated the  $\rho$  and  $\phi$  dependence of  ${}^m B_y$  up to terms of order  $\rho^{m+1}$ . And, to our great pleasure, we see that the the lowest order  $\rho$  and  $\phi$  dependence is exactly that predicted by (3.142)!

Three things remain to be done: Verify the claims (3.193), work out the  $c_{nn}$ , and manipulate the integral on the right side of (3.204) to a more pleasing form, as was done for the dipole case.

### Fourier Expansion of $[\cos(\theta)]^\ell$

For what follows we will need the Fourier expansions of the functions  $[\cos(\theta)]^\ell$ . These functions are even in  $\theta$  and have period  $2\pi$ . They are therefore Fourier expandable in the harmonics  $\cos(m\theta)$ . Let us work out a few examples, first for even  $\ell$  and then for odd  $\ell$ .

The first few even  $\ell$  cases are:

$$[\cos(\theta)]^0 = 1, \quad (16.5.188)$$

$$\begin{aligned} [\cos(\theta)]^2 &= [(1/2)(\exp(i\theta) + \exp(-i\theta))]^2 \\ &= (1/4)[\exp(2i\theta) + 2\exp(i\theta)\exp(-i\theta) + \exp(-2i\theta)] \\ &= (1/2)\cos(2\theta) + (1/2). \end{aligned} \quad (16.5.189)$$

We see that the Fourier coefficients are readily computable from those in the binomial expansion formula and, when  $\ell$  is even, only even values of  $m$  occur with  $m \leq \ell$ . Also, (when  $\ell > 0$ ) the Fourier coefficient of the highest harmonic,  $\cos(\ell\theta)$ , is  $(1/2)^{\ell-1}$ .

The first few odd  $\ell$  cases are:

$$[\cos(\theta)]^1 = \cos(\theta), \quad (16.5.190)$$

$$\begin{aligned} [\cos(\theta)]^3 &= [(1/2)(\exp(i\theta) + \exp(-i\theta))]^3 \\ &= (1/8)[\exp(3i\theta) + 3\exp(2i\theta)\exp(-i\theta) + 3\exp(i\theta)\exp(-2i\theta) + \exp(-3i\theta)] \\ &= (1/4)\cos(3\theta) + (3/4)\cos(\theta). \end{aligned} \quad (16.5.191)$$

Again the Fourier coefficients are readily computable from those in the binomial expansion formula and, when  $\ell$  is odd, only odd values of  $m$  occur with  $m \leq \ell$ . Also, again the Fourier coefficient of the highest harmonic,  $\cos(\ell\theta)$ , is  $(1/2)^{\ell-1}$ .

### Properties of the Coefficients $c_{n\ell}$

The coefficients  $c_{n\ell}$  are given by (3.191), which we repeat below:

$$c_{n\ell} = \int_{-\pi}^{+\pi} d\theta \cos(n\theta) P_\ell^{3/2}[\beta \cos(\theta)]. \quad (16.5.192)$$

Since the  $P_\ell^{3/2}$  are polynomials of degree  $\ell$  and have definite parity, it follows that (for  $\ell \geq 2$ )

$$P_\ell^{3/2}[\beta \cos(\theta)] = \gamma_{\ell\ell}[\beta \cos(\theta)]^\ell + O\{[\beta \cos(\theta)]^{\ell-2}\}. \quad (16.5.193)$$

See (3.177) through (3.190). Also, since  $P_\ell^{3/2}$  is polynomial,  $P_\ell^{3/2}[\beta \cos(\theta)]$  will be an even and periodic function of  $\theta$  with period  $2\pi$ , and therefore Fourier expandable in the harmonics  $\cos(m\theta)$ . What is the highest angular frequency that can occur in such an expansion? We see from (3.212) and relations of the form (3.205) through (3.210) that there is the result

$$\begin{aligned} P_\ell^{3/2}[\beta \cos(\theta)] &= \gamma_{\ell\ell}\beta^\ell(1/2)^{\ell-1} \cos(\ell\theta) \\ &+ \text{terms of angular frequency } (\ell-2) \\ &+ \text{yet lower angular frequency terms if they exist.} \end{aligned} \quad (16.5.194)$$

Since the functions  $\cos(m\theta)$  form an orthogonal set, the integral (3.211) picks out the Fourier coefficient of  $P_\ell^{3/2}[\beta \cos(\theta)]$  for angular frequency  $n$ . From (3.213) we see that there is none if  $n > \ell$ . Therefore we conclude that

$$c_{n\ell} = 0 \text{ for } \ell < n, \quad (16.5.195)$$

in accord with (3.193). What happens if  $\ell = n + 1$ ? Then, from (3.213), we again see that  $P_\ell^{3/2}[\beta \cos(\theta)]$  has no Fourier coefficient for angular frequency  $n$ . Therefore we conclude that

$$c_{n\ell} = 0 \text{ for } \ell = n + 1, \quad (16.5.196)$$

also in accord with (3.193).

We are now prepared to compute some of the  $c_{n\ell}$  and, in particular, the  $c_{nn}$ . At this point we confess to having been a bit cavalier in some of our calculations. To evaluate  $c_{n\ell}$  properly using (3.211), we should consider separately the cases  $n = 0$ ,  $n = 1$ , and  $n \geq 2$ .

Begin with the case  $n = 0$  for which we will compute  $c_{00}$  and  $c_{01}$ . When  $\ell = 0$  we have, from (3.179), the result

$$P_0^{3/2}[\beta \cos(\theta)] = 1 = \gamma_{00}, \quad (16.5.197)$$

and therefore

$$c_{n0} = \int_{-\pi}^{+\pi} d\theta \cos(n\theta) \gamma_{00} = (2\pi)\gamma_{00}\delta_{n0} \quad (16.5.198)$$

so that

$$c_{00} = 2\gamma_{00}\pi, \quad (16.5.199)$$

in accord with (3.193). Observe that the case  $\ell < n$  never arises if  $n = 0$  since we always have  $n \geq 0$  and  $\ell \geq 0$ ; so then one need worry about (3.214). Also, to be on the safe side, we should compute  $c_{01}$ . From (3.180) and (3.188) we find that

$$P_1^{3/2}[\beta \cos(\theta)] = 3\beta \cos(\theta) = \gamma_{11}\beta \cos(\theta), \quad (16.5.200)$$

and therefore

$$c_{01} = \int_{-\pi}^{+\pi} d\theta \ 3\beta \cos(\theta) = 0, \quad (16.5.201)$$

in accord with (3.215).

We now move on to the case  $n = 1$  for which we will compute  $c_{10}$ ,  $c_{11}$ , and  $c_{12}$ . For  $c_{10}$  we have

$$c_{10} = \int_{-\pi}^{+\pi} d\theta \ \cos(\theta) = 0, \quad (16.5.202)$$

in accord with (3.214). Next find  $c_{11}$ . For  $n = \ell = 1$  use of (3.211) and (3.219) gives

$$\begin{aligned} c_{11} &= \int_{-\pi}^{+\pi} d\theta \ \cos(\theta) \gamma_{11} \beta \cos(\theta) = \gamma_{11} \beta \int_{-\pi}^{+\pi} d\theta \ \cos^2(\theta) \\ &= \pi \gamma_{11} \beta, \end{aligned} \quad (16.5.203)$$

in accord with (3.193). What about  $c_{12}$ ? For  $\ell = 2$  use of (3.181) and (3.206) gives

$$\begin{aligned} P_2^{3/2}[\beta \cos(\theta)] &= (15/2)[\beta \cos(\theta)]^2 - 3/2 = \gamma_{22} \beta^2 \cos^2(\theta) - 3/2 \\ &= \gamma_{22} \beta^2 [(1/2) \cos(2\theta) + (1/2)] - 3/2. \end{aligned} \quad (16.5.204)$$

It follows that

$$c_{12} = \int_{-\pi}^{+\pi} d\theta \ \cos(\theta) \{\gamma_{22} \beta^2 [(1/2) \cos(2\theta) + (1/2)] - 3/2\} = 0, \quad (16.5.205)$$

in accord with (3.215).

Moving on yet further, we consider the case  $n = 2$ , for which we will compute  $c_{20}$ ,  $c_{21}$ ,  $c_{22}$ , and  $c_{23}$ . For  $c_{20}$  we find from (3.211) and (3.216) the result

$$c_{20} = \int_{-\pi}^{+\pi} d\theta \ \cos(2\theta) = 0, \quad (16.5.206)$$

in accord with (3.214). For  $c_{21}$  we find from (3.211) and (3.219) the result

$$c_{21} = \int_{-\pi}^{+\pi} d\theta \ \cos(2\theta) \gamma_{11} \beta \cos(\theta) = 0, \quad (16.5.207)$$

again in accord with (3.214). For  $c_{22}$  we find from (3.211) and (3.223) the result

$$\begin{aligned} c_{22} &= \int_{-\pi}^{+\pi} d\theta \ \cos(2\theta) \{\gamma_{22} \beta^2 [(1/2) \cos(2\theta) + (1/2)] - 3/2\} \\ &= \int_{-\pi}^{+\pi} d\theta \ \cos(2\theta) \gamma_{22} \beta^2 [(1/2) \cos(2\theta)] \\ &= \gamma_{22} \beta^2 (1/2) \int_{-\pi}^{+\pi} d\theta \ \cos^2(2\theta) \\ &= \pi (1/2) \gamma_{22} \beta^2, \end{aligned} \quad (16.5.208)$$

again in accord with (3.193). The last  $c_{n\ell}$  to be computed in this series is  $c_{23}$  for which we need  $P_3^{3/2}[\beta \cos(\theta)]$ . According to (3.182) it is given by

$$P_3^{3/2}[\beta \cos(\theta)] = -(15/2)\beta \cos(\theta) + (35/2)\beta^3 \cos^3(\theta). \quad (16.5.209)$$

From (3.208) and (3.209) we see that it contains angular frequencies 1 and 3, but *not* 2. It follows that

$$c_{23} = 0, \quad (16.5.210)$$

in accord with (3.215).

At last we are ready to compute  $c_{nn}$  with  $n \geq 3$ . But, before so doing, we take a small detour: For  $\ell \geq 2$  we have, according to (3.186) and (3.187),

$$\begin{aligned} P_\ell^{3/2}[\beta \cos(\theta)] &= \gamma_{\ell\ell}\beta^\ell \cos^\ell(\theta) + \text{terms of degree } (\ell - 2) \text{ and even lower} \\ &\quad \text{degrees in decrements of 2 if such terms exist.} \end{aligned} \quad (16.5.211)$$

From relations of the form (3.205) through (3.210) we see that (3.230) has the Fourier decomposition

$$\begin{aligned} P_\ell^{3/2}[\beta \cos(\theta)] &= (1/2)^{\ell-1} \gamma_{\ell\ell}\beta^\ell \cos(\ell\theta) \\ &\quad + \text{terms with angular frequency } (\ell - 2) \text{ and even lower} \\ &\quad \text{frequencies in decrements of 2 if such frequencies exist.} \end{aligned} \quad (16.5.212)$$

From this observation the last condition in (3.193) follows.

With this detour ended, and more to the point for our present purposes, it also follows that for  $c_{nn}$  with  $n \geq 3$  there is the result

$$\begin{aligned} c_{nn} &= \int_{-\pi}^{+\pi} d\theta \cos(n\theta) \gamma_{nn}\beta^n [(1/2)^{n-1} \cos(n\theta) + \text{lower frequency terms}] \\ &= \gamma_{nn}\beta^n (1/2)^{n-1} \int_{-\pi}^{+\pi} d\theta \cos^2(2\theta) \\ &= \pi(1/2)^{n-1} \gamma_{nn}\beta^n. \end{aligned} \quad (16.5.213)$$

We also observe that (3.218), (3.222), and (3.227) for  $c_{00}$ ,  $c_{11}$ , and  $c_{22}$  are all particular cases of (3.232), which is therefore actually a general formula for all  $n$ . Finally, we see that the  $c_{nn}$  depend only on  $\beta$ , in accord with (3.193).

## Summing Up

Our last task is to manipulate (3.204) to bring it to a more pleasing form. Begin by observing from (3.175) that

$$\beta^n = a^n [a^2 + (z - z')^2]^{-n/2} \quad (16.5.214)$$

so that (3.232) becomes

$$c_{nn} = \pi(1/2)^{n-1} \gamma_{nn} a^n [a^2 + (z - z')^2]^{-n/2}. \quad (16.5.215)$$

Inserting this result into (3.204) brings it to the form

$$\begin{aligned} {}^m B_y &= [\mu_0/(4\pi)]a(1/2)(1/m)\pi(1/2)^{n-1}\gamma_{nn}a^n \\ &\quad \rho^{m-1} \cos[(m-1)\phi] \int_{-\infty}^{+\infty} dz' \{mf(z') + f'(z')(z-z')\} [a^2 + (z-z')^2]^{-(n/2+n/2+3/2)} \end{aligned} \quad (16.5.216)$$

with, according to (3.196),  $n$  given by

$$n = m - 1. \quad (16.5.217)$$

### Consolidation of Terms

Below are two pieces of algebra:

$$(n/2 + n/2 + 3/2) = n + 3/2 = m + 1/2, \quad (16.5.218)$$

and

$$\begin{aligned} [\mu_0/(4\pi)]a(1/2)(1/m)\pi(1/2)^{n-1}\gamma_{nn}a^n &= [\mu_0](1/m)(1/2)^{n+2}\gamma_{nn}a^{n+1} \\ &= [\mu_0](1/m)(1/2)^{m+1}\gamma_{nn}a^m \\ &= \mu_0 a^m (1/m)(1/2)^{m+1} [(2m-1)!!/(m-1)!]. \end{aligned} \quad (16.5.219)$$

[In (5.277) we have used (5.180).] Therefore we may also write

$$\begin{aligned} {}^m B_y &= \mu_0 a^m (1/m)(1/2)^{m+1} [(2m-1)!!/(m-1)!] \times \\ &\quad \{\rho^{m-1} \cos[(m-1)\phi] \int_{-\infty}^{+\infty} dz' \{mf(z') + f'(z')(z-z')\} [a^2 + (z-z')^2]^{-(m+1/2)} \\ &\quad + O[\rho^{m+1}]\}. \end{aligned} \quad (16.5.220)$$

To facilitate the comparison between (3.142) and (3.203), below we repeat, in slightly modified form, the contents of the second line in (3.242),

$${}^m B_y = C_{m,s}^{[0]}(z)m\rho^{m-1} \cos[(m-1)\phi] + O[\rho^{m+1}]. \quad (16.5.221)$$

Upon equating terms of like power in  $\rho$  in the  ${}^m B_y$  as given by this relation and that given by (5.228), and dividing out like terms, we are able to find the on-axis gradient  $C_{m,s}^{[0]}(z)$  for any  $m$ . It is given by

$$\begin{aligned} C_{m,s}^{[0]}(z) &= \mu_0 a^m (1/m)^2 (1/2)^{m+1} [(2m-1)!!/(m-1)!] \\ &\quad \int_{-\infty}^{+\infty} dz' \{mf(z') + f'(z')(z-z')\} [a^2 + (z-z')^2]^{-(m+1/2)}. \end{aligned} \quad (16.5.222)$$

### Integration by Parts

What remains is to simplify (5.230) by integration by parts. Our approach will be patterned after that employed earlier for the  $m = 1$  normal dipole case. Begin by writing the integral in (5.230) as the sum of two terms depending on  $f$  and  $f'$ , respectively.

Write and manipulate the first term to find the result

$$\begin{aligned} \text{first term} &= \int_{-\infty}^{+\infty} dz' mf(z') [a^2 + (z - z')^2]^{-(m+1/2)} \\ &= \int_{-\infty}^{+\infty} dz' f(z') \{m\} [a^2 + (z - z')^2]^{-(m+1/2)} \\ &= \int_{-\infty}^{+\infty} dz' f(z') \{m\} [a^2 + (z - z')^2] [a^2 + (z - z')^2]^{-(m+3/2)} \\ &= \int_{-\infty}^{+\infty} dz' f(z') [ma^2 + m(z - z')^2] [a^2 + (z - z')^2]^{-(m+3/2)}. \end{aligned} \quad (16.5.223)$$

Write and integrate the second term by parts to find the result

$$\begin{aligned} \text{second term} &= \int_{-\infty}^{+\infty} dz' [f'(z')(z - z')]/[a^2 + (z - z')^2]^{(m+1/2)} \\ &= \{[f(z')(z - z')]/[a^2 + (z - z')^2]^{(m+1/2)}\}|_{z'=-\infty}^{z'=+\infty} \\ &\quad - \int_{-\infty}^{+\infty} dz' f(z') \partial_{z'} \{(z - z')/[a^2 + (z - z')^2]^{(m+1/2)}\} \\ &= - \int_{-\infty}^{+\infty} dz' f(z') \partial_{z'} \{(z - z')/[a^2 + (z - z')^2]^{(m+1/2)}\}. \end{aligned} \quad (16.5.224)$$

Here we have assumed that

$$\{[f(z')(z - z')]/[a^2 + (z - z')^2]^{(m+1/2)}\}|_{z'=-\infty}^{z'=+\infty} = 0, \quad (16.5.225)$$

which will be the case if  $f(z')$  is bounded at  $z' = \pm\infty$ . Moreover, we also have

$$\begin{aligned} \partial_{z'} \{(z - z')/[a^2 + (z - z')^2]^{(m+1/2)}\} &= \\ -1/[a^2 + (z - z')^2]^{(m+1/2)} + 2(m+1/2)(z - z')^2/[a^2 + (z - z')^2]^{(m+3/2)} &= \\ -1/[a^2 + (z - z')^2]^{(m+1/2)} + (2m+1)(z - z')^2/[a^2 + (z - z')^2]^{(m+3/2)}. \end{aligned} \quad (16.5.226)$$

It follows, upon combining (3.241) and (3.243), that

$$\begin{aligned} \text{second term} &= \int_{-\infty}^{+\infty} dz' [f'(z')(z - z')]/[a^2 + (z - z')^2]^{(m+1/2)} \\ &= - \int_{-\infty}^{+\infty} dz' f(z') \partial_{z'} \{(z - z')/[a^2 + (z - z')^2]^{(m+1/2)}\} \\ &= \int_{-\infty}^{+\infty} dz' f(z') \{1/[a^2 + (z - z')^2]^{(m+1/2)} - (2m+1)(z - z')^2/[a^2 + (z - z')^2]^{(m+3/2)}\}. \end{aligned} \quad (16.5.227)$$

Evidently the second term has two parts, The first part can be written and manipulated to give the result

$$\begin{aligned}
 & \text{first part of second term} \\
 &= \int_{-\infty}^{+\infty} dz' f(z') 1/[a^2 + (z - z')^2]^{(m+1/2)} \\
 &= \int_{-\infty}^{+\infty} dz' f(z') [a^2 + (z - z')^2]/[a^2 + (z - z')^2]^{(m+3/2)}. \tag{16.5.228}
 \end{aligned}$$

The second part is evidently given by the relation

$$\begin{aligned}
 & \text{second part of second term} \\
 &= \int_{-\infty}^{+\infty} dz' f(z') [-(2m+1)(z - z')^2]/[a^2 + (z - z')^2]^{(m+3/2)}. \tag{16.5.229}
 \end{aligned}$$

We can now add these two parts together, thereby regaining the full second term, to find

$$\begin{aligned}
 & \text{sum of parts} = \text{full second term} \\
 &= \int_{-\infty}^{+\infty} dz' f(z') \{[a^2 + (z - z')^2] + [-(2m+1)(z - z')^2]\}/[a^2 + (z - z')^2]^{(m+3/2)} \\
 &= \int_{-\infty}^{+\infty} dz' f(z') [a^2 - 2m(z - z')^2]/[a^2 + (z - z')^2]^{(m+3/2)}. \tag{16.5.230}
 \end{aligned}$$

Finally, we can add the first term (3.240) and the second term (3.247) together to find the full result,

$$\begin{aligned}
 & \text{full result} \\
 &= \int_{-\infty}^{+\infty} dz' f(z') \{[ma^2 + m(z - z')^2] + [a^2 - 2m(z - z')^2]\}/[a^2 + (z - z')^2]^{(m+3/2)} \\
 &= \int_{-\infty}^{+\infty} dz' f(z') [(m+1)a^2 - m(z - z')^2]/[a^2 + (z - z')^2]^{(m+3/2)}. \tag{16.5.231}
 \end{aligned}$$

## Final Result

Let us put everything together. Starting with the integral on the right side of (5.230) we have arrived at the result (5.239),

$$\begin{aligned}
 & \int_{-\infty}^{+\infty} dz' \{mf(z') + f'(z')(z - z')\} [a^2 + (z - z')^2]^{-(m+1/2)} \\
 &= \int_{-\infty}^{+\infty} dz' f(z') [(m+1)a^2 - m(z - z')^2]/[a^2 + (z - z')^2]^{(m+3/2)}. \tag{16.5.232}
 \end{aligned}$$

Employing (5.240) in (5.230) produces the final result

$$\begin{aligned}
 C_{m,s}^{[0]}(z) &= \mu_0 a^m (1/m)^2 (1/2)^{m+1} [(2m-1)!!/(m-1)!] \\
 &\quad \int_{-\infty}^{+\infty} dz' f(z') [(m+1)a^2 - m(z - z')^2]/[a^2 + (z - z')^2]^{(m+3/2)} \tag{16.5.233}
 \end{aligned}$$

which can also be written in the form

$$\begin{aligned} C_{m,s}^{[0]}(z) &= -\mu_0 a^m (1/m)^2 (1/2)^{m+1} [(2m-1)!!/(m-1)!] \\ &\int_{-\infty}^{+\infty} dz' f(z') [-(m+1)a^2 + m(z-z')^2] / [a^2 + (z-z')^2]^{(m+3/2)}. \end{aligned} \quad (16.5.234)$$

As a sanity check let us evaluate this result for the dipole case  $m = 1$  to verify that we find the earlier result (3.120). When  $m = 1$  (and with a change in sign) the coefficient in front of the integral in (3.252) becomes

$$-\mu_0 a^m (1/m)^2 (1/2)^{m+1} [(2m-1)!!/(m-1)!] = -\mu_0 (a/4). \quad (16.5.235)$$

And the integral becomes

$$\begin{aligned} &\int_{-\infty}^{+\infty} dz' f(z') [-(m+1)a^2 + m(z-z')^2] / [a^2 + (z-z')^2]^{(m+3/2)}. \\ &= \int_{-\infty}^{+\infty} dz' f(z') [-2a^2 + (z-z')^2] / [a^2 + (z-z')^2]^{5/2}. \end{aligned} \quad (16.5.236)$$

We see that both the coefficient and integral agree with those in (3.120), and therefore the general result (3.252), when evaluated at  $m = 1$ , agrees with the earlier result (3.120).

### Computation of Derivatives of On-Axis Gradients

We have found a formula for the on-axis gradients  $C_{m,s}^{[0]}(z)$ . What about their  $z$  derivatives  $C_{m,s}^{[n]}(z)$ ? Define a function  $\Lambda(m; z, z')$  by the rule

$$\Lambda(m; z, z') = [-(m+1)a^2 + m(z-z')^2] / [a^2 + (z-z')^2]^{(m+3/2)} \quad (16.5.237)$$

so that (3.252) can be written in the form

$$\begin{aligned} C_{m,s}^{[0]}(z) &= -\mu_0 (1/m)^2 (1/2)^{m+1} [(2m-1)!!/(m-1)!] a^m \\ &\int_{-\infty}^{+\infty} dz' f(z') \Lambda(m; z, z'). \end{aligned} \quad (16.5.238)$$

[Here we have used (3.250).] The functions  $\Lambda(m; z, z')$  act as integration *kernels*. Moreover, from (3.259) it is evident that the  $C_{m,s}^{[n]}(z)$  are given by the relations

$$\begin{aligned} C_{m,s}^{[n]}(z) &= -\mu_0 (1/m)^2 (1/2)^{m+1} [(2m-1)!!/(m-1)!] a^m \\ &\int_{-\infty}^{+\infty} dz' f(z') \partial_z^n \Lambda(m; z, z'). \end{aligned} \quad (16.5.239)$$

And, looking at the functional form of the  $\Lambda(m; z, z')$  as given by (3.258), it is clear that the functions  $\partial_z^n \Lambda(m; z, z')$  can be found analytically using, for example, *Mathematica*. Therefore, we also have, in analytic form, kernels for the  $C_{m,s}^{[n]}(z)$ , namely the  $\partial_z^n \Lambda(m; z, z')$ .

### Comparison with Analogous General Result of Walstrom

Consult Equation (49) of the 2 January 2022 paper “Magnetic Field Models for Multipole Magnets in Cylindrical Coordinates” by Peter Walstrom. Equation (49) reads, in effect,

$$C_{m,s}^{[0]}(z) \sim -a[(2m-1)!!]/[2^{m+1}(m-1)!] \int_{-\infty}^{+\infty} dz' f_{ms}(z') \partial_a \{a^m/[a^2 + (z-z')^2]^{m+1/2}\}. \quad (16.5.240)$$

(This result was found using hypergeometric function machinery.) Performing the indicated differentiation gives

$$\begin{aligned} & \partial_a \{a^m/[a^2 + (z-z')^2]^{m+1/2}\} \\ &= ma^{m-1}/[a^2 + (z-z')^2]^{m+1/2} - a^m(2a)(m+1/2)[a^2 + (z-z')^2]^{m+3/2} \\ &= ma^{m-1}[a^2 + (z-z')^2]/[a^2 + (z-z')^2]^{m+3/2} - a^{m+1}(2m_1)/[a^2 + (z-z')^2]^{m+3/2} \\ &= a^{m-1}[ma^2 + m(z-z')^2 - (2m+1)a^2]/[a^2 + (z-z')^2]^{m+3/2} \\ &= a^{m-1}[-(m+1)a^2 + m(z-z')^2]/[a^2 + (z-z')^2]^{m+3/2}. \end{aligned} \quad (16.5.241)$$

Combining (3.255) and (3.256) gives the final result

$$C_{m,s}^{[0]}(z) \sim -a^m[(2m-1)!!]/[2^{m+1}(m-1)!] \int_{-\infty}^{+\infty} dz' f_{ms}(z')[-(m+1)a^2 + m(z-z')^2]/[a^2 + (z-z')^2]^{m+3/2}. \quad (16.5.242)$$

Note that, apart from a factor of  $\mu_0$  which presumably has to do with units, and a factor of  $(1/m)$  which presumably has to do with convention, (3.252) and (3.257) agree!

\*\*\*\*\*

What can be said about the coefficient that precedes the integral in (3.239)? With the aid of (3.190) we find for this coefficient the result

$$\mu_0(1/m)(1/2)^{m+1}\gamma_{nn}a^m = \mu_0a^m(1/m)(1/2)^{m+1}[(2m-1)!!/(m-1)!]. \quad (16.5.243)$$

\*\*\*\*\*

### 16.5.9 Use of Vector Potential as a Stepping Stone

We know that in a source-free region the magnetic field  $\mathbf{B}$  can be obtained from a scalar potential  $\psi$ . We have also found a cylindrical harmonic expansion for  $\psi$ , and therefore also for  $\mathbf{B}$ , in terms of on-axis gradients. In a previous section we exploited the Biot-Savart law, which relates the magnetic field  $\mathbf{B}$  to the current  $\mathbf{J}$ , to find the on-axis gradients, initially in terms of  $\mathbf{J}$ , and ultimately in terms of the shape functions. We also know that there is a law, apparently *unnamed*, see (6.3), that relates the vector potential  $\mathbf{A}$  to the current  $\mathbf{J}$ . The purpose of this section is to show that this unnamed law can also be used to find the on-axis gradients in terms of the shape functions.

## Preliminary Background

### Properties of the Vector Potential

We will begin with a review of some properties of the vector potential  $\mathbf{A}$  and its relation to the current  $\mathbf{J}$ . The vector potential  $\mathbf{A}$  is a vector field with the property

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (16.5.244)$$

where  $\mathbf{B}$  is the underlying magnetic field of physical interest. In the static (no time dependence) case and assuming  $\mathbf{J}$  is divergence free,  $\mathbf{B}$  is given in terms of the current density  $\mathbf{J}(\mathbf{r})$  by the *Biot-Savart* law

$$\mathbf{B}(\mathbf{r}) = [\mu_0/(4\pi)] \int d^3\mathbf{r}' [\mathbf{J}(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}')]/||\mathbf{r} - \mathbf{r}'||^3. \quad (16.5.245)$$

If we define  $\mathbf{A}$  by the unnamed law/rule

$$\mathbf{A}(\mathbf{r}) = [\mu_0/(4\pi)] \int d^3\mathbf{r}' \mathbf{J}(\mathbf{r}')/||\mathbf{r} - \mathbf{r}'||, \quad (16.5.246)$$

then direct computation shows that this  $\mathbf{A}$  satisfies (6.1). Indeed, we find

$$\nabla \times \mathbf{A} = [\mu_0/(4\pi)] \int d^3\mathbf{r}' \nabla \times [\mathbf{J}(\mathbf{r}')/||\mathbf{r} - \mathbf{r}'||]. \quad (16.5.247)$$

Also, vector manipulation formulas give the result

$$\nabla \times [\mathbf{J}(\mathbf{r}')/||\mathbf{r} - \mathbf{r}'||] = -\mathbf{J}(\mathbf{r}') \times \nabla[1/||\mathbf{r} - \mathbf{r}'||] = [\mathbf{J}(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}')]/||\mathbf{r} - \mathbf{r}'||^3. \quad (16.5.248)$$

Therefore use of (6.4), (6.5), and (6.2) in succession yields the result

$$\nabla \times \mathbf{A} = [\mu_0/(4\pi)] \int d^3\mathbf{r}' [\mathbf{J}(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}')]/||\mathbf{r} - \mathbf{r}'||^3 = \mathbf{B}(\mathbf{r}). \quad (16.5.249)$$

We can also verify by direct computation that this  $\mathbf{A}$  is divergence free providing  $\mathbf{J}$  is divergence free,

$$\nabla \cdot \mathbf{J} = 0 \Rightarrow \nabla \cdot \mathbf{A} = 0, \quad (16.5.250)$$

and this  $\mathbf{A}$  will therefore then be in a Coulomb gauge. Beginning with (6.3) we find

$$\nabla \cdot \mathbf{A} = [\mu_0/(4\pi)] \int d^3\mathbf{r}' \nabla \cdot [\mathbf{J}(\mathbf{r}')/||\mathbf{r} - \mathbf{r}'||]. \quad (16.5.251)$$

Also, vector manipulation formulas give the result

$$\nabla \cdot [\mathbf{J}(\mathbf{r}')/||\mathbf{r} - \mathbf{r}'||] = \mathbf{J}(\mathbf{r}') \cdot \nabla[1/||\mathbf{r} - \mathbf{r}'||] = -\mathbf{J}(\mathbf{r}') \cdot \nabla'[1/||\mathbf{r} - \mathbf{r}'||]. \quad (16.5.252)$$

Combining (6.7) and (6.8) gives the result

$$\nabla \cdot \mathbf{A} = -[\mu_0/(4\pi)] \int d^3\mathbf{r}' \mathbf{J}(\mathbf{r}') \cdot \nabla'[1/||\mathbf{r} - \mathbf{r}'||]. \quad (16.5.253)$$

Moreover, there is the formula and result

$$\begin{aligned} & \nabla' \cdot [\mathbf{J}(\mathbf{r}') / \|\mathbf{r} - \mathbf{r}'\|] \\ &= [1/\|\mathbf{r} - \mathbf{r}'\|] \nabla' \cdot \mathbf{J}(\mathbf{r}') + \mathbf{J}(\mathbf{r}') \cdot \nabla' [1/\|\mathbf{r} - \mathbf{r}'\|] \\ &= \mathbf{J}(\mathbf{r}') \cdot \nabla' [1/\|\mathbf{r} - \mathbf{r}'\|] \end{aligned} \quad (16.5.254)$$

where we have assumed that  $\mathbf{J}$  is divergence free. It follows from (6.10) and the divergence theorem that

$$\begin{aligned} \int_V d^3\mathbf{r}' \mathbf{J}(\mathbf{r}') \cdot \nabla' [1/\|\mathbf{r} - \mathbf{r}'\|] &= \int_V d^3\mathbf{r}' \nabla' \cdot [\mathbf{J}(\mathbf{r}') / \|\mathbf{r} - \mathbf{r}'\|] \\ &= \int_S d\mathbf{S}' \cdot [\mathbf{J}(\mathbf{r}') / \|\mathbf{r} - \mathbf{r}'\|] \rightarrow 0 \text{ provided } \mathbf{J}|_{\mathbf{r}' \in S} \rightarrow 0 \\ &\quad \text{sufficiently rapidly as } V \text{ and } S \rightarrow \infty. \end{aligned} \quad (16.5.255)$$

Upon comparing (6.10) and (6.12) we see that

$$\nabla \cdot \mathbf{A} = 0 \quad (16.5.256)$$

under the same hypothesis (6.12) about the behavior of  $\mathbf{J}$  at  $\infty$ .

Since  $\mathbf{A}$  as given by (6.3) is in a Coulomb gauge, one might hope it is in the *symmetric* Coulomb gauge. However, we will see that this is the case only when  $m = 0$ .

Let us complete the circle in our reasoning. From (6.2) it follows that there is (in the static case and for  $\mathbf{J}$  divergence free) the differential relation

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}. \quad (16.5.257)$$

Upon combining (6.1) and (6.15) we see that there is the general result

$$\nabla \times (\nabla \times \mathbf{A}) = -\nabla^2 \mathbf{A} + \nabla(\nabla \cdot \mathbf{A}) = \mu_0 \mathbf{J}. \quad (16.5.258)$$

And, when (6.14) is taken into account, we see that for the  $\mathbf{A}$  given by (6.3) there is the result

$$-\nabla^2 \mathbf{A} = \mu_0 \mathbf{J}. \quad (16.5.259)$$

In particular, in current-free regions, the Cartesian components of this  $\mathbf{A}$  are harmonic functions. Recall the discussion at the beginning of Section 5. We note that the result (6.17) can also be found directly from the definition (6.3),

$$\begin{aligned} -\nabla^2 \mathbf{A}(\mathbf{r}) &= -[\mu_0/(4\pi)] \int d^3\mathbf{r}' \mathbf{J}(\mathbf{r}') \nabla^2 [1/\|(\mathbf{r} - \mathbf{r}')\|] \\ &= -[\mu_0/(4\pi)] \int d^3\mathbf{r}' \mathbf{J}(\mathbf{r}') (-4\pi) \delta_3(\mathbf{r} - \mathbf{r}') \\ &= \mu_0 \mathbf{J}(\mathbf{r}). \end{aligned} \quad (16.5.260)$$

### Plan for use of $\mathbf{A}$

As stated earlier, the overall aim of this section is to obtain the on-axis gradients in terms of  $\mathbf{J}$ . More precisely, since  $\mathbf{J}$  is specified by the shape functions  $g(z)$  and  $f_{ma}$  [see (4.33) through (4.35) and (4.24) through (4.26)], we wish to obtain the on-axis gradients in terms of the shape functions. We have already seen in Section 5 that this can be done using the Biot-Savart law. We will now show that another way to find on-axis gradients in terms of shape functions is as follows:

1. Express the vector potential  $\mathbf{A}$ , initially in terms of  $\mathbf{J}$  and ultimately in terms of the shape functions, using the nameless law (5.246).
2. Express the magnetic field  $\mathbf{B}$  in terms of the shape functions using the resulting  $\mathbf{A}$  in (6.1). Note that this relation does not depend on the choice of gauge for  $\mathbf{A}$ .
3. Express magnetic scalar potentials  $\psi$  in terms of on-axis gradients.
4. Express magnetic fields  $\mathbf{B}$  in terms of magnetic scalar potentials  $\psi$ , and hence in terms of on-axis gradients.
5. Equate the two results for  $\mathbf{B}$  and then perform additional manipulations to obtain on-axis gradients in terms of the shape functions.

Note that steps 3 through 5 above are the same as steps 1, 2, and 4 when the Biot-Savart law is employed. What is the main attribute of this plan? The method of Section 5 employed the less familiar Gegenbauer polynomials to explicate the connection between  $\mathbf{J}$  and  $\mathbf{B}$  given by the Biot-Savart law (5.40). By contrast, for this plan the more familiar Legendre polynomials can be employed to explicate the connection between  $\mathbf{J}$  and  $\mathbf{A}$  given by (5.246).

### Implementation of the Plan

#### Application to a Cylinder for the $m = 0$ Case

Suppose we attempt to compute, in cylindrical coordinates, the vector potential  $\mathbf{A}$  in terms of  $\mathbf{J}$  as given (6.3). As a preliminary exercise, we will first treat the easier  $m = 0$  (solenoid) case, and will call the resulting vector potential  ${}^0\mathbf{A}$ . For a solenoid the magnetic field, which we will denote as  ${}^0\mathbf{B}$ , is primarily in the  $z$  direction. Therefore, since

$$B_z = \partial_x A_y - \partial_y A_x, \quad (16.5.261)$$

we will want to compute  ${}^0A_x$  and  ${}^0A_y$ .

In particular, based on (6.3), let us use  $j_y$  as given by (5.34) to compute  ${}^0A_y$ ,

$$\begin{aligned} {}^0A_y(\mathbf{r}) &= [\mu_0/(4\pi)] \int d^3\mathbf{r}' J_y(\mathbf{r}') / \|\mathbf{r} - \mathbf{r}'\| \\ &= [\mu_0/(4\pi)] a \int_{-\infty}^{\infty} dz' \int_{-\pi}^{\pi} d\phi' j_y(\mathbf{r}') / \|\mathbf{r} - \mathbf{r}'\| \\ &= [\mu_0/(4\pi)] a \int_{-\infty}^{\infty} dz' g(z') \int_{-\pi}^{\pi} d\phi' \cos(\phi') / \|\mathbf{r} - \mathbf{r}'\|. \end{aligned} \quad (16.5.262)$$

Subsequently  $j_x$  as given by (5.33) can be used analogously to find results for  ${}^0 A_x$ . Also, since  $j_z = 0$ , see (5.35), it follows that

$${}^0 A_z = 0. \quad (16.5.263)$$

Concentrate on manipulating the angular integral in (6.19). Use the result (3.92) to write

$$\| \mathbf{r} - \mathbf{r}' \|^{-1} = [\rho^2 - 2a\rho \cos(\phi - \phi') + a^2 + (z - z')^2]^{-1/2}. \quad (16.5.264)$$

As in Section 5, make the change of variables (5.136) through (5.138). When this is done the angular integral becomes

$$\begin{aligned} & \int_{-\pi}^{\pi} d\phi' \cos(\phi') / \| \mathbf{r} - \mathbf{r}' \| \\ &= \int_{-\pi}^{\pi} d\theta \cos(\phi + \theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + (z - z')^2]^{-1/2}. \end{aligned} \quad (16.5.265)$$

Also recall that

$$\cos(\phi + \theta) = \cos(\phi) \cos(\theta) - \sin(\phi) \sin(\theta) \quad (16.5.266)$$

so that we may write

$$\begin{aligned} & \int_{-\pi}^{\pi} d\theta \cos(\phi + \theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + (z - z')^2]^{-1/2} \\ &= \cos(\phi) \int_{-\pi}^{\pi} d\theta \cos(\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + (z - z')^2]^{-1/2}. \end{aligned} \quad (16.5.267)$$

Here we have used the fact that

$$\int_{-\pi}^{\pi} d\theta \sin(\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + (z - z')^2]^{-1/2} = 0 \quad (16.5.268)$$

since the integrand in (6.36) is an odd function of  $\theta$ . Evidently (6.24) isolates the  $\phi$  dependence.

Let us next try to isolate some  $\rho$  dependence. Use (3.166) through (3.168) to write

$$[\rho^2 - 2a\rho \cos(\theta) + a^2 + (z - z')^2] = [a^2 + (z - z')^2][1 - 2wh + h^2] \quad (16.5.269)$$

where

$$h = \rho / [a^2 + (z - z')^2]^{1/2} \quad (16.5.270)$$

and

$$w = \{a / [a^2 + (z - z')^2]^{1/2}\} \cos(\theta) = \beta \cos(\theta) \quad (16.5.271)$$

with

$$\beta = \{a / [a^2 + (z - z')^2]^{1/2}\}. \quad (16.5.272)$$

It follows that

$$[\rho^2 - 2a\rho \cos(\theta) + a^2 + (z - z')^2]^{-1/2} = [a^2 + (z - z')^2]^{-1/2}[1 - 2wh + h^2]^{-1/2}. \quad (16.5.273)$$

Consequently, the angular integral in (6.35) can be rewritten in the form

$$\begin{aligned} & \int_{-\pi}^{\pi} d\theta \cos(\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + (z - z')^2]^{-1/2} \\ &= [a^2 + (z - z')^2]^{-1/2} \int_{-\pi}^{\pi} d\theta \cos(\theta) [1 - 2wh + h^2]^{-1/2}. \end{aligned} \quad (16.5.274)$$

To continue, work on the angular integral in (6.42). For this purpose and further applications, it is useful to consider the more general family of integrals  $K_n$  defined by

$$K_n = \int_{-\pi}^{\pi} d\theta \cos(n\theta) [1 - 2wh + h^2]^{-1/2}. \quad (16.5.275)$$

Recall that there is the Legendre polynomial generating function relation

$$[1 - 2wh + h^2]^{-1/2} = \sum_{\ell=0}^{\infty} h^\ell P_\ell(w). \quad (16.5.276)$$

Consequently, with the aid of (6.38), (6.39), and (6.44), for the angular integrals in (6.43) we may write

$$\begin{aligned} K_n &= \int_{-\pi}^{\pi} d\theta \cos(n\theta) [1 - 2wh + h^2]^{-1/2} = \int_{-\pi}^{+\pi} d\theta \cos(n\theta) \sum_{\ell=0}^{\infty} h^\ell P_\ell(w) \\ &= \sum_{\ell=0}^{\infty} \{\rho/[a^2 + (z - z')^2]^{1/2}\}^\ell \int_{-\pi}^{+\pi} d\theta \cos(n\theta) P_\ell[\beta \cos(\theta)]. \end{aligned} \quad (16.5.277)$$

The stage is now set to analyze the nature of  $K_n$  as given by (6.45). Define coefficients  $c_{n\ell}$  by the rule

$$c_{n\ell} = \int_{-\pi}^{+\pi} d\theta \cos(n\theta) P_\ell[\beta \cos(\theta)]. \quad (16.5.278)$$

With the aid of this definition we may rewrite  $K_n$  in the form

$$K_n = \sum_{\ell=0}^{\infty} \{\rho/[a^2 + (z - z')^2]^{1/2}\}^\ell c_{n\ell}. \quad (16.5.279)$$

Like the analogous coefficients in Section 5, It can be shown show that these coefficients have the properties

$$\begin{aligned} c_{n\ell} &= 0 \text{ for } \ell < n; c_{nn} \neq 0 \text{ and depends only on } \beta; \text{ and } c_{n\ell} = 0 \text{ for } \ell = n + 1, \\ &\text{in fact } c_{n\ell} = 0 \text{ if } n \text{ and } \ell \text{ have different parity.} \end{aligned} \quad (16.5.280)$$

Assuming this to be true, we may rewrite (3.192) in the form

$$K_n = \{\rho/[a^2 + (z - z')^2]^{1/2}\}^n c_{nn} + \sum_{\ell=n+2}^{\infty \text{ by } 2s} \{\rho/[a^2 + (z - z')^2]^{1/2}\}^\ell c_{n\ell}. \quad (16.5.281)$$

Here the text “ $\infty$  by 2s” over the summation sign means that  $\ell$  is supposed to range from  $n + 2$  to  $\infty$  in increments of 2 in accord with (3.193). Observe that the first term on the right side of (3.194) is proportional to  $\rho^n$  and the remaining terms go as  $\rho^{n+2}$  and still higher powers that successively increment by factors of  $\rho^2$ . Thus we may write

$$K_n = \rho^n [a^2 + (z - z')^2]^{-n/2} c_{nn} + O[\rho^{n+2}]. \quad (16.5.282)$$

The  $n$  value determines the lowest power in  $\rho$  that can occur, and the successive higher powers that increment by factors of  $\rho^2$ . That  $K_n$  has such a structure is a good omen, because we observe from (3.125) that  $\psi_m$  has a similar structure.

Let us continue on to evaluate  $K_1$ , which is what is needed for the problem at hand. When  $n = 1$  (4.87) becomes

$$K_1 = \rho [a^2 + (z - z')^2]^{-1/2} c_{11} + O[\rho^3]. \quad (16.5.283)$$

The Legendre polynomial  $P_1$  is simply

$$P_1(w) = w. \quad (16.5.284)$$

Consequently  $c_{11}$  has the value

$$c_{11} = \int_{-\pi}^{+\pi} d\theta \cos(\theta) \beta \cos(\theta) = \beta \int_{-\pi}^{+\pi} d\theta \cos^2(\theta) = \pi\beta. \quad (16.5.285)$$

Correspondingly, looking at (4.77), one sees that

$$K_1 = \pi a \rho [a^2 + (z - z')^2]^{-1} + O[\rho^3]. \quad (16.5.286)$$

Finally, putting all the results \* together, and recalling that

$$\rho \cos(\phi) = x, \quad (16.5.287)$$

yields the net result

$${}^0 A_y(\mathbf{r}) = [\mu_0/4] a^2 x \int_{-\infty}^{\infty} dz' g(z') [a^2 + (z - z')^2]^{-3/2} + O[\rho^3]. \quad (16.5.288)$$

Similarly, it can be shown that

$${}^0 A_x(\mathbf{r}) = -[\mu_0/4] a^2 y \int_{-\infty}^{\infty} dz' g(z') [a^2 + (z - z')^2]^{-3/2} + O[\rho^3]. \quad (16.5.289)$$

Finally, we can compute  ${}^0 B_z$  using (6.18). So doing yields the result

$${}^0 B_z = \partial_x {}^0 A_y - \partial_y {}^0 A_x = \mu_0 (a^2/2) \int_{-\infty}^{\infty} dz' g(z') [a^2 + (z - z')^2]^{-3/2} + O[\rho^2]. \quad (16.5.290)$$

### Net result for $C_{0,c}^{[1]}(z)$

We are ready for the master stroke. Upon comparing (4.25) and \* we see that there must be the relation

$$C_{0,c}^{[1]}(z) = * \int_{-\infty}^{\infty} dz' g(z') [a^2 + (z - z')^2]^{-3/2}. \quad (16.5.291)$$

We have, for the  $m = 0$  case, found the on-axis gradient  $C_{0,c}^{[1]}$  in terms of  $g$ , as desired. As a sanity check, we note that (6.48) agrees with (5.69).

### What is the Gauge of ${}^0\mathbf{A}$ ?

From \* we know that  ${}^0\mathbf{A}$  is in a Coulomb gauge. But is it in the *symmetric* Coulomb gauge? From \* we know that when  $m = 0$  the symmetric Coulomb gauge vector potential is given in terms of on-axis gradients by the relations

$$\hat{A}_x^0 = -(y/2) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! (\ell+1)!} C_{0,c}^{[2\ell+1]}(z) (x^2 + y^2)^{\ell}, \quad (16.5.292)$$

$$\hat{A}_y^0 = (x/2) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! (\ell+1)!} C_{0,c}^{[2\ell+1]}(z) (x^2 + y^2)^{\ell}, \quad (16.5.293)$$

$$\hat{A}_z^0 = 0. \quad (16.5.294)$$

Therefore there are the relations

$$\hat{A}_x^0 = -(y/2) C_{0,c}^{[1]}(z) + O[\rho^3], \quad (16.5.295)$$

$$\hat{A}_y^0 = (x/2) C_{0,c}^{[1]}(z) + O[\rho^3], \quad (16.5.296)$$

$$\hat{A}_z^0 = 0. \quad (16.5.297)$$

By employing (6.48) in (6.20), (6.46), and (6.47) we find that

$${}^0 A_x = -(y/2) C_{0,c}^{[1]}(z) + O[\rho^3], \quad (16.5.298)$$

$${}^0 A_y = (x/2) C_{0,c}^{[1]}(z) + O[\rho^3], \quad (16.5.299)$$

$${}^0 A_z = 0. \quad (16.5.300)$$

Upon comparing (6.55) through (6.57) with (6.52) through (6.54) we see that  ${}^0\mathbf{A}$  is the symmetric coulomb gauge save for possible terms of  $O[\rho^3]$ . We conjecture that it is *exactly* in the symmetric coulomb gauge. See Exercise 6.2.

### Application to a Cylinder for the Normal $m \geq 1$ Cases

For the the normal  $m \geq 1$  cases we need  ${}^m B_y$ , as in Section 5, and observe that

$${}^m B_y = \partial_z {}^m A_x - \partial_x {}^m A_z. \quad (16.5.301)$$

So we need to find  ${}^m A_x$  and  ${}^m A_z$ . Begin the simpler task of finding  ${}^m A_z$ . It involves  $j_z$  as given by (5.26),

$$\begin{aligned} {}^m A_z(\mathbf{r}) &= [\mu_0/(4\pi)] \int d^3 \mathbf{r}' J_z(\mathbf{r}') / \|\mathbf{r} - \mathbf{r}'\| \\ &= [\mu_0/(4\pi)] a \int_{-\infty}^{\infty} dz' \int_{-\pi}^{\pi} d\phi' j_z(\mathbf{r}') / \|\mathbf{r} - \mathbf{r}'\| \\ &= (-m/a)[\mu_0/(4\pi)] a \int_{-\infty}^{\infty} dz' f(z') \int_{-\pi}^{\pi} d\phi' \cos(m\phi') / \|\mathbf{r} - \mathbf{r}'\|. \end{aligned} \quad (16.5.302)$$

Concentrate on manipulating the angular integral in (6.61). As before use the result \* and make the change of variables \*. When this is done the angular integral becomes

$$\begin{aligned} &\int_{-\pi}^{\pi} d\phi' \cos(m\phi') / \|\mathbf{r} - \mathbf{r}'\| \\ &= \int_{-\pi}^{\pi} d\theta \cos[m(\phi + \theta)] [\rho^2 - 2a\rho \cos(\theta) + a^2 + (z - z')^2]^{-1/2}. \end{aligned} \quad (16.5.303)$$

Also again use \* so that we may write

$$\begin{aligned} &\int_{-\pi}^{\pi} d\theta \cos[m(\phi + \theta)] [\rho^2 - 2a\rho \cos(\theta) + a^2 + (z - z')^2]^{-1/2} \\ &= \cos(m\phi) \int_{-\pi}^{\pi} d\theta \cos(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + (z - z')^2]^{-1/2}. \end{aligned} \quad (16.5.304)$$

We have isolated the  $\phi$  dependence.

Let us next try to isolate some  $\rho$  dependence. Observe that the integral appearing on the right side of (6.63) is a generalization of the integral that appears on the left side of (6.42). Reflection on how this latter integral was handled reveals that the integral on the right side of (6.63) can be handled analogously to yield the result

$$\begin{aligned} &\int_{-\pi}^{\pi} d\theta \cos(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + (z - z')^2]^{-1/2} \\ &= [a^2 + (z - z')^2]^{-1/2} \int_{-\pi}^{\pi} d\theta \cos(m\theta) [1 - 2wh + h^2]^{-1/2}. \end{aligned} \quad (16.5.305)$$

Observe further that the integral appearing on the the right side of (6.64) is the same as the integral (6.43), which we have already handled. Reflecting on how it was handled yields the conclusion

$$\begin{aligned} &\int_{-\pi}^{\pi} d\theta \cos(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + (z - z')^2]^{-1/2} \\ &= [a^2 + (z - z')^2]^{-1/2} \int_{-\pi}^{\pi} d\theta \cos(m\theta) [1 - 2wh + h^2]^{-1/2} \\ &= [a^2 + (z - z')^2]^{-1/2} K_m \end{aligned} \quad (16.5.306)$$

with

$$\begin{aligned} K_m &= \{\rho/[a^2 + (z - z')^2]^{1/2}\}^m c_{mm} + \sum_{\ell=m+2}^{\infty \text{ by } 2s} \{\rho/[a^2 + (z - z')^2]^{1/2}\}^\ell c_{m\ell} \\ &= \rho^m [a^2 + (z - z')^2]^{-m/2} c_{mm} + O[\rho^{m+2}]. \end{aligned} \quad (16.5.307)$$

The net result of these observations is that

$$\begin{aligned} &\int_{-\pi}^{\pi} d\theta \cos(m\theta) [\rho^2 - 2a\rho \cos(\theta) + a^2 + (z - z')^2]^{-1/2} \\ &= \rho^m [a^2 + (z - z')^2]^{-(m+1)/2} c_{mm} + O[\rho^{m+2}], \end{aligned} \quad (16.5.308)$$

and

$$\begin{aligned} {}^m A_z(\mathbf{r}) &= \\ &-m[\mu_0/(4\pi)][\rho^m \cos(m\phi)] \int_{-\infty}^{\infty} dz' f(z') [a^2 + (z - z')^2]^{-(m+1)/2} c_{mm} \\ &+ O[\rho^{m+2}]. \end{aligned} \quad (16.5.309)$$

We have isolated the  $\phi$  and leading  $\rho$  behavior.

Upon examining (6.66) we see what remains is to work out  $c_{mm}$  as given by (6.34) in the Legendre polynomial case. Fortunately we have already solved this problem in Section 15.13. There, see (15.13.18), we found

$$\begin{aligned} c_{m,m} &= \{[(2m)!]/[2^m(m!)^2]\}(1/2)^{m-1}\beta^m \int_{-\pi}^{\pi} d\theta \cos^2(m\theta) \\ &= \pi\{[(2m)!]/[2^{2m-1}(m!)^2]\}\beta^m \end{aligned} \quad (16.5.310)$$

where, for our present circumstances,  $\beta$  is given by (6.29). Combining  ${}^*, {}^*$ , and  ${}^*$  gives the result

$$\begin{aligned} &[a^2 + (z - z')^2]^{-(m+1)/2} c_{mm} \\ &= [a^2 + (z - z')^2]^{-(m+1)/2} \pi\{[(2m)!]/[2^{2m-1}(m!)^2]\}\beta^m \\ &= \pi\{[(2m)!]/[2^{2m-1}(m!)^2]\}[a^2 + (z - z')^2]^{-(m+1)/2}\beta^m. \end{aligned} \quad (16.5.311)$$

But, from  ${}^*$ , we also have

$$\begin{aligned} &[a^2 + (z - z')^2]^{-(m+1)/2}\beta^m \\ &= [a^2 + (z - z')^2]^{-(m+1)/2} \{a/[a^2 + (z - z')^2]^{1/2}\}^m \\ &= a^m [a^2 + (z - z')^2]^{-[m+(1/2)]}. \end{aligned} \quad (16.5.312)$$

Consequently,

$$\begin{aligned} &[a^2 + (z - z')^2]^{-(m+1)/2} c_{mm} \\ &= \pi\{[(2m)!]/[2^{2m-1}(m!)^2]\}[a^2 + (z - z')^2]^{-(m+1)/2}\beta^m \\ &= \pi\{[(2m)!]/[2^{2m-1}(m!)^2]\}a^m [a^2 + (z - z')^2]^{-[m+(1/2)]}. \end{aligned} \quad (16.5.313)$$

Upon combining \* and \* we find

$$\begin{aligned} {}^m A_z(\mathbf{r}) = & \\ & - * m[\mu_0/(4\pi)][\rho^m \cos(m\phi)] \int_{-\infty}^{\infty} dz' f(z') [a^2 + (z - z')^2]^{-(m+1/2)} \\ & + O[\rho^{m+2}]. \end{aligned} \quad (16.5.314)$$

Finally, from (6.66) and \*, we find that

$$\begin{aligned} -\partial_x {}^m A_z(\mathbf{r}) = & \\ & m^2 [\mu_0/(4\pi)] \rho^{m-1} \cos[(m-1)\phi] \int_{-\infty}^{\infty} dz' f(z') [a^2 + (z - z')^2]^{-(m+1)/2} c_{mm} \\ & + O[\rho^{m+1}], \end{aligned} \quad (16.5.315)$$

which is a needed ingredient for (6.58).

Enough work on  ${}^m A_z$  has been done for the time being. Let us now turn our attention to  ${}^m A_x$ . It involves  $j_x$  as given by (5.24), and use of \* yields the relation

$${}^m A_x(\mathbf{r}) = [\mu_0/(4\pi)] a \int_{-\infty}^{+\infty} dz' [a^2 + (z - z')^2]^{-1/2} \int_{-\pi}^{+\pi} d\phi' j_x(\phi', z') [1 - 2wh + h^2]^{-1/2}. \quad (16.5.316)$$

At this stage make the change of variables \*, which transforms  $j_x$  to become  $j_x^{tr}(\phi, z'; \theta)$ , so that (7.21) becomes

$${}^m A_x(\mathbf{r}) = [\mu_0/(4\pi)] a \int_{-\infty}^{+\infty} dz' [a^2 + (z - z')^2]^{-1/2} \int_{-\pi}^{+\pi} d\theta j_x^{tr}(\phi, z'; \theta) [1 - 2wh + h^2]^{-1/2}. \quad (16.5.317)$$

Before the change of variables  $j_x$  is given by

$$\begin{aligned} j_x &= -\sin(\phi') \sin(m\phi') f'(z') \\ &= -f'(z') \{(-1/2) \cos[(m+1)\phi'] + (1/2) \cos[(m-1)\phi']\}. \end{aligned} \quad (16.5.318)$$

After the change of variables  $j_x^{tr}$  is given by

$$\begin{aligned} j_x^{tr} &= j_x(\phi', z')|_{\phi'=\phi+\theta} \\ &= -f'(z') \{(-1/2) \cos[(m+1)(\phi+\theta)] + (1/2) \cos[(m-1)(\phi+\theta)]\} \\ &= (1/2) f'(z') \{\cos[(m+1)\phi] \cos[(m+1)\theta] - \sin[(m+1)\phi] \sin[(m+1)\theta]\} \\ &\quad (-1/2) f'(z') \{\cos[(m-1)\phi] \cos[(m-1)\theta] - \sin[(m-1)\phi] \sin[(m-1)\theta]\}. \end{aligned} \quad (16.5.319)$$

Here we have used \* and \*.

Also, observe that  $j_x^{tr}$  can be decomposed into *even* and *odd* parts with respect to its  $\theta$  dependence,

$$j_x^{tr} = j_x^{tre} + j_x^{tro}. \quad (16.5.320)$$

Moreover, we see from \* that  $[1 - 2wh + h^2]$  is an *even* function of  $\theta$ , and therefore only the even part of  $j_x^{tr}$  contributes to \* so that it can be rewritten in the form

$${}^m A_x(\mathbf{r}) = [\mu_0/(4\pi)]a \int_{-\infty}^{+\infty} dz' [a^2 + (z - z')^2]^{-1/2} \int_{-\pi}^{+\pi} d\theta j_x^{tre}(\phi, z'; \theta) [1 - 2wh + h^2]^{-1/2}. \quad (16.5.321)$$

From \* we find the results

$$j_x^{tre} = (1/2)f'(z')\{\cos[(m+1)\phi]\cos[(m+1)\theta] - \cos[(m-1)\phi]\cos[(m-1)\theta]\}. \quad (16.5.322)$$

Define  $\check{j}_x$  to be the portion of  $j_x^{tre}$  that has the smaller angular frequency in  $\theta$ , the term proportional to  $\cos[(m-1)\theta]$ . [As in Section 5, when the time comes, we will want to isolate from  ${}^m A_x(\mathbf{r})$  the term with the lowest order  $\rho$  dependence, and the power of  $\rho$  is governed by the angular frequency in  $\theta$ .] It is given by

$$\check{j}_x = -(1/2)f'(z')\cos[(m-1)\phi]\cos[(m-1)\theta]. \quad (16.5.323)$$

Let  ${}^m \check{A}_x(\mathbf{r})$  be the associated vector potential component produced by  $\check{j}_x$ . It is defined by the relation

$${}^m \check{A}_x(\mathbf{r}) = [\mu_0/(4\pi)]a \int_{-\infty}^{+\infty} dz' [a^2 + (z - z')^2]^{-1/2} \int_{-\pi}^{+\pi} d\theta \check{j}_x(\phi, z'; \theta) [1 - 2wh + h^2]^{-1/2}. \quad (16.5.324)$$

Employing \* in \* yields the result

$$\begin{aligned} {}^m \check{A}_x(\mathbf{r}) &= [\mu_0/(4\pi)](a/2)\cos[(m-1)\phi] \\ &\quad \int_{-\infty}^{+\infty} dz' [a^2 + (z - z')^2]^{-1/2} f'(z') \\ &\quad \int_{-\pi}^{+\pi} d\theta \cos[(m-1)\theta] [1 - 2wh + h^2]^{-1/2}. \end{aligned} \quad (16.5.325)$$

We have isolated the  $\phi$  dependence of  ${}^m \check{A}_x(\mathbf{r})$ .

Let us next try to isolate some  $\rho$  dependence. We have seen a variant of the integral on the far right side of (6.82) before. Indeed, from (6.31) we see that

$$\int_{-\pi}^{+\pi} d\theta \cos[(m-1)\theta] [1 - 2wh + h^2]^{-1/2} = K_{m-1}. \quad (16.5.326)$$

Moreover, it follows from (6.54) that there is the relation

$$K_{m-1} = \rho^{m-1} [a^2 + (z - z')^2]^{-(m-1)/2} c_{m-1,m-1} + O[\rho^{m+1}]. \quad (16.5.327)$$

Upon combining \* and \* we see that

$$\begin{aligned} {}^m \check{A}_x(\mathbf{r})(\mathbf{r}) &= \\ &[\mu_0/(4\pi)](a/2)\rho^{m-1} \cos[(m-1)\phi] \int_{-\infty}^{+\infty} dz' [a^2 + (z - z')^2]^{-m/2} c_{m-1,m-1} f'(z') \\ &+ O[\rho^{m+1}]. \end{aligned} \quad (16.5.328)$$

We have isolated the  $\phi$  dependence and leading  $\rho$  behavior in  ${}^m\check{A}_x(\mathbf{r})$ . Now comes an important observation: Had we included as well in our calculation of  ${}^m\check{A}_x(\mathbf{r})$  the higher angular frequency term  $\cos[(m+1)\theta]$  in (6.79), its contribution would be  $O[\rho^{m+1}]$ . Therefore there is the result

$${}^m\check{A}_x(\mathbf{r}) = {}^m\check{A}_x(\mathbf{r}) + O[\rho^{m+1}] \quad (16.5.329)$$

It follows from \* and \* that

$$\begin{aligned} {}^m\check{A}_x(\mathbf{r}) &= \\ &[\mu_0/(4\pi)](a/2)\rho^{m-1}\cos[(m-1)\phi]\int_{-\infty}^{+\infty} dz'[a^2 + (z - z')^2]^{-m/2}c_{m-1,m-1}f'(z') \\ &+ O[\rho^{m+1}]. \end{aligned} \quad (16.5.330)$$

Upon examining (6.87) we see what remains is to work out  $c_{m-1,m-1}$ . From \* we see that it is given by

$$c_{m-1,m-1} = \pi\{(2m-2)!/[2^{2m-3}[(m-1)!]^2]\}\beta^{m-1}. \quad (16.5.331)$$

It follows that

$$[a^2 + (z - z')^2]^{-m/2}c_{m-1,m-1} = . \quad (16.5.332)$$

Consequently,

$$\begin{aligned} {}^m\check{A}_x(\mathbf{r}) &= \\ &[*\rho^{m-1}\cos[(m-1)\phi]\int_{-\infty}^{+\infty} dz'[a^2 + (z - z')^2]^{-[m-(1/2)]}f'(z')] \\ &+ O[\rho^{m+1}]. \end{aligned} \quad (16.5.333)$$

Similarly, it can be shown that

$$\begin{aligned} {}^m\check{A}_y(\mathbf{r}) &= \\ &[*\rho^{m-1}\cos[(m-1)\phi]\int_{-\infty}^{+\infty} dz'[a^2 + (z - z')^2]^{-[m-(1/2)]}f'(z')] \\ &+ O[\rho^{m+1}]. \end{aligned} \quad (16.5.334)$$

$$\begin{aligned} \partial_z^m A_x(\mathbf{r}) &= \\ &[*\rho^{m-1}\cos[(m-1)\phi]\int_{-\infty}^{+\infty} dz'[a^2 + (z - z')^2]^{-[m+(1/2)]}(z - z')f'(z')] \\ &+ O[\rho^{m+1}]. \end{aligned} \quad (16.5.335)$$

Compare this result with

$$\begin{aligned} C_{m,s}^{[0]}(z) &= \mu_0(1/m)(1/2)^{m+1}\gamma_{nn}a^m \\ &\int_{-\infty}^{+\infty} dz'\{mf(z') + f'(z')(z - z')\}[a^2 + (z - z')^2]^{-[m+(1/2)]}. \end{aligned} \quad (16.5.336)$$

and

$${}^m B_y = C_{m,s}^{[0]}(z) m \rho^{m-1} \cos[(m-1)\phi] + O[\rho^{m+1}].$$

\*\*\*\*\*

### What is the Gauge of ${}^m A$ when $m \geq 1$ ?

From \* we know that  ${}^m A$  is in a Coulomb gauge,. But is it in the *symmetric* Coulomb gauge? From \* we know that when  $m \geq 1$  the symmetric Coulomb gauge vector potential is given in terms of on-axis gradients by the relations

$$\begin{aligned} \hat{A}_x^{m,s} &= (1/2)[(\cos \phi)(\cos m\phi) - (\sin \phi)(\sin m\phi)] \times \\ &\quad \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z) \rho^{2\ell+m+1} \\ &= (1/2) \cos[(m+1)\phi] \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z) \rho^{2\ell+m+1} \\ &= (1/2) \Re[(x+iy)^{m+1}] \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z) (x^2+y^2)^{\ell}, \end{aligned} \tag{16.5.337}$$

$$\begin{aligned} \hat{A}_y^{m,s} &= (1/2)[(\sin \phi)(\cos m\phi) + (\cos \phi)(\sin m\phi)] \times \\ &\quad \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z) \rho^{2\ell+m+1} \\ &= (1/2) \sin[(m+1)\phi] \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z) \rho^{2\ell+m+1} \\ &= (1/2) \Im[(x+iy)^{m+1}] \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z) (x^2+y^2)^{\ell}, \end{aligned} \tag{16.5.338}$$

$$\begin{aligned} \hat{A}_z^{m,s} &= -\cos(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,s}^{[2\ell]}(z) \rho^{2\ell+m} \\ &= -\Re[(x+iy)^m] \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,s}^{[2\ell]}(z) (x^2+y^2)^{\ell}. \end{aligned} \tag{16.5.339}$$

Therefore there are the relations

$$\hat{A}_x^{m,s} = (1/2) \rho^{m+1} \cos[(m+1)\phi] [1/(m+1)] C_{m,s}^{[1]}(z) + O[\rho^{m+3}], \tag{16.5.340}$$

$$\hat{A}_y^{m,s} = (1/2)\rho^{m+1} \sin[(m+1)\phi][1/(m+1)]C_{m,s}^{[1]}(z) + O[\rho^{m+3}], \quad (16.5.341)$$

$$\hat{A}_z^{m,s} = -\rho^m \cos(m\phi) C_{m,s}^{[0]}(z) + O[\rho^{m+2}]. \quad (16.5.342)$$

\*\*\*\*\*

$$\begin{aligned} j_y &= \cos(\phi') \sin(m\phi') f'(z') \\ &= f'(z') \{(1/2) \sin[(m+1)\phi'] + (1/2) \sin[(m-1)\phi']\}, \end{aligned} \quad (16.5.343)$$

$$j_z = -(m/a) \cos(m\phi') f(z'). \quad (16.5.344)$$

$$\begin{aligned}
j_y^{tr} &= j_y(\phi', z')|_{\phi'=\phi+\theta} \\
&= f'(z')\{(1/2)\sin[(m+1)(\phi+\theta)] + (1/2)\sin[(m-1)(\phi+\theta)]\} \\
&= (1/2)f'(z')\{\sin[(m+1)\phi]\cos[(m+1)\theta] + \cos[(m+1)\phi]\sin[(m+1)\theta] \\
&\quad + (1/2)f'(z')\{\sin[(m-1)\phi]\cos[(m-1)\theta] + \cos[(m-1)\phi]\sin[(m-1)\theta], \tag{16.5.345}
\end{aligned}$$

$$\begin{aligned} j_z^{tr} &= j_z(\phi', z')|_{\phi'=\phi+\theta} = -(m/a) \cos[m(\phi + \theta)] f(z') \\ &= -(m/a) f(z') [\cos(m\phi) \cos(m\theta) - \sin(m\phi) \sin(m\theta)]. \end{aligned} \quad (16.5.346)$$

$$j_y^{tre} = (1/2)f'(z')\{\sin[(m+1)\phi]\cos[(m+1)\theta]] + \sin[(m-1)\phi]\cos[(m-1)\theta]\}, \quad (16.5.347)$$

$$j_z^{tre} = -(m/a)f(z') \cos(m\phi) \cos(m\theta). \quad (16.5.348)$$

$$\hat{j}_y = (1/2)f'(z') \sin[(m-1)\phi] \cos[(m-1)\theta]\}, \quad (16.5.349)$$

$$\hat{j}_{\tilde{z}} = 0. \quad (16.5.350)$$

7777777777777\*\*\*\*\*  
\*\*\*\*\*

## Field Fall Off

Our discussion of dipole fringe fields so far has treated the iron-free case, and we have found a  $1/|z|^3$  fall off in all cases susceptible to easy analysis. When iron is present, and the coils are buried in iron or field clamps are employed, the fall off can in principle be much faster including the possibility of essentially exponential fall off.

## Coil Connectors, Wiring, and Power Supplies

## Exercises

### 16.5.1. Verify (6.46).

**16.5.2.** The aim of this exercise is to verify that  ${}^0\mathbf{A}$  is in the symmetric Coulomb gauge.

## 16.6 Rare Earth Cobalt (REC) Pure Multipoles

### 16.6.1 Description of Resulting On-Axis Gradient and On-Axis Field for a REC Quadrupole

A rare earth cobalt (REC) quadrupole typically has a circular annular cross section with outer radius  $r_2$  and inner/bore radius  $r_1$ . The space between  $r_1$  and  $r_2$  is filled with REC material magnetized and arranged so as to produce a pure quadrupole magnetic field within the bore. That is, ideally *only* the  $m = 2$  on-axis gradient is nonzero for the field produced by such an arrangement of REC material.

It can be shown that the on-axis gradient can again be described in terms of a soft-edge bump function which we will call  $\text{bump}(z, r_1, r_2, L)$  where  $L$  is the quadrupole length. That is, the on-axis field gradient  $C_{2,s}^{[0]}(z)$  can be written in the form

$$C_{2,s}^{[0]}(z) = (Q/2)\text{bump}(z, r_1, r_2, L) \quad (16.6.1)$$

where  $Q$  is the strength of the REC quadrupole in the infinite length limit. Moreover, as before, the soft-edge bump function for a REC quadrupole can be written in terms of an associated approximating signum function in the form

$$\text{bump}(z, r_1, r_2, L) = [\text{sgn}(z, r_1, r_2) - \text{sgn}(z - L, r_1, r_2)]/2. \quad (16.6.2)$$

Finally, it can be shown that for the REC quadrupole the approximating signum function  $\text{sgn}(z, r_1, r_2)$  is given by the relation

$$\text{sgn}(z, r_1, r_2) = z[(r_1 + r_2)/(r_1 r_2)][(v_1 v_2)/(v_1 + v_2)][1 + (1/8)v_1 v_2(4 + v_1^2 + v_1 v_2 + v_2^2)] \quad (16.6.3)$$

where  $v_1$  and  $v_2$  are defined by the relations

$$v_1 = 1/\sqrt{1 + (z/r_1)^2}, \quad (16.6.4)$$

$$v_2 = 1/\sqrt{1 + (z/r_2)^2}. \quad (16.6.5)$$

It can be easily checked that the approximating signum function  $\text{sgn}(z, r_1, r_2)$  becomes the true signum function in the limit  $r_1 \rightarrow 0$ . See Exercise 5.4. For example, Figures 5.11 and 5.12 illustrate the behavior of this approximating signum function for two different values of  $r_1$  and fixed values of  $r_2$  and  $L$ .

It follows from (5.34) that the soft-edge bump function for a REC quadrupole also satisfies relations analogous to (1.15) through (1.18) and (5.23) remains true. Figures 5.13 and 5.14 illustrate the properties (1.15) through (1.17) for fixed values of  $r_2$  and  $L$  and two different values of the inner radius  $r_1$ . Evidently the REC soft-edge bump function becomes a hard-edge bump function in the limit  $r_1 \rightarrow 0$ . The inner radius  $r_1$  (as well as  $r_2$ ) plays the role of a characteristic length that controls the rate of fall off. The fringe-field region is large if  $r_1$  is large, and vanishes as  $r_1$  goes to zero. From (5.34) through (5.37) we find the asymptotic behaviors

$$\begin{aligned} \text{sgn}(z, r_1, r_2) &= 1 - (1/16)r_1 r_2[(r_1^5 - r_2^5)/(r_1 - r_2)](1/z)^6 + O(1/z^8) \\ &= 1 - (1/16)r_1 r_2(r_1^4 + r_1^3 r_2 + r_1^2 r_2^2 + r_1 r_2^3 + r_2^4)(1/z^6) + O(1/z^8) \\ &\quad \text{as } z \rightarrow \infty, \end{aligned} \quad (16.6.6)$$

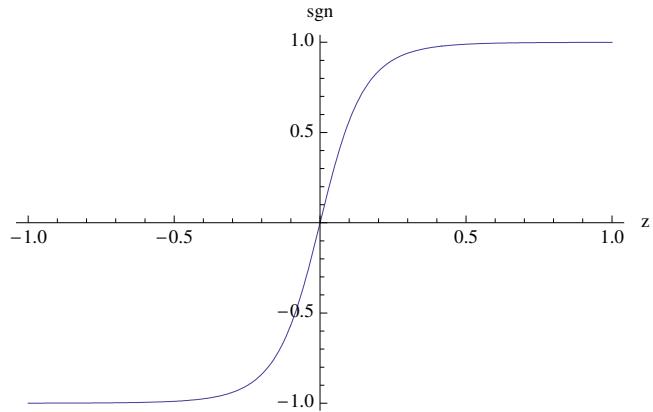


Figure 16.6.1: The approximating signum function (5.35) when  $r_1 = .2$  and  $r_2 = .5$ .

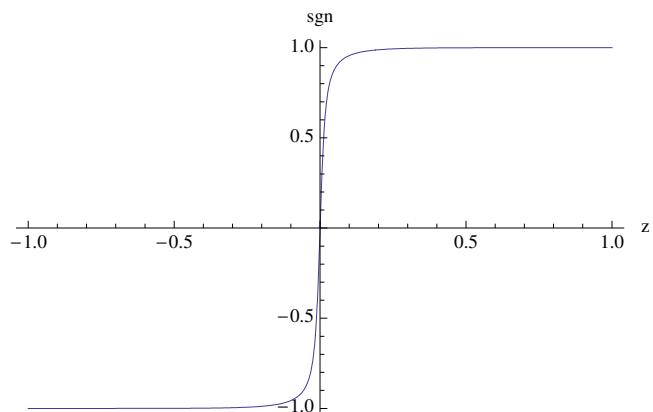


Figure 16.6.2: The approximating signum function (5.35) when  $r_1 = .02$  and  $r_2 = .5$ .

$$\begin{aligned}
\text{bump}(z, r_1, r_2, L) &= (3/16)Lr_1r_2[(r_1^5 - r_2^5)/(r_1 - r_2)](1/|z|^7) + O(1/|z|^8) \\
&= (3/16)Lr_1r_2(r_1^4 + r_1^3r_2 + r_1^2r_2^2 + r_1r_2^3 + r_2^4)(1/|z|^7) + O(1/|z|^8) \\
&\quad \text{as } |z| \rightarrow \infty.
\end{aligned} \tag{16.6.7}$$

Consequently  $C_{2,s}^{[0]}(z)$  falls off for large distances as

$$\begin{aligned}
C_{2,s}^{[0]}(z) &= (3/16)(Q/2)Lr_1r_2[(r_1^5 - r_2^5)/(r_1 - r_2)](1/|z|^7) + O(1/|z|^8) \\
&= (3/16)(Q/2)Lr_1r_2(r_1^4 + r_1^3r_2 + r_1^2r_2^2 + r_1r_2^3 + r_2^4)(1/|z|^7) + O(1/|z|^8).
\end{aligned} \tag{16.6.8}$$

We see that the fall off goes as  $1/|z|^7$ , which is pleasantly rapid. Remarkably, this rate of fall off for a REC quadrupole is two orders higher in  $1/|z|$  than that for an idealized air-core quadrupole. Compare (5.32) and (5.40).

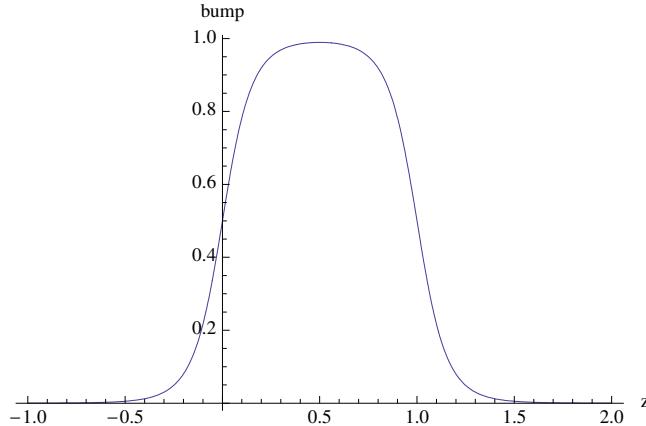


Figure 16.6.3: The soft-edge bump function (5.34) when  $r_1 = .2$ ,  $r_2 = .5$ , and  $L = 1$ .

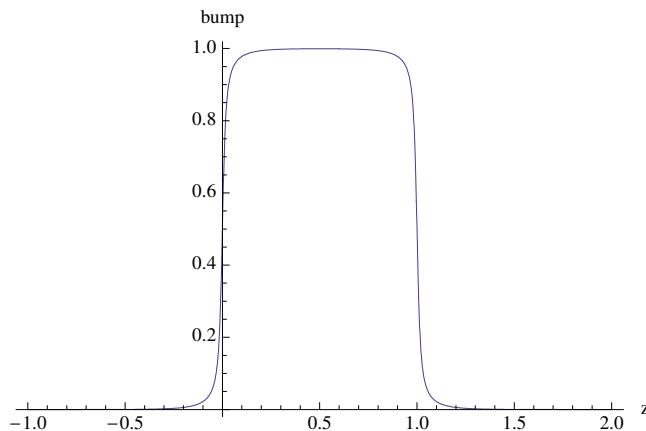


Figure 16.6.4: The soft-edge bump function (5.34) when  $r_1 = .02$ ,  $r_2 = .5$ , and  $L = 1$ .

### 16.6.2 Overlapping Fringe Fields

### 16.6.3 Hard-Edge Quadrupoles

### 16.6.4 Terminating Quadrupole End Fields

#### Preliminaries

For this subsection, as we did in Subsection 4.3, we will use the global coordinates  $\mathbf{R}$  and local coordinates  $\mathbf{r}$  as given by (4.8) and (4.9). Following previous notation, we assume the leading quadrupole fringe field begins at  $Z = Z^{\text{en}}$  and the trailing quadrupole fringe field ends at  $Z = Z^{\text{ex}}$ . In the interval  $[Z^{\text{en}}, Z^{\text{ex}}]$  the design orbit and map integrations will be carried out using the vector potential in the Coulomb gauge of Section 15.5. The entering transition at  $Z^{\text{en}}$  from the leading no-field region to the leading fringe-field region, and the exiting transition at  $Z^{\text{ex}}$  from the trailing fringe-field region to the trailing no-field region, will be made using the minimum vector potential, namely the Poincaré-Coulomb gauge vector potential.

If we wish to make the transition from the leading no-field region to the leading fringe-field region using the minimum vector potential (the vector potential in the Poincaré-Coulomb gauge), and also wish to carry out the design orbit and map integrations using the vector potential in the Coulomb gauge of Section 15.5.2, then we need to find at  $Z = Z^{\text{en}}$  the gauge transformation that relates the Poincaré-Coulomb gauge and the Coulomb gauge of Section 15.5.2. Similarly, If we wish to make the transition from the trailing fringe-field region to the trailing no-field region using the minimum vector potential (the vector potential in the Poincaré-Coulomb gauge), and also wish to carry out the design orbit and map integrations using the vector potential in the Coulomb gauge of Section 15.5.2, then we need to find at  $Z = Z^{\text{ex}}$  the gauge transformation that relates the Poincaré-Coulomb gauge and the Coulomb gauge of Section 15.5.2. For simplicity, we will continue to assume that we are dealing with the case of a *normal* quadrupole; namely  $\alpha = s$ .

For both transitions there is a relation of the form

$${}^P\mathbf{A}^{2,s}(x, y, z; Z^\beta) = \hat{\mathbf{A}}^{2,s}(x, y, Z^\beta + z) + \nabla \chi_{2,s} \quad (16.6.9)$$

where  $\beta = \text{en}$  or  $\beta = \text{ex}$ . See (15.7.5). There is also the relation

$$\begin{aligned} \chi_{2,s}(x, y, z; Z^\beta) &= \cos(2\phi) \sum_{k=0}^{\infty} (-1)^k \frac{2!}{2^{2k} k! (k+2)!} D_{2,s}^{[2k]}(z; Z^\beta) \rho^{2k+2} \\ &= \rho^2 \cos(2\phi) [D_{2,s}^{[0]}(z; Z^\beta) - (1/6)\rho^2 D_{2,s}^{[2]}(z; Z^\beta) + (*)\rho^4 D_{2,s}^{[4]}(z; Z^\beta) + \dots] \\ &= (x^2 - y^2) [D_{2,s}^{[0]}(z; Z^\beta) - (1/6)\rho^2 D_{2,s}^{[2]}(z; Z^\beta) + (*)\rho^4 D_{2,s}^{[4]}(z; Z^\beta) + \dots]. \end{aligned} \quad (16.6.10)$$

See (15.7.20). Moreover we have found, see (15.7.60), the relation

$$\begin{aligned} D_{2,s}^{[0]}(z; Z^\beta) &= \{[1/(2+1)]C_{2,s}^{[0]}(Z^\beta)\}z + \{[1/(2+2)]C_{2,s}^{[1]}(Z^\beta)\}z^2 \\ &\quad + \{[1/(2+3)](1/2!)C_{2,s}^{[2]}(Z^\beta)\}z^3 + \{[1/(2+4)](1/3!)C_{2,s}^{[3]}(Z^\beta)\}z^4 + \dots \\ &= (1/3)C_{2,s}^{[0]}(Z^\beta)z + (1/4)C_{2,s}^{[1]}(Z^\beta)z^2 + (1/10)C_{2,s}^{[2]}(Z^\beta)z^3 + (1/36)C_{2,s}^{[3]}(Z^\beta)z^4 + \dots, \end{aligned} \quad (16.6.11)$$

from which it follows that

$$D_{2,s}^{[0]}(0; Z^\beta) = 0, \quad (16.6.12)$$

$$D_{2,s}^{[2]}(0; Z^\beta) = (1/2)C_{2,s}^{[1]}(Z^\beta), \quad (16.6.13)$$

$$D_{2,s}^{[4]}(0; Z^\beta) = (2/3)C_{2,s}^{[3]}(Z^\beta). \quad (16.6.14)$$

Inserting these results in (5.42) gives the expansion

$$\begin{aligned} \chi_{2,s}(x, y, 0; Z^\beta) &= (x^2 - y^2)[(*)\rho^2 C_{2,s}^{[1]}(Z^\beta) + (*)\rho^4 C_{2,s}^{[3]}(Z^\beta) + \dots] \\ &= (x^4 - y^4)[(*)C_{2,s}^{[1]}(Z^\beta) + (*)\rho^2 C_{2,s}^{[3]}(Z^\beta) + \dots]. \end{aligned} \quad (16.6.15)$$

Finally there is the relation (4.12), which we repeat below:

$$\mathbf{A}^a - \mathbf{A}^b = \nabla \chi, \quad (16.6.16)$$

### Entering a Leading Fringe-Field Region

Compare (5.48) with the relation (5.41) evaluated for the case  $\beta = \text{en}$  and rewritten in the form

$$\hat{\mathbf{A}}^{2,s}(x, y, Z^{\text{en}} + z) - {}^P\mathbf{A}^{2,s}(x, y, z; Z^{\text{en}}) = -\nabla \chi_{2,s}. \quad (16.6.17)$$

We conclude that if we wish to identify  $\hat{\mathbf{A}}^{2,s}$  with  $\mathbf{A}^a$ , and identify  ${}^P\mathbf{A}^{2,s}$  with  $\mathbf{A}^b$ , then we should require the relation

$$\chi = -\chi_{2,s}. \quad (16.6.18)$$

Next define the function  $\chi^{\text{en}}$  by the rule

$$\chi^{\text{en}}(x, y; Z^{\text{en}}) = \chi(x, y, 0; ; Z^{\text{en}}). \quad (16.6.19)$$

With this definition we see from (5.47), (5.50), and (5.51) that there is the result

$$\chi^{\text{en}}(x, y; Z^{\text{en}}) = -(x^4 - y^4)[(*)C_{2,s}^{[1]}(Z^{\text{en}}) + (*)\rho^2 C_{2,s}^{[3]}(Z^{\text{en}}) + \dots]. \quad (16.6.20)$$

We are now ready to invoke the results (1.77) through (1.79). So doing, we find that the canonical coordinates  $(x, y, t; p_x^{\text{can}}, p_y^{\text{can}}, p_t^{\text{can}})$  after and before  $Z^{\text{en}}$  are connected by the symplectic map  $\mathcal{T}^{\text{en}}$ ,

$$\begin{aligned} x^a(Z) &= \mathcal{T}^{\text{en}} x^b(Z) \text{ with } Z = Z^{\text{en}}, \\ y^a(Z) &= \mathcal{T}^{\text{en}} y^b(Z) \text{ with } Z = Z^{\text{en}}, \\ t^a(Z) &= \mathcal{T}^{\text{en}} t^b(Z) \text{ with } Z = Z^{\text{en}}; \end{aligned} \quad (16.6.21)$$

$$\begin{aligned} p_x^{\text{cana}}(Z) &= \mathcal{T}^{\text{en}} p_x^{\text{canb}}(Z) \text{ with } Z = Z^{\text{en}}, \\ p_y^{\text{cana}}(Z) &= \mathcal{T}^{\text{en}} p_y^{\text{canb}}(Z) \text{ with } Z = Z^{\text{en}}, \\ p_t^{\text{cana}}(Z) &= \mathcal{T}^{\text{en}} p_t^{\text{canb}}(Z) \text{ with } Z = Z^{\text{en}}, \end{aligned} \quad (16.6.22)$$

where

$$\mathcal{T}^{\text{en}} = \exp(q : \chi^{\text{en}} :). \quad (16.6.23)$$

### Exiting a Trailing Fringe-Field Region

Compare (5.48) with the relation (5.41) evaluated for the case  $\beta = \mathbf{ex}$  and rewritten in the form

$${}^P\mathbf{A}^{2,s}(x, y, z) - \hat{\mathbf{A}}^{2,s}(x, y, Z^{\text{ex}} + z) = \nabla \chi_{2,s}. \quad (16.6.24)$$

We conclude that if we wish to identify  ${}^P\mathbf{A}^{2,s}$  with  $\mathbf{A}^a$ , and identify  $\hat{\mathbf{A}}^{2,s}$  with  $\mathbf{A}^b$ , then we should now require the relation

$$\chi = \chi_{2,s}. \quad (16.6.25)$$

Next define the function  $\chi^{\text{ex}}$  by the rule

$$\chi^{\text{ex}}(x, y) = \chi(x, y, 0). \quad (16.6.26)$$

With this definition we see from (5.47), (5.57), and (5.58) that there is the result

$$\chi^{\text{ex}}(x, y) = (x^4 - y^4)[(*)C_{2,s}^{[1]}(Z^{\text{ex}}) + (*)\rho^2 C_{2,s}^{[3]}(Z^{\text{ex}}) + \dots]. \quad (16.6.27)$$

We are again ready to invoke the results (1.77) through (1.79). So doing, we find that the canonical coordinates  $(x, y, t; p_x^{\text{can}}, p_y^{\text{can}}, p_t^{\text{can}})$  after and before  $Z^{\text{ex}}$  are connected by the symplectic map  $\mathcal{T}^{\text{ex}}$ ,

$$\begin{aligned} x^a(Z) &= \mathcal{T}^{\text{ex}} x^b(Z) \text{ with } Z = Z^{\text{ex}}, \\ y^a(Z) &= \mathcal{T}^{\text{ex}} y^b(Z) \text{ with } Z = Z^{\text{ex}}, \\ t^a(Z) &= \mathcal{T}^{\text{ex}} t^b(Z) \text{ with } Z = Z^{\text{ex}}; \end{aligned} \quad (16.6.28)$$

$$\begin{aligned} p_x^{\text{cana}}(Z) &= \mathcal{T}^{\text{ex}} p_x^{\text{canb}}(Z) \text{ with } Z = Z^{\text{ex}}, \\ p_y^{\text{cana}}(Z) &= \mathcal{T}^{\text{ex}} p_y^{\text{canb}}(Z) \text{ with } Z = Z^{\text{ex}}, \\ p_t^{\text{cana}}(Z) &= \mathcal{T}^{\text{ex}} p_t^{\text{canb}}(Z) \text{ with } Z = Z^{\text{ex}}, \end{aligned} \quad (16.6.29)$$

where

$$\mathcal{T}^{\text{ex}} = \exp(q : \chi^{\text{ex}} :). \quad (16.6.30)$$

### Behavior of $C_{2,s}^{[1]}(Z)$

According to (5.52) and (5.59) both  $\chi^{\text{en}}$  and  $\chi^{\text{ex}}$  involve  $C_{2,s}^{[1]}(Z)$  and its derivatives. Let us explore the behavior of  $C_{2,s}^{[1]}(Z)$  for the cases of idealized air-core quadrupoles and REC quadrupoles, which are described in Subsections 5.4 and 5.6, respectively. In both cases  $C_{2,s}^{[0]}(Z)$  is proportional to an associated bump function. See (5.27) and (5.33). Therefore, in both cases we are interested in the  $Z$  dependence of  $\text{bump}'$ , the derivative of the bump function. Figures 5.15 and 5.16 display the derivative of the soft-edge bump functions shown in Figures 5.9 and 5.10 for two idealized air-core quadrupoles; and Figures 5.17 and 5.18 display the derivative of the soft-edge bump functions shown in Figures 5.13 and 5.14 for two REC quadrupoles.

Evidently  $\text{bump}'$ , the derivative of the bump function, falls off quite rapidly beyond the quadrupole body. For example we expect, according to (5.31), that in the case of an idealized air-core quadrupole the function  $\text{bump}'$  will fall off like  $1/|z|^6$  as  $z \rightarrow -\infty$ . And, according to (5.39), we expect a fall off like  $1/|z|^8$  for the case of a REC quadrupole.

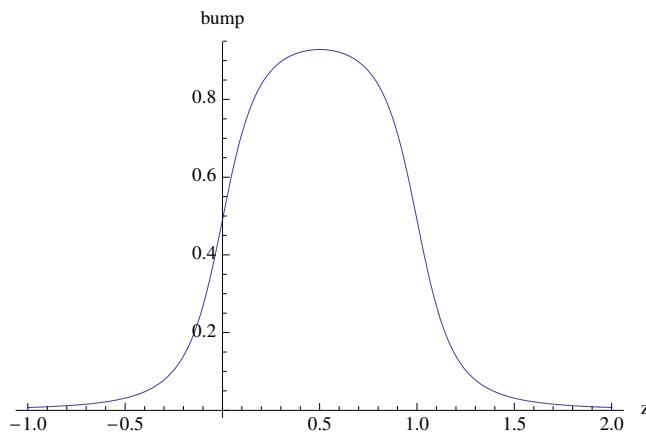


Figure 16.6.5: (Place Holder) Derivative of the soft-edge bump function given by (5.28) and (5.29) when  $a = .2$  and  $L = 1$ , and shown in Figure 5.9.

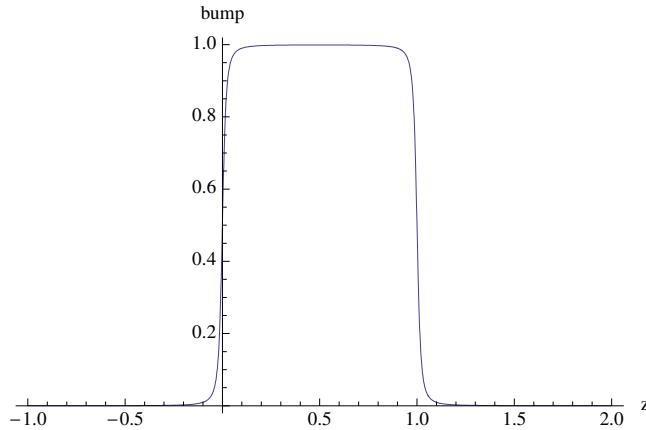


Figure 16.6.6: (Place Holder) Derivative of the soft-edge bump function given by (5.28) and (5.29) when  $a = .02$  and  $L = 1$ , and shown in Figure 5.10.

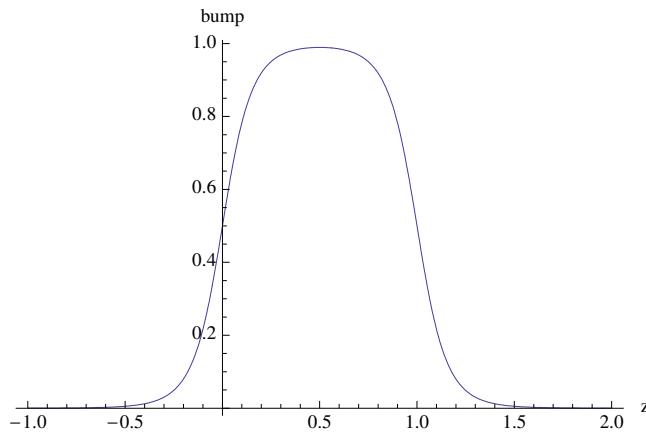


Figure 16.6.7: (Place Holder) Derivative of the soft-edge bump function (5.34) when  $r_1 = .2$ ,  $r_2 = .5$ , and  $L = 1$ , and shown in Figure 5.13.

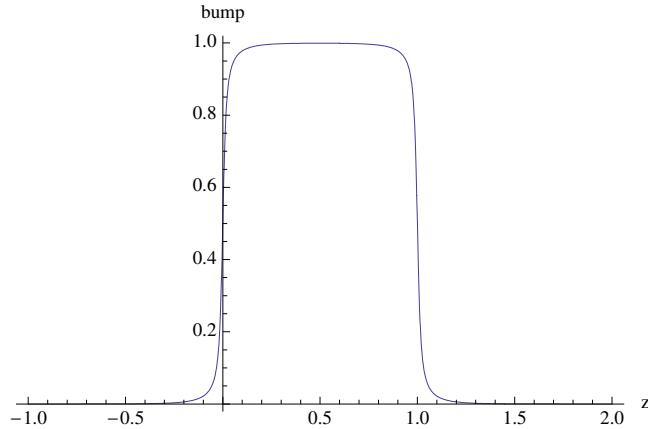


Figure 16.6.8: (Place Holder) Derivative of the soft-edge bump function (5.34) when  $r_1 = .02$ ,  $r_2 = .5$ , and  $L = 1$ , and shown in Figure 5.14.

### Net Total Map

Let  $\mathcal{M}_{\text{en} \rightarrow \text{ex}}$  denote the map obtained by integrating for a quadrupole the design orbit and map equations from  $Z = Z^{\text{en}}$  to  $Z = Z^{\text{ex}}$  using the Coulomb gauge vector potential of Section 15.5. Then the full net map  $\mathcal{M}$  for the quadrupole, including end-field termination effects, is given by the product

$$\mathcal{M} = \mathcal{T}^{\text{en}} \mathcal{M}_{\text{en} \rightarrow \text{ex}} \mathcal{T}^{\text{ex}}. \quad (16.6.31)$$

### Discontinuities in Mechanical Momenta Associated with Termination Approximation

As described in Subsection 1.2, there are discontinuities in the mechanical momenta associated with the use of a symplectic termination procedure. Recall (1.30), (1.31), (1.41), and (1.42). Here we will study for quadrupoles the consequences of terminating end fields using the minimum (Poincaré-Coulomb gauge) vector potential. To do so we will need the  $m = 2$  and  $\alpha = s$  Poincaré-Coulomb gauge vector potential about the expansion point  $(0, 0, Z_0)$ .

For fun let us compute  ${}^P\mathbf{A}^{2,s}(x, y, z)$  from scratch starting with  $\psi_{2,s}(x, y, Z_0 + z)$ . From (5.1) we have the result

$$\psi_{2,s}(x, y, Z_0 + z) = 2xy[C_{2,s}^{[0]}(Z_0 + z) - (1/24)(x^2 + y^2)C_{2,s}^{[2]}(Z_0 + z) + \dots]. \quad (16.6.32)$$

Let  $\mathbf{B}(\mathbf{r}; Z_0)$  be the associated magnet field and employ the notation

$$\mathbf{B}(\mathbf{r}; Z_0) = \mathbf{B}(x, y, z; Z_0) = \mathbf{B}(x, y, Z_0 + z). \quad (16.6.33)$$

We then have the relations

$$B_x(x, y, z; Z_0) = \partial_x \psi_{2,s} = 2yC_{2,s}^{[0]}(Z_0 + z) - (1/12)(3x^2y + y^3)C_{2,s}^{[2]}(Z_0 + z) + \dots, \quad (16.6.34)$$

$$B_y(x, y, z; Z_0) = \partial_y \psi_{2,s} = 2xC_{2,s}^{[0]}(Z_0 + z) - (1/12)(x^3 + 3xy^2)C_{2,s}^{[2]}(Z_0 + z) + \dots, \quad (16.6.35)$$

$$B_z(x, y, z; Z_0) = \partial_z \psi_{2,s} = 2xy[C_{2,s}^{[1]}(Z_0 + z) - (1/24)(x^2 + y^2)C_{2,s}^{[3]}(Z_0 + z) + \dots]. \quad (16.6.36)$$

Next expand  $\mathbf{B}(x, y, Z_0 + z)$  in homogeneous polynomials by writing,

$$\mathbf{B}(\mathbf{r}; Z_0) = \mathbf{B}^1(\mathbf{r}; Z_0) + \mathbf{B}^2(\mathbf{r}; Z_0) + \mathbf{B}^3(\mathbf{r}; Z_0) + \dots. \quad (16.6.37)$$

From (5.65) through (5.67) we see that there are the relations

$$\mathbf{B}^1(\mathbf{r}; Z_0) = 2C_{2,s}^{[0]}(Z_0)(y\mathbf{e}_x + x\mathbf{e}_y), \quad (16.6.38)$$

$$\mathbf{B}^2(\mathbf{r}; Z_0) = 2C_{2,s}^{[1]}(Z_0)(yz\mathbf{e}_x + xz\mathbf{e}_y + xy\mathbf{e}_z). \quad (16.6.39)$$

Now we may use (15.2.111) to find the results

$$\begin{aligned} \mathbf{A}^2(\mathbf{r}; Z_0) &= -(1/3)[\mathbf{r} \times \mathbf{B}^1(\mathbf{r}; Z_0)] \\ &= (-2/3)C_{2,s}^{[0]}(Z_0)[-zx\mathbf{e}_x + zy\mathbf{e}_y + (x^2 - y^2)\mathbf{e}_z], \end{aligned} \quad (16.6.40)$$

$$\begin{aligned} \mathbf{A}^3(\mathbf{r}; Z_0) &= -(1/4)[\mathbf{r} \times \mathbf{B}^2(\mathbf{r}; Z_0)] \\ &= (-1/2)C_{2,s}^{[1]}(Z_0)[(xy^2 - xz^2)\mathbf{e}_x + (yz^2 - yx^2)\mathbf{e}_y + (zx^2 - zy^2)\mathbf{e}_z]. \end{aligned} \quad (16.6.41)$$

[Note that, in view of the relation (5.5), (5.72) agrees with (15.2.165), as it should.] Finally, we write

$${}^P\mathbf{A}^{2,s}(\mathbf{r}; Z_0) = \mathbf{A}^2(\mathbf{r}; Z_0) + \mathbf{A}^3(\mathbf{r}; Z_0) + \dots. \quad (16.6.42)$$

Let us find what this knowledge of  ${}^P\mathbf{A}^{2,s}$  entails for discontinuities in mechanical momenta. Observe that, according to (5.72) and (5.73), there are the results

$$\mathbf{A}^2(x, y, 0; Z_0) = 0, \quad (16.6.43)$$

$$\mathbf{A}^3(x, y, 0; Z_0) = (-1/2)C_{2,s}^{[1]}(Z_0)[(xy^2)\mathbf{e}_x + (-yx^2)\mathbf{e}_y], \quad (16.6.44)$$

so that

$${}^P\mathbf{A}^{2,s}(x, y, 0; Z_0) = (-1/2)C_{2,s}^{[1]}(Z_0)[(xy^2)\mathbf{e}_x + (-yx^2)\mathbf{e}_y] + \dots. \quad (16.6.45)$$

We conclude from \* that upon entry there are the discontinuity results

$$\begin{aligned} \Delta p_x^{\text{mech}} &= q[{}^P A_x^{2,s}(x, y, 0; Z^{\text{en}})] \\ &= q\{[-(1/2)xy^2]C_{2,s}^{[1]}(Z^{\text{en}}) + \dots\}, \end{aligned} \quad (16.6.46)$$

$$\begin{aligned} \Delta p_y^{\text{mech}} &= q[{}^P A_y^{2,s}(x, y, 0; Z^{\text{en}})] \\ &= q\{[(1/2)x^2y]C_{2,s}^{[1]}(Z^{\text{en}}) + \dots\}. \end{aligned} \quad (16.6.47)$$

Similarly, upon exit, we find from \* the discontinuity results

$$\begin{aligned} \Delta p_x^{\text{mech}} &= q[{}^P A_x^{2,s}(x, y, 0; Z^{\text{ex}})] \\ &= q\{[-(1/2)xy^2]C_{2,s}^{[1]}(Z^{\text{ex}}) + \dots\}, \end{aligned} \quad (16.6.48)$$

$$\begin{aligned}\Delta p_y^{\text{mech}} &= q[P A_y^{2,s}(x, y, 0; Z^{\text{ex}})] \\ &= q\{(1/2)x^2y]C_{2,s}^{[1]}(Z^{\text{ex}}) + \dots\}.\end{aligned}\quad (16.6.49)$$

Recall the relations (5.5), (5.27), and (5.33). We see that in all cases the discontinuities are proportional to  $Q'(0, 0, Z)$ , or, equivalently  $\text{bump}'$ , and its derivatives at  $Z = Z^{\text{en}}$  or  $Z = Z^{\text{ex}}$ . We have already seen examples, in Figures 15 through 18, of how these functions behave (fall off) in the cases of idealized air-core and REC quadrupoles. Moreover, the discontinuities also vanish as the spatial deviations from the  $z$  axis (the design orbit) become small.

## 16.7 Lambertson Windings

## 16.8 Limited Utility of Cylindrical Harmonic Expansions for Dipoles

Strictly speaking, and as already alluded to in a previous footnote, cylindrical harmonic expansions for the field of a dipole are of limited use. First, there is this observation: If it is desired that the bore be much smaller than the length of the dipole, as is frequently the case, then the dipole must be bent to accommodate the design orbit. In this case a cylindrical harmonic analysis of the field is no longer possible.<sup>16</sup> Second, cylindrical harmonic expansions are expected to be valid (rapidly convergent) only in the vicinity of the  $z$  axis. But the orbit in a dipole is bent and therefore cannot be confined to the vicinity of the  $z$  axis unless the bend angle is suitably small. Thus there is a conflict between the desire to have a simple model of the design orbit accompanied by a practical dipole design (an essentially circular arc reasonably closely surrounded by coil windings/iron), and the desire for a simple model (cylindrical harmonic expansion) of the dipole field. This conflict occurs both for iron-free dipoles and dipoles with iron. Chapter 22, which does not presuppose a cylindrical harmonic expansion, treats the problem of finding realistic transfer maps for curved beam-line elements with significant sagitta.

At this point we pause to note that there is one area where this conflict does not occur, or at least may be less significant: the modeling of a wiggler/undulator which may be viewed as a string of short dipoles and for which the design orbit throughout the length of the element does not differ much from a straight line. In this case a cylindrical harmonic analysis of the field is appropriate and useful providing the amplitude of the wiggles in the design orbit is modest compared to the half gap of the dipoles. See Section 4 of this chapter. Finally, wigglers/undulators may be treated using the methods of Chapter 22 without the use of cylindrical harmonic expansions.

---

<sup>16</sup>Often, in the case of large storage rings/colliders, long superconducting dipoles with small bores are initially built as straight rectangular magnets with windings (essentially  $\cos \phi$ ) and iron designed in such a way as to produce as far as practical a pure  $m = 1$  field. These magnets are then *mechanically* bent to accommodate a curved design orbit, thereby producing a sector bend with normal entry and exit. The hope, partially verified by experience, is that the transfer map for such a bent dipole will not have unacceptable nonlinearities.

Let us return to the main discussion: instances where the conflict must be addressed. Consider the case of a straight (unbent) rectangular dipole of length  $L$  with equal entry and exit angles. Make the approximation that the design orbit is a circular arc within the dipole and straight lines outside the dipole in the fringe-field regions. Assume the bend angle is  $\theta$ .

The *sagitta*  $s$  of a circular arc of radius  $r$  and having a chord of length  $L$  is given by the relation

$$s = r - [r^2 - (L/2)^2]^{1/2}. \quad (16.8.1)$$

And, if the arc subtends an angle  $\theta$ , the chord length is given by the relation

$$L = 2r \sin \theta/2. \quad (16.8.2)$$

Therefore, in terms of  $L$  and  $\theta$ , the sagitta is given by the relation

$$s = \quad (16.8.3)$$

Then, by simple geometry, in order to *just accommodate* the design trajectory the dipole (without bending) must have a bore radius  $a$  given by the relation

$$a/L = (1 - \cos \theta/2)/(4 \sin \theta/2) = \theta/(16) + O(\theta^3). \quad (16.8.4)$$

See Exercise 3.8. Figure 3.14 displays the ratio  $a/L$  as a function of  $\theta$ .<sup>17</sup> Suppose, for example, that  $\theta = 9$  degrees and  $L = 1$  meter. Then use of (3.44) gives the result  $a/L = .0098$  and therefore  $a = .98$  cm.

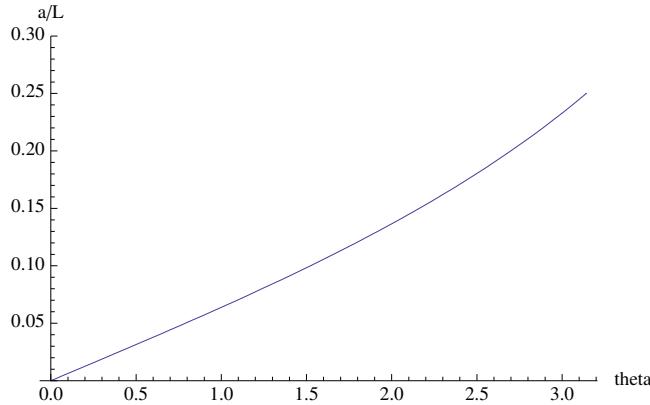


Figure 16.8.1: The ratio  $a/L$  as a function of  $\theta$ .

Next, beyond the simple geometric considerations we have just explored, we must acknowledge that a cylindrical harmonic expansion is expected to be valid only in the vicinity of the  $z$  axis. Suppose we assume that the actual bore should be at least twice the geometrically needed bore in order for the cylindrical harmonic expansion to be reasonably accurate in the region traversed by the design orbit. In this case we should, say for convenience, make the Ansatz  $a = 2$  cm = .02 meters. When  $a = .02$  and  $L = 1$  the field profile is that of

<sup>17</sup>Note that, for an unbent dipole with fixed design-orbit bend angle, the bore is proportional to the length.

Figure 3.13, from which we see that the fringe-field region appears to be relatively small. Specifically, from (3.38) and (3.39), we find the result

$$\text{bump}(z = -.3, a = .02, L = 1) = \text{bump}(z = 1.3, a = .02, L = 1) = -1.04 \times 10^{-3}. \quad (16.8.5)$$

Consequently, in this case one must be a distance of about 30 cm from the ends of the dipole for the on-axis field value to fall to  $10^{-3}$  of its central on-axis value. Thus, in this case and with a  $10^{-3}$  fall-off criterion, the fringe-field region on either end of the dipole is about 1/3 the length of the dipole.

Evidently, a more detailed analysis of this case would involve numerical integration to determine the design orbit accurately. And computation of the transfer map about this design orbit would require integration of the map equations of Section 10.5 using a Hamiltonian based on the vector potential given by (3.6) through (3.8). Note that in this case, according to (3.5) and (3.37) there are the relations

$$C_{1,s}^{[1]}(z) = B \text{ bump}'(z, a, L), \quad (16.8.6)$$

$$C_{1,s}^{[2]}(z) = B \text{ bump}''(z, a, L), \text{ etc.} \quad (16.8.7)$$

These functions are shown in Figures 3.15 and 3.16 below. Like its counterpart shown in Figure 2.7 for a solenoid of the same geometry, the function  $\text{bump}''$  for the ideal air-core dipole of Section 3.5 is quite singular in the case  $a = 0.02$  and  $L = 1$ . We may therefore expect that the transfer map for this ideal air-core dipole will have substantial higher-order aberrations.

### 16.8.1 Terminating Dipole End Fields

As already described, cylindrical harmonic expansions are of limited use for dipoles because the design orbit in a dipole is bent. Correspondingly, it is generally not useful to describe the termination of dipole end fields in terms of cylindrical harmonic expansions. For the special case of wigglers/undulators, where the use of cylindrical harmonic expansions may be appropriate for describing the termination of end fields, see Subsection 4.3. For a treatment of dipole end-field termination in the general case without the use of cylindrical harmonic expansions, see Section 22.8.

### 16.8.2 Limited Utility of Hard-Edge Models for Dipole Fringe Fields

## Exercises

**16.8.1.** Verify (3.44). The Large Hadron Collider (LHC) at CERN has 1232 dipole bending magnets. Each is 15 meters long and each has a circular 56 mm diameter bore. Can such a dipole accommodate the design orbit (assumed to be a circular arc) without being bent? If the dipole is bent so that the design orbit is centered in the beam pipe, can one look down the beam pipe as it enters a dipole and see the other end as it exits the dipole? That is, if a fine laser beam is pointed down the beam pipe as it enters a dipole, will the beam emerge

from the other end of the dipole? Verify that when the dipole is bent the arc, consisting of the design orbit, has a sagitta of 9 mm.

## 16.9 Air-Core Wiggler/Undulator Models

### 16.9.1 Simple Air-Core Wiggler/Undulator Model

The fields of individual monopole doublets or lines of monopole doublets or idealized air-core dipoles may be used to create model fields for wigglers/undulators. We will consider the simplest case where individual monopole doublet fields are employed.

A possible simple three-pole model of a wiggler/undulator may be taken to be a string of three equally spaced monopole doublets having relative strengths  $+1/2, -1, +1/2$ . That is we may define a three-pole wiggler/undulator profile function  $\text{wig}(3, z, a, L)$  by the rule

$$\text{wig}(3, z, a, L) = (1/2)\delta(z + L, a) - \delta(z, a) + (1/2)\delta(z - L, a) \quad (16.9.1)$$

where  $2L$  is the wiggler/undulator period. Here  $\delta(z, a)$  is the approximating delta function given by (3.13). The sum of the pole strengths is zero so that the wiggler/undulator produces no net bending, and the end poles are given half strengths so that the wiggler/undulator produces no net translation in  $x$ . For this profile there is the asymptotic fall off

$$\text{wig}(3, z, a, L) = (3a^2 L^2)/|z|^5 + O(1/|z|^6) \text{ as } |z| \rightarrow \infty. \quad (16.9.2)$$

Figure 4.1 displays the profile function  $\text{wig}(3, z, a, L)$  for the case  $a = .1$  and  $L = .5$ . Evidently in this case, as expected from (4.2), the fringe field falls off quite rapidly. For example, at a distance of one wiggler/undulator period from the end, there is the result

$$\text{wig}(3, 1.5, .1, .5)/\text{wig}(3, 0, .1, .5) = -2.6 \times 10^{-4}. \quad (16.9.3)$$

At this point  $C_{1,s}^{[0]}$  has fallen from its peak value by almost four orders of magnitude.

Another simple model is a string of five equally spaced monopole doublets having relative strengths  $+1/2, -1, +1, -1, +1/2$ . In this model we define a five-pole wiggler/undulator profile function  $\text{wig}(5, z, a, L)$  by the rule

$$\text{wig}(5, z, a, L) = (1/2)\delta(z + 2L, a) - \delta(z + L, a) + \delta(z, a) - \delta(z - L, a) + (1/2)\delta(z - 2L, a). \quad (16.9.4)$$

For this profile there is the asymptotic fall off

$$\text{wig}(5, z, a, L) = (6a^2 L^2)/|z|^5 + O(1/|z|^6) \text{ as } |z| \rightarrow \infty. \quad (16.9.5)$$

Note that the fall off for both the three-pole and five-pole wiggler/undulator goes as  $1/|z|^5$ , which is two orders higher in  $1/|z|$  than that for a single monopole doublet. Compare (3.19), (4.2), and (4.5). This higher fall-off rate arises from cancellations that occur between the doublets because the sum of the pole strengths is zero. That is, we have enforced the relation

$$\int_{-\infty}^{\infty} dz \text{wig}(n, z, a, L) = 0. \quad (16.9.6)$$

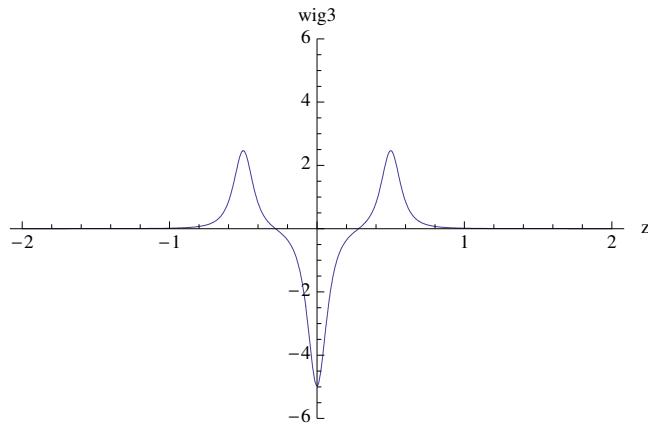


Figure 16.9.1: The three-pole wiggler/undulator profile function (4.1) when  $a = 0.1$  and  $L = 0.5$ .

Figure 4.2 displays the five-pole profile function for the case  $a = .1$  and  $L = .5$ . Evidently the fringe field again falls off quite rapidly. For example, at a distance of one wiggler/undulator period from the end, there is the result

$$\text{wig}(5, 2, .1, .5)/\text{wig}(5, 0, .1, .5) = 2.8 \times 10^{-4}. \quad (16.9.7)$$

At this point  $C_{1,s}^{[0]}$  has again fallen from its peak value by almost four orders of magnitude.

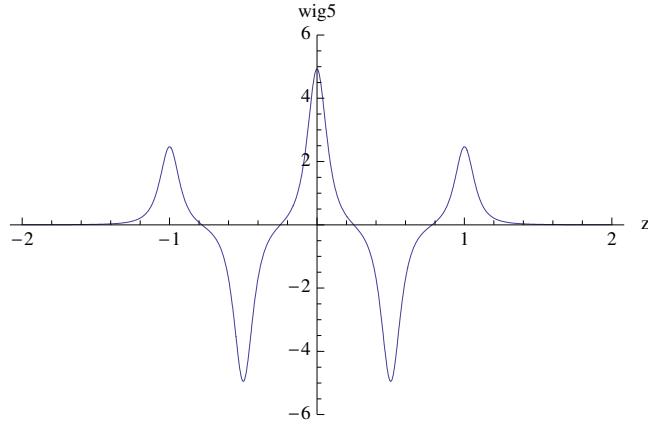


Figure 16.9.2: The five-pole wiggler/undulator profile function (4.4) when  $a = 0.1$  and  $L = 0.5$ .

## 16.9.2 Iron-Free Rare Earth Cobalt (REC) Wiggler/Undulator

### 16.9.3 Terminating Wiggler/Undulator End Fields

#### Preliminaries

There is an application for which expansions employing  $m = 1$  cylindrical harmonics may be useful, namely the case of wigglers/undulators when the excursion of the design orbit

from the axis may be treated as small. That is, it is assumed that the design orbit enters and exits the wiggler/undulator on axis and nearly along the axis, and the excursions of the design orbit from the axis while within the wiggler/undulator may be treated as small.

At this point it is necessary to make a change of notation. In the discussion of most of the preceding sections and subsections, save for that of Sections 15.2, 15.7 and 15.9, we have used the coordinates  $x, y, z$  as global coordinates. For this subsection, as described in Section 15.2, we will use

$$\mathbf{R} = (X, Y, Z) \quad (16.9.8)$$

as global coordinates and

$$\mathbf{r} = (x, y, z) \quad (16.9.9)$$

as local coordinates. In analogy with the notation of Subsection 1.2, we assume the leading wiggler/undulator fringe field begins at  $Z = Z^{\text{en}}$  and the trailing wiggler/undulator fringe field ends at  $Z = Z^{\text{ex}}$ . In the interval  $[Z^{\text{en}}, Z^{\text{ex}}]$  the design orbit and map integrations will be carried out using the vector potential in the Coulomb gauge of Section 15.5. The entering transition at  $Z^{\text{en}}$  from the leading no-field region to the leading fringe-field region, and the exiting transition at  $Z^{\text{ex}}$  from the trailing fringe-field region to the trailing no-field region, will be made using the minimum vector potential, namely the Poincaré-Coulomb gauge vector potential.

### Entering a Leading Fringe-Field Region

If we wish to make the transition from the leading no-field region to the leading fringe-field region using the minimum vector potential (the vector potential in the Poincaré-Coulomb gauge), and also wish to carry out the design orbit and map integrations using the vector potential in the Coulomb gauge of Section 15.5.2, then we need to find at  $Z = Z^{\text{en}}$  the gauge transformation that relates the Poincaré-Coulomb gauge and the Coulomb gauge of Section 15.5.2. The general problem of changing gauges has already been discussed in Subsection 1.3, the general relation between the Poincaré-Coulomb gauge vector potential and the Coulomb gauge vector potential of Section 15.5 has been described in Section 15.7, and the specific  $m = 1$  and  $\alpha = s$  relation has been treated in Section 15.9.2.

We seek relations in the vicinity of the point  $(0, 0, Z^{\text{en}})$ . Let us recapitulate some of what we have learned. In the vicinity of this point, and for the  $\alpha = s$  component of the magnetic field, the Poincaré-Coulomb gauge vector potential and the Coulomb gauge vector potential of Section 15.5 are related by the gauge transformation

$${}^P\mathbf{A}^{1,s}(x, y, z; Z^{\text{en}}) = \hat{\mathbf{A}}^{1,s}(x, y, Z^{\text{en}} + z) + \nabla\chi_{1,s}. \quad (16.9.10)$$

Recall (15.7.5). Also, we found that the gauge term  $\chi_{1,s}$  is related to the  $m = 1$  and  $\alpha = s$  on-axis gradient by the expansion

$$\chi_{1,s}(x, y, z; Z^{\text{en}}) = \{[(1/2)xz]C_{1,s}^{[0]}(Z^{\text{en}})\} + \{[(1/3)xz^2 - (1/12)x\rho^2]C_{1,s}^{[1]}(Z^{\text{en}})\} + \dots \quad (16.9.11)$$

See (15.9.21).

Next, see (1.75), we recall the relation

$$\mathbf{A}^a - \mathbf{A}^b = \nabla\chi, \quad (16.9.12)$$

and compare it with the relation (4.10) rewritten in the form

$$\hat{\mathbf{A}}^{1,s}(x, y, Z^{\text{en}} + z) - {}^P\mathbf{A}^{1,s}(x, y, z; Z^{\text{en}}) = -\nabla\chi_{1,s}. \quad (16.9.13)$$

We conclude that if we wish to identify  $\hat{\mathbf{A}}^{1,s}$  with  $\mathbf{A}^a$ , and identify  ${}^P\mathbf{A}^{1,s}$  with  $\mathbf{A}^b$ , then we should require the relation

$$\chi = -\chi_{1,s}. \quad (16.9.14)$$

Finally, define the function  $\chi^{\text{en}}$  by the rule

$$\chi^{\text{en}}(x, y; Z^{\text{en}}) = \chi(x, y, 0; Z^{\text{en}}). \quad (16.9.15)$$

With this definition we see from (4.11) and (4.14) that there is the result

$$\chi^{\text{en}}(x, y; Z^{\text{en}}) = \{[(1/12)x\rho^2]C_{1,s}^{[1]}(Z^{\text{en}})\} - \dots. \quad (16.9.16)$$

We are now ready to invoke the results (1.77) through (1.79). So doing, we find that the canonical coordinates  $(x, y, t; p_x^{\text{can}}, p_y^{\text{can}}, p_t^{\text{can}})$  *after* and *before*  $Z^{\text{en}}$  are connected by the symplectic map  $\mathcal{T}^{\text{en}}$ ,

$$\begin{aligned} x^a(Z) &= \mathcal{T}^{\text{en}}x^b(Z) \text{ with } Z = Z^{\text{en}}, \\ y^a(Z) &= \mathcal{T}^{\text{en}}y^b(Z) \text{ with } Z = Z^{\text{en}}, \\ t^a(Z) &= \mathcal{T}^{\text{en}}t^b(Z) \text{ with } Z = Z^{\text{en}}; \end{aligned} \quad (16.9.17)$$

$$\begin{aligned} p_x^{\text{cana}}(Z) &= \mathcal{T}^{\text{en}}p_x^{\text{canb}}(Z) \text{ with } Z = Z^{\text{en}}, \\ p_y^{\text{cana}}(Z) &= \mathcal{T}^{\text{en}}p_y^{\text{canb}}(Z) \text{ with } Z = Z^{\text{en}}, \\ p_t^{\text{cana}}(Z) &= \mathcal{T}^{\text{en}}p_t^{\text{canb}}(Z) \text{ with } Z = Z^{\text{en}}, \end{aligned} \quad (16.9.18)$$

where

$$\mathcal{T}^{\text{en}} = \exp(q : \chi^{\text{en}} :). \quad (16.9.19)$$

### Exiting a Trailing Fringe-Field Region

The general considerations for the transition associated with exiting a trailing fringe-field region are similar to those employed earlier for entering a leading fringe-field region. If we wish to make the transition from the trailing fringe-field region to the trailing no-field region using the minimum vector potential (the vector potential in the Poincaré-Coulomb gauge), and also wish to carry out the design orbit and map integrations using the vector potential in the Coulomb gauge of Section 15.5.2, then we need to find at  $Z = Z^{\text{ex}}$  the gauge transformation that relates the Poincaré-Coulomb gauge and the Coulomb gauge of Section 15.5.2.

We now seek relations in the vicinity of the point  $(0, 0, Z^{\text{ex}})$ . Let us again recapitulate some of what we have learned. In the vicinity of this point, and for the  $\alpha = s$  component of the magnetic field, the Poincaré-Coulomb gauge vector potential and the Coulomb gauge vector potential of Section 15.5 are related by the gauge transformation

$${}^P\mathbf{A}^{1,s}(x, y, z; Z^{\text{ex}}) = \hat{\mathbf{A}}^{1,s}(x, y, Z^{\text{ex}} + z) + \nabla\chi_{1,s}. \quad (16.9.20)$$

Also, the gauge term  $\chi_{1,s}$  is related to the  $m = 1$  and  $\alpha = s$  on-axis gradient by the expansion

$$\chi_{1,s}(x, y, z; Z^{\text{ex}}) = \{[(1/2)xz]C_{1,s}^{[0]}(Z^{\text{ex}})\} + \{[(1/3)xz^2 - (1/12)x\rho^2]C_{1,s}^{[1]}(Z^{\text{ex}})\} + \dots \quad (16.9.21)$$

Next we again recall the relation

$$\mathbf{A}^a - \mathbf{A}^b = \nabla\chi, \quad (16.9.22)$$

and compare it with the relation (4.20) rewritten in the form

$${}^P\mathbf{A}^{1,s}(x, y, z; Z^{\text{ex}}) - \hat{\mathbf{A}}^{1,s}(x, y, Z^{\text{ex}} + z) = \nabla\chi_{1,s}. \quad (16.9.23)$$

We conclude that if we wish to identify  ${}^P\mathbf{A}^{1,s}$  with  $\mathbf{A}^a$ , and identify  $\hat{\mathbf{A}}^{1,s}$  with  $\mathbf{A}^b$ , then we should now require the relation

$$\chi = \chi_{1,s}. \quad (16.9.24)$$

Finally, define the function  $\chi^{\text{ex}}$  by the rule

$$\chi^{\text{ex}}(x, y; Z^{\text{ex}}) = \chi(x, y, 0; Z^{\text{ex}}). \quad (16.9.25)$$

With this definition we see from (4.21) and (4.24) that there is the result

$$\chi^{\text{ex}}(x, y; Z^{\text{ex}}) = -\{[(1/12)x\rho^2]C_{1,s}^{[1]}(Z^{\text{ex}})\} + \dots \quad (16.9.26)$$

We are again ready to invoke the results (1.77) through (1.79). So doing, we find that the canonical coordinates  $(x, y, t; p_x^{\text{can}}, p_y^{\text{can}}, p_t^{\text{can}})$  after and before  $Z^{\text{ex}}$  are connected by the symplectic map  $\mathcal{T}^{\text{ex}}$ ,

$$\begin{aligned} x^a(Z) &= \mathcal{T}^{\text{ex}}x^b(Z) \text{ with } Z = Z^{\text{ex}}, \\ y^a(Z) &= \mathcal{T}^{\text{ex}}y^b(Z) \text{ with } Z = Z^{\text{ex}}, \\ t^a(Z) &= \mathcal{T}^{\text{ex}}t^b(Z) \text{ with } Z = Z^{\text{ex}}; \end{aligned} \quad (16.9.27)$$

$$\begin{aligned} p_x^{\text{cana}}(Z) &= \mathcal{T}^{\text{ex}}p_x^{\text{canb}}(Z) \text{ with } Z = Z^{\text{ex}}, \\ p_y^{\text{cana}}(Z) &= \mathcal{T}^{\text{ex}}p_y^{\text{canb}}(Z) \text{ with } Z = Z^{\text{ex}}, \\ p_t^{\text{cana}}(Z) &= \mathcal{T}^{\text{ex}}p_t^{\text{canb}}(Z) \text{ with } Z = Z^{\text{ex}}, \end{aligned} \quad (16.9.28)$$

where

$$\mathcal{T}^{\text{ex}} = \exp(q : \chi^{\text{ex}} :). \quad (16.9.29)$$

### Behavior of $C_{1,s}^{[1]}(Z)$

According to (4.16) and (4.26) both  $\chi^{\text{en}}$  and  $\chi^{\text{ex}}$  are proportional to  $C_{1,s}^{[1]}(Z)$ . Let us explore the behavior of  $C_{1,s}^{[1]}(Z)$  for the simplest wiggler/undulator models described in Subsection 4.1. In accord with (3.13), (4.1), and (4.4) we may write for these models

$$C_{1,s}^{[0]}(Z) = -(4g/a)\text{wig}(3, Z, a, L) \quad (16.9.30)$$

and

$$C_{1,s}^{[0]}(Z) = -(4g/a)\text{wig}(5, Z, a, L) \quad (16.9.31)$$

for the three-pole and five-pole cases, respectively. It follows that for the three-pole model there is the relation

$$C_{1,s}^{[1]}(Z) = -(4g/a)(\partial/\partial Z)\text{wig}(3, Z, a, L) = -(4g/a)\text{wig}'(3, Z, a, L), \quad (16.9.32)$$

and there is an analogous relation for the five-pole model. Figure 4.3 displays the profile function  $\text{wig}'(3, z, a, L)$  for the case  $a = 0.1$  and  $L = 0.5$ . Evidently  $\text{wig}'(3, z, a, L)$  falls off quite rapidly for large  $|z|$ . From (4.2) we expect the asymptotic fall off behavior to go as  $1/|z|^6$  for large  $|z|$ . For example, at a distance of one wiggler/undulator period from the end, there is the result

$$\text{wig}'(3, 1.5, .1, .5)/\text{wig}(3, 0, .1, .5) = *. \quad (16.9.33)$$

And, at a distance of two wiggler/undulator periods from the end, there is the result

$$\text{wig}'(3, 1.5, .1, .5)/\text{wig}(3, 0, .1, .5) = *. \quad (16.9.34)$$

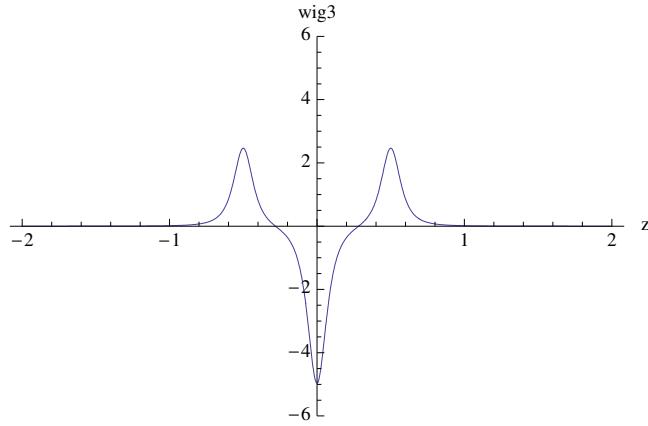


Figure 16.9.3: (Place Holder) The three-pole wiggler/undulator profile function  $\text{wig}'(3, z, a, L)$  when  $a = 0.1$  and  $L = 0.5$ .

### Net Total Map

Let  $\mathcal{M}_{\text{en} \rightarrow \text{ex}}$  denote the map obtained by integrating for a wiggler/undulator the design orbit and map equations from  $Z = Z^{\text{en}}$  to  $Z = Z^{\text{ex}}$  using the Coulomb gauge vector potential of Section 15.5. Then the full net map  $\mathcal{M}$  for the wiggler/undulator, including end-field termination effects, is given by the product

$$\mathcal{M} = \mathcal{T}^{\text{en}} \mathcal{M}_{\text{en} \rightarrow \text{ex}} \mathcal{T}^{\text{ex}}. \quad (16.9.35)$$

### Discontinuities in Mechanical Momenta Associated with Termination Approximation

As described in Subsection 1.2, there are discontinuities in the mechanical momenta associated with the use of a symplectic termination procedure. Recall (1.30), (1.31), (1.41), and (1.42). Here, for our simple wiggler/undulator models, we will study the consequences of terminating end fields using the minimum (Poincaré-Coulomb gauge) vector potential.

The  $m = 1$  and  $\alpha = s$  Poincaré-Coulomb gauge vector potential about the expansion point  $(0, 0, Z_0)$  is given in terms of on-axis gradients by (15.9.25) through (15.9.27). And, for our simple wiggler/undulator models, the on-axis gradients are given by relations of the form (4.30) through (4.32). Combining these relations with (1.30) and (1.31) gives, upon entry and for the 3-pole case, the discontinuity results

$$\begin{aligned}\Delta p_x^{\text{mech}} &= q[P A_x^{1,s}(x, y, 0; Z^{\text{en}})] \\ &= q\{[-(1/3)y^2]C_{1,s}^{[1]}(Z^{\text{en}}) + \dots\} \\ &= q\{[-(1/3)y^2](-4g/a)\text{wig}'(3, Z^{\text{en}}, a, L) + \dots\},\end{aligned}\tag{16.9.36}$$

$$\begin{aligned}\Delta p_y^{\text{mech}} &= q[P A_y^{1,s}(x, y, 0; Z^{\text{en}})] \\ &= q\{[(1/3)xy]C_{1,s}^{[1]}(Z^{\text{en}}) + \dots\} \\ &= q\{[(1/3)xy](-4g/a)\text{wig}'(3, Z^{\text{en}}, a, L) + \dots\}.\end{aligned}\tag{16.9.37}$$

Similarly, upon exit, we find from (15.9.25) through (15.9.27), (1.40), and (1.41) the discontinuity results

$$\begin{aligned}\Delta p_x^{\text{mech}} &= q[P A_x^{1,s}(x, y, 0; Z^{\text{ex}})] \\ &= q\{[-(1/3)y^2]C_{1,s}^{[1]}(Z^{\text{ex}}) + \dots\} \\ &= q\{[-(1/3)y^2](-4g/a)\text{wig}'(3, Z^{\text{en}}, a, L) + \dots\},\end{aligned}\tag{16.9.38}$$

$$\begin{aligned}\Delta p_y^{\text{mech}} &= q[P A_y^{1,s}(x, y, 0; Z^{\text{ex}})] \\ &= q\{[(1/3)xy]C_{1,s}^{[1]}(Z^{\text{ex}}) + \dots\} \\ &= q\{[(1/3)xy](-4g/a)\text{wig}'(3, Z^{\text{ex}}, a, L) + \dots\}.\end{aligned}\tag{16.9.39}$$

Recall the relation (15.9.35) which we rewrite in the form

$$C_{1,s}^{[0]}(Z) = B_y(0, 0, Z).\tag{16.9.40}$$

We see from (4.36) through (4.39) and (4.40) that in all cases the discontinuities are proportional to  $B_y'(0, 0, Z)$  and its derivatives at  $Z = Z^{\text{en}}$  or  $Z = Z^{\text{ex}}$ . Recall (4.32) and see Figure 4.3 for an example of how these functions behave (fall off) in the case of the simplest 3-pole wiggler/undulator model. Moreover, the discontinuities also vanish as the spatial deviations from the  $z$  axis (the design orbit) become small.

## Exercises

- 16.9.1.** Verify (4.11) and (4.12). Verify that (4.12) also follows from (15.8.21) and (15.8.33).
- 16.9.2.** Verify (4.14) through (4.17).
- 16.9.3.** Verify (4.20) through (4.22).
- 16.9.4.** Verify (4.24) and (4.25).
- 16.9.5.** Verify ....
- 16.9.6.** Verify the relation (4.33).

### Field Fall Off

Our discussion of quadrupole fringe fields so far has treated the iron-free case, and we have found a  $1/|z|^5$  fall off in all cases. When iron is present, and the coils are buried in iron or field clamps are employed, the fall off can in principle be much faster including the possibility of essentially exponential fall off.

## Exercises

- 16.9.7.** Verify the relations (7.32) through (7.39).
  - 16.9.8.** Evidently the second-order portion of  $\psi^e(\mathbf{r}^d; x_0, z_0)$  as given in (7.61) is composed of the monomials  $\xi\eta$  and  $\eta\zeta$ . Show that these are the only monomials allowed at this order based on symmetry considerations. Verify that each monomial is an harmonic polynomial. Indeed, making the usual correspondence between  $\xi, \eta, \zeta$  and  $x, y, z$  show, following the harmonic polynomial labeling scheme (U.2.9), that there are the relations
- $$\xi\eta = [1/(4i)][\sqrt{32\pi/15}][H_2^2(\mathbf{r}) - H_2^{-2}(\mathbf{r})], \quad (16.9.41)$$
- $$\eta\zeta = [-1/(2i)][\sqrt{8\pi/15}][H_2^1(\mathbf{r}) + H_2^{-1}(\mathbf{r})]. \quad (16.9.42)$$
- Would these relations have been simpler had the polar axis, used to set up spherical polar coordinates, been taken to be the  $y$  axis instead of the  $z$  axis?
- 16.9.9.** Verify (6.13) through (6.16).
  - 16.9.10.** Verify (6.19) through (6.24).
  - 16.9.11.** Verify that  $\text{sgn}(z, r)$  as given by (6.24) becomes the true signum function in the limit  $r \rightarrow 0$ . Verify the relations (6.27) through (6.29).
  - 16.9.12.** The purpose of this exercise is to verify that the approximating signum function  $\text{sgn}(z, r_1, r_2)$  becomes the true signum function in the limit that  $r_1$  goes to zero,

$$\lim_{r_1 \rightarrow 0} \text{sgn}(z, r_1, r_2) = \text{sgn}(z). \quad (16.9.43)$$

Along the way we will also verify some other expected properties of  $\text{sgn}(z, r_1, r_2)$ .

- 16.9.13.** Verify (6.41) through (6.43).

Moreover, the general pattern is now clear. We expect that the on-axis gradient for an idealized air-core  $2m$ -pole magnet will fall off for large  $|z|$  as  $1/|z|^{2m+1}$ .

## 16.10 Lithium Lenses

### Acknowledgment

We have benefitted greatly from many conversations with Peter Walstrom, the reading of some of his internal Los Alamos Technical Notes, and the use of his Bessel function routines. We are also grateful to Thomas Mottershead, Filippo Neri, Peter Walstrom, Marco Venturini, and Robert Ryne for their many contributions to MaryLie 5.0.



# Bibliography

## General References

- [1] M. Venturini, “Lie Methods, Exact Map computation, and the Problem of Dispersion in space Charge Dominated Beams”, University of Maryland College Park Physics Department Ph.D. thesis (1998).
- [2] M. Venturini and A. Dragt, “Accurate Computation of Transfer Maps from Magnetic Field Data”, *Nuclear Instruments and Methods* **A427**, p. 387 (1999).
- [3] É. Forest, *Beam Dynamics: A New Attitude and Framework*, Harwood Academic Publishers (1998).
- [4] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, Dover (1972). Also available on the Web by Googling “abramowitz and stegun 1972”.
- [5] G. Szegö, *Orthogonal Polynomials*, American Mathematical Society (1959). See Section 4.7 on ultraspherical (Gegenbauer) polynomials.

## Solenoids

- [6] A. El-Kareh and J. El-Kareh, *Electron Beams, Lenses, and Optics*, Vols. 1 and 2, Academic Press (1970).
- [7] A. Dragt, “Numerical third-order transfer map for solenoid”, *Nuclear Instruments and Methods in Physics Research* **A298**, p. 441 (1990).
- [8] P. W. Hawkes and E. Kasper, *Principles of Electron Optics*, Vols. 1 through 3, Academic Press (1996).
- [9] W. Smythe, *Static and Dynamic Electricity*, McGraw-Hill (1939).

## Dipoles

- [10] P. L. Walstrom, “Dipole-magnet field models based on a conformal map”, *Physical Review Special Topics-Accelerators and Beams* **15**, 102401 (2012).
- [11] L. N. Brouwer, “Canted-Cosine-Theta Superconducting Accelerator Magnets for High Energy Physics and Ion Beam Cancer Therapy”, Ph.D. Thesis, University of California, Berkeley (2015). <https://escholarship.org/uc/item/8jp4g75g>

- [12] M. Bassetti and C. Biscari, “Analytical Formulae for Magnetic Multipoles”, *Particle Accelerators* **52**, pp. 221-250 (1996). <http://cds.cern.ch/record/1120230/files/p221.pdf>
- [13] M. Bassetti and C. Biscari, “Cylinder Model of Multipoles”, *Handbook of Accelerator Physics and Engineering*, First Edition, Section 2.2.2, A. Chao and M. Tigner Edit., World Scientific (1999). Unfortunately, subsequent editions of this book no longer contain this material.

Wigglers

Air-Core and Lambertson Windings

- [14] R.P. Avery, B.R. Lambertson, C.D. Pike, PAC 1971 Proceedings, p. 885.

- [15] T.G. Godlove, S. Bernal, M. Reiser, Printed Circuit Quadrupole Design, PAC 1995 Proceedings, p. 2117

- [16] M. Venturini, Transfer Map for Printed-Circuit Magnetic Quadrupoles, Technical Note, Dept. of Physics, Univ. of Maryland (1995).

Rare Earth Cobalt Magnets

- [17] K. Halbach, “Physical and Optical Properties of Rare Earth Cobalt Magnets”, *Nuclear Instruments and Methods* **187**, pp. 109-117 (1981).

Computation of Charged-Particle Beam Transport

- [18] A. Dragt et al., *MaryLie 3.0 Users’ Manual* (2003). See [www.physics.umd.edu/dsat/](http://www.physics.umd.edu/dsat/).

# Chapter 17

## Surface Methods for General Straight Beam-Line Elements

### 17.1 Introduction

Section 15.1 described the need for Taylor expansions of the vector potential  $\mathbf{A}$  in order to determine the transfer map  $\mathcal{M}$ . As illustrated in Chapter 16, there are cases in which these Taylor expansions can be found analytically. However, for most cases, all that we can hope to have are magnetic field values determined numerically at points on some regular 3-dimensional grid with the aid of some electromagnetic code.<sup>1</sup> This places us in what might appear to be a hopeless position: it is well known that it is generally difficult to extract reliable information about derivatives from numerical data on a grid. And we want to know about high derivatives! Hildebrand, author of *Introduction to Numerical Analysis*, writes

*Once an interpolating polynomial  $y(x)$  has been determined so that it satisfactorily approximates a given function  $f(x)$  over a certain interval  $I$ , it may be hoped that the results of differentiating  $y(x) \dots$  will also satisfactorily approximate the corresponding derivative  $\dots$  of  $f(x)$ . However  $\dots$  we may anticipate the fact that, even though the deviation between  $y(x)$  and  $f(x)$  will be small throughout the interval, still the slopes of the two curves representing them may differ quite appreciably. Further, it is seen that roundoff errors (or errors of observation) of alternating sign in consecutive ordinates could affect the calculation of the derivative quite strongly if those ordinates were fairly closely spaced  $\dots$ . In particular, numerical differentiation should be avoided whenever possible, particularly when the data are empirical and subject to appreciable errors of observation.*

Remarkably, we will find that this problem can be overcome to some satisfactory aberration order with the use of *surface* data.<sup>2</sup> We will fit field data onto some surface, and then use

---

<sup>1</sup>Alternatively, see Section 17.2, we may have numerically-determined values of the magnetic scalar potential.

<sup>2</sup>The determination of the solution of Laplace's equation in terms of surface data is called the *Dirichlet* (1805-1859) problem. Dirichlet was the thesis advisor of, among others, Kronecker and Lipschitz. It is also interesting to note that he married Rebecka Mendelssohn, one of the sisters of Felix Mendelssohn.

this surface data to compute interior fields. Specifically, and in summary, we will find that surface methods have the following virtues:

- Only functions with known (orthonormal) completeness properties and known (optimal) convergence properties are employed.
- The Maxwell equations are exactly satisfied.
- The results are manifestly analytic in all variables.
- The error is globally controlled. Harmonic fields, fields that satisfy the Laplace equation, take their extrema on boundaries. Both the exact and computed fields satisfy the Laplace equation. Therefore their difference, the error field, also satisfies the Laplace equation, and must take its extrema on the boundary. But this is precisely where a controlled fit is made. Thus, the error on the boundary is controlled, and the interior error must be even smaller.
- Because harmonic fields take their extrema on boundaries, interior values inferred from surface data are relatively insensitive to errors/noise in the surface data. Put another way, the inverse Laplacian (Laplace Green function), which relates interior data to surface data, is *smoothing*. It is this smoothing that we seek to exploit. We will find that the sensitivity to noise in the data decreases rapidly (as some high inverse power of distance) with increasing distance from the surface, and this property improves the accuracy of the high-order interior derivatives needed to compute high-order transfer maps.

In this chapter, devoted to the case of straight beam-line elements, we will develop methods for computing high-order transfer maps based on data provided on a 3-dimensional grid. See Figure 1.1. These methods make it possible to compute realistic transfer maps for real (straight) beam-line elements including all fringe-field and higher-order multipole effects. In Chapter 15 we learned how to characterize magnetic fields in terms of cylindrical harmonics described by on-axis gradients, and also how to determine vector potentials in terms of on-axis gradients. In this chapter we will see how on-axis gradients can in turn be computed from numerical data provided on a 3-dimensional grid. Chapters 18 through 21 will elaborate on these methods and apply them to a variety of straight beam-line elements. In Chapters 22 through 25 we will consider realistic transfer maps for curved beam-line elements.

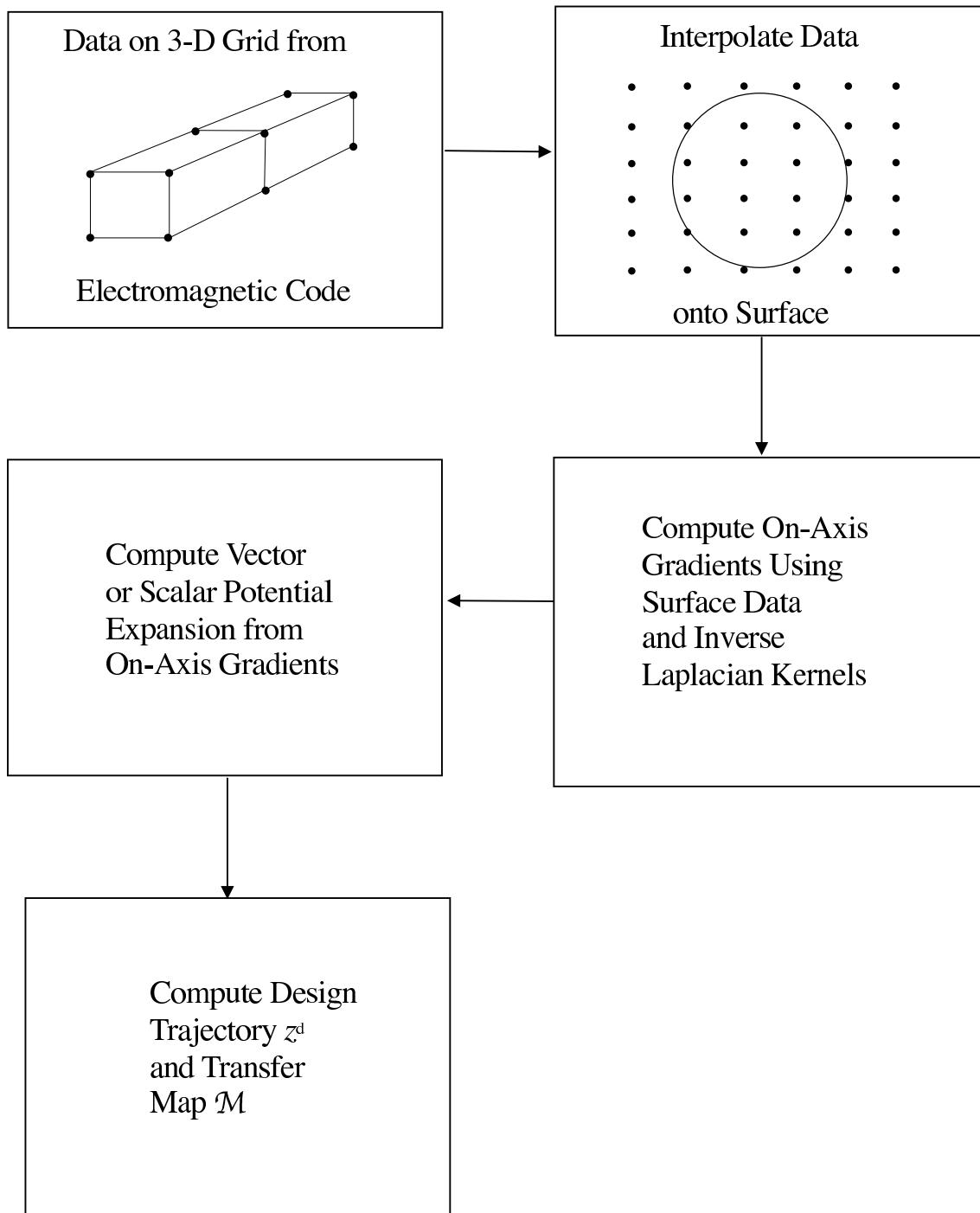


Figure 17.1.1: Calculation of realistic design trajectory  $z^d$  and its associated realistic transfer map  $\mathcal{M}$  based on data provided on a 3-dimensional grid for a real beam-line element. Only a few points on the 3-dimensional grid are shown. In this illustration, data from the 3-dimensional grid is interpolated onto the surface of a cylinder with circular cross section, and this surface data is then processed to compute the design trajectory and the transfer map. The use of other surfaces is also possible, and may offer various advantages.

At this point one might wonder about a seemingly simpler approach: Suppose some fine grid of possible initial conditions is laid out in phase space. Next suppose the final conditions associated with these initial conditions are computed numerically, say by integrating Newton's equations with a Lorentz force, using volume magnetic field data interpolated off a three-dimensional grid. Based on the collection of initial and final conditions, make a polynomial expansion of the final conditions in terms of the initial conditions. Announce that these truncated Taylor series, generally six in number assuming a six-dimensional phase space, constitute a (Taylor) transfer map. What could be wrong with that?

There are three reasons why such an approach is problematic:

1. After some reflection, we see that this procedure essentially amounts to high-order numerical differentiation, and therefore Hildebrand's warning still holds. The problem of error associated with high-order numerical differentiation remains.
2. If Newton's equations are integrated, the symplectic symmetry inherent in a Hamiltonian formulation cannot be exploited.
3. Again assuming Newton's equations are integrated, and even in the absence of the error associated with high-order numerical differentiation, the result will not be symplectic if there is a residual magnetic field at the beginning or end of the integration region. See Exercise 6.4.11. Therefore the result may not be suitable for long-term tracking. Perhaps this possible lack of symplecticity could in principle be handled by factorizing the resulting Taylor map into symplectic and nonsymplectic parts, and then using only the symplectic part for any subsequent calculations. See Section 29.1. As a bonus, examination of the size of the nonsymplectic part might give some indication of the error involved in the calculation.

There are also some other approaches that have sometimes been attempted to obtain transfer maps based on 3-d field data on a grid. They are described in Section 17.6. They too involve high-order numerical differentiation, and therefore are unlikely to succeed beyond modest order, at best.

Finally, we mention that there are two other possible ways of determining on-axis gradients that warrant exploration. The first is to infer on-axis gradients from experimental spinning coil data. The second, applicable in the case of air-core magnets, is to compute on-axis gradients based on data describing coil winding geometry and currents flowing in the windings. See Appendix K.

## Exercises

**17.1.1.** This exercise explores some aspects of the Laplace/Poisson equation. We will consider solutions  $\psi(x, y, z)$  about some point which, without loss of generality, may be taken to be the origin  $\mathbf{r} = 0$ .

Suppose that  $\psi$  is analytic in the Cartesian components of  $\mathbf{r}$ , has at the origin the value

$$\psi(0) = \psi_0, \quad (17.1.1)$$

and is harmonic in some volume  $V$  surrounding the origin,

$$\nabla^2\psi(\mathbf{r}) = 0. \quad (17.1.2)$$

Introduce spherical coordinates in the usual way and let  $Q_{\ell,m,c}(\theta, \phi)$  and  $Q_{\ell,m,s}(\theta, \phi)$  denote the functions

$$Q_{\ell,m,c}(\theta, \phi) = \Re[Y_{\ell,m}(\theta, \phi)] = O_{\ell,m}(\theta) \cos(m\phi), \quad (17.1.3)$$

$$Q_{\ell,m,s}(\theta, \phi) = \Im[Y_{\ell,m}(\theta, \phi)] = O_{\ell,m}(\theta) \sin(m\phi). \quad (17.1.4)$$

Here we impose the requirement  $m \geq 0$  and make the definitions

$$Q_{\ell,0,s}(\theta, \phi) = 0, \quad (17.1.5)$$

$$O_{\ell,m}(\theta) = (-1)^m \sqrt{[(2\ell+1)(\ell-m)!]/[(4\pi)(\ell+m)!]} P_{\ell}^m(\cos \theta). \quad (17.1.6)$$

Expand  $\psi(x, y, z)$  in a Taylor series about the origin and group terms of like degree so as to yield an expansion in homogeneous polynomials. Show that rewriting this expansion in spherical coordinates gives the result

$$\psi = \psi_0 + \sum_{\ell=1}^{\infty} \sum_{m=0}^{\ell} \sum_{\alpha=c,s} d_{\ell,m,\alpha} [r^{\ell} Q_{\ell,m,\alpha}(\theta, \phi)]. \quad (17.1.7)$$

Here the quantities  $d_{\ell,m,\alpha}$  are arbitrary coefficients and we have enforced the conditions (1.1) and (1.2).

Next, integrate  $\psi$  over the surface of a sphere of radius  $R$  centered on the origin. Show, recalling the orthogonality properties of the  $Y_{\ell,m}$ , that using the expansion (1.7) yields the result

$$\int_S \psi dS = 4\pi R^2 \psi_0. \quad (17.1.8)$$

Consequently, there is the relation

$$[1/(4\pi R^2)] \int_S \psi dS = \psi_0. \quad (17.1.9)$$

The average of  $\psi$  over the surface of a sphere equals its value at the center of the sphere. It follows that if  $\psi > \psi_0$  at some point on the surface of the sphere, then it must be the case that  $\psi < \psi_0$  at some other point on the surface of the sphere, and vice versa, in order for (1.9) to hold. Finally, an analogous result must be true for any expansion point within  $V$ . Consequently,  $\psi$  has no local minima or maxima, and must take its extrema on the boundary of  $V$ .

Suppose we replace the harmonic requirement (1.2) by the condition

$$\nabla^2\psi|_{\mathbf{r}=0} = \rho_0. \quad (17.1.10)$$

What happens now? In this case expand  $\psi(x, y, z)$  in a Taylor series about the origin through terms of degree 2 and group terms of like degree so as to again yield an expansion

in homogeneous polynomials. Show that rewriting this expansion in spherical coordinates gives, through terms of degree 2, the result

$$\psi = \psi_0 + \sum_{m=0}^1 \sum_{\alpha=c,s} d_{1,m,\alpha} [r Q_{1,m,\alpha}(\theta, \phi)] + (\rho_0/6) r^2 + \sum_{m=0}^2 \sum_{\alpha=c,s} d_{2,m,\alpha} [r^2 Q_{2,m,\alpha}(\theta, \phi)] + \dots . \quad (17.1.11)$$

Here the quantities  $d_{\ell,m,\alpha}$  are again arbitrary coefficients and we have enforced the conditions (1.1) and (1.10).

Again integrate  $\psi$  over the surface of a sphere of radius  $R$  centered on the origin. Show that using the expansion (1.11) yields the result

$$\int_S \psi dS = 4\pi R^2 \psi_0 + 4\pi (\rho_0/6) R^4 + O(R^6). \quad (17.1.12)$$

Consequently, there is the relation

$$[1/(4\pi R^2)] \int_S \psi dS = \psi_0 + (\rho_0/6) R^2 + O(R^4). \quad (17.1.13)$$

The average of  $\psi$  over the surface of a small sphere equals its value at the center of the sphere, plus a correction of order  $R^2$  that involves  $\rho_0$ , plus corrections of order  $R^4$ . In lowest order, the difference between the spherical average of  $\psi$  and its central value  $\psi_0$  involves  $\rho_0$ . For this reason, the quantity  $\rho_0$  is called the *concentration* of  $\psi$  at  $\mathbf{r} = 0$ .

## 17.2 Use of Potential Data on Surface of Circular Cylinder

We will begin our discussion with the use of the surface of a cylinder with circular cross section, and the use of scalar potential data on this surface. This is conceptually the simplest case, and will give us opportunity to develop various needed concepts. Moreover, some electromagnetic codes calculate directly the scalar potential on some regular three-dimensional grid, and this data can be interpolated onto the surface of a cylinder. Therefore, this method can also be of practical use.

Consider a circular cylinder of radius  $R$ , centered on the  $z$ -axis, fitting within the bore of the beam-line element in question, and extending beyond the fringe-field regions at the ends of the beam-line element. The beam-line element could be any straight element such as a solenoid, quadrupole, sextupole, octupole, etc., or it could be wiggler with no net bending. See Figure 2.1, which illustrates the case of a wiggler. Write

$$\psi(x, y, z) = \psi(\rho, \phi, z), \quad (17.2.1)$$

and suppose  $\psi(R, \phi, z)$  is known. Here we have used the coordinates (15.2.12) through (15.2.16). In general, determination of  $\psi(R, \phi, z)$  will require interpolation onto a circle of data on a square (or rectangular) grid in  $x$  and  $y$  for each  $z$  value on the grid. See the second frame of Figure 1.1 which depicts a square or rectangular grid in the  $x,y$  plane for a fixed  $z$

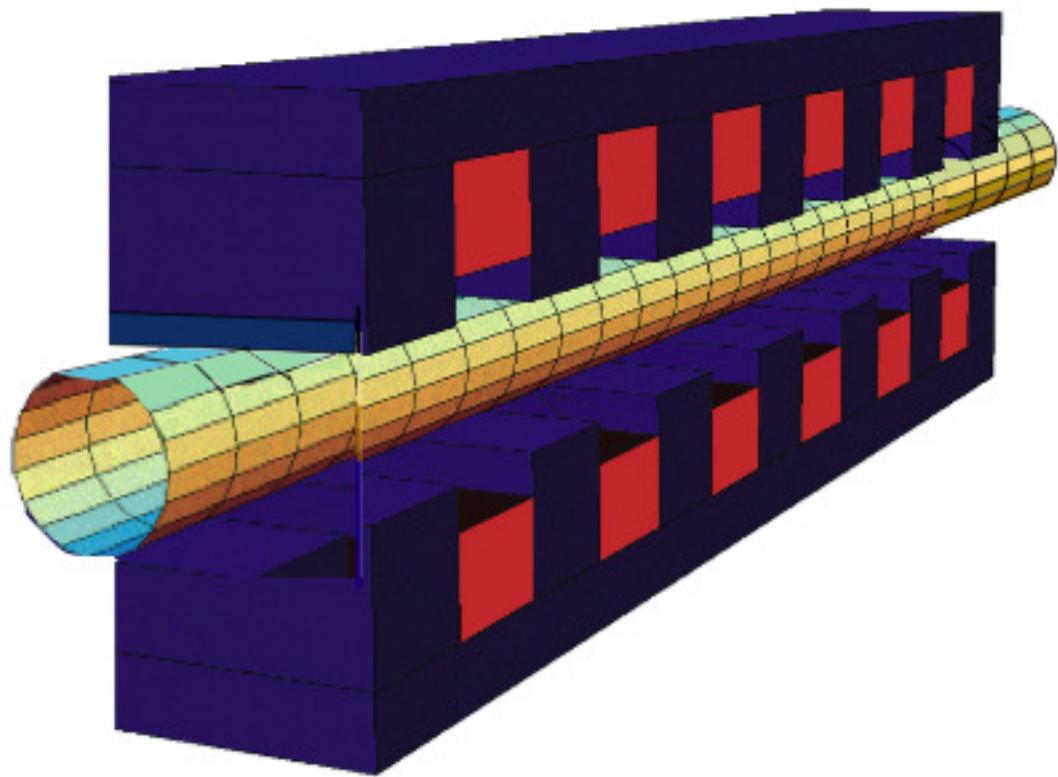


Figure 17.2.1: A circular cylinder of radius  $R$ , centered on the  $z$ -axis, fitting within the bore of a beam-line element, in this case a wiggler, and extending beyond the fringe-field regions at the ends of the beam-line element.

value on the 3-dimensional grid. Values at data points near the circle are to be interpolated onto the circle.

From this given function  $\psi(R, \phi, z)$ , obtained by interpolation, form/define the function  $\tilde{\psi}(R, m', k')$  by the rule

$$\tilde{\psi}(R, m', k') = [1/(2\pi)]^2 \int_{-\infty}^{\infty} dz \exp(-ik'z) \int_0^{2\pi} d\phi \exp(-im'\phi) \psi(R, \phi, z). \quad (17.2.2)$$

Here we pause a moment to describe our nomenclature and notation: We will refer to the operation of Fourier transforming over the *line*  $[-\infty, \infty]$  as performing a *linear* Fourier transform, and the result of this transform will be labeled by a continuous variable usually called  $k$ . We will refer to the operation of Fourier transforming over the angular domain  $[0, 2\pi]$  as performing an *angular* Fourier transform.<sup>3</sup> Moreover, the result of performing an angular Fourier transform will be called a Fourier coefficient, and these coefficients will be labeled by integers such as  $m$  and  $n$ . Finally, we have used the symbol  $\tilde{\cdot}$  to denote a linear or angular Fourier transform, and the symbol  $\tilde{\tilde{\cdot}}$  to denote that both have been performed.

To continue, we know from (15.3.7) that  $\psi(R, \phi, z)$  has the representation

$$\psi(R, \phi, z) = \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} dk G_m(k) \exp(ikz) \exp(im\phi) I_m(kR). \quad (17.2.3)$$

Employing this representation in (2.2) and performing the indicated integrations give the result

$$\tilde{\psi}(R, m', k') = G_{m'}(k') I_{m'}(k'R), \quad (17.2.4)$$

from which we conclude that

$$G_m(k) = \tilde{\psi}(R, m, k) / I_m(kR). \quad (17.2.5)$$

This relation for  $G_m(k)$  can now be employed in (15.3.15) to give the result

$$C_m^{[n]}(z) = i^n (1/2)^{|m|} (1/|m|!) \int_{-\infty}^{\infty} dk [k^{n+|m|} / I_m(kR)] \tilde{\psi}(R, m, k) \exp(ikz). \quad (17.2.6)$$

We have found an expression for the on-axis gradients in terms of potential data on the surface of the cylinder. Equation (2.6) may be viewed as the convolution of Fourier surface data  $\tilde{\psi}(R, m, k)$  with the *inverse Laplacian* kernel  $[k^{n+|m|} / I_m(kR)]$ . Moreover, this kernel has a very desirable property. The Bessel functions  $I_m(kR)$  have the asymptotic behavior

$$|I_m(kR)| \sim \exp(|k|R) / \sqrt{2\pi|R|} \text{ as } |k| \rightarrow \infty. \quad (17.2.7)$$

<sup>3</sup>Joseph Fourier (1768-1830) was a student of Lagrange. Fourier was the first to make extensive use of the trigonometric series that bear his name, and to make the claim that they could be used to represent arbitrary functions. This claim his elders and contemporaries found hard to believe. In reviewing one of his fundamental papers on the theory of heat that employed these series the referees Lagrange, Laplace, Legendre, and others complained that *... the manner in which the author arrives at these equations is not exempt of difficulties and that his analysis to integrate them still leaves something to be desired on the score of generality and even rigor.* As a result, the paper was not published.

Since  $I_m(kR)$  appears in the denominator of (2.6), we see that the integrand is exponentially damped for large  $|k|$ . Now suppose there is uncorrelated point-to-point noise in the surface data. Such noise will result in anomalously large  $|k|$  contributions to the  $\tilde{\psi}(R, m, k)$ . But, because of the exponential damping arising from  $I_m(kR)$  in the denominator, the effect of this noise is effectively filtered out. Moreover, this filtering action is improved by making  $R$  as large as possible. This filtering, or *smoothing*, feature will be discussed in more detail in Chapter 18.

### 17.3 Use of Field Data on Surface of Circular Cylinder

All three-dimensional electromagnetic codes calculate all three components of the field on some three-dimensional grid. Also, such data is in principle available from actual field measurements. In this section we will describe how to compute the on-axis gradients from field data.

Again we will employ a cylinder of radius  $R$  centered on the  $z$  axis. Suppose the magnetic field  $\mathbf{B}(x, y, z)$  is interpolated onto the surface of the cylinder using values at the grid points near the surface. Next, from the values on the surface, compute  $B_\rho(x, y, z) = B_\rho(R, \phi, z)$ , the component of  $\mathbf{B}(x, y, z)$  *normal* to the surface. We will now see how to compute the on-axis gradients from a knowledge of  $B_\rho(x, y, z) = B_\rho(R, \phi, z)$ .

From this known function form the functions  $\tilde{B}_\rho(R, m', z)$  and  $\tilde{B}_\rho(R, m', k')$  by the rules

$$\tilde{B}_\rho(R, m', z) = [1/(2\pi)] \int_0^{2\pi} d\phi \exp(-im'\phi) B_\rho(R, \phi, z), \quad (17.3.1)$$

$$\tilde{B}_\rho(R, m', k') = [1/(2\pi)] \int_{-\infty}^{\infty} dz \exp(-ik'z) \tilde{B}_\rho(R, m', z). \quad (17.3.2)$$

Note that we may also directly write that

$$\tilde{B}_\rho(R, m', k') = [1/(2\pi)]^2 \int_{-\infty}^{\infty} dz \exp(-ik'z) \int_0^{2\pi} d\phi \exp(-im'\phi) B_\rho(R, \phi, z), \quad (17.3.3)$$

and the indicated integrations may be performed in either order. We also know that

$$B_\rho(R, \phi, z) = [\partial_\rho \psi(\rho, \phi, z)]|_{\rho=R}, \quad (17.3.4)$$

from which it follows, using the representation (15.3.7), that

$$B_\rho(R, \phi, z) = \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} dk G_m(k) \exp(ikz) \exp(im\phi) k I'_m(kR). \quad (17.3.5)$$

Now substitute (3.5) into the right side of (3.3) and perform the indicated integrations to get the result

$$\tilde{B}_\rho(R, m', k') = G_{m'}(k') k' I'_{m'}(k'R), \quad (17.3.6)$$

from which it follows that

$$G_m(k) = \tilde{B}_\rho(R, m, k) / [k I'_m(kR)]. \quad (17.3.7)$$

This relation for  $G_m(k)$  can be employed in (15.3.15) to give the result

$$C_m^{[n]}(z) = i^n (1/2)^{|m|} (1/|m|!) \int_{-\infty}^{\infty} dk [k^{n+|m|-1} / I'_m(kR)] \tilde{B}_\rho(R, m, k) \exp(ikz). \quad (17.3.8)$$

We have found an expression for the on-axis gradients in terms of field data (normal component) on the surface of the cylinder. Moreover, this expression again has the smoothing property since the denominator functions  $I'_m(kR)$  also have the asymptotic behavior (2.7) and therefore also provide exponential damping,

$$|I'_m(kR)| \sim \exp(|k|R) / \sqrt{2\pi|k|R} \text{ as } |k| \rightarrow \infty. \quad (17.3.9)$$

For future use it is also convenient to have explicit formulas for the  $C_{m,\alpha}^{[n]}(z)$ . Motivated by (15.3.28) and (15.3.31), define quantities  $\tilde{B}_\rho^\alpha(R, m', k')$  and  $\tilde{B}_\rho^\alpha(R, m', z)$  with  $m' \geq 1$  by the rules

$$\tilde{B}_\rho^s(R, m', k') = i[\tilde{B}_\rho(R, m', k') - \tilde{B}_\rho(R, -m', k')], \quad (17.3.10)$$

$$\tilde{B}_\rho^c(R, m', k') = [\tilde{B}_\rho(R, m', k') + \tilde{B}_\rho(R, -m', k')], \quad (17.3.11)$$

$$\tilde{B}_\rho^s(R, m', z) = i[\tilde{B}_\rho(R, m', z) - \tilde{B}_\rho(R, -m', z)], \quad (17.3.12)$$

$$\tilde{B}_\rho^c(R, m', z) = [\tilde{B}_\rho(R, m', z) + \tilde{B}_\rho(R, -m', z)]. \quad (17.3.13)$$

Then we have the results

$$\tilde{B}_\rho^\alpha(R, m', k') = [1/(2\pi)] \int_{-\infty}^{\infty} dz \exp(-ik'z) \tilde{B}_\rho^\alpha(R, m', z) \quad (17.3.14)$$

with

$$\tilde{B}_\rho^s(R, m', z) = (1/\pi) \int_0^{2\pi} d\phi \sin(m'\phi) B_\rho(R, \phi, z), \quad (17.3.15)$$

$$\tilde{B}_\rho^c(R, m', z) = (1/\pi) \int_0^{2\pi} d\phi \cos(m'\phi) B_\rho(R, \phi, z). \quad (17.3.16)$$

And, in accord with (15.3.35) and (15.3.36), for  $m' = 0$  make the definitions

$$\tilde{B}_\rho^s(R, m' = 0, k') = 0, \quad (17.3.17)$$

$$\tilde{B}_\rho^c(R, m' = 0, k') = \tilde{B}_\rho(R, m' = 0, k'), \quad (17.3.18)$$

$$\tilde{B}_\rho^s(R, m' = 0, z) = 0, \quad (17.3.19)$$

$$\tilde{B}_\rho^c(R, m' = 0, z) = \tilde{B}_\rho(R, m' = 0, z). \quad (17.3.20)$$

Then we have the further results

$$\tilde{B}_\rho^c(R, m' = 0, k') = [1/(2\pi)] \int_{-\infty}^{\infty} dz \exp(-ik'z) \tilde{B}_\rho^c(R, m' = 0, z), \quad (17.3.21)$$

$$\tilde{B}_\rho^c(R, m' = 0, z) = \tilde{B}_\rho(R, m' = 0, z) = [1/(2\pi)] \int_0^{2\pi} d\phi B_\rho(R, \phi, z). \quad (17.3.22)$$

Note that the quantities  $\tilde{B}_\rho^\alpha(R, m', z)$  are real. Correspondingly, we see from (3.14) that the real part of  $\tilde{\tilde{B}}_\rho^\alpha(R, m', k')$  is even in  $k$  and the imaginary part is odd in  $k$ .

With these definitions in hand, we are ready to state the final results:

$$C_{m,\alpha}^{[n]}(z) = i^n (1/2)^m (1/m!) \int_{-\infty}^{\infty} dk [k^{n+m-1} / I'_m(kR)] \tilde{\tilde{B}}_\rho^\alpha(R, m, k) \exp(ikz) \quad (17.3.23)$$

for  $m > 0$ , and

$$C_{m=0,s}^{[n]}(z) = 0, \quad (17.3.24)$$

$$C_{m=0,c}^{[n]}(z) = C_0^{[n]}(z) = i^n \int_{-\infty}^{\infty} dk [k^{n-1} / I'_0(kR)] \tilde{\tilde{B}}_\rho^c(R, m = 0, k) \exp(ikz). \quad (17.3.25)$$

We close this section with the remark that if one wishes to extract the  $C_0^{[n]}(z)$  (monopole) on-axis gradients from field data, it may be preferable to use the longitudinal component  $B_z(R, \phi, z)$  on the surface of the cylinder rather than the normal component  $B_\rho(R, \phi, z)$ .<sup>4</sup> See Section 19.2.

## 17.4 Use of Field Data on Surface of Elliptical Cylinder

### 17.4.1 Background

In the previous two sections we employed a cylinder with circular cross section, and observed mathematically that it is desirable for error insensitivity to use a cylinder with a large radius  $R$ . Physically, this is because we want the data points to be as far from the axis as possible since the effect of inhomogeneities (noise) in the data decays with distance from the inhomogeneity. Evidently the use of a large circular cylinder is optimal for beam-line elements with a circular bore. However, for dipoles or wigglers with small gaps and wide pole faces, use of a cylinder with elliptical cross section should give improved error insensitivity. See Figure 4.1. In this section we will set up the machinery required for the use of elliptical cylinders, and apply it to the calculation of on-axis gradients based on field data.

We will see that the use of elliptic cylinders requires a knowledge of Mathieu functions. Since these functions may well be relatively unfamiliar to the reader, considerable effort will be devoted to describing their properties.

For brevity, we will omit treatment of the related case where potential data is used on the surface of the elliptic cylinder. The reader should be able to solve this simpler problem based on the work of the current section and what was done in Section 14.2.

---

<sup>4</sup>Note that in any case we only need the  $C_0^{[n]}(z)$  with  $n \geq 1$  because they are what is required to compute the vector potential. See (15.5.32) through (15.5.34). Thus, (3.6) is well defined for all values of  $m$  and  $n$  of physical interest.

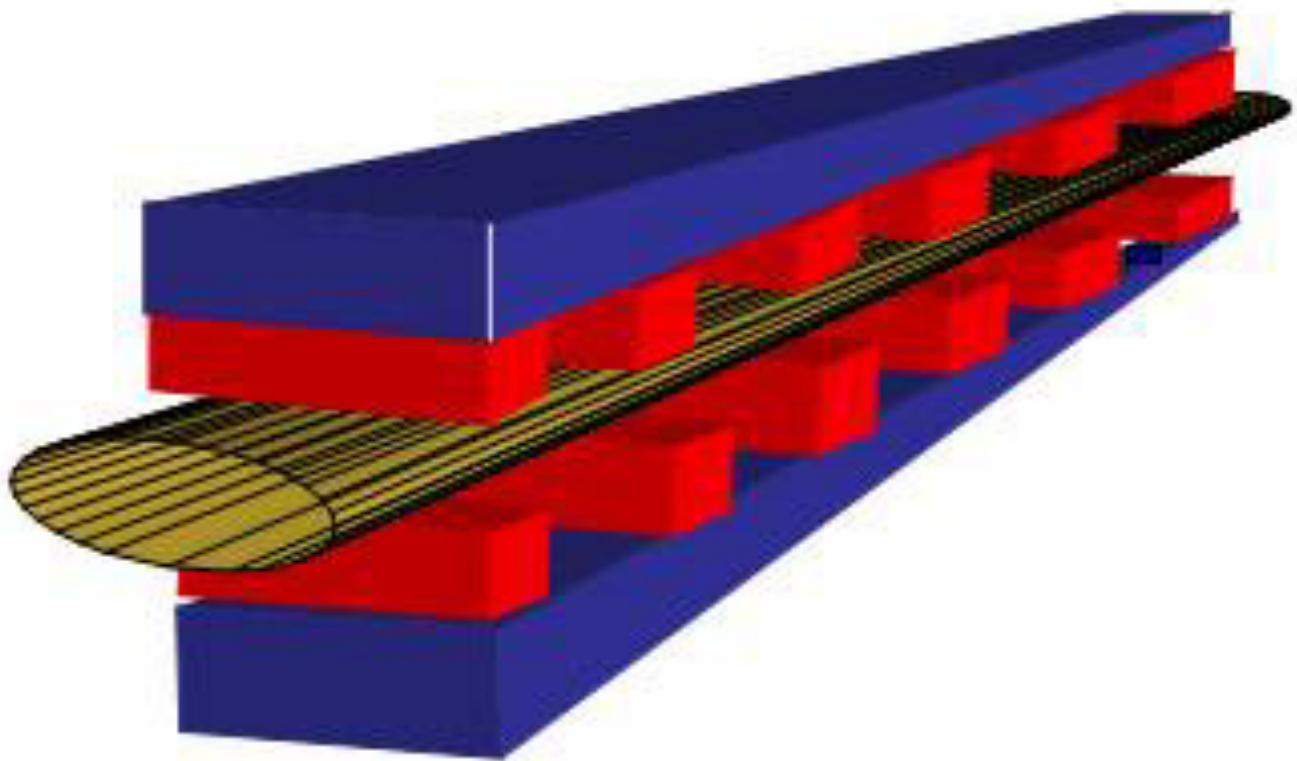


Figure 17.4.1: An elliptical cylinder, centered on the  $z$ -axis, fitting within the bore of a wiggler, and extending beyond the fringe-field regions at the ends of the wiggler.

### 17.4.2 Elliptic Coordinates

Elliptic coordinates in the  $x, y$  plane are described by the relations

$$x = f \cosh(u) \cos(v), \quad (17.4.1)$$

$$y = f \sinh(u) \sin(v). \quad (17.4.2)$$

Contours of constant  $u$ , with  $u \in [0, \infty]$ , are nested ellipses with common foci located at  $(x; y) = (\pm f; 0)$ . Contours of constant  $v$ , with  $v \in [0, 2\pi]$ , are hyperbolae. Together these contours form an orthogonal coordinate system. See Figure 4.2. Data is to be interpolated onto the ellipse whose cross section is that of the elliptical cylinder of Figure 4.1. See Figure 4.3.

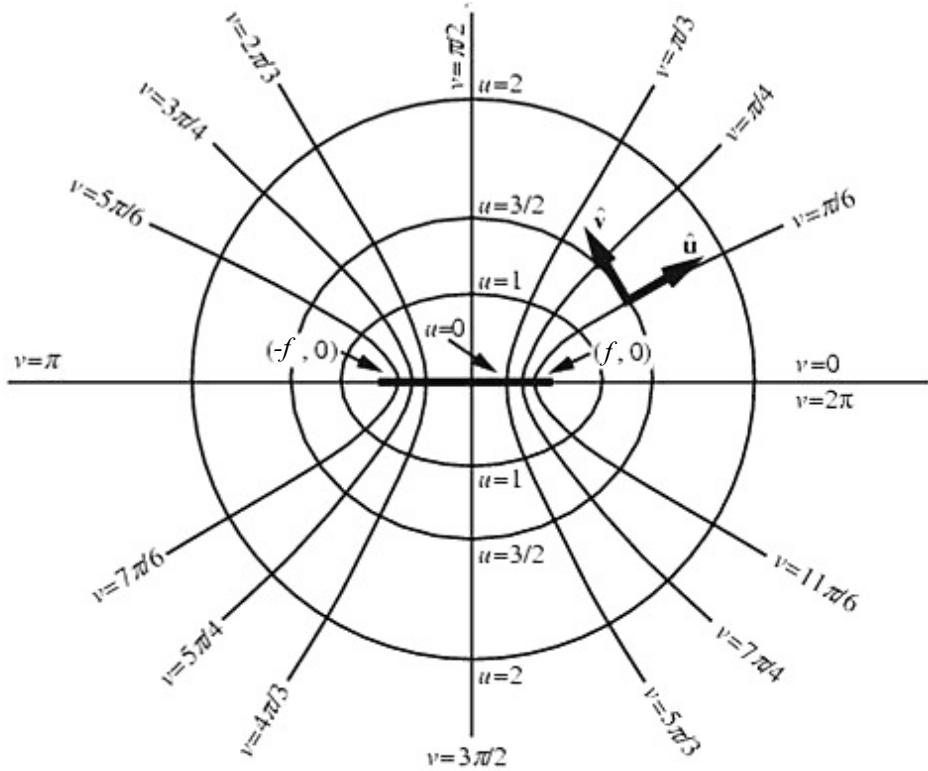


Figure 17.4.2: Elliptical coordinates showing contours of constant  $u$  and constant  $v$ .

For our work we will need the unit vector  $\hat{\mathbf{e}}_u$ , the unit vector (outwardly) normal to the surface of the elliptical cylinder. Write

$$\begin{aligned} \mathbf{r} &= x\hat{\mathbf{e}}_x + y\hat{\mathbf{e}}_y + z\hat{\mathbf{e}}_z \\ &= f \cosh(u) \cos(v)\hat{\mathbf{e}}_x + f \sinh(u) \sin(v)\hat{\mathbf{e}}_y + z\hat{\mathbf{e}}_z. \end{aligned} \quad (17.4.3)$$

Then, by definition, we have the result

$$\begin{aligned} \hat{\mathbf{e}}_u &= (\partial \mathbf{r} / \partial u) / \|(\partial \mathbf{r} / \partial u)\| \\ &= [\sinh(u) \cos(v)\hat{\mathbf{e}}_x + \cosh(u) \sin(v)\hat{\mathbf{e}}_y] / [\cosh^2(u) - \cos^2(v)]^{1/2}. \end{aligned} \quad (17.4.4)$$

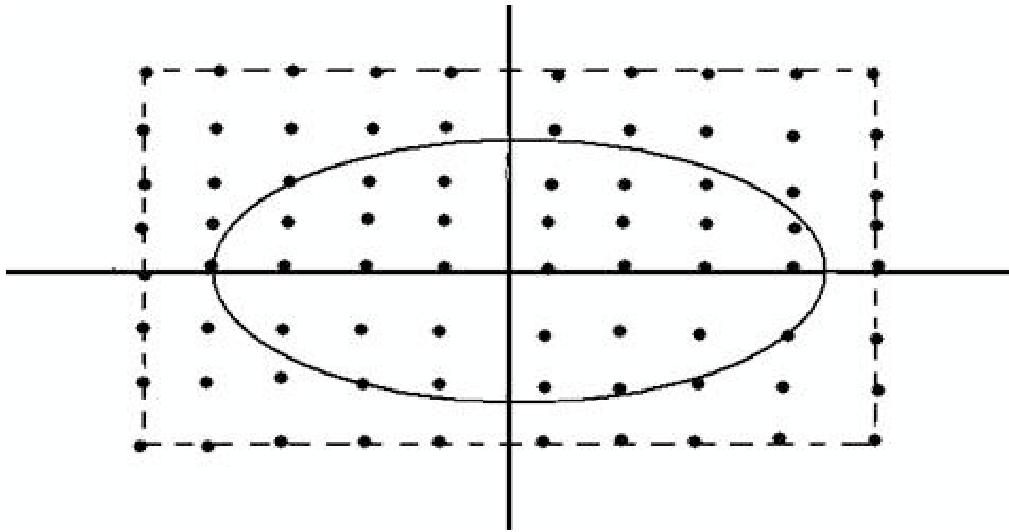


Figure 17.4.3: A square or rectangular grid in the  $x,y$  plane for a fixed  $z$  value on the 3-dimensional grid. Values at data points near the ellipse are to be interpolated onto the ellipse.

It is also convenient to employ the complex variables

$$\zeta = x + iy, \quad (17.4.5)$$

and

$$w = u + iv. \quad (17.4.6)$$

In these variables, the relations (4.1) and (4.2) can be written in the more compact form

$$\zeta = f \cosh(w). \quad (17.4.7)$$

[For a discussion of the analytic properties of  $\zeta(w)$  and its inverse  $w(\zeta)$ , see Exercise 4.2.] Form differentials of both sides of (4.7). Doing so gives the result

$$dx + idy = f \sinh(w)(du + idv) \quad (17.4.8)$$

and the complex conjugate result

$$dx - idy = f \sinh(\bar{w})(du - idv). \quad (17.4.9)$$

Now form the product of (4.8) and (4.9) to get the transverse line-element relation

$$\begin{aligned} ds_{\perp}^2 &= dx^2 + dy^2 = f^2 \sinh(u + iv) \sinh(u - iv)(du^2 + dv^2) \\ &= f^2[\cosh^2(u) - \cos^2(v)](du^2 + dv^2). \end{aligned} \quad (17.4.10)$$

From this relation we infer the results

$$B_u = \hat{\mathbf{e}}_u \cdot \mathbf{B} = (\nabla\psi)_u = (1/f)[\cosh^2(u) - \cos^2(v)]^{-1/2}(\partial\psi/\partial u), \quad (17.4.11)$$

$$\nabla^2\psi = (1/f^2)[\cosh^2(u) - \cos^2(v)]^{-1}[(\partial_u)^2 + (\partial_v)^2]\psi + (\partial_z)^2\psi. \quad (17.4.12)$$

### 17.4.3 Mathieu Equations

Let us seek to construct harmonic functions of the form

$$\psi \sim P(u)Q(v) \exp(ikz) \quad (17.4.13)$$

where the functions  $P$  and  $Q$  are yet to be determined. Employing the Ansatz (4.13) in Laplace's equation and use of (4.12) yields the requirement

$$[(\partial_u)^2 + (\partial_v)^2][P(u)Q(v)] = k^2 f^2 [\cosh^2(u) - \cos^2(v)] P(u)Q(v). \quad (17.4.14)$$

We also observe that there is the trigonometric identity

$$\cosh^2(u) - \cos^2(v) = (1/2)[\cosh(2u) - \cos(2v)] \quad (17.4.15)$$

so that the requirement (4.14) can be rewritten in the form

$$[(\partial_u)^2 + (\partial_v)^2][P(u)Q(v)] = (k^2 f^2/4)[2 \cosh(2u) - 2 \cos(2v)] P(u)Q(v). \quad (17.4.16)$$

Upon dividing both sides by  $PQ$ , (4.16) becomes

$$(1/P)(\partial_u)^2 P + (1/Q)(\partial_v)^2 Q = (k^2 f^2/4)[2 \cosh(2u) - 2 \cos(2v)], \quad (17.4.17)$$

from which it follows that

$$(1/P)(\partial_u)^2 P - (k^2 f^2/4)[2 \cosh(2u)] = -(1/Q)(\partial_v)^2 Q - (k^2 f^2/4)[2 \cos(2v)]. \quad (17.4.18)$$

Therefore, there is a common *separation* constant  $a$  such that

$$(1/P)(\partial_u)^2 P - (k^2 f^2/4)[2 \cosh(2u)] = a \quad (17.4.19)$$

and

$$-(1/Q)(\partial_v)^2 Q - (k^2 f^2/4)[2 \cos(2v)] = a. \quad (17.4.20)$$

Correspondingly,  $P$  and  $Q$  must satisfy the ordinary and linear differential equations

$$d^2 P / du^2 - [a - 2q \cosh(2u)] P = 0, \quad (17.4.21)$$

$$d^2 Q / dv^2 + [a - 2q \cos(2v)] Q = 0, \quad (17.4.22)$$

where

$$q = -k^2 f^2/4. \quad (17.4.23)$$

Equation (4.22) for  $Q$  is called the *Mathieu* equation, and Equation (4.21) for  $P$  is called the *modified Mathieu* equation.<sup>5</sup>

---

<sup>5</sup>We remark that many, and probably the majority, of the special functions ordinarily encountered in Mathematical Physics are particular cases of the hypergeometric function. The Mathieu functions do not fall in this category. In some sense, they are *more transcendental* than the hypergeometric function.

### 17.4.4 Periodic Mathieu Functions and Separation Constants

For our purposes, we will need solutions  $Q(v)$  of (4.22) that are *periodic* with period  $2\pi$ . See Figure 4.2. Such solutions exist only for certain specific values of the separation constant  $a$ . These values are called  $a_n(q)$  for  $n = 0, 1, 2, 3, \dots$  and  $b_n(q)$  for  $n = 1, 2, 3, \dots$ <sup>6</sup> The functions  $a_n(q)$  and  $b_n(q)$  are all *real* for real values of  $q$ . As the notation indicates, their values depend on  $q$  (and on  $n$ ). For small  $q$  they have expansions of the form

$$a_0(q) = -(1/2)q^2 + (7/128)q^4 + \dots, \quad (17.4.24)$$

$$a_1(q) = 1 + q - (1/8)q^2 - (1/64)q^3 - (1/1536)q^4 + \dots, \quad (17.4.25)$$

$$a_2(q) = 4 + (5/12)q^2 - (763/13824)q^4 + \dots, \text{ etc.}; \quad (17.4.26)$$

$$b_1(q) = 1 - q - (1/8)q^2 + (1/64)q^3 - (1/1536)q^4 + \dots, \quad (17.4.27)$$

$$b_2(q) = 4 - (1/12)q^2 + (5/13824)q^4 + \dots, \text{ etc.} \quad (17.4.28)$$

In each case the leading (the  $q$  independent) term is  $n^2$ .

Note that, according to (4.23), for our purposes we are interested in negative, and possibly quite negative, values of  $q$ .<sup>7</sup> Figures 4.4 and 4.5 display the first few  $a_n(q)$  and  $b_n(q)$  for negative values of  $q$ . Observe that, as  $q \rightarrow -\infty$ , the quantities  $a_{2m}(q)$  and  $a_{2m+1}(q)$ , for  $m = 0, 1, 2, 3, \dots$ , tend to agree. Similarly, for large negative  $q$ , the quantities  $b_{2m+1}(q)$  and  $b_{2m+2}(q)$ , for  $m = 0, 1, 2, 3, \dots$ , tend to agree. Indeed, it can be shown that there is the asymptotic behavior

$$\begin{aligned} a_{2m}(q) &\sim a_{2m+1}(q) \\ &\sim 2q + (8m+2)(-q)^{1/2} - (1/4)(8m^2 + 4m + 1) \\ &\quad - (1/32)(4m^2 + 2m + 1)(4m + 1)(-q)^{-1/2} + O(1/q) \\ &\quad \text{as } q \rightarrow -\infty \text{ for } m = 0, 1, 2, 3, \dots, \end{aligned} \quad (17.4.29)$$

$$\begin{aligned} b_{2m+1}(q) &\sim b_{2m+2}(q) \\ &\sim 2q + (8m+6)(-q)^{1/2} - (1/4)(8m^2 + 12m + 5) \\ &\quad - (1/32)(4m^2 + 6m + 3)(4m + 3)(-q)^{-1/2} + O(1/q) \\ &\quad \text{as } q \rightarrow -\infty \text{ for } m = 0, 1, 2, 3, \dots. \end{aligned} \quad (17.4.30)$$

We also remark, in passing, that there are the relations

$$a_n(-q) = a_n(q) \text{ for } n \text{ even}, \quad (17.4.31)$$

$$b_n(-q) = b_n(q) \text{ for } n \text{ even}, \quad (17.4.32)$$

$$a_n(-q) = b_n(q) \text{ for } n \text{ odd}, \quad (17.4.33)$$

$$b_n(-q) = a_n(q) \text{ for } n \text{ odd}. \quad (17.4.34)$$

---

<sup>6</sup>The reader might find confusing the use of the symbols  $a$ ,  $a_n$ , and  $b_n$  to denote separation constants. We agree, but it is standard in the Mathieu-equation literature.

<sup>7</sup>Unfortunately for our purposes, the Mathieu function literature treats primarily the  $q > 0$  case because this is the case that arises in the solution of the wave equation. See Exercise 4.1.

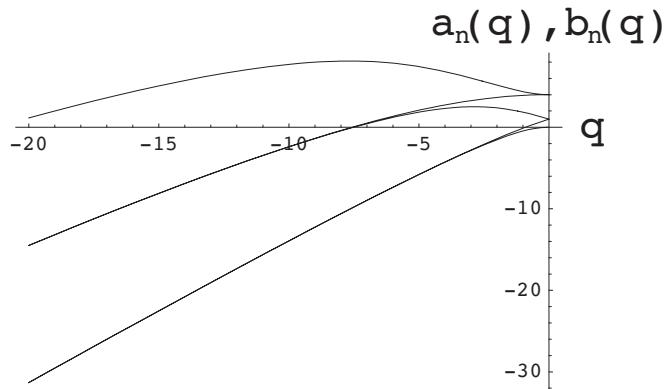


Figure 17.4.4: The functions  $a_0(q)$  through  $a_2(q)$  and  $b_1(q)$  and  $b_2(q)$  for negative values of  $q$ .

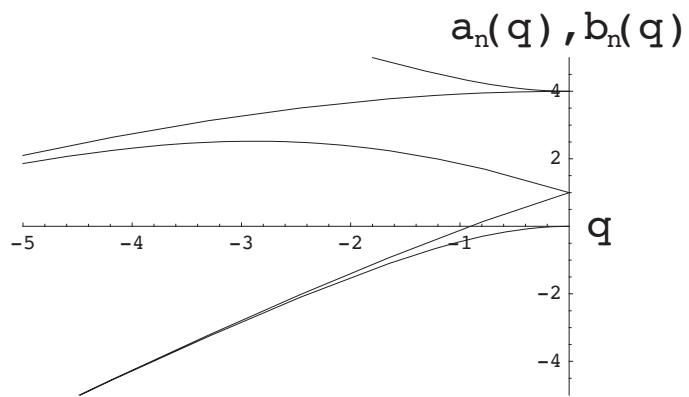


Figure 17.4.5: An enlargement of a portion of Figure 4.4. For  $q$  fixed and slightly negative, the curves, in order of increasing value, are  $a_0(q)$ ,  $a_1(q)$ ,  $b_1(q)$ ,  $b_2(q)$ , and  $a_2(q)$ . See (4.24) through (4.28). Note that the pair  $a_0(q)$  and  $a_1(q)$  tends to merge for large negative  $q$ , as does the pair  $b_1(q)$  and  $b_2(q)$ . Similarly, although not shown in this figure, the pair  $a_2(q)$  and  $a_3(q)$  tends to merge as does the pair  $b_3(q)$  and  $b_4(q)$ , etc. See (4.29) and (4.30).

The solutions associated with the separation constants  $a = a_n(q)$  are called  $\text{ce}_0(v, q)$ ,  $\text{ce}_1(v, q)$ ,  $\text{ce}_2(v, q)$ ,  $\text{ce}_3(v, q) \dots$ . They are *even* functions of  $v$  and, in the small  $q$  limit, are proportional to the functions 1,  $\cos(v)$ ,  $\cos(2v)$ ,  $\cos(3v)$ ,  $\dots$ . The solutions associated with the separation constants  $a = b_n(q)$  are called  $\text{se}_1(v, q)$ ,  $\text{se}_2(v, q)$ ,  $\text{se}_3(v, q)$ ,  $\dots$ . They are *odd* functions of  $v$  and, in the small  $q$  limit, are proportional to the functions  $\sin(v)$ ,  $\sin(2v)$ ,  $\sin(3v)$ ,  $\dots$ <sup>8</sup> Indeed, the  $\text{ce}_n(v, q)$  and  $\text{se}_n(v, q)$  are normalized so that in the limit  $q \rightarrow 0$  there are the relations

$$\text{ce}_0(v, 0) = 1/\sqrt{2}, \quad (17.4.35)$$

$$\text{ce}_n(v, 0) = \cos(nv) \text{ for } n \geq 1, \quad (17.4.36)$$

$$\text{se}_n(v, 0) = \sin(nv) \text{ for } n \geq 1. \quad (17.4.37)$$

Moreover, like their trigonometric counterparts, the functions  $\text{ce}_n(v, q)$  and  $\text{se}_n(v, q)$  form a complete set over the interval  $[0, 2\pi]$ . In fact, they form a complete orthogonal set and are normalized so that

$$\int_0^{2\pi} dv \text{ce}_m(v, q) \text{ce}_n(v, q) = \pi\delta_{mn}, \quad (17.4.38)$$

$$\int_0^{2\pi} dv \text{se}_m(v, q) \text{se}_n(v, q) = \pi\delta_{mn}, \quad (17.4.39)$$

$$\int_0^{2\pi} dv \text{ce}_m(v, q) \text{se}_n(v, q) = 0. \quad (17.4.40)$$

Apart from  $\text{ce}_0(v, q)$ , this normalization is like that of their trigonometric counterparts. See (4.35) through (4.37). Finally, again like their trigonometric counterparts, it can be shown that the functions  $\text{ce}_n(v, q)$  and  $\text{se}_n(v, q)$  have  $n$  zeroes in the half-open interval  $v \in [0, \pi]$ .

As noted earlier, we are primarily interested in the case  $q \leq 0$ . However we note for the record that, in concert with the relations (4.31) through (4.34), there are the relations

$$\text{ce}_{2n}(v, -q) = (-1)^n \text{ce}_{2n}(\pi/2 - v, q), \quad (17.4.41)$$

$$\text{ce}_{2n+1}(v, -q) = (-1)^n \text{se}_{2n+1}(\pi/2 - v, q), \quad (17.4.42)$$

$$\text{se}_{2n+1}(v, -q) = (-1)^n \text{ce}_{2n+1}(\pi/2 - v, q), \quad (17.4.43)$$

$$\text{se}_{2n+2}(v, -q) = (-1)^n \text{se}_{2n+2}(\pi/2 - v, q). \quad (17.4.44)$$

We will shortly present figures that display the first few  $\text{ce}_n(v, q)$  and  $\text{se}_n(v, q)$  as functions of  $v$ . Before doing so it is useful to look more closely at the terms appearing in the Mathieu equations. Inspired by both the analogy to Schrödinger's equation and the harmonic oscillator, rewrite (4.22) in the form

$$d^2Q/dv^2 - \lambda(v, q)Q = 0 \quad (17.4.45)$$

where

$$\lambda(v, q) = -[a - 2q \cos(2v)]. \quad (17.4.46)$$

---

<sup>8</sup>Note that, unlike their trigonometric counterparts  $\cos(nv)$  and  $\sin(nv)$ , the functions  $\text{ce}_n(v, q)$  and  $\text{se}_n(v, q)$  do not satisfy the *same* differential equation. This is because  $a_n(q) \neq b_n(q)$ .

From a Schrödinger perspective, we may view  $Q(v)$  as the wave function and  $\lambda(v, q)$  as the ‘potential’. In the harmonic oscillator analogy, we may view  $Q$  as the oscillator coordinate and  $-\lambda(v, q)$  as the instantaneous square of the time ( $v$ ) dependent frequency. With this background in mind, Figure 4.6 shows  $\lambda(v, q = -2)$  for various  $n$  values with  $a = a_n(q)$ . These are the potentials appropriate to the  $ce_n(v, q)$ . Similarly, Figure 4.7 shows  $\lambda(v, q = -2)$  for various  $n$  values with  $a = b_n(q)$ . These are the potentials appropriate to the  $se_n(v, q)$ . According to (4.45), understood in the harmonic oscillator analogy, when  $\lambda < 0$  we expect oscillatory behavior; and when  $\lambda > 0$  we expect exponentially growing or decaying behavior. From the Schrödinger perspective, the region where  $\lambda < 0$  is an allowed region, and the region where  $\lambda > 0$  is a forbidden or tunneling region. Inspection of Figures 4.6 and 4.7 shows that (when  $q = -2$ ) part of the  $v$  axis is forbidden for small  $n$  values, but that all of it is allowed once  $n$  becomes sufficiently large.

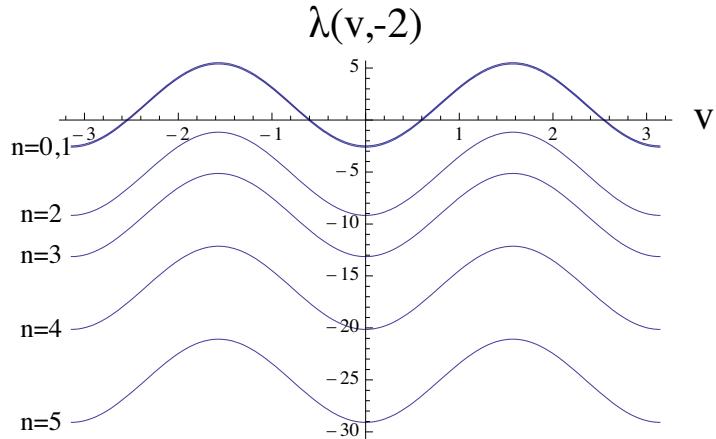


Figure 17.4.6: The effective potentials  $\lambda(v, q)$  for the  $ce_n(v, q)$  in the case  $q = -2$ . They are displayed as a function of  $v$ , over the interval  $[-\pi, \pi]$ , for various  $n$  values with  $a = a_n(q)$ . The top two curves, which very nearly coincide so as to almost look identical on the scale of the figure, are for the cases  $n = 0$  and  $n = 1$ . According to Figure 4.5, the curve for  $n = 0$  lies just slightly above that for  $n = 1$ . The bottom curve is that for  $n = 5$ . The curves in between are for  $n = 2, 3, 4$  in that order.

Figures 4.8 through 4.10 display the first few  $ce_n(v, q)$  as a function of  $v$  for  $q = -2$ . Figures 4.11 and 4.12 do the same for  $se_1(v, q)$  and  $se_2(v, q)$ . It can be shown, as a consequence of Poincaré’s theorem (see Section 1.3), that the  $ce_n(v, q)$  and  $se_n(v, q)$  are *entire* functions (analytic everywhere in the complex plane except at infinity) of  $v$ . Also, since the differential equation (4.22) has real coefficients for  $q$  real, the solutions  $ce_n(v, q)$  and  $se_n(v, q)$  are taken to be real for real  $q$  and real  $v$ .<sup>9</sup>

Observe from Figure 4.10 that  $ce_2(v, q)$  is freely oscillating. This is to be expected from Figure 4.6 because we see that for  $n \geq 2$  all of the  $v$  axis allowed. By contrast, Figure

<sup>9</sup>Since (4.22) is a second-order differential equation, there will also be second solutions that are linearly independent of the  $ce_n(v, q)$  when  $a = a_n(q)$ , and second solutions that are linearly independent of the  $se_n(v, q)$  when  $a = b_n(q)$ . Since the differential equation is invariant under parity (it is even in  $v$ ), these solutions could, for example, be taken to have the opposite parity of the  $ce_n(v, q)$  and the  $se_n(v, q)$ . They will not have period  $2\pi$ .

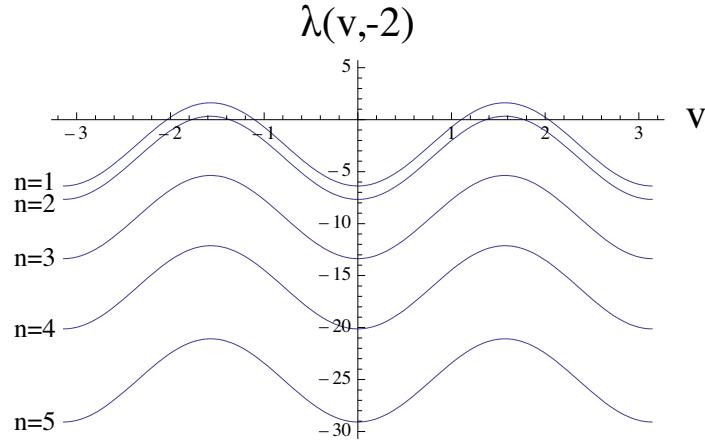


Figure 17.4.7: The effective potentials  $\lambda(v, q)$  for the  $se_n(v, q)$  in the case  $q = -2$ . They are displayed as a function of  $v$ , over the interval  $[-\pi, \pi]$ , for various  $n$  values with  $a = b_n(q)$ . The top curve is that for  $n = 1$ , and the bottom that for  $n = 5$ . The curves in between are for  $n = 2, 3, 4$  in that order.

4.8 shows that  $ce_0(v, q)$  does not change sign. This is because (as  $v$  increases) this solution enters a forbidden region for  $v \approx .6$  and at this point the function begins to decay. Again see Figure 4.6. Moreover, in the forbidden region, there is also a small exponentially growing part, with *positive* coefficient, that eventually dominates the solution by the time  $v = \pi/2$  so that the solution begins to grow beyond this point. Finally, for  $v > \approx 2.5$  the solution again enters an allowed region and begins to oscillate so that it has zero slope by the time  $v = \pi$ .

The case of  $ce_1(v, q)$  is more delicate. As has already been noted in the caption to Figure 4.6, when  $q = -2$  the potentials  $\lambda$  for  $a = a_0$  and for  $a = a_1$  are almost the same. Yet, inspection of Figures 4.8 and 4.9 shows that  $ce_0(v, q)$  and  $ce_1(v, q)$  are very different! Because  $\lambda|_{a_1} < \lambda|_{a_0}$ , the forbidden region for  $ce_1(v, q)$  is somewhat smaller than for  $ce_0(v, q)$ . Therefore  $ce_1(v, q)$  ‘oscillates’ a bit more before entering the forbidden region, and does so in such a way that the exponentially growing part in the forbidden region now has a negative sign. This exponentially growing part, although initially small in magnitude, eventually dominates at  $v = \pi/2$  so that  $ce_1(v, q)$  crosses through zero and continues on to become negative. Eventually  $v$  again reaches an allowed region and  $ce_1(v, q)$  begins to oscillate so that it has zero slope by the time  $v = \pi$ .

What about the behavior of the  $se_n(v, q)$ ? Figure 4.7 shows their effective potentials for the case  $q = -2$ . Evidently these potentials are all completely negative when  $n \geq 3$  and therefore the  $se_n(v, q)$  will be freely oscillatory when  $n \geq 3$ . Moreover, for  $n = 1$  the forbidden region is small, and for  $n = 2$  it is smaller yet. Therefore we expect the effects of the forbidden regions will be small. For example, the dips in  $se_1(v, q)$  at  $v = \pm\pi/2$ , see Figure 4.11, arise from the solution momentarily tunneling in forbidden regions. And examination of Figure 4.12 shows that, for  $se_2(v, q)$ , passage through the forbidden regions has little noticeable effect.

It is also instructive to examine the behavior of the  $ce_n(v, q)$  and  $se_n(v, q)$  when  $q$  has a much more negative value. Figures 4.13 and 4.14 show their effective potentials for the case

$\text{ce}_0(v, -2)$

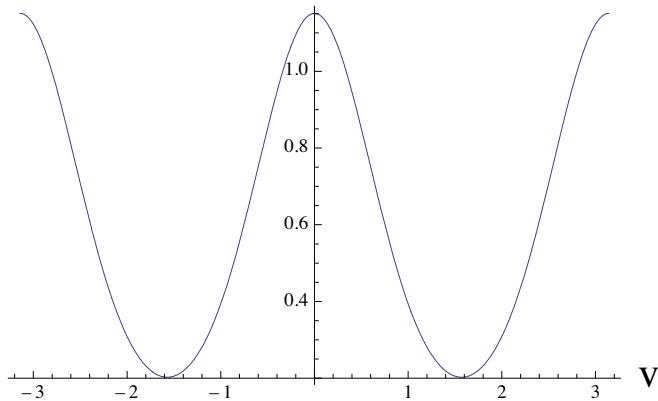


Figure 17.4.8: The function  $\text{ce}_0(v, q)$  as a function of  $v$ , over the interval  $[-\pi, \pi]$ , for  $q = -2$ . High magnification of this figure would reveal that the graph of  $\text{ce}_0(v, q)$  never touches or crosses, but always lies above, the  $v$  axis so that  $\text{ce}_0(v, q)$  has no zeroes.

$\text{ce}_1(v, -2)$

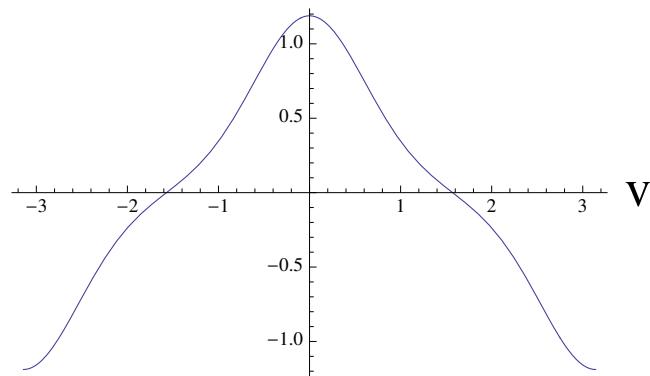


Figure 17.4.9: The function  $\text{ce}_1(v, q)$  as a function of  $v$ , over the interval  $[-\pi, \pi]$ , for  $q = -2$ .

$\text{ce}_2(v, -2)$

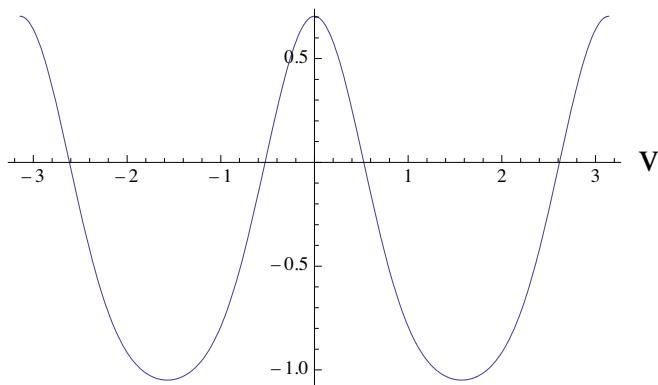


Figure 17.4.10: The function  $\text{ce}_2(v, q)$  as a function of  $v$ , over the interval  $[-\pi, \pi]$ , for  $q = -2$ .

$\text{se}_1(v, -2)$

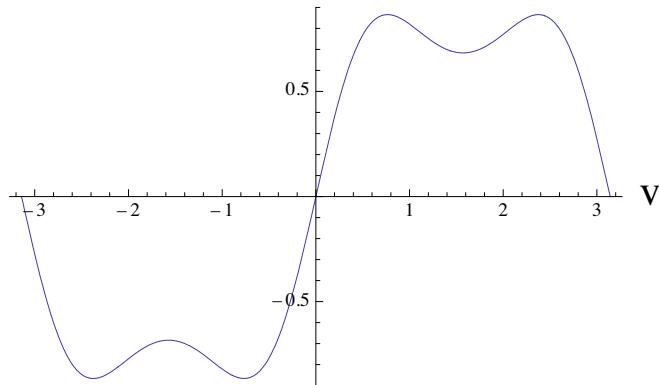


Figure 17.4.11: The function  $\text{se}_1(v, q)$  as a function of  $v$ , over the interval  $[-\pi, \pi]$ , for  $q = -2$ . The small dips at  $v = \pm\pi/2$  arise from passage through forbidden regions.

$\text{se}_2(v, -2)$

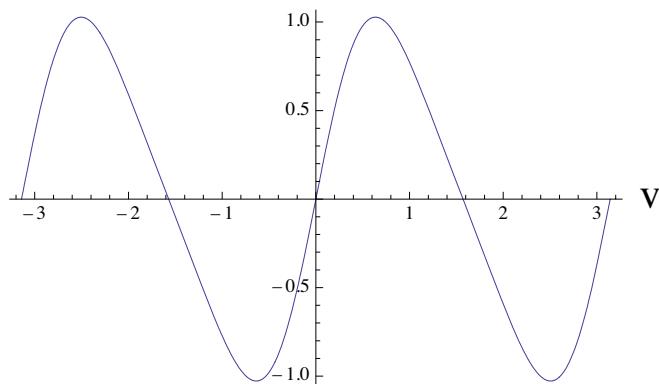


Figure 17.4.12: The function  $\text{se}_2(v, q)$  as a function of  $v$ , over the interval  $[-\pi, \pi]$ , for  $q = -2$ .

$q = -300$ . Now we see that, for modest values of  $n$ , the potentials are *positive* for most values of  $v$  save for small intervals where they are negative. Therefore large portions of the  $v$  axis are forbidden regions. Consequently, for modest  $n$  values, the functions  $\text{ce}_n(v, q)$  and  $\text{se}_n(v, q)$  contain exponentially decaying terms for most values of  $v$ , and are oscillatory only over small intervals. On the other hand, as  $n$  is increased, the forbidden regions become smaller and the allowed regions become larger until for sufficiently large  $n$  the entire  $v$  axis becomes an allowed region. Therefore, for sufficiently large  $n$ , the functions  $\text{ce}_n(v, q)$  and  $\text{se}_n(v, q)$  are fully oscillatory.

As an illustration of this expected behavior, Figures 4.15 through 4.17 display the  $\text{ce}_n(v, q)$  for  $q = -300$  and  $n = 0, 1, 2$ . We see that these functions begin bravely in the small allowed region about  $v = 0$ , rapidly decay to very nearly zero values in the forbidden regions centered about  $v = \pm\pi/2$ , and then rapidly revive in the allowed region centered about the (equivalent, due to periodicity) points  $v = \pm\pi$ . Compare these figures with their  $q = -2$  counterparts, Figures 4.8 through 4.10. By contrast, Figure 4.18 shows  $\text{ce}_{22}(v, q)$  for  $q = -300$ . It can be shown that in this case the effective potential  $\lambda(v, q = -300)$  is negative for all  $v$ . Therefore, in accord with Figure 4.18,  $\text{ce}_{22}(v, q)$  is fully oscillatory.

Similarly, Figures 4.19 and 4.20 display the  $\text{se}_n(v, q)$  for  $q = -300$  and  $n = 1, 2$ . Again we see these functions are very nearly zero in the forbidden regions. For example, the dips in Figure 4.11 have become, in Figure 4.19, *canyons* with very steep walls and very flat floors. By contrast, Figure 4.21 shows  $\text{se}_{23}(v, q)$  for  $q = -300$ . It can be shown that in this case the effective potential  $\lambda(v, q = -300)$  is negative for all  $v$ . Therefore, in accord with Figure 4.21,  $\text{se}_{23}(v, q)$  is fully oscillatory.

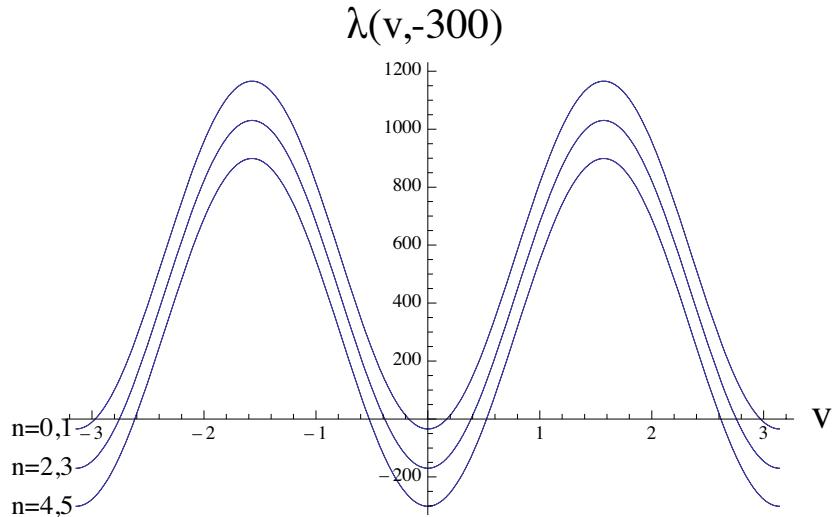


Figure 17.4.13: The effective potentials  $\lambda(v, q)$  for the  $\text{ce}_n(v, q)$  in the case  $q = -300$ . They are displayed as a function of  $v$ , over the interval  $[-\pi, \pi]$ , for the  $n$  values  $n = 0, 1, 2, 3, 4, 5$  with  $a = a_n(q)$ . The top two curves, which very nearly coincide so as to almost look identical on the scale of the figure, are for the cases  $n = 0$  and  $n = 1$ . The next two curves, which also nearly coincide, are for  $n = 2$  and  $n = 3$ . Finally, the bottom two curves also nearly coincide and are for the cases  $n = 4$  and  $n = 5$ . As in Figure 4.6, the higher the  $n$  value, the lower the curve.

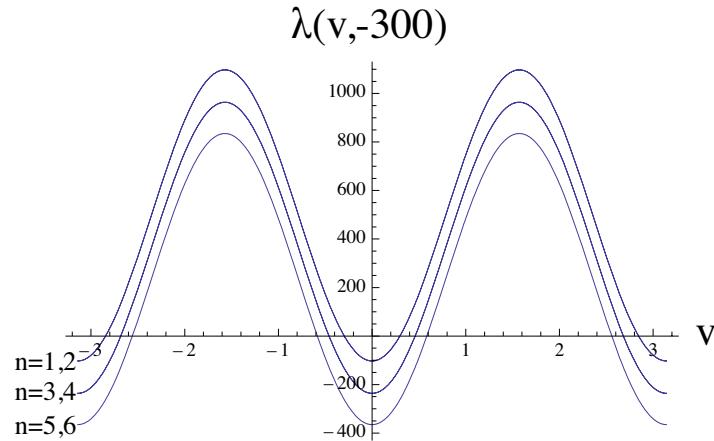


Figure 17.4.14: The effective potentials  $\lambda(v, q)$  for the  $se_n(v, q)$  in the case  $q = -300$ . They are displayed as a function of  $v$ , over the interval  $[-\pi, \pi]$ , for the  $n$  values  $n = 1, 2, 3, 4, 5, 6$  with  $a = b_n(q)$ . The top two curves, which very nearly coincide so as to almost look identical on the scale of the figure, are for the cases  $n = 1$  and  $n = 2$ . The next two curves, which also nearly coincide, are for  $n = 3$  and  $n = 4$ . Finally, the bottom two curves also nearly coincide and are for the cases  $n = 5$  and  $n = 6$ . As in Figure 4.7, the higher the  $n$  value, the lower the curve.

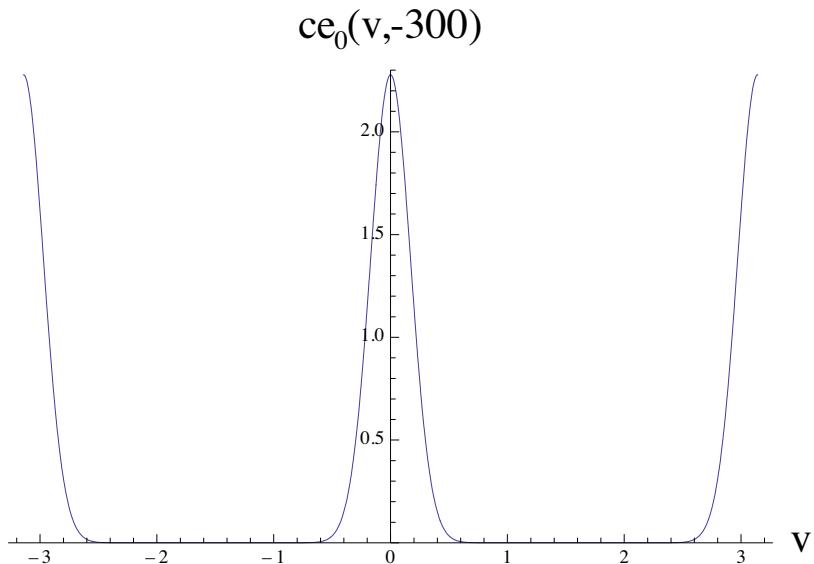


Figure 17.4.15: The function  $ce_0(v, q)$  as a function of  $v$ , over the interval  $[-\pi, \pi]$ , for  $q = -300$ . Most of the  $v$  axis is forbidden.

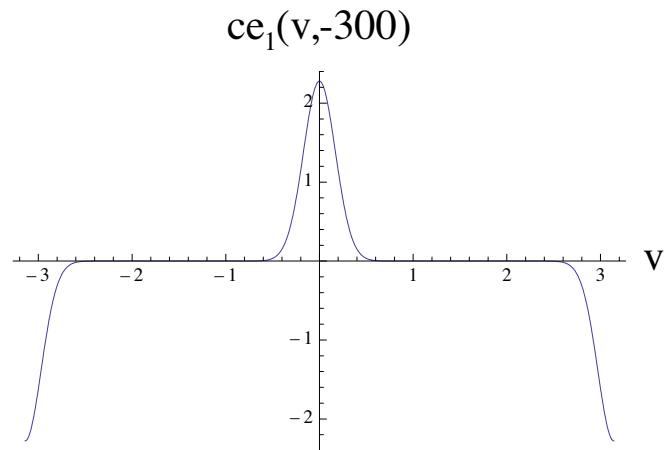


Figure 17.4.16: The function  $\text{ce}_1(v, q)$  as a function of  $v$ , over the interval  $[-\pi, \pi]$ , for  $q = -300$ . Most of the  $v$  axis is forbidden.

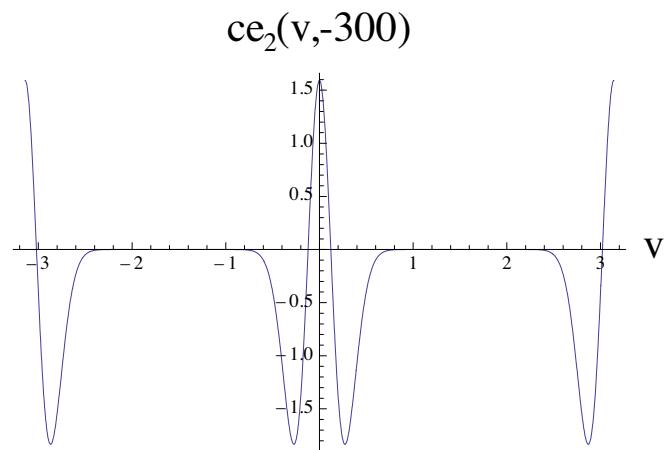


Figure 17.4.17: The function  $\text{ce}_2(v, q)$  as a function of  $v$ , over the interval  $[-\pi, \pi]$ , for  $q = -300$ . Most of the  $v$  axis is forbidden.

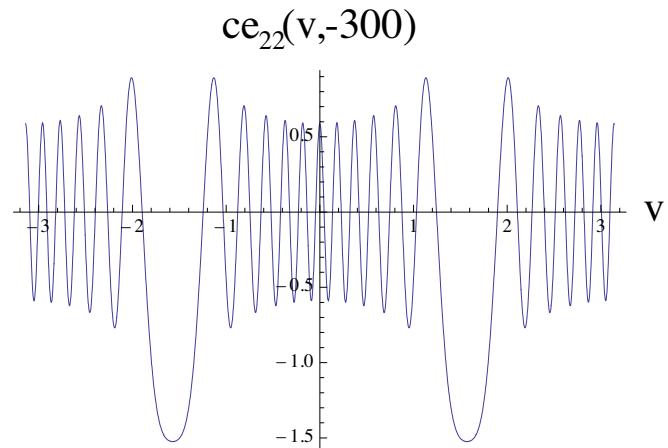


Figure 17.4.18: The function  $ce_{22}(v, q)$  as a function of  $v$ , over the interval  $[-\pi, \pi]$ , for  $q = -300$ . For these  $q$  and  $n$  values all of the  $v$  axis is allowed, and the function is fully oscillatory.

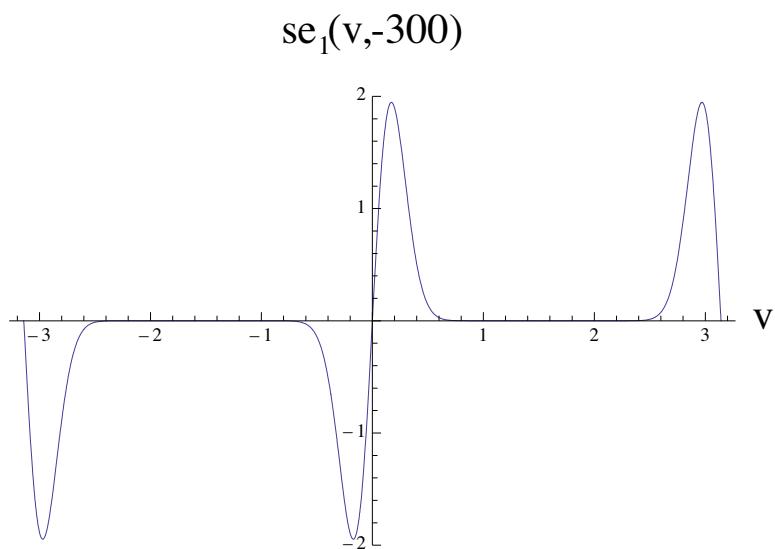


Figure 17.4.19: The function  $se_1(v, q)$  as a function of  $v$ , over the interval  $[-\pi, \pi]$ , for  $q = -300$ . Most of the  $v$  axis is forbidden.

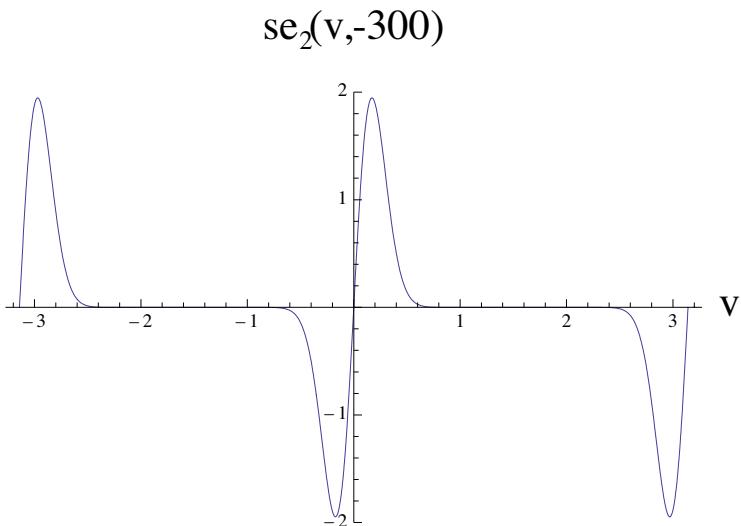


Figure 17.4.20: The function  $\text{se}_2(v, q)$  as a function of  $v$ , over the interval  $[-\pi, \pi]$ , for  $q = -300$ . Most of the  $v$  axis is forbidden.

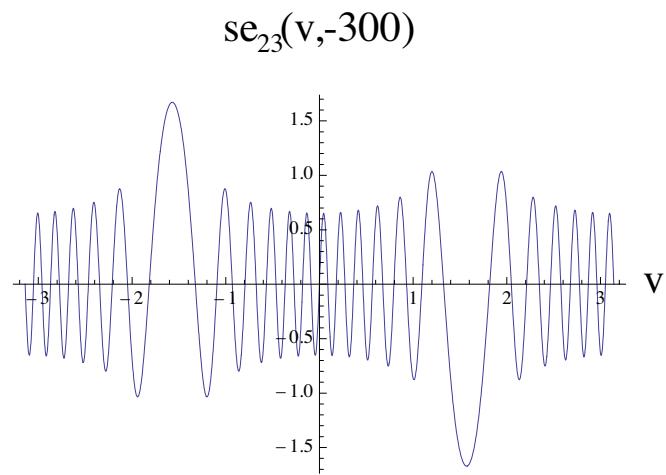


Figure 17.4.21: The function  $\text{se}_{23}(v, q)$  as a function of  $v$ , over the interval  $[-\pi, \pi]$ , for  $q = -300$ . For these  $q$  and  $n$  values all of the  $v$  axis is allowed, and the function is fully oscillatory.

Look again at the potentials  $\lambda(v, q)$  shown in Figures 4.6, 4.7, 4.13, and 4.14. We see, as is also evident from (4.46), that they have maxima at  $v = \pm\pi/2$ , and at these points they have the maximum values

$$\lambda_{\max}(q) = \lambda(\pm\pi/2, q) = -a - 2q. \quad (17.4.47)$$

Figure 4.22 shows  $\lambda_{\max}(q)$  for various  $n$  values in the case  $a = a_n(q)$ , and Figure 4.23 does the same in the case  $a = b_n(q)$ . From (4.29) and (4.30) we have the asymptotic formulas

$$\begin{aligned} \lambda_{\max}(q) &\sim -4q - (8m + 2)(-q)^{1/2} + \dots \\ \text{as } q \rightarrow -\infty \text{ for } a &= a_{2m}(q) \sim a_{2m+1}(q) \text{ and } m = 0, 1, 2, 3, \dots, \end{aligned} \quad (17.4.48)$$

$$\begin{aligned} \lambda_{\max}(q) &\sim -4q - (8m + 6)(-q)^{1/2} + \dots \\ \text{as } q \rightarrow -\infty \text{ for } a &= b_{2m+1}(q) \sim b_{2m+2}(q) \text{ and } m = 0, 1, 2, 3, \dots. \end{aligned} \quad (17.4.49)$$

We know that all of the interval  $v \in [-\pi, \pi]$  is allowed when  $\lambda_{\max}(q) < 0$ , and part of it becomes forbidden when  $\lambda_{\max}(q) > 0$ . Thus, for each  $n$  value and each alternative  $a = a_n(q)$  or  $a = b_n(q)$ , there is a *critical* value  $q_{\text{cr}}$  such that  $\lambda_{\max}(q_{\text{cr}}) = 0$ . From (4.47) we see that these critical values, in the two alternatives, are given (implicitly) by the relations

$$q_{\text{cr}}(n) = -(1/2)a_n[q_{\text{cr}}(n)], \quad (17.4.50)$$

$$q_{\text{cr}}(n) = -(1/2)b_n[q_{\text{cr}}(n)]. \quad (17.4.51)$$

These critical values, which can be read off from the ‘ $x$ ’ intercepts of the curves in Figures 4.22 and 4.23, are listed in Table 4.1. For a given value of  $n$ , all of the  $v$  axis is allowed and  $\text{ce}_n(v, q)$  is fully oscillatory if  $q > q_{\text{cr}}(n)$ , and otherwise part of the  $v$  axis is forbidden. Here  $q_{\text{cr}}(n)$  is to be calculated using (4.50). An analogous statement holds for  $\text{se}_n(v, q)$  where now  $q_{\text{cr}}(n)$  is to be calculated using (4.51).

At this point we are prepared to comment on the symmetry properties of the  $\text{ce}_n(v, q)$  and  $\text{se}_n(v, q)$ . We begin with the  $\text{ce}_n(v, q)$ . We know they are even and periodic with period  $2\pi$ . They therefore have Fourier series expansions consisting only of cosine terms. Also, consistent with the behavior of  $\text{ce}_0(v, q)$  and  $\text{ce}_2(v, q)$  displayed in Figures 4.8, 4.10, 4.15, and 4.17, it can be shown that the  $\text{ce}_n(v, q)$  for *even*  $n$  are *symmetric* about the point  $v = \pi/2$ . Specifically, the  $\text{ce}_n(v, q)$  for even  $n$  have Fourier expansions of the form

$$\text{ce}_n(v, q) = *1 + *\cos(2v) + *\cos(4v) + \dots \text{ for even } n \quad (17.4.52)$$

where the \*’s denote  $q$  and  $n$  dependent coefficients. That is, there is the relation

$$\text{ce}_n(\pi/2 + \Delta, q) = \text{ce}_n(\pi/2 - \Delta, q) \text{ for even } n. \quad (17.4.53)$$

It follows that the  $\text{ce}_n(v, q)$  for even  $n$  have vanishing first derivative at  $v = \pi/2$ ,

$$\text{ce}'_n(\pi/2, q) = 0 \text{ for even } n. \quad (17.4.54)$$

By contrast, as illustrated in Figures 4.9 and 4.16, the  $\text{ce}_n(v, q)$  for *odd*  $n$  are *antisymmetric* about the point  $v = \pi/2$  and have Fourier expansions of the form

$$\text{ce}_n(v, q) = *\cos(v) + *\cos(3v) + *\cos(5v) + \dots \text{ for odd } n. \quad (17.4.55)$$

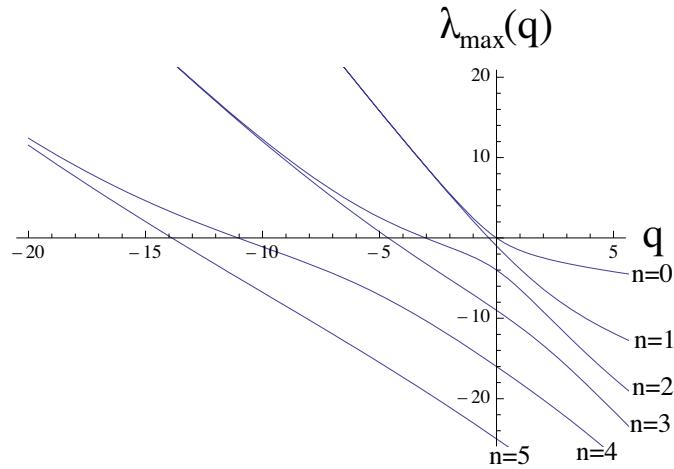


Figure 17.4.22: The function  $\lambda_{\max}(q)$  for the  $n$  values 0 through 5 in the case  $a = a_n(q)$ . When  $\lambda_{\max}(q) < 0$ , all of the  $v$  axis is allowed, and the function  $ce_n(v, q)$  is fully oscillatory. When  $\lambda_{\max}(q) > 0$ , part of the  $v$  axis is forbidden. The higher the  $n$  value, the lower the curve. Note that the ‘ $y$ ’ intercepts have the values  $-n^2$  in accord with (4.24) through (4.26) and (4.47). The ‘ $x$ ’ intercepts are the values  $q_{\text{cr}}(n)$ . Note also that the values of  $\lambda_{\max}(q)$  for  $n = 0$  and  $n = 1$  tend to merge for large negative  $q$ , as do the values for  $n = 2$  and  $n = 3$ , etc. See Figure 4.5 and (4.48).

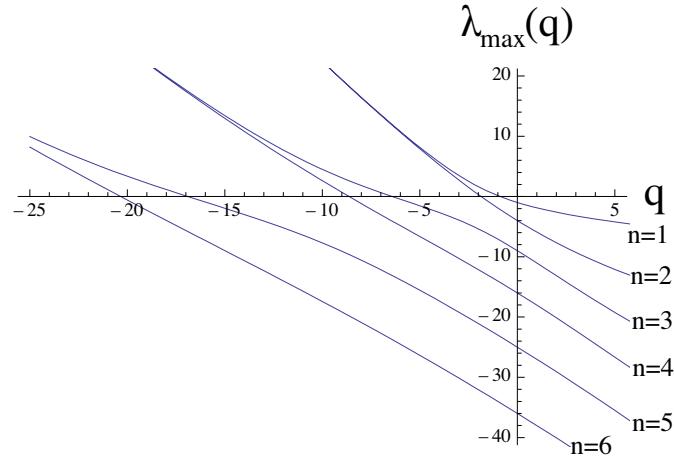


Figure 17.4.23: The function  $\lambda_{\max}(q)$  for the  $n$  values 1 through 6 in the case  $a = b_n(q)$ . When  $\lambda_{\max}(q) < 0$ , all of the  $v$  axis is allowed, and the function  $se_n(v, q)$  is fully oscillatory. When  $\lambda_{\max}(q) > 0$ , part of the  $v$  axis is forbidden. The higher the  $n$  value, the lower the curve. Note that the ‘ $y$ ’ intercepts have the values  $-n^2$  in accord with (4.27), (4.28), and (4.47). The ‘ $x$ ’ intercepts are the values  $q_{\text{cr}}(n)$ . Note also that the values of  $\lambda_{\max}(q)$  for  $n = 1$  and  $n = 2$  tend to merge for large negative  $q$ , as do the values for  $n = 3$  and  $n = 4$ , etc. See Figure 4.5 and (4.49).

Table 17.4.1: The quantity  $q_{\text{cr}}(n)$  for various values of  $n$ .

$n$	$q_{\text{cr}}(n)$ when $a = a_n(q)$	$q_{\text{cr}}(n)$ when $a = b_n(q)$
0	0	*
1	-0.329005727826915	-0.889819993831662
2	-3.039073671630782	-1.8582116914842934
3	-4.626950799904568	-6.425863307211811
4	-11.047992936386709	-8.6316091625993501
5	-13.871128836603399	-16.904741557017985
6	-23.995780230075020	-20.345062417526364
7	-28.0531793998642485	-32.320930596434941
8	-41.880084880521011	-36.995345508020719
9	-47.171475670427398	-52.673172894843788
10	-64.7001463432813892	-58.581512590132760

That is, there is the relation

$$\text{ce}_n(\pi/2 + \Delta, q) = -\text{ce}_n(\pi/2 - \Delta, q) \text{ for odd } n. \quad (17.4.56)$$

It follows that the  $\text{ce}_n(v, q)$  for odd  $n$  vanish at  $v = \pi/2$ ,

$$\text{ce}_n(\pi/2, q) = 0 \text{ for odd } n. \quad (17.4.57)$$

Next consider the symmetry properties of the  $\text{se}_n(v, q)$ . We know they are odd and periodic with period  $2\pi$ . They therefore have Fourier series expansions consisting only of sine terms. Also, consistent with the behavior of  $\text{se}_2(v, q)$  displayed in Figures 4.12 and 4.20, it can be shown that the  $\text{se}_n(v, q)$  for even  $n$  are antisymmetric about the point  $v = \pi/2$ . Specifically, the  $\text{se}_n(v, q)$  for even  $n$  have Fourier expansions of the form

$$\text{se}_n(v, q) = * \sin(2v) + * \sin(4v) + * \sin(6v) + \dots \text{ for even } n. \quad (17.4.58)$$

That is, there is the relation

$$\text{se}_n(\pi/2 + \Delta, q) = -\text{se}_n(\pi/2 - \Delta, q) \text{ for even } n, \quad (17.4.59)$$

from which it follows that

$$\text{se}_n(\pi/2, q) = 0 \text{ for even } n. \quad (17.4.60)$$

By contrast, as illustrated in Figures 4.11 and 4.19 for  $\text{se}_1(v, q)$ , the  $\text{se}_n(v, q)$  for odd  $n$  are symmetric about the point  $v = \pi/2$  and have Fourier expansions of the form

$$\text{se}_n(v, q) = * \sin(v) + * \sin(3v) + * \sin(5v) + \dots \text{ for odd } n. \quad (17.4.61)$$

That is, there is the relation

$$\text{se}_n(\pi/2 + \Delta, q) = \text{se}_n(\pi/2 - \Delta, q) \text{ for odd } n, \quad (17.4.62)$$

from which it follows that

$$\text{se}'_n(\pi/2, q) = 0 \text{ for odd } n. \quad (17.4.63)$$

Finally, it follows from (4.52) and (4.58) that

$$\text{ce}_n(v + \pi, q) = \text{ce}_n(v, q) \text{ for even } n \quad (17.4.64)$$

and

$$\text{se}_n(v + \pi, q) = \text{se}_n(v, q) \text{ for even } n. \quad (17.4.65)$$

Thus the  $\text{se}_n(v, q)$  and  $\text{se}_n(v, q)$  for even  $n$  have period  $\pi$  as well as period  $2\pi$ . By contrast, we see from (4.55) and (4.61) that there are the relations

$$\text{ce}_n(v + \pi, q) = -\text{ce}_n(v, q) \text{ for odd } n \quad (17.4.66)$$

and

$$\text{se}_n(v + \pi, q) = -\text{se}_n(v, q) \text{ for odd } n. \quad (17.4.67)$$

The relations (4.64) through (4.67) can be written more succinctly in the form

$$\text{ce}_n(v + \pi, q) = (-1)^n \text{ce}_n(v, q), \quad (17.4.68)$$

$$\text{se}_n(v + \pi, q) = (-1)^n \text{se}_n(v, q). \quad (17.4.69)$$

From a computational perspective, an important consequence of these symmetry properties of the  $\text{ce}_n(v, q)$  and  $\text{se}_n(v, q)$  is that they only need to be computed over the interval  $v \in [0, \pi/2]$ . Their values elsewhere are then determined by their symmetry properties. Moreover, if  $q$  and  $n$  are such that a value of  $v$  is deep within a strongly forbidden region, then we may set the associated value of  $\text{ce}_n(v, q)$  or  $\text{se}_n(v, q)$  to zero for these values of  $v$ . Recall that the forbidden regions are centered about the values  $v = \pm\pi/2$ . Thus, if there are such strongly forbidden regions, we only need to compute  $\text{ce}_n(v, q)$  or  $\text{se}_n(v, q)$  over the smaller interval  $v \in [0, v_{\text{deep}}]$  where  $v_{\text{deep}}$  is the smallest  $v$  value deep within the strongly forbidden region.

### 17.4.5 Modified Mathieu Functions

Now that the possible values of the separation constant  $a$  have been determined by the periodicity requirement, these values of  $a$  can be employed in (4.21) to determine the functions  $P(u)$ . The so-called solutions of the *first kind* for (4.21), when  $a = a_n(q)$ , are denoted as  $\text{Ce}_n(u, q)$ ; and the solutions of the first kind, when  $a = b_n(q)$ , are denoted as  $\text{Se}_n(u, q)$ . The functions  $\text{Ce}_n(u, q)$  are even functions of  $u$  and the functions  $\text{Se}_n(u, q)$  are odd functions of  $u$ . They can be conveniently arranged to satisfy the relations

$$\text{Ce}_n(u, q) = \text{ce}_n(iu, q), \quad (17.4.70)$$

$$\text{Se}_n(u, q) = -i\text{se}_n(iu, q). \quad (17.4.71)$$

Evidently, they are also entire functions, and they are also real for  $q$  and  $u$  real. Since the  $\text{ce}_n(v, q)$  and  $\text{se}_n(v, q)$  are analogous to cosines and sines, see (4.52), (4.55), (4.58), and

(4.61), the relations (4.70) and (4.71) indicate that the  $\text{Ce}_n(u, q)$  and  $\text{Se}_n(u, q)$  are analogous to hyperbolic cosines and hyperbolic sines.

Suppose we write (4.21) in the form

$$d^2P/du^2 - \Lambda(u, q)P = 0 \quad (17.4.72)$$

where

$$\Lambda(u, q) = a - 2q \cosh(2u). \quad (17.4.73)$$

Figure 4.24 shows these  $\Lambda(u, q)$  for various  $n$  values with  $a = a_n(q)$  and  $q = -2$ . These are the potentials appropriate to the  $\text{Ce}_n(u, q)$ . The potentials for the  $\text{Se}_n(u, q)$  computed with  $a = b_n(q)$  are similar. Inspection of these  $\text{Ce}_n(u, q)$  potentials shows that they are all *positive* for all  $u$ . The same can be shown to be true for the  $\text{Se}_n(u, q)$  potentials.

In fact, more can be said. From (4.73) it is evident that, when  $q < 0$  (which is what we have assumed),  $\Lambda$  has a minimum at  $u = 0$  and, at this point has the value

$$\Lambda(0, q) = a - 2q. \quad (17.4.74)$$

Therefore, if we can show that

$$\Lambda(0, q) = a - 2q > 0 \quad (17.4.75)$$

for all  $q < 0$  and all  $n$  with  $a = a_n(q)$  or  $a = b_n(q)$ , then we will have shown that all  $\Lambda$  are positive for all  $u$ . Figure 4.25 displays  $\Lambda(0, q)$  for various  $n$  values when  $a = a_n(q)$ , and Figure 4.26 does the same for the case  $a = b_n(q)$ . Note also that from (4.29) and (4.30) there is the asymptotic behavior

$$\begin{aligned} \Lambda(0, q) &\sim (8m + 2)(-q)^{1/2} + \dots \\ \text{as } q \rightarrow -\infty \text{ for } a &= a_{2m}(q) \sim a_{2m+1}(q) \text{ and } m = 0, 1, 2, 3, \dots, \end{aligned} \quad (17.4.76)$$

$$\begin{aligned} \Lambda(0, q) &\sim (8m + 6)(-q)^{1/2} + \dots \\ \text{as } q \rightarrow -\infty \text{ for } a &= b_{2m+1}(q) \sim b_{2m+2}(q) \text{ and } m = 0, 1, 2, 3, \dots. \end{aligned} \quad (17.4.77)$$

Evidently (4.75) is always satisfied when  $q < 0$ . We conclude that for  $q \leq 0$  the  $\text{Ce}_n(u, q)$  and  $\text{Se}_n(u, q)$  are non-oscillatory and must all be exponentially growing.<sup>10</sup>

As examples, Figures 4.27 and 4.28 display the first few  $\text{Ce}_n(u, q)$  and  $\text{Se}_n(u, q)$  as a function of  $u$  for  $q = -2$ .<sup>11</sup> We see that, as predicted, they are non-oscillatory and their magnitudes do indeed become large for large values of  $|u|$ .

---

<sup>10</sup>Also, as Figures 4.25 and 4.26 illustrate, oscillatory behavior for the  $\text{Ce}_n(u, q)$  and  $\text{Se}_n(u, q)$  is possible if  $q > 0$ . See Exercise 4.1.

<sup>11</sup>Since the modified Mathieu equation (4.21) is also of second order, it will also have second solutions that are linearly independent of the  $\text{Ce}_n(u, q)$  and  $\text{Se}_n(u, q)$ . Because it too is invariant under parity, these solutions could be constructed to have parities opposite to those of the  $\text{Ce}_n(u, q)$  and  $\text{Se}_n(u, q)$ .

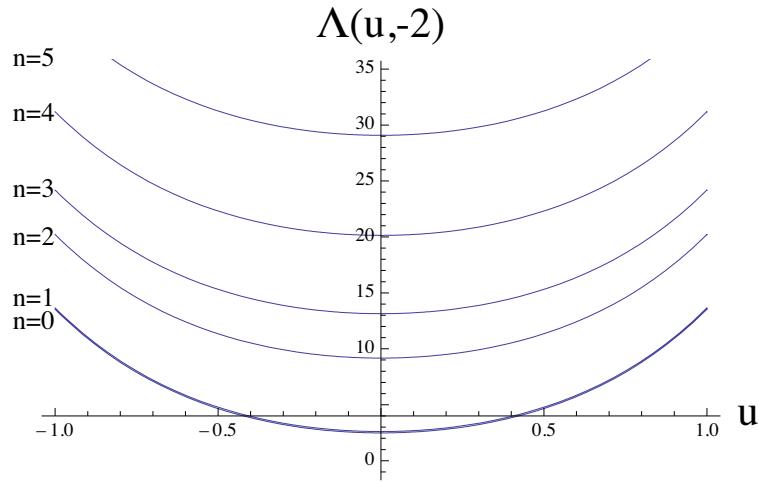


Figure 17.4.24: The effective potentials  $\Lambda(u, q)$  for the  $Ce_n(v, q)$  in the case  $q = -2$ . They are displayed as a function of  $u$  for the  $n$  values  $n = 0, 1, 2, 3, 4, 5$  with  $a = a_n(q)$ . As in Figure 4.6, the curves for  $n = 0$  and  $n = 1$  nearly coincide. Now, because of the difference in sign between (4.46) and (4.73), the higher the  $n$  value the higher the curve.

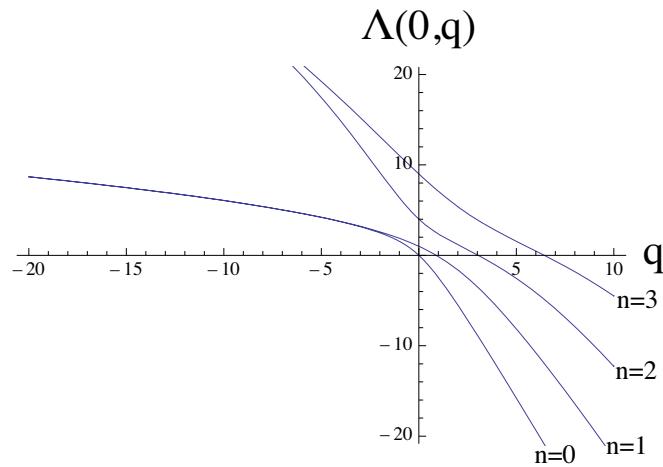


Figure 17.4.25: The quantities  $\Lambda(0, q)$  with  $a = a_n(q)$  and  $n = 0, 1, 2, 3$ . The higher the  $n$  value, the higher the curve. Note that the ‘ $y$ ’ intercepts have the values  $n^2$  in agreement with (4.24) through (4.26). Also, values of  $\Lambda(0, q)$  for  $n = 0$  and  $n = 1$  tend to merge for large negative  $q$ , as do the values for  $n = 2$  and  $n = 3$ , etc. See Figure 4.5 and (4.76).

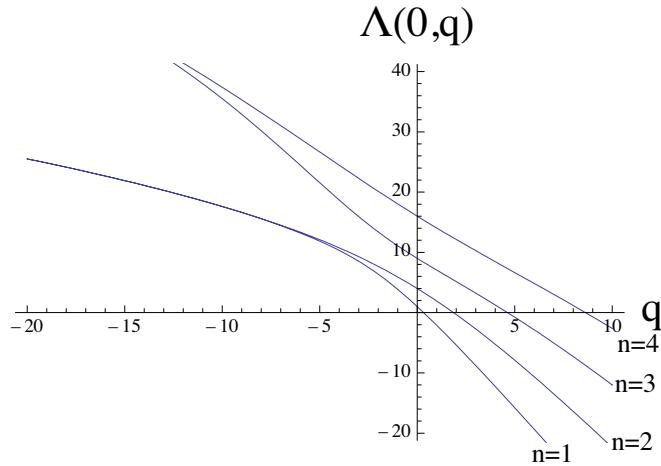


Figure 17.4.26: The quantities  $\Lambda(0, q)$  with  $a = b_n(q)$  and  $n = 1, 2, 3, 4$ . The higher the  $n$  value, the higher the curve. Note that the ‘ $y$ ’ intercepts have the values  $n^2$  in agreement with (4.27) and (4.28). Also, values of  $\Lambda(0, q)$  for  $n = 1$  and  $n = 2$  tend to merge for large negative  $q$ , as do the values for  $n = 3$  and  $n = 4$ , etc. See Figure 4.5 and (4.77).

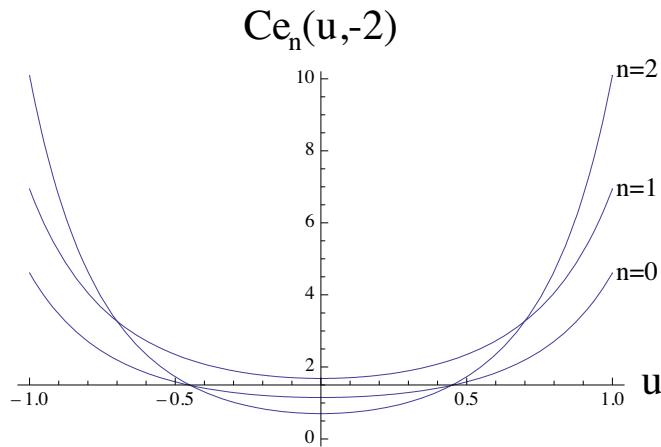


Figure 17.4.27: The functions  $Ce_0(u, q)$  through  $Ce_2(u, q)$ , as a function of  $u$ , for  $q = -2$ . At  $u = 1$  they satisfy the inequalities  $Ce_0(1, -2) < Ce_1(1, -2) < Ce_2(1, -2)$ .

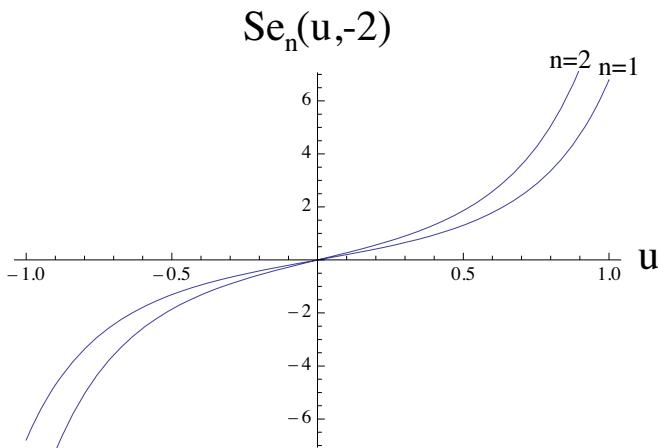


Figure 17.4.28: The functions  $\text{Se}_1(u, q)$  and  $\text{Se}_2(u, q)$ , as a function of  $u$ , for  $q = -2$ . At  $u = 1$  they satisfy the inequality  $\text{Se}_1(1, -2) < \text{Se}_2(1, -2)$ .

#### 17.4.6 Analyticity in $x$ and $y$

So far we have described the functions  $\text{ce}_n(v, q)$  and  $\text{se}_n(v, q)$ , and the functions  $\text{Ce}_n(u, q)$  and  $\text{Se}_n(u, q)$ , and have seen that they are entire functions of the variables  $u$  and  $v$ . But this does not mean that they are entire functions of the variables  $x$  and  $y$  because the relation (4.7) that connects  $u, v$  to  $x, y$  has singularities. See Exercise 4.2. Remarkably, however, the products  $[\text{Ce}_n(u, q) \times \text{ce}_n(v, q)]$  and  $[\text{Se}_n(u, q) \times \text{se}_n(v, q)]$ , which according to (4.13) is what we hope to use to construct solutions of the Laplace equation, are entire functions of the variables  $x$  and  $y$ . This situation is analogous to the case of cylindrical coordinates where the functions  $\exp(im\phi)$  and  $I_m(k\rho)$  are entire functions of  $\phi$  and  $\rho$ , respectively, but are not entire functions of  $x$  and  $y$ . However, the products  $[\exp(im\phi) \times I_m(k\rho)]$  are entire functions of  $x$  and  $y$ . In the case of the Mathieu functions, for example, there is an integral representation of the form

$$\text{Ce}_{2n}(u, q) \text{ ce}_{2n}(v, q) = p_{2n}(q) \int_0^{\pi/2} d\tau \text{ ce}_{2n}(\tau, q) \cosh[kx \cos(\tau)] \cosh[ky \sin(\tau)] \quad (17.4.78)$$

where  $p_{2n}(q)$  is some  $q$ -dependent coefficient. There are similar representations for the other relevant products. The right side of (4.78) is manifestly an entire function of  $x$  and  $y$ .

#### 17.4.7 Elliptic Cylinder Harmonic Expansion and On-Axis Gradients

The stage is now set to describe the expansion of any harmonic function  $\psi$  in terms of Mathieu functions. The general harmonic function that is analytic in  $x$  and  $y$  near the

origin can be written in the form

$$\begin{aligned}\psi(x, y, z) = \psi(u, v, z) &= \sum_{n=0}^{\infty} \int_{-\infty}^{\infty} dk c_n(k) \exp(ikz) \text{Ce}_n(u, q) \text{ce}_n(v, q) \\ &+ \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} dk s_n(k) \exp(ikz) \text{Se}_n(u, q) \text{se}_n(v, q)\end{aligned}\quad (17.4.79)$$

where the functions  $c_n(k)$  and  $s_n(k)$  are arbitrary. We will call (4.79) an *elliptic cylinder harmonic expansion*.

To exploit this expansion, suppose the magnetic field  $\mathbf{B}(x, y, z)$  is interpolated onto the surface  $u = U$  of an elliptic cylinder using values at the grid points near the surface. See Figure 4.3. Let us employ the notation  $\mathbf{B}(x, y, z) = \mathbf{B}(u, v, z)$  so that the magnetic field on the surface can be written as  $\mathbf{B}(U, v, z)$ . Next, from the values on the surface, compute  $B_u(U, v, z)$ , the component of  $\mathbf{B}(x, y, z)$  *normal* to the surface. Our aim will be to determine the on-axis gradients from a knowledge of  $B_u(U, v, z)$ .

Let us begin by solving (4.11) for  $(\partial\psi/\partial u)$ . We find, using (4.4), the result,

$$\begin{aligned}(\partial\psi/\partial u) &= f[\cosh^2(u) - \cos^2(v)]^{1/2} B_u \\ &= f(\sinh u \cos v) B_x(u, v, z) + f(\cosh u \sin v) B_y(u, v, z).\end{aligned}\quad (17.4.80)$$

We see that the right side of (4.80) is a well-behaved function  $F(u, v, z)$  whose values are known for  $u = U$ ,

$$F(U, v, z) = f(\sinh U \cos v) B_x(U, v, z) + f(\cosh U \sin v) B_y(U, v, z). \quad (17.4.81)$$

Moreover, using the representation (4.79) in (4.80) and (4.81), we may also write

$$\begin{aligned}F(U, v, z) &= \sum_{n=0}^{\infty} \int_{-\infty}^{\infty} dk c_n(k) \exp(ikz) \text{Ce}'_n(U, q) \text{ce}_n(v, q) \\ &+ \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} dk s_n(k) \exp(ikz) \text{Se}'_n(U, q) \text{se}_n(v, q).\end{aligned}\quad (17.4.82)$$

Next multiply both sides of (4.82) by  $\exp(-ik'z)$  and integrate over  $z$ . So doing gives the result

$$\begin{aligned}(1/2\pi) \int_{-\infty}^{\infty} dz \exp(-ik'z) F(U, v, z) &= \\ \sum_{n=0}^{\infty} c_n(k') \text{Ce}'_n(U, q') \text{ce}_n(v, q') &+ \sum_{n=1}^{\infty} s_n(k') \text{Se}'_n(U, q') \text{se}_n(v, q').\end{aligned}\quad (17.4.83)$$

Now, employ the orthogonality properties (4.38) through (4.40) to obtain the relations

$$c_r(k') \text{Ce}'_r(U, q') = [1/(2\pi^2)] \int_0^{2\pi} dv \int_{-\infty}^{\infty} dz \exp(-ik'z) \text{ce}_r(v, q') F(U, v, z), \quad (17.4.84)$$

$$s_r(k') \text{Se}'_r(U, q') = [1/(2\pi^2)] \int_0^{2\pi} dv \int_{-\infty}^{\infty} dz \exp(-ik'z) \text{se}_r(v, q') F(U, v, z). \quad (17.4.85)$$

In view of (4.84) and (4.85), define the function  $\tilde{F}(v, k)$  by the rule

$$\tilde{F}(v, k) = [1/(2\pi)] \int_{-\infty}^{\infty} dz \exp(-ikz) F(U, v, z), \quad (17.4.86)$$

and define functions  $\tilde{F}_r^c(k)$  and  $\tilde{F}_r^s(k)$  by the rules

$$\begin{aligned} \tilde{F}_r^c(k) &= (1/\pi) \int_0^{2\pi} dv \text{ce}_r(v, q) \tilde{F}(v, k) \\ &= [1/(2\pi^2)] \int_0^{2\pi} dv \int_{-\infty}^{\infty} dz \exp(-ikz) \text{ce}_r(v, q) F(U, v, z), \end{aligned} \quad (17.4.87)$$

$$\begin{aligned} \tilde{F}_r^s(k) &= (1/\pi) \int_0^{2\pi} dv \text{se}_r(v, q) \tilde{F}(v, k) \\ &= [1/(2\pi^2)] \int_0^{2\pi} dv \int_{-\infty}^{\infty} dz \exp(-ikz) \text{se}_r(v, q) F(U, v, z). \end{aligned} \quad (17.4.88)$$

[Here we have extended the use of the  $\sim$  notation to include angular *Mathieu* transforms, such as those in (4.87) and (4.88), where  $\cos(r\phi)$  and  $\sin(r\phi)$  are replaced by  $\text{ce}_r(v, q)$  and  $\text{se}_r(v, q)$ .] We will call the functions  $\tilde{F}_r^\alpha(k)$  *Mathieu coefficient* functions in analogy to the Fourier coefficients that arise in Fourier analysis. Note that, because  $F$  and the Mathieu functions are real, the real parts of the functions  $\tilde{F}_r^\alpha(k)$  are even in  $k$ , and the imaginary parts are odd in  $k$ .

With these definitions, the relations (4.84) and (4.85) can be rewritten in the form

$$c_r(k) = \tilde{F}_r^c(k)/\text{Ce}'_r(U, q), \quad (17.4.89)$$

$$s_r(k) = \tilde{F}_r^s(k)/\text{Se}'_r(U, q). \quad (17.4.90)$$

Finally, employ (4.89) and (4.90) in (4.79). So doing gives the result

$$\begin{aligned} \psi(x, y, z) &= \sum_{r=0}^{\infty} \int_{-\infty}^{\infty} dk \exp(ikz) [\tilde{F}_r^c(k)/\text{Ce}'_r(U, q)] \text{Ce}_r(u, q) \text{ce}_r(v, q) \\ &\quad + \sum_{r=1}^{\infty} \int_{-\infty}^{\infty} dk \exp(ikz) [\tilde{F}_r^s(k)/\text{Se}'_r(U, q)] \text{Se}_r(u, q) \text{se}_r(v, q). \end{aligned} \quad (17.4.91)$$

We have obtained an elliptical cylinder harmonic expansion for  $\psi$  in terms of surface field data.

Of course, what we really want are the on-axis gradients. They can be found by employing two remarkable *connections* (identities) between elliptic and circular cylinder functions of the form

$$\text{Ce}_r(u, q) \text{ ce}_r(v, q) = \sum_{m=0}^{\infty} \alpha_m^r(k) I_m(k\rho) \cos(m\phi), \quad (17.4.92)$$

$$\text{Se}_r(u, q) \text{ se}_r(v, q) = \sum_{m=1}^{\infty} \beta_m^r(k) I_m(k\rho) \sin(m\phi). \quad (17.4.93)$$

For further reference, we will call the quantities  $\alpha_m^r(k)$  and  $\beta_m^r(k)$  *Mathieu-Bessel connection coefficients*. Let us employ these identities in (4.91) to find the results

$$\begin{aligned} & \sum_{r=0}^{\infty} \int_{-\infty}^{\infty} dk \exp(ikz) [\tilde{F}_r^c(k)/\text{Ce}'_r(U, q)] \text{Ce}_r(u, q) \text{ ce}_r(v, q) \\ &= \sum_{m=0}^{\infty} \int_{-\infty}^{\infty} dk \exp(ikz) I_m(k\rho) \cos(m\phi) \sum_{r=0}^{\infty} \alpha_m^r(k) [\tilde{F}_r^c(k)/\text{Ce}'_r(U, q)], \end{aligned} \quad (17.4.94)$$

and

$$\begin{aligned} & \sum_{r=1}^{\infty} \int_{-\infty}^{\infty} dk \exp(ikz) [\tilde{F}_r^s(k)/\text{Se}'_r(U, q)] \text{Se}_r(u, q) \text{ se}_r(v, q) \\ &= \sum_{m=1}^{\infty} \int_{-\infty}^{\infty} dk \exp(ikz) I_m(k\rho) \sin(m\phi) \sum_{r=1}^{\infty} \beta_m^r(k) [\tilde{F}_r^s(k)/\text{Se}'_r(U, q)]. \end{aligned} \quad (17.4.95)$$

Using these results, (4.91) can be rewritten in the form

$$\begin{aligned} \psi(x, y, z) &= \sum_{m=0}^{\infty} \int_{-\infty}^{\infty} dk \exp(ikz) I_m(k\rho) \cos(m\phi) \sum_{r=0}^{\infty} \alpha_m^r(k) [\tilde{F}_r^c(k)/\text{Ce}'_r(U, q)] \\ &+ \sum_{m=1}^{\infty} \int_{-\infty}^{\infty} dk \exp(ikz) I_m(k\rho) \sin(m\phi) \sum_{r=1}^{\infty} \beta_m^r(k) [\tilde{F}_r^s(k)/\text{Se}'_r(U, q)]. \end{aligned} \quad (17.4.96)$$

Upon comparing (4.96) with (14.2.54), we conclude that there are the relations

$$G_{m,c}(k) = \sum_{r=0}^{\infty} \alpha_m^r(k) [\tilde{F}_r^c(k)/\text{Ce}'_r(U, q)], \quad (17.4.97)$$

and

$$G_{m,s}(k) = \sum_{r=1}^{\infty} \beta_m^r(k) [\tilde{F}_r^s(k)/\text{Se}'_r(U, q)]. \quad (17.4.98)$$

We remark that it can be shown that the real parts of the  $G_{m,\alpha}$  are even in  $k$ , and the imaginary parts are odd in  $k$ . See Exercise 4.2. Finally, in view of (14.2.55) and (14.2.56), we have the desired results

$$\begin{aligned} C_{m,c}^{[n]}(z) &= i^n (1/2)^m (1/m!) \int_{-\infty}^{\infty} dk \exp(ikz) k^{n+m} \sum_{r=0}^{\infty} \alpha_m^r(k) [\tilde{F}_r^c(k)/\text{Ce}'_r(U, q)] \\ &= i^n (1/2)^m (1/m!) \int_{-\infty}^{\infty} dk \exp(ikz) k^{n+m} G_{m,c}(k), \end{aligned} \quad (17.4.99)$$

$$\begin{aligned} C_{m,s}^{[n]}(z) &= i^n (1/2)^m (1/m!) \int_{-\infty}^{\infty} dk \exp(ikz) k^{n+m} \sum_{r=1}^{\infty} \beta_m^r(k) [\tilde{F}_r^s(k)/\text{Se}'_r(U, q)] \\ &= i^n (1/2)^m (1/m!) \int_{-\infty}^{\infty} dk \exp(ikz) k^{n+m} G_{m,s}(k). \end{aligned} \quad (17.4.100)$$

We have found expressions for the on-axis gradients in terms of field data (normal component) on the surface of an elliptic cylinder. These results hold for the cases  $m \geq 1$ . When  $m = 0$  there are the results

$$\begin{aligned} C_{m=0,c}^{[n]}(z) &= C_0^{[n]}(z) = i^n \int_{-\infty}^{\infty} dk \exp(ikz) k^n \sum_{r=0}^{\infty} \alpha_0^r(k) [\tilde{F}_r^c(k)/\text{Ce}'_r(U, q)] \\ &= i^n \int_{-\infty}^{\infty} dk \exp(ikz) k^n G_{0,c}(k), \end{aligned} \quad (17.4.101)$$

$$C_{m=0,s}^{[n]}(z) = 0. \quad (17.4.102)$$

Just as in the  $m = 0$  case for the circular cylinder, so too here it may be better to derive and employ formulas based on the tangential component  $B_z$  rather than the normal component. See Section 19.2.

## Exercises

**17.4.1.** Verify that (4.1) and (4.2) can be written in the form (4.7).

**17.4.2.** The purpose of this exercise is to study the analytic properties of elliptic coordinates. Our discussion is based on the relation (4.7). According to (4.7), the function  $\zeta(w)$  is an entire function of  $w$ . What can be said about its inverse  $w(\zeta)$ ? Verify that (4.7) has the inverse

$$w = \cosh^{-1}(\zeta/f) = \log[(\zeta/f) + \sqrt{(\zeta/f)^2 - 1}]. \quad (17.4.103)$$

Evidently  $w(\zeta)$  has branch points at  $\zeta = \pm f$ . Verify that  $w(\zeta)$  is analytic in the cut  $\zeta$  plane with a cut consisting of a straight line extending from  $\zeta = -f$  to  $\zeta = f$  as illustrated in Figure 4.2.

**17.4.3.** Exercise on wave equation.

## 17.5 Use of Field Data on Surface of Rectangular Cylinder

### 17.5.1 Finding the Magnetic Scalar Potential $\psi(x, y, z)$

Consider the domain  $x \in [-W/2, W/2]$ ,  $y \in [-H/2, H/2]$ ,  $z \in [-\infty, \infty]$ . This domain is the interior and surface of a cylinder of infinite extent in the  $\pm z$  direction, centered on the  $z$  axis, and having rectangular cross section with width  $W$  and height  $H$ . In this section we will describe the use of field data on the surface of this cylinder. This cylinder has 4 surfaces (sides) which we will call  $t$  and  $b$  for *top* and *bottom*, and  $\ell$  and  $r$  for *left* and *right*.

Suppose that we are given the normal component of the magnetic field on the top, bottom, left, and right surfaces. That is, we are given the field data

$$B_y^t(x, z) = B_y(x, y, z)|_{y=H/2} \text{ with } x \in [-W/2, W/2] \text{ and } z \in [-\infty, \infty], \quad (17.5.1)$$

$$B_y^b(x, z) = B_y(x, y, z)|_{y=-H/2} \text{ with } x \in [-W/2, W/2] \text{ and } z \in [-\infty, \infty], \quad (17.5.2)$$

$$B_x^\ell(y, z) = B_x(x, y, z)|_{x=-W/2} \text{ with } y \in [-H/2, H/2] \text{ and } z \in [-\infty, \infty], \quad (17.5.3)$$

$$B_x^r(y, z) = B_x(x, y, z)|_{x=W/2} \text{ with } y \in [-H/2, H/2] \text{ and } z \in [-\infty, \infty]. \quad (17.5.4)$$

Our goal is to find the scalar magnetic potential  $\psi(x, y, z)$  in terms of this field data.

For purposes of Fourier analysis, we will *extend* the normal field surface data beyond the ranges indicated above as follows:

- Extend  $B_y^t(x, z)$  and  $B_y^b(x, z)$  from the interval  $x \in [-W/2, W/2]$  to the extended interval  $x \in [-W/2, 3W/2]$  by requiring that the extension be *even* in  $x$  about the value  $x = W/2$ . Note that the extended interval  $x \in [-W/2, 3W/2]$  can be written in the form

$$[-W/2, 3W/2] = [W/2 - W, W/2 + W]. \quad (17.5.5)$$

- Extend  $B_x^\ell(y, z)$  and  $B_x^r(y, z)$  from the interval  $y \in [-H/2, H/2]$  to the extended interval  $y \in [-H/2, 3H/2]$  by requiring that the extension be *even* in  $y$  about the value  $y = H/2$ . Note that the extended interval  $[-H/2, 3H/2]$  can be written in the form

$$[-H/2, 3H/2] = [H/2 - H, H/2 + H]. \quad (17.5.6)$$

So doing will produce functions that are continuous over the extended intervals

$$x \in [W/2 - W, W/2 + W] \text{ and } y \in [H/2 - H, H/2 + H]. \quad (17.5.7)$$

Moreover, by construction, the extended functions will take the *same* value at both ends of each extended interval. Therefore they can be *further* extended beyond their extended intervals, both to the left and the right, in a *continuous* way by requiring that their further extensions be periodic with periods  $2W$  and  $2H$ , respectively. (Moreover, these extensions will also be even about the values  $x = -W/2$  and  $y = -H/2$ , respectively.) In summary, we have produced extensions that are even about  $\pm W/2$  or  $\pm H/2$ , are periodic with periods

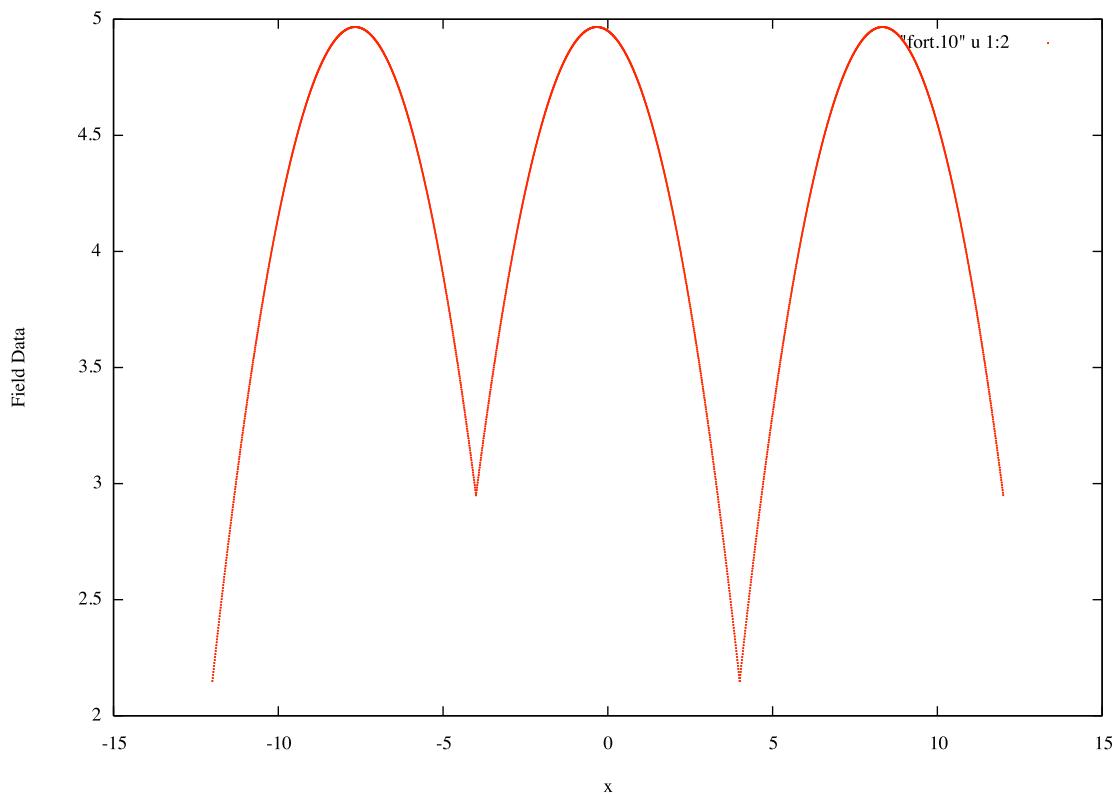


Figure 17.5.1: Hypothetical  $B_y^t(x, z)$  data in the interval  $x \in [-W/2, W/2]$  and its extension to the full  $x$  axis to facilitate Fourier analysis. In this example,  $W = 8$  so that  $[-W/2, W/2] = [-4, 4]$ . The extension has period  $2W = 16$ , is even about the points  $x = \pm W/2 = \pm 4$  and their periodic counterparts, and is continuous. Generally, the first derivative is discontinuous at the points  $x = \pm W/2 = \pm 4$  and their periodic counterparts.

$2W$  or  $2H$ , and are continuous. See, for example, Figure 5.1. They are therefore ideally suited to Fourier analysis over the extended intervals (5.7).

Now consider the functions

$$\cos[(x + W/2)(n\pi/W)] \text{ with } n = 0, 1, 2, \dots \quad (17.5.8)$$

and the functions

$$\sin[(x + W/2)(n\pi/W)] \text{ with } n = 1, 2, \dots \quad (17.5.9)$$

They have period  $2W$  and form a complete orthogonal set over the interval (5.5). Therefore  $B_y^t(x, z)$  and  $B_y^b(x, z)$ , when extended as described above, can be expanded in terms of them. Note also that the cosine functions (5.8) are even about  $x = W/2$  and the sine functions (5.9) are odd. Since (by construction) the extended  $B_y^t(x, z)$  and  $B_y^b(x, z)$  are even about  $x = W/2$ , it follows that only the cosine terms will appear in the Fourier expansion. Thus, we have the Fourier representations

$$B_y^t(x, z) = \sum_{n=0}^{\infty} \tilde{B}_y^t(n, z) \cos[(x + W/2)(n\pi/W)], \quad (17.5.10)$$

$$B_y^b(x, z) = \sum_{n=0}^{\infty} \tilde{B}_y^b(n, z) \cos[(x + W/2)(n\pi/W)], \quad (17.5.11)$$

where

$$\tilde{B}_y^t(0, z) = [1/(2W)] \int_{-W/2}^{3W/2} dx B_y^t(x, z) = (1/W) \int_{-W/2}^{W/2} dx B_y^t(x, z), \quad (17.5.12)$$

$$\begin{aligned} \tilde{B}_y^t(n, z) &= (1/W) \int_{-W/2}^{3W/2} dx B_y^t(x, z) \cos[(x + W/2)(n\pi/W)] \\ &= (2/W) \int_{-W/2}^{W/2} dx B_y^t(x, z) \cos[(x + W/2)(n\pi/W)] \text{ for } n > 0, \end{aligned} \quad (17.5.13)$$

$$\tilde{B}_y^b(0, z) = [1/(2W)] \int_{-W/2}^{3W/2} dx B_y^b(x, z) = (1/W) \int_{-W/2}^{W/2} dx B_y^b(x, z), \quad (17.5.14)$$

$$\begin{aligned} \tilde{B}_y^b(n, z) &= (1/W) \int_{-W/2}^{3W/2} dx B_y^b(x, z) \cos[(x + W/2)(n\pi/W)] \\ &= (2/W) \int_{-W/2}^{W/2} dx B_y^b(x, z) \cos[(x + W/2)(n\pi/W)] \text{ for } n > 0. \end{aligned} \quad (17.5.15)$$

Here we have used the fact that  $B_y^t(x, z)$  and  $B_y^b(x, z)$  and the cosine functions are even about  $x = W/2$ .

We have already seen that the fully extended  $B_y^t(x, z)$  and  $B_y^b(x, z)$  are continuous and periodic with period  $2W$ . However, they will generally not have continuous first derivatives

across the joins at  $x = \pm W/2$ ,  $x = \pm 3W/2$ , etc. It follows from standard Fourier analysis theory that the coefficients  $\tilde{B}_y^t(n, z)$  and  $\tilde{B}_y^b(n, z)$  must, in general, fall off like  $1/n^2$ ,

$$\tilde{B}_y^t(n, z) \sim (1/n^2) \text{ as } n \rightarrow \infty, \text{ etc.} \quad (17.5.16)$$

Therefore the series (5.10) and (5.11) are point-wise absolutely, but not wonderfully, convergent. The slow falloff (5.16) is the price to be paid for working with a bounding surface that has sharp corners. By contrast, it can be shown that the analogous coefficients in the cases of circular and elliptic cylinders fall off much more rapidly, namely as  $(1/\Lambda)^{|n|}$  for some  $\Lambda > 1$ . See Exercises 16.2.3 and 16.2.4.

In a similar way, we have the Fourier representations

$$B_x^\ell(y, z) = \sum_{n=0}^{\infty} \tilde{B}_y^\ell(n, z) \cos[(y + H/2)(n\pi/H)], \quad (17.5.17)$$

$$B_x^r(y, z) = \sum_{n=0}^{\infty} \tilde{B}_y^r(n, z) \cos[(y + H/2)(n\pi/H)], \quad (17.5.18)$$

where

$$\tilde{B}_x^\ell(0, z) = [1/(2H)] \int_{-H/2}^{3H/2} dy B_x^\ell(y, z) = (1/H) \int_{-H/2}^{H/2} dx B_x^\ell(y, z), \quad (17.5.19)$$

$$\begin{aligned} \tilde{B}_x^\ell(n, z) &= (1/H) \int_{-H/2}^{3H/2} dy B_x^\ell(y, z) \cos[(y + H/2)(n\pi/H)] \\ &= (2/H) \int_{-H/2}^{H/2} dy B_x^\ell(y, z) \cos[(y + H/2)(n\pi/H)] \text{ for } n > 0, \end{aligned} \quad (17.5.20)$$

$$\tilde{B}_x^r(0, z) = [1/(2H)] \int_{-H/2}^{3H/2} dy B_x^r(y, z) = (1/H) \int_{-H/2}^{H/2} dx B_x^r(y, z), \quad (17.5.21)$$

$$\begin{aligned} \tilde{B}_x^r(n, z) &= (1/H) \int_{-H/2}^{3H/2} dy B_x^r(y, z) \cos[(y + H/2)(n\pi/H)] \\ &= (2/H) \int_{-H/2}^{H/2} dy B_x^r(y, z) \cos[(y + H/2)(n\pi/H)] \text{ for } n > 0. \end{aligned} \quad (17.5.22)$$

To proceed further, we perform Fourier transforms in  $z$ . Thus, we make the definitions

$$\tilde{\tilde{B}}_y^t(n, k) = [1/(2\pi)] \int_{-\infty}^{\infty} dz \exp(-ikz) \tilde{B}_y^t(n, z), \quad (17.5.23)$$

$$\tilde{\tilde{B}}_y^b(n, k) = [1/(2\pi)] \int_{-\infty}^{\infty} dz \exp(-ikz) \tilde{B}_y^b(n, z), \quad (17.5.24)$$

$$\tilde{\tilde{B}}_x^\ell(n, k) = [1/(2\pi)] \int_{-\infty}^{\infty} dz \exp(-ikz) \tilde{B}_x^\ell(n, z), \quad (17.5.25)$$

$$\tilde{\tilde{B}}_x^r(n, k) = [1/(2\pi)] \int_{-\infty}^{\infty} dz \exp(-ikz) \tilde{B}_x^r(n, z), \quad (17.5.26)$$

Note that the various  $\tilde{B}$  terms on the right sides of (5.23) through (5.26) are real. It follows that the real parts of the various  $\tilde{B}$  terms on the left sides of (5.23) through (5.26) are even in  $k$ , and the imaginary parts are odd in  $k$ .

With these definitions in hand, we are ready to determine the scalar potential  $\psi(x, y, z)$  in terms of surface field values. First we note, as is easily checked, that functions of the form

$$\exp(ikz) \cos[(x + W/2)(n\pi/W)] \cosh[\sigma_n(y \pm H/2)] \quad (17.5.27)$$

and

$$\exp(ikz) \cos[(y + H/2)(n\pi/H)] \cosh[\tau_n(x \pm W/2)], \quad (17.5.28)$$

where

$$\sigma_n = [k^2 + (n\pi/W)^2]^{1/2} \quad (17.5.29)$$

and

$$\tau_n = [k^2 + (n\pi/H)^2]^{1/2}, \quad (17.5.30)$$

satisfy Laplace's equation. Next define functions  $\psi^t(x, y, z)$ ,  $\psi^b(x, y, z)$ ,  $\psi^\ell(x, y, z)$ , and  $\psi^r(x, y, z)$  by the relations

$$\psi^t(x, y, z) = \int_{-\infty}^{\infty} dk \exp(ikz) \sum_{n=0}^{\infty} \frac{\tilde{\tilde{B}}_y^t(n, k) \cos[(x + W/2)(n\pi/W)] \cosh[\sigma_n(y + H/2)]}{\sigma_n \sinh(H\sigma_n)}, \quad (17.5.31)$$

$$\psi^b(x, y, z) = \int_{-\infty}^{\infty} dk \exp(ikz) \sum_{n=0}^{\infty} \frac{\tilde{\tilde{B}}_y^b(n, k) \cos[(x + W/2)(n\pi/W)] \cosh[\sigma_n(y - H/2)]}{\sigma_n \sinh(H\sigma_n)}, \quad (17.5.32)$$

$$\psi^\ell(x, y, z) = \int_{-\infty}^{\infty} dk \exp(ikz) \sum_{n=0}^{\infty} \frac{\tilde{\tilde{B}}_y^\ell(n, k) \cos[(y + H/2)(n\pi/H)] \cosh[\tau_n(x + W/2)]}{\tau_n \sinh(W\tau_n)}, \quad (17.5.33)$$

$$\psi^r(x, y, z) = \int_{-\infty}^{\infty} dk \exp(ikz) \sum_{n=0}^{\infty} \frac{\tilde{\tilde{B}}_y^r(n, k) \cos[(y + H/2)(n\pi/H)] \cosh[\tau_n(x - W/2)]}{\tau_n \sinh(W\tau_n)}. \quad (17.5.34)$$

Evidently by construction they all satisfy Laplace's equation.

Now watch closely: From the definition (5.31) we have the result

$$\begin{aligned} \partial\psi^t(x, y, z)/\partial x = \\ - \int_{-\infty}^{\infty} dk \exp(ikz) \sum_{n=0}^{\infty} \frac{\tilde{\tilde{B}}_y^t(n, k)(n\pi/W) \sin[(x + W/2)(n\pi/W)] \cosh[\sigma_n(y + H/2)]}{\sigma_n \sinh(H\sigma_n)}. \end{aligned} \quad (17.5.35)$$

[Note that we may interchange summation and differentiation: the series (5.31) converges absolutely and uniformly due to the  $\sigma_n \sinh(H\sigma_n)$  denominator.] It follows that

$$[\partial\psi^t(x, y, z)/\partial x]|_{x=\pm W/2} = 0. \quad (17.5.36)$$

Similarly, we find that

$$[\partial\psi^b(x, y, z)/\partial x]|_{x=\pm W/2} = 0. \quad (17.5.37)$$

Also from (5.31) we find that

$$\begin{aligned} \partial\psi^t(x, y, z)/\partial y &= \\ \int_{-\infty}^{\infty} dk \exp(ikz) \sum_{n=0}^{\infty} \frac{\tilde{\tilde{B}}_y^t(n, k) \cos[(x + W/2)(n\pi/W)] \sinh[\sigma_n(y + H/2)]}{\sinh(H\sigma_n)}. \end{aligned} \quad (17.5.38)$$

It follows that

$$[\partial\psi^t(x, y, z)/\partial y]|_{y=-H/2} = 0 \quad (17.5.39)$$

and

$$\begin{aligned} [\partial\psi^t(x, y, z)/\partial y]|_{y=H/2} &= \int_{-\infty}^{\infty} dk \exp(ikz) \sum_{n=0}^{\infty} \tilde{\tilde{B}}_y^t(n, k) \cos[(x + W/2)(n\pi/W)] \\ &= \sum_{n=0}^{\infty} \tilde{B}_y^t(n, z) \cos[(x + W/2)(n\pi/W)] \\ &= B_y^t(x, z). \end{aligned} \quad (17.5.40)$$

Similarly, we find that

$$[\partial\psi^b(x, y, z)/\partial y]|_{y=H/2} = 0 \quad (17.5.41)$$

and

$$\begin{aligned} [\partial\psi^b(x, y, z)/\partial y]|_{y=-H/2} &= \int_{-\infty}^{\infty} dk \exp(ikz) \sum_{n=0}^{\infty} \tilde{\tilde{B}}_y^b(n, k) \cos[(x + W/2)(n\pi/W)] \\ &= \sum_{n=0}^{\infty} \tilde{B}_y^b(n, z) \cos[(x + W/2)(n\pi/W)] \\ &= B_y^b(x, z). \end{aligned} \quad (17.5.42)$$

Finally, analogous results hold for  $\psi^\ell(x, y, z)$  and  $\psi^r(x, y, z)$ . They satisfy the relations

$$[\partial\psi^\ell(x, y, z)/\partial y]|_{y=\pm H/2} = 0, \quad (17.5.43)$$

$$[\partial\psi^r(x, y, z)/\partial y]|_{y=\pm H/2} = 0, \quad (17.5.44)$$

$$[\partial\psi^\ell(x, y, z)/\partial x]|_{x=W/2} = 0, \quad (17.5.45)$$

$$[\partial\psi^\ell(x, y, z)/\partial x]|_{x=-W/2} = B_x^\ell(y, z), \quad (17.5.46)$$

$$[\partial\psi^r(x, y, z)/\partial x]|_{x=-W/2} = 0, \quad (17.5.47)$$

$$[\partial\psi^r(x, y, z)/\partial x]|_{x=W/2} = B_x^r(y, z). \quad (17.5.48)$$

At last we are ready to construct  $\psi(x, y, z)$ . We make the definition

$$\psi(x, y, z) = \psi^t(x, y, z) + \psi^b(x, y, z) + \psi^\ell(x, y, z) + \psi^r(x, y, z). \quad (17.5.49)$$

Evidently this  $\psi$  satisfies Laplace's equation. Also we find that

$$\begin{aligned} & [\partial\psi(x, y, z)/\partial y]|_{y=H/2} = \\ & [\partial\psi^t(x, y, z)/\partial y]|_{y=H/2} + [\partial\psi^b(x, y, z)/\partial y]|_{y=H/2} \\ & + [\partial\psi^\ell(x, y, z)/\partial y]|_{y=H/2} + [\partial\psi^r(x, y, z)/\partial y]|_{y=H/2} \\ & = B_y^t(x, z). \end{aligned} \quad (17.5.50)$$

Here we have used (5.35) through (5.48). Similarly we find that

$$[\partial\psi(x, y, z)/\partial y]|_{y=-H/2} = B_y^b(x, z), \quad (17.5.51)$$

$$[\partial\psi(x, y, z)/\partial x]|_{x=-W/2} = B_x^\ell(y, z), \quad (17.5.52)$$

$$[\partial\psi(x, y, z)/\partial x]|_{x=W/2} = B_x^r(y, z). \quad (17.5.53)$$

Thus  $\psi$  satisfies the required boundary conditions on all four surfaces. Finally, observe that the quantities appearing on the right sides of (5.31) through (5.34) are *entire* functions of  $x$ ,  $y$ , and  $z$ . Now suppose that  $x$ ,  $y$  have values corresponding to a point *inside* the cylinder. Then, thanks to the denominators appearing on the right sides of (5.31) through (5.34), the integrals and sums are rapidly convergent. It follows that  $\psi(x, y, z)$  is *analytic* in  $x, y, z$  for all points within the cylinder.

### 17.5.2 Finding the On-Axis Gradients

The remaining task is to determine the on-axis gradients by finding the cylindrical harmonic expansion for  $\psi$  as given by (5.31) through (5.34) and (5.49). In analogy to the case of the elliptical cylinder, this will be done with the aid of what may be called *Fourier-Bessel connection coefficients*. Namely, there are the formulas

$$\begin{aligned} & \cos[(x + W/2)(j\pi/W)] \cosh[\sigma_j(y + H/2)] = \\ & \sum_{m=0}^{\infty} d_{mj}^{tc}(k) I_m(k\rho) \cos(m\phi) + \sum_{m=1}^{\infty} d_{mj}^{ts}(k) I_m(k\rho) \sin(m\phi), \end{aligned} \quad (17.5.54)$$

$$\begin{aligned} & \cos[(x + W/2)(j\pi/W)] \cosh[\sigma_j(y - H/2)] = \\ & \sum_{m=0}^{\infty} d_{mj}^{bc}(k) I_m(k\rho) \cos(m\phi) + \sum_{m=1}^{\infty} d_{mj}^{bs}(k) I_m(k\rho) \sin(m\phi), \end{aligned} \quad (17.5.55)$$

$$\begin{aligned} & \cos[(y + H/2)(j\pi/H)] \cosh[\tau_j(x + W/2)] = \\ & \sum_{m=0}^{\infty} d_{mj}^{\ell c}(k) I_m(k\rho) \cos(m\phi) + \sum_{m=1}^{\infty} d_{mj}^{\ell s}(k) I_m(k\rho) \sin(m\phi), \end{aligned} \quad (17.5.56)$$

$$\begin{aligned} & \cos[(y + H/2)(j\pi/H)] \cosh[\tau_j(x - W/2)] = \\ & \sum_{m=0}^{\infty} d_{mj}^{rc}(k) I_m(k\rho) \cos(m\phi) + \sum_{m=1}^{\infty} d_{mj}^{rs}(k) I_m(k\rho) \sin(m\phi). \end{aligned} \quad (17.5.57)$$

We will derive them shortly. Before doing so, we will use them to find the on-axis gradients. Suppose we employ (5.54) in (5.31). Doing so gives the result

$$\begin{aligned} \psi^t(x, y, z) = & \sum_{m=0}^{\infty} \int_{-\infty}^{\infty} dk \exp(ikz) I_m(k\rho) \cos(m\phi) \sum_{j=0}^{\infty} \frac{d_{mj}^{tc}(k) \tilde{\tilde{B}}_y^t(j, k)}{\sigma_j \sinh(H\sigma_j)} \\ & + \sum_{m=1}^{\infty} \int_{-\infty}^{\infty} dk \exp(ikz) I_m(k\rho) \sin(m\phi) \sum_{j=0}^{\infty} \frac{d_{mj}^{ts}(k) \tilde{\tilde{B}}_y^t(j, k)}{\sigma_j \sinh(H\sigma_j)}. \end{aligned} \quad (17.5.58)$$

Similarly, we find the relations

$$\begin{aligned} \psi^b(x, y, z) = & \sum_{m=0}^{\infty} \int_{-\infty}^{\infty} dk \exp(ikz) I_m(k\rho) \cos(m\phi) \sum_{j=0}^{\infty} \frac{d_{mj}^{bc}(k) \tilde{\tilde{B}}_y^b(j, k)}{\sigma_j \sinh(H\sigma_j)} \\ & + \sum_{m=1}^{\infty} \int_{-\infty}^{\infty} dk \exp(ikz) I_m(k\rho) \sin(m\phi) \sum_{j=0}^{\infty} \frac{d_{mj}^{bs}(k) \tilde{\tilde{B}}_y^b(j, k)}{\sigma_j \sinh(H\sigma_j)}, \end{aligned} \quad (17.5.59)$$

$$\begin{aligned} \psi^\ell(x, y, z) = & \sum_{m=0}^{\infty} \int_{-\infty}^{\infty} dk \exp(ikz) I_m(k\rho) \cos(m\phi) \sum_{j=0}^{\infty} \frac{d_{mj}^{\ell c}(k) \tilde{\tilde{B}}_y^\ell(j, k)}{\tau_j \sinh(W\tau_j)} \\ & + \sum_{m=1}^{\infty} \int_{-\infty}^{\infty} dk \exp(ikz) I_m(k\rho) \sin(m\phi) \sum_{j=0}^{\infty} \frac{d_{mj}^{\ell s}(k) \tilde{\tilde{B}}_y^\ell(j, k)}{\tau_j \sinh(W\tau_j)}, \end{aligned} \quad (17.5.60)$$

$$\begin{aligned} \psi^r(x, y, z) = & \sum_{m=0}^{\infty} \int_{-\infty}^{\infty} dk \exp(ikz) I_m(k\rho) \cos(m\phi) \sum_{j=0}^{\infty} \frac{d_{mj}^{rc}(k) \tilde{\tilde{B}}_y^r(j, k)}{\tau_j \sinh(W\tau_j)} \\ & + \sum_{m=1}^{\infty} \int_{-\infty}^{\infty} dk \exp(ikz) I_m(k\rho) \sin(m\phi) \sum_{j=0}^{\infty} \frac{d_{mj}^{rs}(k) \tilde{\tilde{B}}_y^r(j, k)}{\tau_j \sinh(W\tau_j)}. \end{aligned} \quad (17.5.61)$$

It follows that  $\psi$  can be written in the form

$$\begin{aligned} \psi(x, y, z) = & \sum_{m=0}^{\infty} \int_{-\infty}^{\infty} dk \exp(ikz) I_m(k\rho) \cos(m\phi) G_{m,c}(k) \\ & + \sum_{m=1}^{\infty} \int_{-\infty}^{\infty} dk \exp(ikz) I_m(k\rho) \sin(m\phi) G_{m,s}(k) \end{aligned} \quad (17.5.62)$$

where

$$G_{m,c}(k) = \sum_{j=0}^{\infty} \left[ \frac{d_{mj}^{tc}(k) \tilde{\tilde{B}}_y^t(j, k) + d_{mj}^{bc}(k) \tilde{\tilde{B}}_y^b(j, k)}{\sigma_j \sinh(H\sigma_j)} + \frac{d_{mj}^{\ell c}(k) \tilde{\tilde{B}}_y^\ell(j, k) + d_{mj}^{rc}(k) \tilde{\tilde{B}}_y^r(j, k)}{\tau_j \sinh(W\tau_j)} \right], \quad (17.5.63)$$

$$G_{m,s}(k) = \sum_{j=0}^{\infty} \left[ \frac{d_{mj}^{ts}(k) \tilde{\tilde{B}}_y^t(j, k) + d_{mj}^{bs}(k) \tilde{\tilde{B}}_y^b(j, k)}{\sigma_j \sinh(H\sigma_j)} + \frac{d_{mj}^{\ell s}(k) \tilde{\tilde{B}}_y^\ell(j, k) + d_{mj}^{rs}(k) \tilde{\tilde{B}}_y^r(j, k)}{\tau_j \sinh(W\tau_j)} \right]. \quad (17.5.64)$$

We remark that it can be shown that the real parts of the  $G_{m,\alpha}$  are even in  $k$ , and the imaginary parts are odd in  $k$ . See Exercise 5.2. Finally, in view of (14.2.55) and (14.2.56), we have the desired results

$$C_{m,c}^{[n]}(z) = i^n (1/2)^m (1/m!) \int_{-\infty}^{\infty} dk \exp(ikz) k^{n+m} G_{m,c}(k), \quad (17.5.65)$$

$$C_{m,s}^{[n]}(z) = i^n (1/2)^m (1/m!) \int_{-\infty}^{\infty} dk \exp(ikz) k^{n+m} G_{m,s}(k). \quad (17.5.66)$$

We have found expressions for the on-axis gradients in terms of field data (normal component) on the surface of a rectangular cylinder. These results hold for the cases  $m \geq 1$ . When  $m = 0$  there are the results

$$C_{m=0,c}^{[n]}(z) = C_0^{[n]}(z) = i^n \int_{-\infty}^{\infty} dk \exp(ikz) k^n G_{0,c}(k), \quad (17.5.67)$$

$$C_{m=0,s}^{[n]}(z) = 0. \quad (17.5.68)$$

Just as in the  $m = 0$  case for the circular and elliptical cylinder, in the rectangular case it may also be better to derive and employ formulas based on the tangential component  $B_z$  rather than the normal component. See Section 18.2.

### 17.5.3 Fourier-Bessel Connection Coefficients

The purpose of this subsection is to derive the Fourier-Bessel connection coefficients postulated in (5.54) through (5.57). We will do so in pieces. First, recall the Bessel generating-function relation

$$\begin{aligned} \exp[z \cos(\theta)] &= I_0(z) + 2 \sum_{m=1}^{\infty} \cos(m\theta) I_m(z) \\ &= \sum_{m=-\infty}^{\infty} \exp(im\theta) I_m(z). \end{aligned} \quad (17.5.69)$$

Here we have again employed (14.2.12). Second, we have been taught from the cradle that circular and hyperbolic functions, which appear on the left sides of (5.54) through (5.57), are made of exponential functions. We will see that the combinations of these exponential functions that occur in (5.54) through (5.57) can in turn be written in a form that enables the use of (5.69).

Begin with (5.54), whose left side can be written in the expanded form

$$\begin{aligned} \cos[(x + W/2)(j\pi/W)] \cosh[\sigma_j(y + H/2)] &= (1/4) \times \\ \{\exp[i(x + W/2)(j\pi/W)] + \exp[-i(x + W/2)(j\pi/W)]\} \times \\ \{\exp[\sigma_j(y + H/2)] + \exp[-\sigma_j(y + H/2)]\}. \end{aligned} \quad (17.5.70)$$

Multiplying out the factors that occur on the right side of (5.70) produces a sum of four terms:

$$\begin{aligned} \cos[(x + W/2)(j\pi/W)] \cosh[\sigma_j(y + H/2)] &= (1/4) \times \\ [(i)^j \exp(\sigma_j H/2) \exp(ij\pi x/W) \exp(\sigma_j y) + \\ (i)^j \exp(-\sigma_j H/2) \exp(ij\pi x/W) \exp(-\sigma_j y) + \\ (-i)^j \exp(\sigma_j H/2) \exp(-ij\pi x/W) \exp(\sigma_j y) + \\ (-i)^j \exp(-\sigma_j H/2) \exp(-ij\pi x/W) \exp(-\sigma_j y)]. \end{aligned} \quad (17.5.71)$$

Here we have used the the result

$$\exp[(\pm iW/2)(j\pi/W)] = \exp(\pm ij\pi/2) = (\pm i)^j. \quad (17.5.72)$$

We see that we have to deal with products of the form  $\exp(\pm ij\pi x/W) \exp(\pm \sigma_j y)$  where the signs are to be taken independently. We will treat each of these four possibilities separately.

For the ++ possibility we write

$$\exp(ij\pi x/W) \exp(\sigma_j y) = \exp(ij\pi x/W + \sigma_j y) = \exp[(ij\pi\rho/W) \cos\phi + (\rho\sigma_j) \sin\phi]. \quad (17.5.73)$$

Here we have made the substitutions (14.2.3) and (14.2.4). Next recall the identity

$$\cos(\phi + \psi) = \cos\psi \cos\phi - \sin\psi \sin\phi. \quad (17.5.74)$$

Let us write the argument appearing on the right side of (5.73) in the form

$$(ij\pi\rho/W) \cos\phi + (\rho\sigma_j) \sin\phi = \lambda \cos(\phi + \psi) = \lambda \cos\psi \cos\phi - \lambda \sin\psi \sin\phi \quad (17.5.75)$$

where  $\lambda$ ,  $\sin\psi$ , and  $\cos\psi$  are yet to be determined. Equating like terms in  $\phi$  yields the relations

$$\lambda \cos\psi = ij\pi\rho/W, \quad (17.5.76)$$

$$\lambda \sin\psi = -\rho\sigma_j. \quad (17.5.77)$$

Now square both sides of (5.76) and (5.77) and add the results to obtain the relation

$$\lambda^2 = \rho^2[-(j\pi/W)^2 + \sigma_j^2] = (k\rho)^2. \quad (17.5.78)$$

Here we have used (5.29). It follows that

$$\lambda = k\rho, \quad (17.5.79)$$

and we may also write

$$\cos \psi = ij\pi/(kW), \quad (17.5.80)$$

$$\sin \psi = -\sigma_j/k. \quad (17.5.81)$$

For future use, we invoke Euler to write

$$\exp(i\psi) = \cos \psi + i \sin \psi = ij\pi/(kW) - i\sigma_j/k = (i/k)(j\pi/W - \sigma_j), \quad (17.5.82)$$

$$\exp(-i\psi) = \cos \psi - i \sin \psi = ij\pi/(kW) + i\sigma_j/k = (i/k)(j\pi/W + \sigma_j). \quad (17.5.83)$$

Finally employing (5.79), first in (5.75) and then in (5.73), yields the results

$$(ij\pi\rho/W) \cos \phi + (\rho\sigma_j) \sin \phi = k\rho \cos(\phi + \psi) \quad (17.5.84)$$

and

$$\exp(ij\pi x/W) \exp(\sigma_j y) = \exp[k\rho \cos(\phi + \psi)]. \quad (17.5.85)$$

We next deal with the other sign possibilities. Consider the substitutions  $\phi \rightarrow -\phi$ ,  $\phi \rightarrow \phi + \pi$ , and  $\phi \rightarrow \pi - \phi$ . It is readily verified from (14.2.3) and (14.2.4) that these substitutions correspond to the following substitutions in  $x$  and  $y$ :

$$\phi \rightarrow -\phi \iff x \rightarrow x, y \rightarrow -y; \quad (17.5.86)$$

$$\phi \rightarrow \phi + \pi \iff x \rightarrow -x, y \rightarrow -y; \quad (17.5.87)$$

$$\phi \rightarrow \pi - \phi \iff x \rightarrow -x, y \rightarrow y. \quad (17.5.88)$$

Therefore, from (5.85), we find the results

$$\exp(ij\pi x/W) \exp(-\sigma_j y) = \exp[k\rho \cos(-\phi + \psi)] = \exp[k\rho \cos(\phi - \psi)], \quad (17.5.89)$$

$$\exp(-ij\pi x/W) \exp(\sigma_j y) = \exp[k\rho \cos(\pi - \phi + \psi)] = \exp[k\rho \cos(\phi - \psi - \pi)], \quad (17.5.90)$$

$$\exp(-ij\pi x/W) \exp(-\sigma_j y) = \exp[k\rho \cos(\phi + \psi + \pi)]. \quad (17.5.91)$$

Let us now see how the results for the products  $\exp(\pm ij\pi x/W) \exp(\pm \sigma_j y)$ , as given by (5.85) and (5.89) through (5.91), can be used in conjunction with (5.69). For the ++ case use of (5.69) and (5.85) gives the result

$$\begin{aligned} \exp(ij\pi x/W) \exp(\sigma_j y) &= \exp[k\rho \cos(\phi + \psi)] = \\ &\sum_{m=-\infty}^{\infty} \exp[im(\phi + \psi)] I_m(k\rho) = \\ &\sum_{m=-\infty}^{\infty} \exp(im\psi) \exp(im\phi) I_m(k\rho) = \\ &\sum_{m=-\infty}^{\infty} (i/k)^m (j\pi/W - \sigma_j)^m \exp(im\phi) I_m(k\rho). \end{aligned} \quad (17.5.92)$$

Here we have also used (5.82). Similarly, for the remaining cases, we find the results

$$\begin{aligned} \exp(ij\pi x/W) \exp(-\sigma_j y) &= \exp[k\rho \cos(\phi - \psi)] = \\ , \quad \sum_{m=-\infty}^{\infty} \exp(-im\psi) \exp(im\phi) I_m(k\rho) &= \\ \sum_{m=-\infty}^{\infty} (i/k)^m (j\pi/W + \sigma_j)^m \exp(im\phi) I_m(k\rho), \end{aligned} \quad (17.5.93)$$

$$\begin{aligned} \exp(-ij\pi x/W) \exp(\sigma_j y) &= \exp[k\rho \cos(\phi - \psi - \pi)] = \\ \sum_{m=-\infty}^{\infty} \exp(-im\pi) \exp(-im\psi) \exp(im\phi) I_m(k\rho) &= \\ \sum_{m=-\infty}^{\infty} (-1)^m (i/k)^m (j\pi/W + \sigma_j)^m \exp(im\phi) I_m(k\rho), \end{aligned} \quad (17.5.94)$$

$$\begin{aligned} \exp(-ij\pi x/W) \exp(-\sigma_j y) &= \exp[k\rho \cos(\phi + \psi + \pi)] = \\ \sum_{m=-\infty}^{\infty} \exp(im\pi) \exp(im\psi) \exp(im\phi) I_m(k\rho) &= \\ \sum_{m=-\infty}^{\infty} (-1)^m (i/k)^m (j\pi/W - \sigma_j)^m \exp(im\phi) I_m(k\rho). \end{aligned} \quad (17.5.95)$$

Bessel expansions have now been obtained for all the various pieces that result from expanding in exponentials the terms on the left side of (5.54). We now combine them to find a Bessel expansion for the left side of (5.54). From (5.71) and (5.92) through (5.95), we find the result

$$\begin{aligned} \cos[(x + W/2)(j\pi/W)] \cosh[\sigma_j(y + H/2)] &= \\ \sum_{m=-\infty}^{\infty} d_{mj}^t(k) I_m(k\rho) \exp(im\phi) \end{aligned} \quad (17.5.96)$$

where

$$\begin{aligned} d_{mj}^t(k) &= (1/4)(i)^{j+m} \{ [(j\pi/W - \sigma_j)/k]^m + (-1)^{j+m} [(j\pi/W + \sigma_j)/k]^m \} \exp(\sigma_j H/2) \\ &+ (1/4)(i)^{j+m} \{ [(j\pi/W + \sigma_j)/k]^m + (-1)^{j+m} [(j\pi/W - \sigma_j)/k]^m \} \exp(-\sigma_j H/2). \end{aligned} \quad (17.5.97)$$

Upon comparing (5.54) and (5.96) we see that for  $m \geq 1$  there are the relations

$$d_{mj}^{tc} = d_{mj}^t + d_{-mj}^t, \quad (17.5.98)$$

$$d_{mj}^{ts} = id_{mj}^t - id_{-mj}^t. \quad (17.5.99)$$

And for  $m = 0$  there is the result

$$d_{0j}^{tc} = d_{0j}^t, \quad (17.5.100)$$

$$d_{0j}^{ts} = 0. \quad (17.5.101)$$

Let us first deal with the simplest case, that for  $m = 0$ . Use of (5.97) then gives the result

$$\begin{aligned} d_{0j}^t(k) &= (1/4)(i)^j \{ [1 + (-1)^j] \exp(\sigma_j H/2) \\ &\quad + (1/4)(i)^j \{ [1 + (-1)^j] \exp(-\sigma_j H/2). \end{aligned} \quad (17.5.102)$$

Therefore, for  $m = 0$ , we conclude that

$$d_{0j}^{tc} = 0 \text{ for } j \text{ odd}, \quad (17.5.103)$$

$$d_{0j}^{tc} = (-1)^{j/2} \cosh(\sigma_j H/2) \text{ for } j \text{ even}. \quad (17.5.104)$$

Now tackle the more complicated case  $m \geq 1$ . Note that there is the relation

$$[(j\pi/W - \sigma_j)/k][(j\pi/W + \sigma_j)/k] = [(j\pi/W)^2 - (\sigma_j^2)]/k^2 = -k^2/k^2 = -1. \quad (17.5.105)$$

Here we have again used (5.29). Consequently, we find that

$$\begin{aligned} d_{-mj}^t(k) &= (1/4)(i)^{j-m} \{ [(j\pi/W - \sigma_j)/k]^{-m} + (-1)^{j-m} [(j\pi/W + \sigma_j)/k]^{-m} \} \exp(\sigma_j H/2) \\ &\quad + (1/4)(i)^{j-m} \{ [(j\pi/W + \sigma_j)/k]^{-m} + (-1)^{j-m} [(j\pi/W - \sigma_j)/k]^{-m} \} \exp(-\sigma_j H/2) \\ &= (1/4)(i)^{j-m} (-1)^m \{ [(j\pi/W + \sigma_j)/k]^m + (-1)^{j-m} [(j\pi/W - \sigma_j)/k]^m \} \exp(\sigma_j H/2) \\ &\quad + (1/4)(i)^{j-m} (-1)^m \{ [(j\pi/W - \sigma_j)/k]^m + (-1)^{j-m} [(j\pi/W + \sigma_j)/k]^m \} \exp(-\sigma_j H/2). \end{aligned} \quad (17.5.106)$$

We also note that

$$(i)^{j-m} (-1)^m = (i)^{j-m} (i)^{2m} = (i)^{j+m} \quad (17.5.107)$$

and

$$(-1)^{j-m} = (-1)^{2m} (-1)^{j-m} = (-1)^{j+m}. \quad (17.5.108)$$

Therefore, we may also write

$$\begin{aligned} d_{-mj}^t(k) &= (1/4)(i)^{j+m} \{ [(j\pi/W + \sigma_j)/k]^m + (-1)^{j+m} [(j\pi/W - \sigma_j)/k]^m \} \exp(\sigma_j H/2) \\ &\quad + (1/4)(i)^{j+m} \{ [(j\pi/W - \sigma_j)/k]^m + (-1)^{j+m} [(j\pi/W + \sigma_j)/k]^m \} \exp(-\sigma_j H/2). \end{aligned} \quad (17.5.109)$$

Let us now compute  $d_{mj}^{tc}(k)$ . It follows from (5.97), (5.98), and (5.109) that for  $m \geq 1$  there is the result

$$\begin{aligned} d_{mj}^{tc}(k) &= (1/4)(i)^{j+m} [1 + (-1)^{j+m}] \times \\ &\quad \{ [(j\pi/W - \sigma_j)/k]^m + [(j\pi/W + \sigma_j)/k]^m \} \exp(\sigma_j H/2) \\ &\quad + (1/4)(i)^{j+m} [1 + (-1)^{j+m}] \times \\ &\quad \{ [(j\pi/W + \sigma_j)/k]^m + [(j\pi/W - \sigma_j)/k]^m \} \exp(-\sigma_j H/2) \\ &= (1/2)(i)^{j+m} [1 + (-1)^{j+m}] \times \\ &\quad \{ [(j\pi/W - \sigma_j)/k]^m + [(j\pi/W + \sigma_j)/k]^m \} \cosh(\sigma_j H/2). \end{aligned} \quad (17.5.110)$$

Therefore, when  $(j + m)$  is odd,

$$d_{mj}^{tc} = 0. \quad (17.5.111)$$

And, when  $(j + m)$  is even,

$$d_{mj}^{tc} = (-1)^{(j+m)/2} \{ [(j\pi/W - \sigma_j)/k]^m + [(j\pi/W + \sigma_j)/k]^m \} \cosh(\sigma_j H/2). \quad (17.5.112)$$

Finally, let us compute  $d_{mj}^{ts}$ . It follows from (5.97), (5.99), and (5.109) that for  $m \geq 1$  there is the result

$$\begin{aligned} d_{mj}^{ts}(k) &= (1/4)i(i)^{j+m}[1 - (-1)^{j+m}] \times \\ &\quad \{ [(j\pi/W - \sigma_j)/k]^m - [(j\pi/W + \sigma_j)/k]^m \} \exp(\sigma_j H/2) \\ &+ (1/4)i(i)^{j+m}[1 - (-1)^{j+m}] \times \\ &\quad \{ [(j\pi/W + \sigma_j)/k]^m - [(j\pi/W - \sigma_j)/k]^m \} \exp(-\sigma_j H/2) \\ &= (1/2)i(i)^{j+m}[1 - (-1)^{j+m}] \times \\ &\quad \{ [(j\pi/W - \sigma_j)/k]^m - [(j\pi/W + \sigma_j)/k]^m \} \sinh(\sigma_j H/2). \end{aligned} \quad (17.5.113)$$

Therefore, when  $(j + m)$  is even,

$$d_{mj}^{ts} = 0. \quad (17.5.114)$$

And, when  $(j + m)$  is odd,

$$d_{mj}^{ts} = (-1)^{(j+m+1)/2} \{ [(j\pi/W - \sigma_j)/k]^m - [(j\pi/W + \sigma_j)/k]^m \} \sinh(\sigma_j H/2). \quad (17.5.115)$$

We have found the Fourier-Bessel coefficients  $d_{mj}^{ta}(k)$ . Next observe that the left sides of (5.54) and (5.55) are interchanged under the substitution  $H \leftrightarrow -H$ . Therefore, for  $m \geq 1$ , there are also the relations

$$d_{mj}^{bc} = d_{mj}^{tc}, \quad (17.5.116)$$

$$d_{mj}^{bs} = -d_{mj}^{ts}, \quad (17.5.117)$$

And for  $m = 0$  there are the results

$$d_{0j}^{bc} = d_{0j}^{tc}, \quad (17.5.118)$$

$$d_{0j}^{bs} = 0. \quad (17.5.119)$$

Analogous calculations can be made to find Bessel expansions for the right sides of (5.56) and (5.57). Instead, for variety, we will take a different approach that utilizes results already obtained. Consider the relation (5.54). Using (14.2.3) and (14.2.4) we may rewrite it in the form

$$\begin{aligned} \cos\{[\rho \cos(\phi) + W/2][j\pi/W]\} \cosh\{\sigma_j[\rho \sin(\phi) + H/2]\} &= \\ \sum_{m=0}^{\infty} d_{mj}^{tc}(k) I_m(k\rho) \cos(m\phi) + \sum_{m=1}^{\infty} d_{mj}^{ts}(k) I_m(k\rho) \sin(m\phi). \end{aligned} \quad (17.5.120)$$

Now make the substitution  $\phi \rightarrow \phi + \pi/2$ . So doing gives the result

$$\begin{aligned} \cos\{[\rho \cos(\phi + \pi/2) + W/2][j\pi/W]\} \cosh\{\sigma_j[\rho \sin(\phi + \pi/2) + H/2]\} &= \\ \sum_{m=0}^{\infty} d_{mj}^{tc}(k) I_m(k\rho) \cos[m(\phi + \pi/2)] + \sum_{m=1}^{\infty} d_{mj}^{ts}(k) I_m(k\rho) \sin[m(\phi + \pi/2)]. \end{aligned} \quad (17.5.121)$$

Next employ the identities

$$\rho \cos(\phi + \pi/2) = -\rho \sin(\phi) = -y, \quad (17.5.122)$$

$$\rho \sin(\phi + \pi/2) = \rho \cos(\phi) = x, \quad (17.5.123)$$

$$\cos[m(\phi + \pi/2)] = \cos(m\phi) \cos(m\pi/2) - \sin(m\phi) \sin(m\pi/2), \quad (17.5.124)$$

$$\sin[m(\phi + \pi/2)] = \sin(m\phi) \cos(m\pi/2) + \cos(m\phi) \sin(m\pi/2). \quad (17.5.125)$$

We see that (5.120) can be rewritten in the form

$$\begin{aligned} & \cos\{-y + W/2][j\pi/W]\} \cosh\{\sigma_j[x + H/2]\} = \\ & \sum_{m=0}^{\infty} D_{mj}^c(k) I_m(k\rho) \cos(m\phi) + \sum_{m=1}^{\infty} D_{mj}^s(k) I_m(k\rho) \sin(m\phi) \end{aligned} \quad (17.5.126)$$

where

$$D_{mj}^c(k) = d_{mj}^{tc}(k) \cos(m\pi/2) + d_{mj}^{ts}(k) \sin(m\pi/2), \quad (17.5.127)$$

$$D_{mj}^s(k) = -d_{mj}^{tc}(k) \sin(m\pi/2) + d_{mj}^{ts}(k) \cos(m\pi/2). \quad (17.5.128)$$

Now employ the already known results (5.101), (5.103), (5.104), (5.111), (5.112), (5.114), and (5.115). First, for  $m = 0$ , we find the relations

$$D_{0j}^c(k) = d_{0j}^{tc} = (-1)^{j/2} \cosh(\sigma_j H/2) \text{ for } j \text{ even}, \quad (17.5.129)$$

$$D_{0j}^c(k) = d_{0j}^{tc} = 0 \text{ for } j \text{ odd}, \quad (17.5.130)$$

$$D_{0j}^s(k) = d_{0j}^{ts} = 0. \quad (17.5.131)$$

For the remaining  $D_{mj}^{\alpha}(k)$  we must distinguish the cases  $(j+m)$  odd and even. When  $(j+m)$  is odd we find the results

$$\begin{aligned} & D_{mj}^c(k) = d_{mj}^{ts}(k) \sin(m\pi/2) \\ & = (-1)^{(j+m+1)/2} \sin(m\pi/2) \{[(j\pi/W - \sigma_j)/k]^m - [(j\pi/W + \sigma_j)/k]^m\} \sinh(\sigma_j H/2), \end{aligned} \quad (17.5.132)$$

$$\begin{aligned} & D_{mj}^s(k) = d_{mj}^{ts}(k) \cos(m\pi/2) \\ & = (-1)^{(j+m+1)/2} \cos(m\pi/2) \{[(j\pi/W - \sigma_j)/k]^m - [(j\pi/W + \sigma_j)/k]^m\} \sinh(\sigma_j H/2). \end{aligned} \quad (17.5.133)$$

And, when  $(j+m)$  is even, we find the results

$$\begin{aligned} & D_{mj}^c(k) = d_{mj}^{tc}(k) \cos(m\pi/2) \\ & = (-1)^{(j+m)/2} \cos(m\pi/2) \{[(j\pi/W - \sigma_j)/k]^m + [(j\pi/W + \sigma_j)/k]^m\} \cosh(\sigma_j H/2), \end{aligned} \quad (17.5.134)$$

$$\begin{aligned}
 D_{mj}^s(k) &= -d_{mj}^{tc}(k) \sin(m\pi/2) \\
 &= (-1)^{(j+m)/2} \sin(m\pi/2) \{ [(j\pi/W - \sigma_j)/k]^m + [(j\pi/W + \sigma_j)/k]^m \} \cosh(\sigma_j H/2).
 \end{aligned} \tag{17.5.135}$$

For the penultimate step, compare the left side of (5.126) with the left side of (5.56). We see that the first is transformed into the second under the substitutions  $W \rightarrow -H$ ,  $\sigma_j \rightarrow \tau_j$ , and  $H \rightarrow W$ . It follows that there are the relations

$$d_{0j}^{\ell c} = (-1)^{j/2} \cosh(\tau_j W/2) \text{ for } j \text{ even,} \tag{17.5.136}$$

$$d_{0j}^{\ell c} = 0 \text{ for } j \text{ odd,} \tag{17.5.137}$$

$$d_{0j}^{\ell s} = 0. \tag{17.5.138}$$

For the remaining  $d_{mj}^{\ell \alpha}$  we must again distinguish the cases  $(j+m)$  odd and even. When  $(j+m)$  is odd we find the results

$$d_{mj}^{\ell c}(k) = (-1)^{(j+m+1)/2} (-1)^m \sin(m\pi/2) \{ [(j\pi/H + \tau_j)/k]^m - [(j\pi/H - \tau_j)/k]^m \} \sinh(\tau_j W/2), \tag{17.5.139}$$

$$d_{mj}^{\ell s}(k) = (-1)^{(j+m+1)/2} (-1)^m \cos(m\pi/2) \{ [(j\pi/H + \tau_j)/k]^m - [(j\pi/H - \tau_j)/k]^m \} \sinh(\tau_j W/2). \tag{17.5.140}$$

And, when  $(j+m)$  is even, we find the results

$$d_{mj}^{\ell c}(k) = (-1)^{(j+m)/2} (-1)^m \cos(m\pi/2) \{ [(j\pi/H + \tau_j)/k]^m + [(j\pi/H - \tau_j)/k]^m \} \cosh(\tau_j W/2), \tag{17.5.141}$$

$$d_{mj}^{\ell s}(k) = -(-1)^{(j+m)/2} (-1)^m \sin(m\pi/2) \{ [(j\pi/H + \tau_j)/k]^m + [(j\pi/H - \tau_j)/k]^m \} \cosh(\tau_j W/2). \tag{17.5.142}$$

Finally, observe that the left sides of (5.56) and (5.57) are interchanged under the substitution  $W \leftrightarrow -W$ . It follows that for  $m = 0$  there are the relations

$$d_{0j}^{rc} = d_{0j}^{\ell c}, \tag{17.5.143}$$

$$d_{0j}^{rs} = 0. \tag{17.5.144}$$

For the remaining  $d_{mj}^{r\alpha}$ , when  $(j+m)$  is odd, we find the results

$$d_{mj}^{rc}(k) = -d_{mj}^{\ell c}(k), \tag{17.5.145}$$

$$d_{mj}^{rs}(k) = -d_{mj}^{\ell s}(k). \tag{17.5.146}$$

And, when  $(j+m)$  is even, we find the results

$$d_{mj}^{rc}(k) = d_{mj}^{\ell c}(k), \tag{17.5.147}$$

$$d_{mj}^{rs}(k) = d_{mj}^{\ell s}(k). \tag{17.5.148}$$

## Exercises

**17.5.1.** How could one have known that Fourier-Bessel expansions of the form (5.54) through (5.57) must exist? Consider, for example, (5.54). Multiply both sides by  $\exp(ikz)$ . Show that both sides then become harmonic functions. Moreover, the left side is analytic in the vicinity of the  $z$  axis. But we know from Section 14.2.1 that such functions must have an expansion of the form (14.2.11).

**17.5.2.** Verify the relations (5.116) through (5.119).

**17.5.3.** Check the consistency of the relations (5.92) through (5.95) by verifying that they are transformed among themselves by the substitutions  $W \leftrightarrow -W$  and  $\sigma_j \leftrightarrow -\sigma_j$ .

**17.5.4.** Verify that the functions  $\sigma_j \sinh(H\sigma_j)$  and  $\tau_j \sinh(W\tau_j)$  and the Fourier-Bessel coefficients  $d_{mj}^{\beta\alpha}$  are even functions of  $k$ . Use these facts to show that the real parts of the  $G_{m,\alpha}$  are even in  $k$ , and the imaginary parts are odd in  $k$ .

**17.5.5.** Consider the functions  $\psi_j(x, y, z)$  for  $j = 1, 2, 3$  defined by the relations

$$\psi_1(x, y, z) = a \cos(k_x x) \sinh(k_y y) \cos(kz + \chi) \quad (17.5.149)$$

with

$$-k_x^2 + k_y^2 = k^2; \quad (17.5.150)$$

$$\psi_2(x, y, z) = a \cosh(k_x x) \sinh(k_y y) \cos(kz + \chi) \quad (17.5.151)$$

with

$$k_x^2 + k_y^2 = k^2; \quad (17.5.152)$$

$$\psi_3(x, y, z) = a \cosh(k_x x) \sin(k_y y) \cos(kz + \chi) \quad (17.5.153)$$

with

$$k_x^2 - k_y^2 = k^2. \quad (17.5.154)$$

Verify that each  $\psi_j$  satisfies Laplace's equation, is analytic everywhere, and in particular is analytic in  $x, y$  near the  $z$  axis. Verify that each  $\psi_j$  can be written in the form (14.2.11). Verify that each  $\psi_j$  produces a vertical ( $\pm y$  direction) field in the midplane  $y = 0$  that oscillates in  $z$ , and therefore is some approximation (at least near the  $z$  axis) to the field of an infinitely long wiggler.

**17.5.6.** Consider a  $\psi$  of the form

$$\psi(x, y, z) = (a + by) \exp(kx) \exp(ikz). \quad (17.5.155)$$

Verify that this  $\psi$  satisfies Laplace's equation, is analytic everywhere, and in particular is analytic in  $x, y$  near the  $z$  axis. Verify that this  $\psi$  can be written in the form (14.2.11). In particular, verify that this  $\psi$  can be written in the form

$$\begin{aligned} \psi &= a[I_0(k\rho) + 2 \sum_{m=1}^{\infty} \cos(m\phi) I_m(k\rho)] \exp(ikz) \\ &+ [(2b/k) \sum_{m=1}^{\infty} m \sin(m\phi) I_m(k\rho)] \exp(ikz). \end{aligned} \quad (17.5.156)$$

## 17.6 Attempted Use of Nearly On-Axis and Midplane Field Data

As promised at the end of Section 14, here we examine other attempted approaches. All will be seen to involve what, in essence, is high-order numerical differentiation. Therefore, they are unlikely to yield reliable results beyond modest order, at best.

### 17.6.1 Use of Nearly On-Axis Data

Let us begin with the cylindrical multipole expansion

$$\begin{aligned} \psi(\rho, \phi, z) &= \psi(x, y, z) = \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! \ell!} C_0^{[2\ell]}(z) \rho^{2\ell} \\ &+ \sum_{m=1}^{\infty} \cos(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,c}^{[2\ell]}(z) \rho^{2\ell+m} \\ &+ \sum_{m=1}^{\infty} \sin(m\phi) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,s}^{[2\ell]}(z) \rho^{2\ell+m}. \end{aligned} \quad (17.6.1)$$

Suppose (6.1) is multiplied by factors of  $\cos(m\phi)$  or  $\sin(m\phi)$  and the integrated over  $\phi$ . Doing so gives the results

$$\tilde{\psi}(\rho, 0, z) = \int_0^{2\pi} d\phi \psi(\rho, \phi, z) = \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{2\pi}{2^{2\ell} \ell! \ell!} C_0^{[2\ell]}(z) \rho^{2\ell}, \quad (17.6.2)$$

$$\tilde{\psi}_c(\rho, m, z) = \int_0^{2\pi} d\phi \cos(m\phi) \psi(\rho, \phi, z) = \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{\pi m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,c}^{[2\ell]}(z) \rho^{2\ell+m}, \quad (17.6.3)$$

$$\tilde{\psi}_s(\rho, m, z) = \int_0^{2\pi} d\phi \sin(m\phi) \psi(\rho, \phi, z) = \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{\pi m!}{2^{2\ell} \ell! (\ell+m)!} C_{m,s}^{[2\ell]}(z) \rho^{2\ell+m}. \quad (17.6.4)$$

If  $\psi(\rho, \phi, z)$  is known, and it is in fact provided at grid points by some three-dimensional codes, then the integrals in (6.2) through (6.4) can be computed. Moreover, we have the relations

$$C_0^{[0]}(z) = [1/(2\pi)] \lim_{\rho \rightarrow 0} \tilde{\psi}(\rho, 0, z), \quad (17.6.5)$$

$$C_{m,c}^{[0]}(z) = (1/\pi) \lim_{\rho \rightarrow 0} (1/\rho^m) \tilde{\psi}_c(\rho, m, z), \quad (17.6.6)$$

$$C_{m,s}^{[0]}(z) = (1/\pi) \lim_{\rho \rightarrow 0} (1/\rho^m) \tilde{\psi}_s(\rho, m, z). \quad (17.6.7)$$

It is also in principle possible to compute the on-axis gradients from field data. Let  $B_\rho(\rho, \phi, z)$  be the  $\rho$  component of  $\mathbf{B}$ . It is defined by the relation

$$B_\rho(\rho, \phi, z) = B_\rho(x, y, z) = \mathbf{e}_\rho \cdot \mathbf{B} = (x/\rho) B_x + (y/\rho) B_y = \cos(\phi) B_x + \sin(\phi) B_y. \quad (17.6.8)$$

In terms of  $B_\rho(\rho, \phi, z)$ , define the quantities

$$\tilde{B}_\rho(\rho, 0, z) = \int_0^{2\pi} d\phi B_\rho(\rho, \phi, z), \quad (17.6.9)$$

$$\tilde{B}_{\rho c}(\rho, m, z) = \int_0^{2\pi} d\phi \cos(m\phi) B_\rho(\rho, \phi, z), \quad (17.6.10)$$

$$\tilde{B}_{\rho s}(\rho, m, z) = \int_0^{2\pi} d\phi \sin(m\phi) B_\rho(\rho, \phi, z). \quad (17.6.11)$$

We also know that

$$B_\rho(\rho, \phi, z) = (\partial/\partial\rho)\psi(\rho, \phi, z). \quad (17.6.12)$$

It follows that there are the relations

$$\tilde{B}_\rho(\rho, 0, z) = \sum_{\ell=0}^{\infty} (-1)^\ell \frac{4\pi\ell}{2^{2\ell}\ell!\ell!} C_0^{[2\ell]}(z) \rho^{2\ell-1}, \quad (17.6.13)$$

$$\tilde{B}_{\rho c}(\rho, m, z) = \sum_{\ell=0}^{\infty} (-1)^\ell \frac{\pi(2\ell+m)m!}{2^{2\ell}\ell!(\ell+m)!} C_{m,c}^{[2\ell]}(z) \rho^{2\ell+m-1}, \quad (17.6.14)$$

$$\tilde{B}_{\rho s}(\rho, m, z) = \sum_{\ell=0}^{\infty} (-1)^\ell \frac{\pi(2\ell+m)m!}{2^{2\ell}\ell!(\ell+m)!} C_{m,s}^{[2\ell]}(z) \rho^{2\ell+m-1}. \quad (17.6.15)$$

Consequently, there are the relations

$$C_0^{[2]}(z) = (1/\pi) \lim_{\rho \rightarrow 0} (1/\rho^m) \tilde{B}_\rho(\rho, 0, z), \quad (17.6.16)$$

$$C_{m,c}^{[0]}(z) = (1/\pi) \lim_{\rho \rightarrow 0} (1/\rho^?) \tilde{B}_{\rho c}(\rho, m, z), \quad (17.6.17)$$

$$C_{m,s}^{[0]}(z) = (1/\pi) \lim_{\rho \rightarrow 0} (1/\rho^m) \tilde{B}_{\rho s}(\rho, m, z). \quad (17.6.18)$$

Also, from \*, we have the relation.

$$C_0^{[1]}(z) = \lim_{x,y \rightarrow 0} B_z(x, y, z) = B_z(0, 0, z). \quad (17.6.19)$$

Let us now examine the feasibility of carrying out the indicated calculations based on numerical data provided on a 3-d grid. We see that, apart from (6.5) and (6.19), which can be reliably estimated, particularly if the grid is chosen so that there are grid points on the  $z$  axis, it is necessary to perform a limiting process in which both a numerator and denominator approach zero. Such a limiting process is akin to numerical differentiation. Also, to compute the  $C_{m,\alpha}^{[n]}(z)$  for larger values of  $n$ , it is necessary to repeatedly differentiate the above relations with respect to  $z$ . When based on grid data, this again involves multiple numerical differentiation.

### 17.6.2 Use of Midplane Field Data

In place of nearly on-axis data, one might consider the use of Midplane Field Data. Use of the relations (H.1.17) through (H.1.19) gives for the nearly midplane field the expansions

$$\begin{aligned} B_x(x, y, z) &= \partial_x \psi = C_{1,c}^{[0]}(z) + x[2C_{2,c}^{[0]}(z) - (1/2)C_0^{[2]}(z)] + 2yC_{2,s}^{[0]}(z) \\ &+ 3x^2[C_{3,c}^{[0]}(z) - (1/8)C_{1,c}^{[2]}(z)] - y^2[3C_{3,c}^{[0]}(z) + (1/8)C_{1,c}^{[2]}(z)] \\ &+ 2xy[3C_{3,s}^{[0]}(z) - (1/8)C_{1,s}^{[2]}(z)] \dots, \end{aligned} \quad (17.6.20)$$

$$\begin{aligned} B_y(x, y, z) &= \partial_y \psi = +C_{1,s}^{[0]}(z) - y[2C_{2,c}^{[0]}(z) + (1/2)C_0^{[2]}(z)] + 2xC_{2,s}^{[0]}(z) \\ &- 3y^2[C_{3,s}^{[0]}(z) + (1/8)C_{1,s}^{[2]}(z)] + x^2[3C_{3,s}^{[0]}(z) - (1/8)C_{1,s}^{[2]}(z)] \\ &- 2xy[3C_{3,c}^{[0]}(z) + (1/8)C_{1,c}^{[2]}(z)] + \dots, \end{aligned} \quad (17.6.21)$$

$$\begin{aligned} B_z(x, y, z) &= \partial_z \psi = C_0^{[1]}(z) + xC_{1,c}^{[1]}(z) + yC_{1,s}^{[1]}(z) \\ &+ (x^2 - y^2)C_{2,c}^{[1]}(z) + 2xyC_{2,s}^{[1]}(z) - (1/4)(x^2 + y^2)C_0^{[3]}(z) + \dots. \end{aligned} \quad (17.6.22)$$

Evaluating these expansions in the midplane gives the results

$$\begin{aligned} B_x(x, y = 0, z) &= C_{1,c}^{[0]}(z) + x[2C_{2,c}^{[0]}(z) - (1/2)C_0^{[2]}(z)] \\ &+ 3x^2[C_{3,c}^{[0]}(z) - (1/8)C_{1,c}^{[2]}(z)] \dots, \end{aligned} \quad (17.6.23)$$

$$\begin{aligned} B_y(x, y = 0, z) &= C_{1,s}^{[0]}(z) + 2xC_{2,s}^{[0]}(z) \\ &+ x^2[3C_{3,s}^{[0]}(z) - (1/8)C_{1,s}^{[2]}(z)] + \dots, \end{aligned} \quad (17.6.24)$$

$$\begin{aligned} B_z(x, y = 0, z) &= C_0^{[1]}(z) + xC_{1,c}^{[1]}(z) \\ &+ x^2[C_{2,c}^{[1]}(z) - (1/4)C_0^{[3]}(z)] + \dots. \end{aligned} \quad (17.6.25)$$

The relations (6.4) through (6.6) express the midplane fields in terms of on-axis gradients. These relations can be inverted to determine the on-axis gradients in terms of the midplane fields. By repeatedly differentiating them with respect to  $x$  and  $z$  and then setting  $x = 0$ , one finds the results

$$C_0^{[1]}(z) = B_z(x = 0, y = 0, z), \quad (17.6.26)$$

$$C_{1,c}^{[0]}(z) = B_x(x = 0, y = 0, z), \quad (17.6.27)$$

$$C_{1,s}^{[0]}(z) = B_y(x = 0, y = 0, z), \quad (17.6.28)$$

$$C_{2,c}^{[0]}(z) = (1/2)(\partial B_x / \partial x) \Big|_{(0,0,z)} + (1/4)(\partial B_z / \partial z) \Big|_{(0,0,z)}, \quad (17.6.29)$$

$$C_{2,s}^{[0]}(z) = (1/2)(\partial B_y / \partial x) \Big|_{(0,0,z)}, \quad (17.6.30)$$

$$C_{3,c}^{[0]}(z) = (1/6)(\partial^2 B_x / \partial x^2) \Big|_{(0,0,z)} + (1/8)(\partial^2 B_x / \partial z^2) \Big|_{(0,0,z)}, \quad (17.6.31)$$

$$C_{3,s}^{[0]}(z) = (1/6)(\partial^2 B_y / \partial x^2) \Big|_{(0,0,z)} + (1/24)(\partial^2 B_y / \partial z^2) \Big|_{(0,0,z)}, \text{ etc.} \quad (17.6.32)$$

See Exercise 6.1. Finally, by repeatedly differentiating these relations with respect to  $z$ , one can obtain the  $C_{m,\alpha}^{[n]}(z)$  for  $n > 0$ . In general, the computation of the  $C_{m,\alpha}^{[n]}(z)$  requires  $m+n-1$  differentiations. Again, when based on grid data, this involves multiple numerical differentiation, and therefore is expected to be unreliable.

## Exercises

**17.6.1.** The aim of this exercise is to verify the relations (6.26) through (6.32). Begin by setting  $x = 0$  in the relations (6.4) through (6.6). Show that so doing yields the results

$$B_x(x = 0, y = 0, z) = C_{1,c}^{[0]}(z), \quad (17.6.33)$$

$$B_y(x = 0, y = 0, z) = C_{1,s}^{[0]}(z), \quad (17.6.34)$$

$$B_z(x = 0, y = 0, z) = C_0^{[1]}(z). \quad (17.6.35)$$

Next, differentiate (6.4) through (6.6) with respect to  $x$  and then set  $x = 0$ . Show that so doing gives the results

$$(\partial B_x / \partial x)|_{0,0,z} = -(1/2)C_0^{[2]}(z) + 2C_{2,c}^{[0]}(z), \quad (17.6.36)$$

$$(\partial B_y / \partial x)|_{0,0,z} = 2C_{2,s}^{[0]}(z), \quad (17.6.37)$$

$$(\partial B_z / \partial x)|_{0,0,z} = C_{1,c}^{[1]}(z). \quad (17.6.38)$$

Show that solving (6.7) through (6.12) for the on-axis gradients gives, so far, the results

$$C_{1,c}^{[0]}(z) = B_x(x = 0, y = 0, z), \quad (17.6.39)$$

$$C_{1,s}^{[0]}(z) = B_y(x = 0, y = 0, z), \quad (17.6.40)$$

$$C_0^{[1]}(z) = B_z(x = 0, y = 0, z), \quad (17.6.41)$$

$$C_{2,c}^{[0]}(z) = (1/2)(\partial B_x / \partial x)|_{0,0,z} + (1/4)(\partial B_z / \partial z)|_{0,0,z}, \quad (17.6.42)$$

$$C_{2,s}^{[0]}(z) = (1/2)(\partial B_y / \partial x)|_{0,0,z}, \quad (17.6.43)$$

$$C_{1,c}^{[1]}(z) = (\partial B_z / \partial x)|_{0,0,z}. \quad (17.6.44)$$

Verify that (6.18) is redundant because from (6.13) we also have the relation

$$C_{1,c}^{[1]}(z) = (\partial B_x / \partial z)|_{0,0,z}. \quad (17.6.45)$$

Alternatively, (6.19) serves as a consistency check on (6.18).<sup>12</sup> Next differentiate the mid-plane fields twice with respect to  $x$  and then set  $x = 0$ . Show that so doing yields the relations

$$(\partial^2 B_x / \partial x^2)|_{0,0,z} = 6[C_{3,c}^{[0]}(z) - (1/8)C_{1,c}^{[2]}(z)], \quad (17.6.46)$$

$$(\partial^2 B_y / \partial x^2)|_{0,0,z} = 2[3C_{3,s}^{[0]}(z) - (1/8)C_{1,s}^{[2]}(z)], \quad (17.6.47)$$

$$(\partial^2 B_z / \partial x^2)|_{0,0,z} = 2[C_{2,c}^{[1]}(z) - (1/4)C_0^{[3]}(z)]. \quad (17.6.48)$$

Show that these relations, with the aid of the previous relations, can be solved for the next set of on-axis gradients to give the results

$$(\partial^2 B_x / \partial x^2)|_{0,0,z} = 6[C_{3,c}^{[0]}(z) - (1/8)C_{1,c}^{[2]}(z)], \quad (17.6.49)$$

$$(\partial^2 B_y / \partial x^2)|_{0,0,z} = 2[3C_{3,s}^{[0]}(z) - (1/8)C_{1,s}^{[2]}(z)], \quad (17.6.50)$$

$$(\partial^2 B_z / \partial x^2)|_{0,0,z} = 2[C_{2,c}^{[1]}(z) - (1/4)C_0^{[3]}(z)], \quad (17.6.51)$$

$$[C_{3,c}^{[0]}(z) - (1/8)C_{1,c}^{[2]}(z)] = (1/6)(\partial^2 B_x / \partial x^2)|_{0,0,z}, \quad (17.6.52)$$

$$C_{3,c}^{[0]}(z) = (1/6)(\partial^2 B_x / \partial x^2)|_{0,0,z} + (1/8)C_{1,c}^{[2]}(z), \quad (17.6.53)$$

$$C_{3,c}^{[0]}(z) = (1/6)(\partial^2 B_x / \partial x^2)|_{0,0,z} + (1/8)(\partial^2 B_x / \partial z^2)\Big|_{(0,0,z)}, \quad (17.6.54)$$

$$(\partial^2 B_y / \partial x^2)|_{0,0,z} = 2[3C_{3,s}^{[0]}(z) - (1/8)C_{1,s}^{[2]}(z)]. \quad (17.6.55)$$

## 17.7 Terminating End Fields

In principle, the fringe field of an individual beam-line element at either end of the element has infinite extent. But in practice in many instances we may wish to regard a beam line as a collection of separated/isolated elements. To do this it is necessary to make an approximation in which leading and trailing end fields are “terminated” in some way. The crucial problem is how to relate canonical coordinates in the absence of a magnetic field with canonical coordinates in the presence of a magnetic field.

### 17.7.1 Preliminary Observations

We begin with some preliminary observations. In Cartesian coordinates the Hamiltonian describing charged-particle motion with  $z$  as the independent variable is given by the relation

$$K = -[(p_t^{\text{can}})^2/c^2 - m^2 c^2 - (p_x^{\text{can}} - qA_x)^2 - (p_y^{\text{can}} - qA_y)^2]^{1/2} - qA_z. \quad (17.7.1)$$

Here we have assumed that the electric scalar potential  $\psi$  vanishes and  $\mathbf{A}$  is static so that there is no electric field. Also, we have used the notation  $p_x^{\text{can}}$ ,  $p_y^{\text{can}}$ , and  $p_t^{\text{can}}$  to indicate that it is the components of the *canonical* momenta that are involved in a Hamiltonian description of motion. See (1.6.16).

---

<sup>12</sup>This check arises from the requirement  $\nabla \times \mathbf{B} = 0$ .

According to Hamilton's equations of motion, the change of a coordinate, say  $x(z)$ , with  $z$  is given by

$$\begin{aligned} dx/dz &= \partial K/\partial p_x^{\text{can}} \\ &= (p_x^{\text{can}} - qA_x)/[(p_t^{\text{can}})^2/c^2 - m^2c^2 - (p_x^{\text{can}} - qA_x)^2 - (p_y^{\text{can}} - qA_y)^2]^{1/2} \\ &= (p_x^{\text{can}} - qA_x)/[-(K + qA_z)]. \end{aligned} \quad (17.7.2)$$

Let us verify that this result agrees with what we already know. Recall that

$$K = -p_z^{\text{can}}. \quad (17.7.3)$$

See (1.6.6). It follows that (7.2) can be rewritten in the form

$$dx/dz = (p_x^{\text{can}} - qA_x)/(p_z^{\text{can}} - qA_z). \quad (17.7.4)$$

According to (1.5.27) through (1.5.30) there is the relation

$$\mathbf{p}^{\text{can}} - q\mathbf{A} = \mathbf{p}^{\text{mech}} \quad (17.7.5)$$

where  $\mathbf{p}^{\text{mech}}$  is the *mechanical* momentum given by

$$\mathbf{p}^{\text{mech}} = \gamma m \mathbf{v}. \quad (17.7.6)$$

Consequently, (7.4) can be rewritten in the form

$$dx/dz = p_x^{\text{mech}}/p_z^{\text{mech}} = \gamma mv_x/(\gamma mv_z) = v_x/v_z = \frac{dx/dt}{dz/dt}. \quad (17.7.7)$$

Evidently, the far left and far right sides of (7.7) agree. It is also easy to see that results analogous to those just found also hold for  $y(z)$ .

To complete the story we need to examine also the equation of motion for  $t(z)$ . In this case application of the standard Hamiltonian rules gives the result

$$\begin{aligned} dt/dz &= \partial K/\partial p_t^{\text{can}} \\ &= (-p_t^{\text{can}}/c^2)/[(p_t^{\text{can}})^2/c^2 - m^2c^2 - (p_x^{\text{can}} - qA_x)^2 - (p_y^{\text{can}} - qA_y)^2]^{1/2} \\ &= (-p_t^{\text{can}}/c^2)/[-(K + qA_z)]. \end{aligned} \quad (17.7.8)$$

Now use of (7.3), (7.5), and (1.6.17) yields the relation

$$dt/dz = (-p_t^{\text{can}}/c^2)/p_z^{\text{mech}} = \gamma m/(\gamma mv_z) = \frac{1}{dz/dt} \quad (17.7.9)$$

so that the far left and far right sides of (7.9) also agree.

### 17.7.2 Changing Gauge

It may be useful to change gauges at various points during the course of integrating a trajectory and computing an associated transfer map. Suppose the gauge is to be *changed* at the point  $z = z^c$ . Let  $x^b$ ,  $y^b$ , and  $t^b$  denote coordinate functions *before* the change, and let  $x^a$ ,  $y^a$ , and  $t^a$  denote coordinate functions *after* the change. Also, let  $\mathbf{A}^b(x^b, y^b; z)$  and  $\mathbf{A}^a(x^a, y^a; z)$  be the vector potentials before ( $z < z^c$ ) and after ( $z > z^c$ ) the change point  $z^c$ . Finally, let  $p_x^{\text{canb}}$ ,  $p_y^{\text{canb}}$ ,  $p_t^{\text{canb}}$  be the canonical momentum functions before the change, and let  $p_x^{\text{cana}}$ ,  $p_y^{\text{cana}}$ ,  $p_t^{\text{cana}}$  be the canonical momentum functions after the change. In terms of these quantities, the before and after Hamiltonians  $K^b$  and  $K^a$  are given by the relations

$$K^b = -[(p_t^{\text{canb}})^2/c^2 - m^2 c^2 - (p_x^{\text{canb}} - qA_x^b)^2 - (p_y^{\text{canb}} - qA_y^b)^2]^{1/2} - qA_z^b \text{ for } z < z^c, \quad (17.7.10)$$

$$K^a = -[(p_t^{\text{cana}})^2/c^2 - m^2 c^2 - (p_x^{\text{cana}} - qA_x^a)^2 - (p_y^{\text{cana}} - qA_y^a)^2]^{1/2} - qA_z^a \text{ for } z > z^c. \quad (17.7.11)$$

What should be the matching relations between the phase-space quantities before and after? Since the choice of gauge should have no physical effect, there is the immediate requirement that the coordinate functions be continuous:

$$\begin{aligned} x^a(z) &= x^b(z) \text{ when } z = z^c, \\ y^a(z) &= y^b(z) \text{ when } z = z^c, \\ t^a(z) &= t^b(z) \text{ when } z = z^c. \end{aligned} \quad (17.7.12)$$

For the same reason, we require that the velocities, and hence the mechanical momenta, be continuous. From (7.5) and (7.6) we see that this requirement is equivalent to the relations

$$\mathbf{p}^{\text{cana}} - q\mathbf{A}^a = \mathbf{p}^{\text{canb}} - q\mathbf{A}^b \text{ when } z = z^c. \quad (17.7.13)$$

In terms of components, (7.13) yields the matching relations

$$\begin{aligned} p_x^{\text{cana}} &= p_x^{\text{canb}} + q(A_x^a - A_x^b) \text{ when } z = z^c, \\ p_y^{\text{cana}} &= p_y^{\text{canb}} + q(A_y^a - A_y^b) \text{ when } z = z^c. \end{aligned} \quad (17.7.14)$$

Finally, the total energy cannot change under a gauge transformation and therefore, since we have assumed that the scalar potential  $\psi$  vanishes, there is the matching relation

$$p_t^{\text{cana}} = p_t^{\text{canb}} \text{ when } z = z^c. \quad (17.7.15)$$

We note that this relation also follows from (1.6.17), (7.5), (7.6), and (7.13).

We assume there is some common overlap region where both  $\mathbf{A}^b$  and  $\mathbf{A}^a$  are defined. Since they both give rise to the same magnetic field, there is the relation

$$\nabla \times (\mathbf{A}^a - \mathbf{A}^b) = 0. \quad (17.7.16)$$

It follows that there is a function  $\chi$  such that

$$\mathbf{A}^a - \mathbf{A}^b = \nabla \chi. \quad (17.7.17)$$

Consequently, the relations (7.14) can be rewritten in the form

$$\begin{aligned} p_x^{\text{cana}} &= p_x^{\text{canb}} + q(\partial/\partial x)\chi \text{ when } z = z^c, \\ p_y^{\text{cana}} &= p_y^{\text{canb}} + q(\partial/\partial y)\chi \text{ when } z = z^c. \end{aligned} \quad (17.7.18)$$

There is one last step. Let  $\mathcal{T}^c$  be the symplectic *transformation* map defined by the relation

$$\mathcal{T}^c = \exp(q : \chi :). \quad (17.7.19)$$

With aid of this map it is easily verified that the relations (7.12), (7.14), and (7.15) can be rewritten in the form

$$\begin{aligned} x^a(z) &= \exp(q : \chi :)x^b(z) \text{ with } z = z^c, \\ y^a(z) &= \exp(q : \chi :)y^b(z) \text{ with } z = z^c, \\ t^a(z) &= \exp(q : \chi :)t^b(z) \text{ with } z = z^c; \end{aligned} \quad (17.7.20)$$

$$\begin{aligned} p_x^{\text{cana}}(z) &= \exp(q : \chi :)p_x^{\text{canb}}(z) \text{ with } z = z^c, \\ p_y^{\text{cana}}(z) &= \exp(q : \chi :)p_y^{\text{canb}}(z) \text{ with } z = z^c, \\ p_t^{\text{cana}}(z) &= \exp(q : \chi :)p_t^{\text{canb}}(z) \text{ with } z = z^c. \end{aligned} \quad (17.7.21)$$

We have determined that a change in gauge amounts to making a symplectic transformation. Review Exercises 6.2.8 and 6.5.3.

### 17.7.3 Finding the Minimal Vector Potential

The goal of this subsection is, given  $\mathbf{B}(\mathbf{r})$  in some region, to find an associated vector potential  $\mathbf{A}^s$  that is as *small*/minimal as possible in the sense that  $\mathbf{A}^s$  is small if  $\mathbf{B}(\mathbf{r})$  is small. The reason for this goal will become apparent in following subsections.

Our plan is as follows: Make Taylor expansions, with initially unknown coefficients, for the Cartesian components of  $\mathbf{A}^s$ , organize these expansions into homogeneous polynomials, and then further organize them as spherical polynomial vector fields. Then use this representation to compute and organize  $\nabla \times \mathbf{A}^s$  in terms of spherical polynomial vector fields. At the same time parameterize  $\mathbf{B}(\mathbf{r})$  in terms of a scalar potential  $\psi$  expanded in harmonic polynomials. Finally, compare the two expansions for  $\mathbf{B}(\mathbf{r})$  given by  $\mathbf{B} = \nabla\psi$  and  $\mathbf{B} = \nabla \times \mathbf{A}^s$ , equate coefficients of like terms, and thereby determine the coefficients in the Taylor expansion for the components of  $\mathbf{A}^s$  in terms of the coefficients in the expansion for  $\psi$ . For the notation and machinery required for the execution of this plan, see Appendix U.

We begin with an expansion for  $\mathbf{B}(\mathbf{r})$  based on the use of a scalar potential. Without loss of generality, we may take the region of interest to be centered at the origin. We also assume that  $\mathbf{B}(\mathbf{r})$  has a Taylor expansion in the components of  $\mathbf{r}$  and is divergence and curl free. In this case there is a harmonic magnetic scalar potential  $\psi(\mathbf{r})$  such that

$$\mathbf{B} = \nabla\psi. \quad (17.7.22)$$

Recall the beginning of Section 15.2. Employing the notation of Appendix U, we may assume without loss of generality that  $\psi$  has a spherical polynomial expansion of the form

$$\psi(\mathbf{r}) = \sum_{n=1}^{n_{\max}} \sum_m d_{nm} S_{nn}^m(\mathbf{r}) \quad (17.7.23)$$

where the quantities  $d_{nm}$  are arbitrary coefficients. Here we assume an expansion through terms of degree  $n_{\max}$  and omit  $n = 0$  terms since constant terms make no contribution to  $\mathbf{B}$  as given by (7.22).

For the associated vector potential  $\mathbf{A}^s$  we make the spherical polynomial vector field expansion

$$\mathbf{A}^s(\mathbf{r}) = \sum_{n=1}^{n_{\max}} \sum_{\ell} \sum_J \sum_M c_{n\ell JM} \mathbf{S}_{n\ell J}^M(\mathbf{r}). \quad (17.7.24)$$

Again see Appendix U. Given the coefficients  $d_{nm}$ , our task is to use the equality

$$\nabla \times \mathbf{A}^s(\mathbf{r}) = \nabla \times \sum_{n=1}^{n_{\max}} \sum_{\ell} \sum_J \sum_M c_{n\ell JM} \mathbf{S}_{n\ell J}^M(\mathbf{r}) = \nabla \sum_{n=1}^{n_{\max}} \sum_m d_{nm} S_{nn}^m(\mathbf{r}) = \nabla \psi(\mathbf{r}) \quad (17.7.25)$$

to find the coefficients  $c_{n\ell JM}$ .

Let us begin by evaluating the right side of (7.25). We find the results

$$\mathbf{B}(r) = \nabla \psi(\mathbf{r}) = \nabla \sum_{n=1}^{n_{\max}} \sum_m d_{nm} S_{nn}^m(\mathbf{r}) = \sum_{n=1}^{n_{\max}} \sum_m d_{nm} \sqrt{n(2n+1)} \mathbf{S}_{n-1,n-1,n}^m(\mathbf{r}). \quad (17.7.26)$$

Here we have used (U.5.3).

Next work on evaluating the left side of (7.25). This is a more complicated task. In accord with the range rules (U.3.7) and (U.3.8) we decompose the expansion into the sum of four pieces with each containing a particular kind of term:

- a) All terms for which  $\ell = 0$  and hence  $J = 1$ . Also, therefore,  $n = 2k$  with  $k > 0$ . The associated spherical polynomial vectors are of the form  $\mathbf{S}_{2k,0,1}^M(\mathbf{r})$ .
- b) All terms for which  $\ell > 0$  and  $J = \ell + 1$ . The associated spherical polynomial vectors are of the form  $\mathbf{S}_{n,\ell,\ell+1}^M(\mathbf{r})$ .
- c) All terms for which  $\ell > 0$  and  $J = \ell$ . The associated spherical polynomial vectors are of the form  $\mathbf{S}_{n,\ell,\ell}^M(\mathbf{r})$ .
- d) All terms for which  $\ell > 0$  and  $J = \ell - 1$ . The associated spherical polynomial vectors are of the form  $\mathbf{S}_{n,\ell,\ell-1}^M(\mathbf{r})$ .

Thus, we write

$$\mathbf{A}^s = \mathbf{A}^{sa} + \mathbf{A}^{sb} + \mathbf{A}^{sc} + \mathbf{A}^{sd} \quad (17.7.27)$$

where

$$\mathbf{A}^{sa}(\mathbf{r}) = \sum_{k=1}^{k_{\max}} \sum_M c_{2k,0,1,M} \mathbf{S}_{2k,0,1}^M(\mathbf{r}), \quad (17.7.28)$$

$$\mathbf{A}^{sa}(\mathbf{r}) = \sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M c_{n,\ell,\ell+1,M} \mathbf{S}_{n,\ell,\ell+1}^M(\mathbf{r}), \quad (17.7.29)$$

$$\mathbf{A}^{sc}(\mathbf{r}) = \sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M c_{n,\ell,\ell,M} \mathbf{S}_{n,\ell,\ell}^M(\mathbf{r}), \quad (17.7.30)$$

$$\mathbf{A}^{sd}(\mathbf{r}) = \sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M c_{n,\ell,\ell-1,M} \mathbf{S}_{n,\ell,\ell-1}^M(\mathbf{r}). \quad (17.7.31)$$

We are now ready to proceed. For the  $\mathbf{A}^{sa}$  term we find, using (U.5.20), the result

$$\nabla \times \mathbf{A}^{sa}(\mathbf{r}) = \nabla \times \sum_{k=1}^{k_{\max}} \sum_M c_{2k,0,1,M} \mathbf{S}_{2k,0,1}^M(\mathbf{r}) = \sum_{k=1}^{k_{\max}} \sum_M c_{2k,0,1,M} [i(\sqrt{2/3})(2k)] \mathbf{S}_{2k-1,1,1}^M(\mathbf{r}). \quad (17.7.32)$$

For the  $\mathbf{A}^{sb}$  term we find, using (U.5.17), the result

$$\begin{aligned} \nabla \times \mathbf{A}^{sb}(\mathbf{r}) &= \nabla \times \sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M c_{n,\ell,\ell+1,M} \mathbf{S}_{n,\ell,\ell+1}^M(\mathbf{r}) = \\ &\sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M c_{n,\ell,\ell+1,M} [i\sqrt{(\ell+2)/(2\ell+3)}(n-\ell)] \mathbf{S}_{n-1,\ell+1,\ell+1}^M(\mathbf{r}). \end{aligned} \quad (17.7.33)$$

For the  $\mathbf{A}^{sc}$  term we find, using (U.5.18), the result

$$\begin{aligned} \nabla \times \mathbf{A}^{sc}(\mathbf{r}) &= \nabla \times \sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M c_{n,\ell,\ell,M} \mathbf{S}_{n,\ell,\ell}^M(\mathbf{r}) = \\ &\sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M c_{n,\ell,\ell,M} [i\sqrt{(\ell+1)/(2\ell+1)}(n+\ell+1)] \mathbf{S}_{n-1,\ell-1,\ell}^M(\mathbf{r}) \\ &+ \sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M c_{n,\ell,\ell,M} [i\sqrt{\ell/(2\ell+1)}(n-\ell)] \mathbf{S}_{n-1,\ell+1,\ell}^M(\mathbf{r}). \end{aligned} \quad (17.7.34)$$

Finally, for the  $\mathbf{A}^{sd}$  term we find, using (U.5.19), the result

$$\begin{aligned} \nabla \times \mathbf{A}^{sd}(\mathbf{r}) &= \nabla \times \sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M c_{n,\ell,\ell-1,M} \mathbf{S}_{n,\ell,\ell-1}^M(\mathbf{r}) = \\ &\sum_{n=1}^{n_{\max}} \sum_{\ell>0} \sum_M c_{n,\ell,\ell-1,M} [i\sqrt{(\ell-1)/(2\ell-1)}(n+\ell+1)] \mathbf{S}_{n-1,\ell-1,\ell-1}^M(\mathbf{r}). \end{aligned} \quad (17.7.35)$$

We are now prepared to equate coefficients of like terms. Let us begin with the first few corresponding to small values of  $n$ . The first of these, corresponding to  $n = 0$ , is  $\mathbf{S}_{0,0,1}^M$ . From (7.26) we see that

$$\text{coefficient of } \mathbf{S}_{0,0,1}^M \text{ in } \nabla \psi = \sqrt{3} d_{1M}. \quad (17.7.36)$$

We next examine the terms in  $\nabla \times \mathbf{A}^s$ : From (7.32) we see that there are no terms of the desired kind, namely terms involving  $\mathbf{S}_{0,0,1}^M$ , in  $\nabla \times \mathbf{A}^{sa}$ . From (7.33) we see that there are no terms of the desired kind in  $\nabla \times \mathbf{A}^{sb}$ . From (7.34) we see that there are terms of the desired kind in  $\nabla \times \mathbf{A}^{sc}$ , and find the relation

$$\text{coefficient of } \mathbf{S}_{0,0,1}^M \text{ in } \nabla \times \mathbf{A}^{sc} = i\sqrt{6} c_{1,1,1,M}. \quad (17.7.37)$$

Finally, from (7.35) we see that there are no terms of the desired kind in  $\nabla \times \mathbf{A}^{sd}$ .

Upon comparing (7.36) and (7.37) we conclude that there must be the relation

$$i\sqrt{6} c_{1,1,1,M} = \sqrt{3} d_{1M}, \quad (17.7.38)$$

and therefore

$$c_{1,1,1,M} = -i\sqrt{1/2} d_{1M}. \quad (17.7.39)$$

Note that this relation is consistent with (U.6.39). Moreover, we conclude that the six remaining  $n = 1$  coefficients in  $\mathbf{A}^s$ , namely  $c_{1,1,0,0}$  and the  $c_{1,1,2,M}$ , can be anything since there are the relations (U.6.38) and (U.6.40). For simplicity, we set these coefficients to zero. Then, so far, we have the result

$$\mathbf{A}^s(\mathbf{r}) = \sum_M (-i)\sqrt{1/2} d_{1M} \mathbf{S}_{111}^M(\mathbf{r}) + \text{terms of degree } > 1. \quad (17.7.40)$$

In terms of Cartesian components, (7.40) yields the relation

$$\mathbf{A}^s(\mathbf{r}) = -(1/2)\mathbf{r} \times \mathbf{B}(0) + \text{terms of degree } > 1. \quad (17.7.41)$$

Here we have used (7.22), (7.23), and (U.6.25) evaluated for  $n = 1$ . We observe that this choice for the leading term in  $\mathbf{A}^s$  is in the symmetric/Poincaré/Coulomb gauge. See Exercise 28.2.7.

Let us push on to the case  $n = 1$ ; in which case there are the spherical polynomial vector fields  $\mathbf{S}_{110}^0$ ,  $\mathbf{S}_{111}^M$  with  $-1 \leq M \leq 1$ , and  $\mathbf{S}_{112}^M$  with  $-2 \leq M \leq 2$ . First see where/how they occur in  $\nabla\psi$ . Examination of (7.26) shows that the only such term in  $\nabla\psi$  is  $\mathbf{S}_{112}^M$ , and we have the relation

$$\text{coefficient of } \mathbf{S}_{1,1,2}^M \text{ in } \nabla\psi = \sqrt{10} d_{2M}. \quad (17.7.42)$$

We next examine the terms in  $\nabla \times \mathbf{A}^s$ : From (7.32) we see that there are no terms of the desired kind, namely terms involving  $\mathbf{S}_{1,1,2}^M$ , in  $\nabla \times \mathbf{A}^{sa}$ . From (7.33) we see that there are no terms of the desired kind in  $\nabla \times \mathbf{A}^{sb}$ . From (7.34) we see that there are terms of the desired kind in  $\nabla \times \mathbf{A}^{sc}$ , and find the relation

$$\text{coefficient of } \mathbf{S}_{1,1,2}^M \text{ in } \nabla \times \mathbf{A}^{sc} = i\sqrt{15} c_{2,2,2,M}. \quad (17.7.43)$$

Finally, from (7.35) we see that there are no terms of the desired kind in  $\nabla \times \mathbf{A}^{sd}$ .

Upon comparing (7.42) and (7.43) we conclude that there must be the relation

$$i\sqrt{15} c_{2,2,2,M} = \sqrt{10} d_{2M}, \quad (17.7.44)$$

and therefore

$$c_{2,2,2,M} = -i\sqrt{2/3} d_{2M}. \quad (17.7.45)$$

What can be said about the thirteen remaining  $n = 2$  coefficients in  $\mathbf{A}^s$ , namely the  $c_{201M}$ ,  $c_{2,2,3,M}$ , and  $c_{2,2,1,M}$ ? It can be shown that  $\nabla \times \mathbf{S}_{223}^M(\mathbf{r}) = 0$ , and therefore the terms with coefficients  $c_{2,2,3,M}$  make no contribution to  $\mathbf{B}(\mathbf{r})$ . See Exercise (U.6.21). For simplicity, we set these coefficients to zero. It can be shown that terms with the coefficients  $c_{201M}$  and  $c_{2,2,1,M}$  produce terms in  $\mathbf{B}(\mathbf{r})$  having nonzero curl. Again see Exercise (U.6.21). We also set these coefficients to zero to ensure that  $\mathbf{B}(\mathbf{r})$  is curl free. Then, so far, we have the result

$$\mathbf{A}^s(\mathbf{r}) = \sum_M (-i) \sqrt{1/2} d_{1M} \mathbf{S}_{111}^M(\mathbf{r}) + \sum_M (-i) \sqrt{2/3} d_{2M} \mathbf{S}_{222}^M(\mathbf{r}) + \text{terms of degree } > 2. \quad (17.7.46)$$

The pattern should now be clear. There are the general relations

$$\nabla S_{nn}^M(\mathbf{r}) = \sqrt{n(2n+1)} \mathbf{S}_{n-1,n-1,n}^M(\mathbf{r}) \quad (17.7.47)$$

and

$$\nabla \times \mathbf{S}_{n,n,n}^M(\mathbf{r}) = i \sqrt{(n+1)(2n+1)} \mathbf{S}_{n-1,n-1,n}^M(\mathbf{r}). \quad (17.7.48)$$

Therefore there is the general relation

$$\mathbf{A}^s(\mathbf{r}) = \sum_{n=1}^{n_{\max}} \sum_{M=-n}^n (-i) \sqrt{n/(n+1)} d_{nM} \mathbf{S}_{nnn}^M(\mathbf{r}). \quad (17.7.49)$$

It can be verified that this particular choice of  $\mathbf{A}^s(\mathbf{r})$  has the two properties

$$\nabla \cdot \mathbf{A}^s(\mathbf{r}) = 0 \quad (17.7.50)$$

and

$$\mathbf{r} \cdot \mathbf{A}^s(\mathbf{r}) = 0. \quad (17.7.51)$$

See (U.5.11) and (U.6.9). Thus this vector potential is in both a Coulomb and Poincaré gauge.

The relation (7.49) can be further manipulated using (U.6.25). Doing so gives the result

$$\mathbf{A}^s(\mathbf{r}) = - \sum_{n=1}^{n_{\max}} \sum_{M=-n}^n [1/(n+1)] d_{nM} [\mathbf{r} \times \nabla S_{nn}^M(\mathbf{r})]. \quad (17.7.52)$$

We observe that this result agrees with that found by Ansatz in Exercises (15.5.8) and (15.5.9). See (15.5.81) and (15.5.82).

Have we achieved our goal of finding a “minimal vector potential”? We have, in the following sense: Inspection of (7.26) shows that it provides an expansion of  $\mathbf{B}(r)$  in terms of spherical polynomial vector fields  $\mathbf{S}_{n-1,n-1,n}^m(\mathbf{r})$  with expansion coefficients proportional to the  $d_{nm}$ . Inspection of (7.49) shows that it provides an expansion of  $\mathbf{A}^s(r)$  in terms of spherical polynomial vector fields  $\mathbf{S}_{nnn}^M(\mathbf{r})$  with expansion coefficients again proportional to the  $d_{nM}$ . The vector potential  $\mathbf{A}^s(\mathbf{r})$  has no constant part, and its non-constant parts are directly proportional to the coefficients  $d_{nm}$  that describe the constant and non-constant parts of  $\mathbf{B}(r)$ . Moreover, there is an order-by-order relation. Terms of order  $n$  in  $\mathbf{A}^s(\mathbf{r})$  are proportional to terms of order  $n-1$  in  $\mathbf{B}(r)$ . Thus,  $\mathbf{A}^s(\mathbf{r})$  is small if  $\mathbf{B}(r)$  is small. In particular, if high-order terms in  $\mathbf{B}(r)$  are negligible, they will also be negligible in  $\mathbf{A}^s(\mathbf{r})$ .

There is yet another sense in which the vector potential we have found is minimal. Suppose, for example, that we confine our attention to the case of a vector potential that is homogeneous of degree 1, which is the case we need to produce a constant magnetic field. When  $n = 1$  we see from Table U.3.1 that  $\ell = 1$  and  $J = 0, 1, 2$ . Therefore, such a vector potential, call it  $\mathbf{A}^{[1]}$ , can be written in the form

$$\mathbf{A}^{[1]}(\mathbf{r}) = \sum_J \sum_M c_{11JM} \mathbf{S}_{11J}^M(\mathbf{r}). \quad (17.7.53)$$

Recall (7.24). Let us compute the *norm* of  $\mathbf{A}^{[1]}$  as defined by the relation

$$\|\mathbf{A}^{[1]}(\mathbf{r})\|^2 = \int d\Omega [\mathbf{A}^{[1]}(\mathbf{r})]^* \cdot \mathbf{A}^{[1]}(\mathbf{r}). \quad (17.7.54)$$

Since the  $\mathbf{S}_{11J}^M(\mathbf{r})$  are mutually orthogonal under angular integration, we find from (7.53), (U.3.18), and (U.4.3) the result

$$\|\mathbf{A}^{[1]}(\mathbf{r})\|^2 = r^2 \sum_J \sum_M |c_{11JM}|^2. \quad (17.7.55)$$

We know the value of  $c_{111M}$  is fixed by (7.39), and we have chosen to set the remaining  $c_{11JM}$  to zero. We now see, since (7.55) is a sum of squares, that doing so *minimizes*  $\|\mathbf{A}^{[1]}(\mathbf{r})\|$ . Similar computations may be made for other values of  $n$ . The result is that the choice we have made for  $\mathbf{A}^s$  minimizes  $\|\mathbf{A}^{s[n]}(\mathbf{r})\|$  for each value of  $n$ .

#### 17.7.4 The $m = 0$ Case: Solenoid Example

In this subsection we will explore the fringe fields for a solenoid. Our aim will be to compare the vector potential in the symmetric Coulomb gauge as given in Section 15.4 and the vector potential in the minimum gauge.<sup>13</sup> Reference to Section 20.1.2 shows that, for a simple air-core solenoid,  $C_0^{[1]}(z)$  falls off as  $1/|z|^3$  for large  $|z|$ . See (20.1.28) and Figures 20.1.3 and 20.1.4. Correspondingly, in this case, the  $C_0^{[n+1]}(z)$  will fall off as  $1/|z|^{n+3}$  for large  $|z|$ . We expect the simple air-core to be representative of the worst scenario in the sense that the fringe fields for other kinds of solenoids will fall off at this same rate or *faster*.

According to Subsection 15.2.3, the scalar potential for the  $m = 0$  case is given by the relation

$$\psi_0(x, y, z) = C_0^{[0]}(z) - (1/4)(x^2 + y^2)C_0^{[2]}(z) + \dots \quad (17.7.56)$$

Let us expand  $\psi$  about the point  $(0, 0, z_0)$ . To do so, introduce local deviation variables  $\xi, \eta$ , and  $\zeta$  by making the definitions

$$x = 0 + \xi, \quad (17.7.57)$$

$$y = 0 + \eta, \quad (17.7.58)$$

$$z = z_0 + \zeta. \quad (17.7.59)$$

---

<sup>13</sup>Note that according to Section 15.5 there are a variety of Coulomb gauges including vertical-free and horizontal-free Coulomb gauges. Here we treat the case where there is the greatest symmetry between the vertical and horizontal components of  $\mathbf{A}$ .

Also define a *deviation* vector  $\mathbf{r}^d$  by writing

$$\mathbf{r}^d = \xi \mathbf{e}_x + \eta \mathbf{e}_y + \zeta \mathbf{e}_z. \quad (17.7.60)$$

We can then define a scalar potential  $\psi^e$  suitable for *expansion* by writing the relation

$$\psi^e(\mathbf{r}^d; z_0) = \psi(\xi, \eta, z_0 + \zeta). \quad (17.7.61)$$

Indeed, making use of (7.56) yields for  $\psi^e$  the expansion

$$\begin{aligned} \psi^e(\mathbf{r}^d; z_0) &= C_0^{[0]}(z_0 + \zeta) - (1/4)(\xi^2 + \eta^2)C_0^{[2]}(z_0 + \zeta) + \text{terms of order 3 and higher} \\ &= C_0^{[0]}(z_0) + C_0^{[1]}(z_0)\zeta \\ &\quad + C_0^{[2]}(z_0)(\zeta^2/2) - (1/4)(\xi^2 + \eta^2)C_0^{[2]}(z_0) \\ &\quad + \text{terms of order 3 and higher} \\ &= \psi^{e[0]} + \psi^{e[1]} + \psi^{e[2]} + \text{terms of order 3 and higher}. \end{aligned} \quad (17.7.62)$$

Here the upper index in square brackets on a quantity denotes its degree.<sup>14</sup> And, from (7.22), the magnetic field associated with this expansion is given by the expansion

$$\mathbf{B} = \mathbf{B}^{[0]} + \mathbf{B}^{[1]} + \text{terms of order 2 and higher} \quad (17.7.63)$$

where

$$\mathbf{B}^{[0]} = C_0^{[1]}(z_0)\mathbf{e}_z \quad (17.7.64)$$

and

$$\begin{aligned} \mathbf{B}^{[1]} &= -(1/2)C_0^{[2]}(z_0)(\xi \mathbf{e}_x + \eta \mathbf{e}_y) + C_0^{[2]}(z_0)\zeta \mathbf{e}_z \\ &= -(1/2)C_0^{[2]}(z_0)(\xi \mathbf{e}_x + \eta \mathbf{e}_y - 2\zeta \mathbf{e}_z) \\ &= -(1/2)C_0^{[2]}(z_0)(\xi \mathbf{e}_x + \eta \mathbf{e}_y + \zeta \mathbf{e}_z - 3\zeta \mathbf{e}_z) \\ &= -(1/2)C_0^{[2]}(z_0)(\mathbf{r}^d - 3\zeta \mathbf{e}_z). \end{aligned} \quad (17.7.65)$$

Let us find the associated minimum vector potential. According to (7.52), we expect that  $\mathbf{A}^s$  will be of the form

$$\mathbf{A}^s(\mathbf{r}^d) = \mathbf{A}^{s[1]}(\mathbf{r}^d) + \mathbf{A}^{s[2]}(\mathbf{r}^d) \quad (17.7.66)$$

with

$$\mathbf{A}^{s[1]}(\mathbf{r}^d) = -(1/2)\mathbf{r}^d \times \mathbf{B}^{[0]}(\mathbf{r}^d) \quad (17.7.67)$$

and

$$\mathbf{A}^{s[2]}(\mathbf{r}) = -(1/3)\mathbf{r}^d \times \mathbf{B}^{[1]}(\mathbf{r}^d). \quad (17.7.68)$$

Working out the indicated cross products in (7.67) and (7.68) gives the results

$$\mathbf{A}^{s[1]}(\mathbf{r}^d) = -(1/2)C_0^{[1]}(z_0)(\eta \mathbf{e}_x - \xi \mathbf{e}_y), \quad (17.7.69)$$

---

<sup>14</sup>Note that according to Section U.7 the  $\psi^{e[n]}$  will be homogenous harmonic polynomials of degree  $n$ .

$$\mathbf{A}^{s[2]}(\mathbf{r}^d) = -(1/2)C_0^{[2]}(z_0)(\zeta\eta\mathbf{e}_x - \zeta\xi\mathbf{e}_y). \quad (17.7.70)$$

Simple calculation verifies that there are indeed the relations

$$\nabla \times \mathbf{A}^{s[1]}(\mathbf{r}^d) = \mathbf{B}^{[0]}(\mathbf{r}^d), \quad (17.7.71)$$

$$\nabla \times \mathbf{A}^{s[2]}(\mathbf{r}^d) = \mathbf{B}^{[1]}(\mathbf{r}^d), \quad (17.7.72)$$

as desired.

How do the results given by (7.69) and (7.70) compare with those provided by the symmetric Coulomb gauge? According to Section 15.4, the vector potential in the symmetric Coulomb gauge for the  $m = 0$  case is given by the expressions

$$\begin{aligned} \hat{A}_x^0(\mathbf{r}) &= -(y/2) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{1}{2^{2\ell}\ell!(\ell+1)!} C_0^{[2\ell+1]}(z)(x^2 + y^2)^\ell \\ &= -(y/2)[C_0^{[1]}(z) - (1/8)C_0^{[3]}(z)(x^2 + y^2) + \dots], \end{aligned} \quad (17.7.73)$$

$$\begin{aligned} \hat{A}_y^0(\mathbf{r}) &= (x/2) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{1}{2^{2\ell}\ell!(\ell+1)!} C_0^{[2\ell+1]}(z)(x^2 + y^2)^\ell \\ &= (x/2)[C_0^{[1]}(z) - (1/8)C_0^{[3]}(z)(x^2 + y^2) + \dots], \end{aligned} \quad (17.7.74)$$

$$\hat{A}_z^0(\mathbf{r}) = 0. \quad (17.7.75)$$

In terms of the expansion variables (7.57) through (7.59) these expressions become

$$\begin{aligned} \hat{A}_x^0(\xi, \eta, z_0 + \zeta) &= -(\eta/2)[C_0^{[1]}(z_0 + \zeta) - (1/8)C_0^{[3]}(z_0 + \zeta)(\xi^2 + \eta^2) + \dots] \\ &= -(\eta/2)[C_0^{[1]}(z_0) + C_0^{[2]}(z_0)\zeta] + \dots \\ &= -(1/2)C_0^{[1]}(z_0)\eta - (1/2)C_0^{[2]}(z_0)\zeta\eta + \text{terms of order 3 and higher}, \end{aligned} \quad (17.7.76)$$

$$\begin{aligned} \hat{A}_y^0(\xi, \eta, z_0 + \zeta) &= (\xi/2)[C_0^{[1]}(z_0 + \zeta) - (1/8)C_0^{[3]}(z_0 + \zeta)(\xi^2 + \eta^2) + \dots] \\ &= (\xi/2)[C_0^{[1]}(z_0) + C_0^{[2]}(z_0)\zeta] + \dots \\ &= (1/2)C_0^{[1]}(z_0)\xi + (1/2)C_0^{[2]}(z_0)\zeta\xi + \text{terms of order 3 and higher}, \end{aligned} \quad (17.7.77)$$

$$\hat{A}_z^0(\xi, \eta, z_0 + \zeta) = 0. \quad (17.7.78)$$

Comparison of (7.69) and (7.70) with (7.76) through (7.78) shows that for the  $m = 0$  case, at least for the orders computed, the minimum vector potential constructed from field expansions about on-axis points *agrees* with the symmetric Coulomb gauge vector potential constructed from on-axis field data. In retrospect, this result should not be surprising. We should expect agreement through *all* orders because, according to Subsection 15.6.2, the  $m = 0$  symmetric Coulomb gauge vector potential constructed from on-axis field data is, in fact, in the Poincaré-Coulomb gauge.

### 17.7.5 The $m = 1$ Case: Magnetic Monopole Doublet and Wiggler Examples

In this subsection we will first study the behavior of the leading and trailing fringe fields for a magnetic monopole doublet. Subsequently we will examine the case of a wiggler.

#### Magnetic Monopole Doublet Example

The doublet will be located at the origin  $(0, 0, 0)$  as in Subsection 15.8.1 and we will expand the scalar potential  $\psi(x, y, z)$  given by (15.8.3) about the mid-plane point  $(x_0, 0, z_0)$ . Consequently, if  $z_0 \ll 0$ , we will obtain an expansion in the leading region, and if  $z_0 \gg 0$ , we will obtain an expansion in the trailing region. Moreover, if  $x_0 = 0$ , the expansion will be on axis; and setting  $x_0 \neq 0$  allows for expansion about a point on the design orbit. See Subsection 21.5.1 and Figures 21.5.1 and 21.5.6.

As before, introduce local deviation variables  $\xi, \eta$ , and  $\zeta$  and a deviation vector  $\mathbf{r}^d$  by making the definitions

$$x = x_0 + \xi, \quad (17.7.79)$$

$$y = \eta, \quad (17.7.80)$$

$$z = z_0 + \zeta, \quad (17.7.81)$$

and (7.60). We can then define a scalar potential  $\psi^e$  by writing the relation

$$\psi^e(\mathbf{r}^d; x_0, z_0) = \psi(x_0 + \xi, \eta, z_0 + \zeta). \quad (17.7.82)$$

Indeed, making use of (15.8.3) and (7.82) yields the expansion

$$\begin{aligned} \psi^e(\mathbf{r}^d; x_0, z_0) &= [-2ga/(x_0^2 + z_0^2 + a^2)^{3/2}]\eta \\ &\quad + [6ga/(x_0^2 + z_0^2 + a^2)^{5/2}][\eta(x_0\xi + z_0\zeta)] \\ &\quad + \text{terms of order 3 and higher.} \end{aligned} \quad (17.7.83)$$

Note that  $\psi(x_0, 0, z_0)$  vanishes so that there is no constant term in the expansion (7.83). We observe that the first term in (7.83) falls off as  $(1/x_0)^3$  or  $(1/z_0)^3$  for large  $x_0$  or  $z_0$ , and the second falls off as  $(1/x_0)^4$  or  $(1/z_0)^4$ . In general, successive terms fall off with ever increasing powers of  $(1/x_0)$  or  $(1/z_0)$ .

Let us compute the magnetic field  $\mathbf{B}$  associated with the first two terms in (7.83). We find the result

$$\begin{aligned} \mathbf{B}(\mathbf{r}^d; x_0, z_0) &= -[2ga/(x_0^2 + z_0^2 + a^2)^{3/2}]\mathbf{e}_y \\ &\quad + [6ga/(x_0^2 + z_0^2 + a^2)^{5/2}](x_0\xi + z_0\zeta)\mathbf{e}_y \\ &\quad + [6ga/(x_0^2 + z_0^2 + a^2)^{5/2}][\eta(x_0\mathbf{e}_x + z_0\mathbf{e}_z)]. \end{aligned} \quad (17.7.84)$$

Next let us find the minimum vector potential  $\mathbf{A}^s$  associated with the first two terms in (7.83). Begin by decomposing  $\mathbf{B}$  into homogeneous polynomials by rewriting (7.84) in the form (7.63) with

$$\mathbf{B}^{[0]}(\mathbf{r}^d) = -[2ga/(x_0^2 + z_0^2 + a^2)^{3/2}]\mathbf{e}_y \quad (17.7.85)$$

and

$$\mathbf{B}^{[1]}(\mathbf{r}^d) = [6ga/(x_0^2 + z_0^2 + a^2)^{5/2}][(x_0\xi + z_0\zeta)\mathbf{e}_y + \eta(x_0\mathbf{e}_x + z_0\mathbf{e}_z)]. \quad (17.7.86)$$

The minimum vector potential associated with this magnetic field will again be given by the relations (7.66) through (7.68). Working out the indicated cross products yields the results

$$\mathbf{A}^{s[1]}(\mathbf{r}^d) = [ga/(x_0^2 + z_0^2 + a^2)^{3/2}](-\zeta\mathbf{e}_x + \xi\mathbf{e}_z), \quad (17.7.87)$$

$$\begin{aligned} \mathbf{A}^{s[2]}(\mathbf{r}^d) &= [-2ga/(x_0^2 + z_0^2 + a^2)^{5/2}] \times \\ &[(z_0\eta^2 - z_0\zeta^2 - x_0\xi\zeta)\mathbf{e}_x + (x_0\eta\zeta - z_0\xi\eta)\mathbf{e}_y + (x_0\xi^2 + z_0\xi\zeta - x_0\eta^2)\mathbf{e}_z]. \end{aligned} \quad (17.7.88)$$

Simple calculation verifies that there are indeed the relations (7.71) and (7.72) as desired.

At this point it is instructive to compare the minimum vector potential with other possible vector potentials. We note that the design orbit for a dipole field is *curved*, and therefore for the most part does not lie on axis. Consequently we must generally compare the minimum vector potential with other possible vector potentials at off-axis points. By construction, the minimum vector potential *vanishes* at every expansion point. In contrast, other vector potentials (for example those based on employing on-axis expansions of a scalar potential or the use of Dirac strings) generally not have this property. We conclude that problems involving curved design orbits are more complicated than those for straight beam-line elements, and their treatment requires special care. This treatment is deferred to Chapter 21.

### On-axis Entry and Exit Wiggler Example

There is an application for which expansions of  $m = 1$  cylindrical harmonics may be useful, namely the case of wigglers when the excursion of the design orbit from the axis may be treated as small. That is, it is assumed that the design orbit enters and exits the wiggler on axis and nearly along the axis, and the excursions of the design orbit from the axis while within the wiggler may be treated as small.

According to (15.2.61) the scalar potential for a (normal) dipole is given by the relation

$$\psi_{1,s}(x, y, z) = y[C_{1,s}^{[0]}(z) - (1/8)(x^2 + y^2)C_{1,s}^{[2]}(z) + \dots]. \quad (17.7.89)$$

Upon invoking the definitions (7.57) through (7.60), the expansion (7.89) yields the expansion

$$\begin{aligned} \psi^e(\mathbf{r}^d; z_0) &= \eta[C_{1,s}^{[0]}(z_0 + \zeta) - (1/8)(\xi^2 + \eta^2)C_{1,s}^{[2]}(z_0 + \zeta) + \dots] \\ &= \eta C_{1,s}^{[0]}(z_0) + \eta\zeta C_{1,s}^{[1]}(z_0) + \text{terms of order 3 and higher.} \end{aligned} \quad (17.7.90)$$

The magnetic field associated with the scalar potential (7.90) has an expansion of the form (7.63) with

$$\mathbf{B}^{[0]} = C_{1,s}^{[0]}(z_0)\mathbf{e}_y \quad (17.7.91)$$

and

$$\mathbf{B}^{[1]} = \zeta C_{1,s}^{[1]}(z_0) \mathbf{e}_y + \eta C_{1,s}^{[1]}(z_0) \mathbf{e}_z. \quad (17.7.92)$$

The first two terms in minimum vector potential expansion associated with this magnetic field will again be given by the relations (7.66) through (7.68). Working out the indicated cross products now yields the results

$$\mathbf{A}^{s[1]}(\mathbf{r}^d) = (1/2)C_{1,s}^{[0]}(z_0)(\zeta \mathbf{e}_x - \xi \mathbf{e}_z), \quad (17.7.93)$$

$$\mathbf{A}^{s[2]}(\mathbf{r}^d) = (1/3)C_{1,s}^{[1]}(z_0)[(\zeta^2 - \eta^2)\mathbf{e}_x + \xi\eta\mathbf{e}_y - \xi\zeta\mathbf{e}_z]. \quad (17.7.94)$$

And again simple calculation verifies that there are indeed the relations (7.71) and (7.72) as desired.

How does the the vector potential for a normal dipole in the Coulomb gauge compare with the minimum vector potential just found? From (15.4.95) through (15.4.97) we find, in the Coulomb gauge, that  $\hat{\mathbf{A}}^{1,s}$  has the expansion

$$\begin{aligned} \hat{A}_x^{1,s}(\xi, \eta, z_0 + \zeta) &= (1/4)(\xi^2 - \eta^2)C_{1,s}^{[1]}(z_0 + \zeta) + \dots \\ &= (1/4)(\xi^2 - \eta^2)C_{1,s}^{[1]}(z_0) + \text{terms of order 3 and higher}, \end{aligned} \quad (17.7.95)$$

$$\begin{aligned} \hat{A}_y^{1,s}(\xi, \eta, z_0 + \zeta) &= (1/2)\xi\eta C_{1,s}^{[1]}(z_0 + \zeta) + \dots \\ &= (1/2)\xi\eta C_{1,s}^{[1]}(z_0) + \text{terms of order 3 and higher}, \end{aligned} \quad (17.7.96)$$

$$\begin{aligned} \hat{A}_z^{1,s}(\xi, \eta, z_0 + \zeta) &= -\xi C_{1,s}^{[0]}(z_0 + \zeta) + \dots \\ &= -\xi C_{1,s}^{[0]}(z_0) - \xi\zeta C_{1,s}^{[1]}(z_0) + \text{terms of order 3 and higher}. \end{aligned} \quad (17.7.97)$$

Comparison of (7.93) and (7.94) with (7.95) through (7.97) shows that for the  $m = 1$  case the minimum vector potential constructed from field expansions about on-axis points *differs* from the Coulomb gauge vector potential constructed from on-axis field data.

What can be said about the  $m = 1$  azimuthal-free gauge vector potential? From (15.3.31) through (15.3.33), we see that  $\mathbf{A}^{1,s}$  has the expansion

$$A_x^{1,s}(\xi, \eta, z_0 + \zeta) = \xi^2 C_{1,s}^{[1]}(z_0) + \dots, \quad (17.7.98)$$

$$A_y^{1,s}(\xi, \eta, z_0 + \zeta) = \xi\eta C_{1,s}^{[1]}(z_0) + \dots, \quad (17.7.99)$$

$$\begin{aligned} A_z^{1,s}(\xi, \eta, z_0 + \zeta) &= -\xi C_{1,s}^{[0]}(z_0 + \zeta) + \dots \\ &= -\xi C_{1,s}^{[0]}(z_0) - \xi\zeta C_{1,s}^{[1]}(z_0) + \text{terms of order 3 and higher}. \end{aligned} \quad (17.7.100)$$

Comparison of (7.93) and (7.94) with (7.98) through (7.100) shows that for the  $m = 1$  case the minimum vector potential constructed from field expansions about on-axis points also *differs* the azimuthal-free gauge vector potential constructed from on-axis field data. In retrospect, this difference should be less surprising because the minimum vector potential satisfies the Coulomb gauge condition, and the azimuthal-free gauge vector potential does not.

### 17.7.6 The $m = 2$ Case

**Text to be worked on:**

As a second example of a vector potential in the azimuthal-free gauge, suppose all terms in (2.37) vanish save for the ‘pure’ quadrupole terms  $C_{2,s}^{[n]}(z)$ . Then, again using (3.28) through (3.30), we find through terms of degree four that  $\mathbf{A}^{2,s}$  has the expansion

$$A_x^{2,s} = (1/2)(x^3 - xy^2)C_{2,s}^{[1]}(z) + \dots, \quad (17.7.101)$$

$$A_y^{2,s} = -(1/2)(y^3 - yx^2)C_{2,s}^{[1]}(z) + \dots, \quad (17.7.102)$$

$$A_z^{2,s} = -(x^2 - y^2)C_{2,s}^{[0]}(z) + (1/6)(x^4 - y^4)C_{2,s}^{[2]}(z) + \dots. \quad (17.7.103)$$

Note that the results (3.34) through (3.36) agree with (1.5.59) if we make the identification  $Q/2 = C_{2,s}^{[0]}$ . However, we know that  $C_{2,s}^{[0]}(z)$  must depend on  $z$  because the on-axis gradients must vanish far outside any magnet. Therefore the functions  $C_{2,s}^{[1]}(z)$ ,  $C_{2,s}^{[2]}(z)$ , etc. must be nonzero, at least near the end and fringe-field regions of any quadrupole magnet. We conclude again that, as a consequence of Maxwell’s equations, the vector potential must contain terms beyond degree two in the variables  $x, y$ . Correspondingly, the transfer map for any real quadrupole must contain nonlinear terms.

As a second example of the use of these relations, let us compute  $\hat{\mathbf{A}}^{2,s}$  for the quadrupole case  $m = 2$ . As before, suppose all terms in (2.37) vanish save for the quadrupole terms  $C_{2,s}^{[n]}(z)$ . Then, again using (4.92) through (4.94), we find, through terms of degree four, that  $\hat{\mathbf{A}}^{2,s}$  has the expansion

$$\hat{A}_x^{2,s} = (1/6)(x^3 - 3xy^2)C_{2,s}^{[1]}(z) + \dots, \quad (17.7.104)$$

$$\hat{A}_y^{2,s} = -(1/6)(y^3 - 3x^2y)C_{2,s}^{[1]}(z) + \dots, \quad (17.7.105)$$

$$\hat{A}_z^{2,s} = -(x^2 - y^2)C_{2,s}^{[0]}(z) + (1/12)(x^4 - y^4)C_{2,s}^{[2]}(z) + \dots. \quad (17.7.106)$$

This expansion should be compared with the azimuthal-free gauge expansion given by (3.34) through (3.36). Direct calculation again verifies that (4.1) and (4.4) are satisfied by  $\hat{\mathbf{A}}^{2,s}$  through the order of the terms that have been retained in the expansion

### 17.7.7 The $m = 3$ Case

#### Exercises

**17.7.1.** Verify the relations (7.32) through (7.39).

**17.7.2.** Evidently the second-order portion of  $\psi^e(\mathbf{r}^d; x_0, z_0)$  as given in (7.61) is composed of the monomials  $\xi\eta$  and  $\eta\zeta$ . Show that these are the only monomials allowed at this order based on symmetry considerations. Verify that each monomial is an harmonic polynomial. Indeed, making the usual correspondence between  $\xi, \eta, \zeta$  and  $x, y, z$  show, following the harmonic polynomial labeling scheme (U.2.9), that there are the relations

$$\xi\eta = [1/(4i)][\sqrt{32\pi/15}][H_2^2(\mathbf{r}) - H_2^{-2}(\mathbf{r})], \quad (17.7.107)$$

$$\eta\zeta = [-1/(2i)][\sqrt{8\pi/15}][H_2^1(\mathbf{r}) + H_2^{-1}(\mathbf{r})]. \quad (17.7.108)$$

Would these relations have been simpler had the polar axis, used to set up spherical polar coordinates, been taken to be the  $y$  axis instead of the  $z$  axis?

### 17.7.8 More Text

To proceed further it is useful to introduce some notation. Let  $z^{\text{en}}$  denote the  $z$  value where a transition is to be made from a region where the magnetic field is *taken* to vanish to the beginning of the leading fringe-field region. That is, the charged particle in question *enters* the leading fringe-field region when  $z = z^{\text{en}}$ . We will also use the notation  $z^{\text{ben}}$  and  $z^{\text{aen}}$  to denote  $z$  values just *before* and just *after* entry. Similarly, let  $z^{\text{ex}}$  denote the  $z$  value where a transition is to be made from the end of a trailing fringe-field region to a region where the magnetic field is again taken to vanish. That is, the charged particle in question *exits* the trailing fringe-field region when  $z = z^{\text{ex}}$ .

#### Entering a Leading Fringe-Field Region

Suppose we begin with a consideration of the transition between a field-free region and a leading-fringe field region. Let  $K^{\text{ben}}$  be the Hamiltonian before entry into the fringe-field region, and let  $K^{\text{aen}}$  be the Hamiltonian after entry into the fringe-field region. Then, since the magnetic field and its associated vector potential are assumed to vanish before entry, we have the relation

$$K^{\text{ben}} = -[(p_t^{\text{canben}})^2/c^2 - m^2c^2 - (p_x^{\text{canben}})^2 - (p_y^{\text{canben}})^2]^{1/2}. \quad (17.7.109)$$

And, since the magnetic field (and therefore also the vector potential) is nonzero after entry, we have the relation

$$K^{\text{aen}} = -[(p_t^{\text{canaen}})^2/c^2 - m^2c^2 - (p_x^{\text{canaen}} - qA_x)^2 - (p_y^{\text{canean}} - qA_y)^2]^{1/2} - qA_z. \quad (17.7.110)$$

Here we have added the suffixes *ben* and *aen* to the phase-space coordinates to denote their values before and after entry. Our task is to relate these phase-space coordinates.

As a first step, we naturally require that the coordinates be continuous at  $z^{\text{en}}$ ,

$$x^{\text{ben}} = x^{\text{aen}}, \quad (17.7.111)$$

$$y^{\text{ben}} = y^{\text{aen}}, \quad (17.7.112)$$

$$t^{\text{ben}} = t^{\text{aen}}, \quad (17.7.113)$$

when  $z = z^{\text{en}}$ . The next step is specify what is to be done with the momenta.

One possibility is to require that the slopes/“velocities”  $dx/dz$ ,  $dy/dz$ , and  $dt/dz$  be continuous at  $z^{\text{en}}$ . Let us work out the consequences of such a requirement. Before entry we have the result

$$\begin{aligned} dx/dz &= \partial K^{\text{ben}} / \partial p_x^{\text{canben}} = \\ &(p_x^{\text{canben}} / [(p_t^{\text{canben}})^2 / c^2 - m^2 c^2 - (p_x^{\text{canben}})^2 - (p_y^{\text{canben}})^2]^{1/2}), \end{aligned} \quad (17.7.114)$$

and after entry there is the result

$$\begin{aligned} dx/dz &= \partial K^{\text{aen}} / \partial p_x^{\text{canaen}} = \\ &(p_x^{\text{canaen}} - qA_x) / [(p_t^{\text{canaen}})^2 / c^2 - m^2 c^2 - (p_x^{\text{canaen}} - qA_x)^2 - (p_y^{\text{canaen}} - qA_y)^2]^{1/2}. \end{aligned} \quad (17.7.115)$$

An analogous result holds for  $dy/dz$ . Finally, for  $dt/dz$  there is the before entry result

$$\begin{aligned} dt/dz &= \partial K^{\text{ben}} / \partial p_t^{\text{canben}} = \\ &(-p_t^{\text{canben}} / c^2) / [(p_t^{\text{canben}})^2 / c^2 - m^2 c^2 - (p_x^{\text{canben}})^2 - (p_y^{\text{canben}})^2]^{1/2}, \end{aligned} \quad (17.7.116)$$

and the after entry result

$$\begin{aligned} dt/dz &= \partial K^{\text{aen}} / \partial p_t^{\text{canaen}} = \\ &(-p_t^{\text{canaen}} / c^2) / [(p_t^{\text{canaen}})^2 / c^2 - m^2 c^2 - (p_x^{\text{canaen}} - qA_x)^2 - (p_y^{\text{canaen}} - qA_y)^2]^{1/2}. \end{aligned} \quad (17.7.117)$$

Now equate the far right sides of (7.16) and (7.17), the far right sides of there  $dy/dz$  counterparts, and the far right sides of (7.18) and (7.19). So doing yields the transition matching relations

$$p_x^{\text{canben}} = p_x^{\text{canaen}} - qA_x, \quad (17.7.118)$$

$$p_y^{\text{canben}} = p_y^{\text{canaen}} - qA_y, \quad (17.7.119)$$

$$p_t^{\text{canben}} = p_t^{\text{canaen}}. \quad (17.7.120)$$

In view of (7.3) and (7.5) these relations can also be written in the form

$$p_x^{\text{mechben}} = p_x^{\text{mechaen}}, \quad (17.7.121)$$

$$p_y^{\text{mechben}} = p_y^{\text{mechaen}}, \quad (17.7.122)$$

$$p_z^{\text{mechben}} = p_z^{\text{mechaen}}, \quad (17.7.123)$$

The relation (7.22) is satisfactory because magnetic forces do not change the energy. Recall (1.6.17). However, we also desire that the phase-space transformation given by (7.12) through (7.14) and (7.20) through (7.22) be symplectic. Calculation shows that it is not. Compute the Poisson bracket of the right sides of (7.20) and (7.21) to find the result

$$\begin{aligned} [p_x^{\text{canaen}} - qA_x, p_y^{\text{canaen}} - qA_y] &= [p_x^{\text{canaen}}, -qA_y] + [-qA_x, p_y^{\text{canaen}}] \\ &= q\{\partial A_y/\partial x^{\text{aen}} - \partial A_x/\partial y^{\text{aen}}\} = qB_z. \end{aligned} \quad (17.7.124)$$

While hopefully small, generally  $B_z(x, y, z^{\text{en}})$  differs from zero at the beginning of the leading fringe-field region. On the other hand, the Poisson bracket of the left sides of (7.20) and (7.21) must vanish since  $p_x^{\text{canben}}$  and  $p_y^{\text{canben}}$  are supposed to be canonical momenta. Therefore the the phase-space transformation given by (7.12) through (7.14) and (7.20) through (7.22) is generally not symplectic.

We expect that neglect of the magnetic field in the region  $z < z^{\text{en}}$  will lead to some error in trajectories. However, we do not want this error to violate the symplectic condition. The simplest way to maintain the symplectic condition is to retain the relations (7.12) through (7.14) and replace the relations (7.20) through (7.22) by the relations

$$p_x^{\text{canben}} = p_x^{\text{canaen}}, \quad (17.7.125)$$

$$p_y^{\text{canben}} = p_y^{\text{canaen}}, \quad (17.7.126)$$

$$p_t^{\text{canben}} = p_t^{\text{canaen}}. \quad (17.7.127)$$

In this case the transition matching relations (7.12) through (7.14) and (7.20) through (7.22) amount to the identity map  $\mathcal{I}$ , and the symplectic condition is trivially satisfied. Now, however, the error in trajectories manifests itself in that the slopes/“velocities”  $dx/dz$ ,  $dy/dz$ , and  $dt/dz$  may be expected to be discontinuous at at  $z^{\text{en}}$ . Inspection of (7.16) and (7.17), their  $dy/dz$  counterparts, and (7.18) and (7.19) shows that, in lowest approximation, these discontinuities are proportional to  $A_x(x, y, z^{\text{en}})$  and  $A_y(x, y, z^{\text{en}})$ . Indeed, again in view of (7.3) and (7.5), the transition relations (7.26) through (7.28) can be written in the form

$$\Delta \mathbf{p}^{\text{mech}} = \mathbf{p}^{\text{mechaen}} - \mathbf{p}^{\text{mechben}} = q\mathbf{A}(x, y, z^{\text{en}}). \quad (17.7.128)$$

It is therefore desirable, where feasible, to work in a gauge where  $\mathbf{A}(x, y, z^{\text{en}})$  is as small as possible.

One way to view the symplectic matching relations (7.12) and (7.14) and (7.24) through (7.26) is to replace the Hamiltonian (7.1) by a modified Hamiltonian  $K^{\text{mod}}$  given by

$$K^{\text{mod}} = -[(p_t^{\text{can}})^2/c^2 - m^2 c^2 - (p_x^{\text{can}} - qA_x^{\text{mod}})^2 - (p_y^{\text{can}} - qA_y^{\text{mod}})^2]^{1/2} - qA_z^{\text{mod}} \quad (17.7.129)$$

where

$$\mathbf{A}^{\text{mod}}(x, y, z) = \theta(z - z^{\text{en}})\mathbf{A}(x, y, z). \quad (17.7.130)$$

That is, the vector potential is taken to vanish for  $z < z^{\text{en}}$  and turns on at  $z = z^{\text{en}}$ . A little thought shows that integrating the equations of motion associated with this modified Hamiltonian automatically produces the matching relations (7.12) and (7.14) and (7.24) through (7.26).

What is the modified magnetic field  $\mathbf{B}^{\text{mod}}$  associated with this modified vector potential? Evaluation of  $\nabla \times \mathbf{A}^{\text{mod}}$  gives the relations

$$B_x^{\text{mod}}(x, y, z) = \theta(z - z^{\text{en}})B_x(x, y, z) - \delta(z - z^{\text{en}})A_y(x, y, z), \quad (17.7.131)$$

$$B_y^{\text{mod}}(x, y, z) = \theta(z - z^{\text{en}})B_y(x, y, z) + \delta(z - z^{\text{en}})A_x(x, y, z), \quad (17.7.132)$$

$$B_z^{\text{mod}}(x, y, z) = \theta(z - z^{\text{en}})B_z(x, y, z). \quad (17.7.133)$$

Calculation shows that  $\mathbf{B}^{\text{mod}}$  has divergence

$$\nabla \cdot \mathbf{B}^{\text{mod}} = 0, \quad (17.7.134)$$

as desired. What current produces this modified magnetic field? The modified magnetic field satisfies the curl relation

$$\nabla \times \mathbf{B}^{\text{mod}} = \mathbf{j}^{\text{mod}} \quad (17.7.135)$$

where

$$\begin{aligned} j_x^{\text{mod}} &= (\partial/\partial y)B_z^{\text{mod}} - (\partial/\partial z)B_y^{\text{mod}} \\ &= -\delta(z - z^{\text{en}})[B_y(x, y, z) + (\partial/\partial z)A_x(x, y, z)] - \delta'(z - z^{\text{en}})A_x(x, y, z) \\ &= -\delta(z - z^{\text{en}})[2(\partial/\partial z)A_x(x, y, z) - (\partial/\partial x)A_z(x, y, z)] - \delta'(z - z^{\text{en}})A_x(x, y, z), \end{aligned} \quad (17.7.136)$$

$$\begin{aligned} j_y^{\text{mod}} &= (\partial/\partial z)B_x^{\text{mod}} - (\partial/\partial x)B_z^{\text{mod}} \\ &= -\delta(z - z^{\text{en}})[-B_x(x, y, z) + (\partial/\partial z)A_y(x, y, z)] - \delta'(z - z^{\text{en}})A_y(x, y, z) \\ &= -\delta(z - z^{\text{en}})[2(\partial/\partial z)A_y(x, y, z) - (\partial/\partial y)A_z(x, y, z)] - \delta'(z - z^{\text{en}})A_y(x, y, z), \end{aligned} \quad (17.7.137)$$

$$\begin{aligned} j_z^{\text{mod}} &= (\partial/\partial x)B_y^{\text{mod}} - (\partial/\partial y)B_x^{\text{mod}} \\ &= \delta(z - z^{\text{en}})[(\partial/\partial x)A_x(x, y, z) + (\partial/\partial y)A_y(x, y, z)]. \end{aligned} \quad (17.7.138)$$

Evidently terminating the vector potential at  $z = z^{\text{en}}$  is equivalent to introducing sheet (corresponding to the  $\delta$  function) and double-sheet (corresponding to the  $\delta'$  function) currents at  $z = z^{\text{en}}$ . And the strengths of these currents are proportional to the values of  $\mathbf{A}$  and its first derivatives at  $z = z^{\text{en}}$ .

### Exiting a Trailing Fringe-Field Region



# Bibliography

## Fitting Based on Use of Surface Data

- [1] M. Venturini and A. Dragt, “Accurate Computation of Transfer Maps from Magnetic Field Data”, *Nuclear Instruments and Methods* **A427**, p. 387 (1991).
- [2] M. Venturini, “Lie Methods, Exact Map Computation, and the Problem of Dispersion in Space Charge Dominated Beams”, University of Maryland Physics Department Ph.D. Thesis (1998).
- [3] C. Mitchell, “Calculation of Realistic Charged-Particle Transfer Maps”, University of Maryland Physics Department Ph.D. Thesis (2007).
- [4] C.E. Mitchell and A.J. Dragt, “Accurate Transfer Maps for Realistic Beamline Elements: Part I, Straight Elements”, 19 pages, *Phys. Rev. ST Accel. Beams* **13**, 064001 (2010).

## Mathieu Functions

See also the references on Krein-Moser Theory and Periodic Linear Systems at the end of Chapter 3.

- [5] N.W. McLachlan, *Theory and Application of Mathieu Functions*, Dover (1964).
- [6] M. Strutt, *Lamésche-Mathieusche-und Verwandte Functionen in Physik und Technik*, Chelsea (1967).
- [7] A. Erdélyi, Edit., *Higher Transcendental Functions*, Volume III, Chapter XVI, McGraw-Hill (1955).
- [8] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, Chapter 20, Dover (1972). Also available on the Web by Googling “abramowitz and stegun 1972”.
- [9] 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/>.
- [10] W. Magnus and S. Winkler, *Hill’s Equation*, Dover (1979).

## Other Fitting Methods

- [11] L. Teng, “Expanded Form of Magnetic Field with Median Plane”, Argonne National Laboratory Report ANL-LCT-28 (15 December 1962). <https://www.osti.gov/scitech/servlets/purl/7364341>.
- [12] E. Akeley, “The Vector Potential of the Magnetic Field in the Mark V Accelerator”, Midwestern Universities Research Association (MURA) Report MURA-70 ESA(MURA-3) (28 May 1955). <http://lss.fnal.gov/archive/other/mura/MURA-070.pdf>.

# Chapter 18

## Tools for Numerical Implementation

This chapter develops the tools that are necessary for the numerical implementation of the methods of Chapter 14. These tools include splines, bicubic interpolation, spline-based Fourier transforms, and routines for the calculation of Bessel and Mathieu functions.

### 18.1 Third-Order Splines

For our purposes splines are piecewise polynomial fits where various continuity conditions are imposed at the points the pieces join. There are two common possibilities: Either a fit is desired over some interval that may be viewed a portion of the real line; or a fit is desired over a full angular interval, in which case periodicity is to be imposed.

#### 18.1.1 Fitting Over an Interval

Let  $y = f(x)$  be a function of a single variable. Suppose its values  $y_j$  are specified at  $N + 1$  equally spaced points  $x_0, x_1, \dots, x_N$  over the interval  $[x_0, x_N]$ . (See Figure 2.1.1 for a similar setup employing the variable  $t$ .) Also suppose that on each subinterval  $[x_j, x_{j+1}]$  we want to approximate  $f$  by a cubic polynomial with cubic polynomials on adjacent subintervals matched in such a way that  $f$  has continuous first and second derivatives at each *interior* point  $x_j$ . Such an approximation will be called a *cubic* or *third-order spline*. We will use these splines both for interpolation and for the calculation of direct and inverse Fourier transforms. See Sections 15.2 and 15.3.5.

Let us see what information is required to construct such a sequence of third-order polynomials (one for each interval). On the first subinterval,  $[x_0, x_1]$ , write

$$y = f_0(x) = a_0 + b_0(x - x_0) + c_0(x - x_0)^2 + d_0(x - x_0)^3 \quad (18.1.1)$$

where the coefficients  $a_0$  through  $d_0$  are to be determined. Then the condition

$$f_0(x_0) = y_0 \quad (18.1.2)$$

yields the relation

$$a_0 = y_0. \quad (18.1.3)$$

Next, for the moment, suppose we further require that

$$f'_0(x_0) = \beta_0 \quad (18.1.4)$$

and

$$f''_0(x_0) = \gamma_0. \quad (18.1.5)$$

These requirements yield the further relations

$$b_0 = \beta_0, \quad (18.1.6)$$

$$c_0 = \gamma_0/2. \quad (18.1.7)$$

Finally, the condition

$$f_0(x_1) = y_1 \quad (18.1.8)$$

yields the relation

$$y_1 = y_0 + \beta_0(x_1 - x_0) + (\gamma_0/2)(x_1 - x_0)^2 + d_0(x_1 - x_0)^3, \quad (18.1.9)$$

which can be solved to yield the value of  $d_0$ . The conditions (1.2) and (1.8), plus the requirements (1.4) and (1.5), have completely specified the first cubic polynomial (1.1).

Let us now move on to the second subinterval  $[x_1, x_2]$ . On this subinterval we assume that there is the cubic polynomial representation

$$y = f_1(x) = a_1 + b_1(x - x_1) + c_1(x - x_1)^2 + d_1(x - x_1)^3, \quad (18.1.10)$$

and we find from the condition

$$f_1(x_1) = y_1 \quad (18.1.11)$$

the relation

$$a_1 = y_1. \quad (18.1.12)$$

Also, since  $f_0(x)$  has already been determined, the values  $f'_0(x_1)$  and  $f''_0(x_1)$  are already known. The relations (1.8) and (1.11) already guarantee continuity of  $f_0$  and  $f_1$  across the join at  $x_1$ . Next, as set forth in our initial statement of intent, let us require that

$$f'_1(x_1) = \beta_1 = f'_0(x_1) \quad (18.1.13)$$

and

$$f''_1(x_1) = \gamma_1 = f''_0(x_1). \quad (18.1.14)$$

That is, we also require continuity in the first and second derivatives. From (1.13) and (1.14) we conclude that

$$b_1 = \beta_1 = f'_0(x_1) \quad (18.1.15)$$

and

$$c_1 = \gamma_1/2 = f''_0(x_1)/2. \quad (18.1.16)$$

Finally, the condition

$$f_1(x_2) = y_2 \quad (18.1.17)$$

yields the relation

$$y_2 = y_1 + \beta_1(x_2 - x_1) + (\gamma_1/2)(x_2 - x_1)^2 + d_1(x_2 - x_1)^3, \quad (18.1.18)$$

which can be solved to yield the value of  $d_1$ . We see that the condition (1.17) plus the continuity requirements have completely specified the second cubic polynomial (1.10).

It is clear that this process can be continued for the subsequent subintervals  $[x_2, x_3] \dots [x_{N-1}, x_N]$  so that all the cubic polynomials are completely specified in terms of the  $y_0 \dots y_N$  and the two numbers  $\beta_0$  and  $\gamma_0$ . Now we come to a subtle point. Since the cubic polynomials are all completely specified, the value  $f'_{N-1}(x_N)$  is also specified in terms of the  $y_1 \dots y_N$  and the two numbers  $\beta_0$  and  $\gamma_0$ . In fact, there will be a relation of the form

$$f'_{N-1}(x_N) = \delta + \epsilon\gamma_0 \quad (18.1.19)$$

where  $\delta(\beta_0, y_0, \dots, y_N)$  is some (linear) function of  $\beta_0, y_0, \dots, y_N$ , and  $\epsilon$  is some *nonzero* coefficient. Therefore, we may adjust  $\gamma_0$  in such a way as to give  $f'_{N-1}(x_N)$  any desired value. Put another way, we may replace a knowledge of  $\gamma_0$  with a specification of  $f'_{N-1}(x_N)$ . Let us write

$$f'(x_0) = f'_0(x_0) = \beta_0 \quad (18.1.20)$$

and

$$f'(x_N) = f'_{N-1}(x_N). \quad (18.1.21)$$

With this notation in mind, we may view a cubic spline as being completely specified by the values  $y_0 \dots y_N$  and the two end-point derivatives  $f'(x_0)$  and  $f'(x_N)$ .<sup>1</sup>

Of course, in general the end-point derivatives are unknown. Many users of cubic splines simply set end-point derivatives (either first or second) to zero on the grounds of convenience and the fact (to be demonstrated shortly) that their values actually have little effect on the spline approximation once one is a few grid points away from the ends.<sup>2</sup> For our purposes, we prefer to use the first few data points near the end points to estimate the end-point first derivatives. For example, upon deciding to use the first three points to estimate  $f'(x_0)$  and the last three to estimate  $f'(x_N)$ , we use the approximations

$$\begin{aligned} f'(x_0) &= (1/h)[-(3/2)y_0 + 2y_1 - (1/2)y_2] + O(h^2), \\ f'(x_N) &= (1/h)[(3/2)y_N - 2y_{N-1} + (1/2)y_{N-2}] + O(h^2) \end{aligned} \quad (18.1.22)$$

where  $h$  is the spacing between successive grid points,

$$h = x_1 - x_0. \quad (18.1.23)$$

If we choose to employ the first and last four points, we use the approximations

$$\begin{aligned} f'(x_0) &= (1/h)[-(11/6)y_0 + 3y_1 - (3/2)y_2 + (1/3)y_3] + O(h^3), \\ f'(x_N) &= (1/h)[(11/6)y_N - 3y_{N-1} + (3/2)y_{N-2} - (1/3)y_{N-3}] + O(h^3). \end{aligned} \quad (18.1.24)$$

---

<sup>1</sup>Evidently, an alternate procedure is to specify the values  $y_0 \dots y_N$  and the two second-order end-point derivatives  $f''(x_0)$  and  $f''(x_N)$ .

<sup>2</sup>If the second derivatives at each end point are set to zero, such a cubic spline is said to be *natural*.

See Exercise 1.1.

At this juncture we must remark that the algorithm we have been describing for computing cubic splines, while pedagogically instructive, is not numerically stable against roundoff errors. A stable spline routine is given in Appendix L.

As already alluded to, one of the advantages of spline fits is *localization* in that the fits over any subinterval in  $x$  depend primarily on the  $y_j$  values whose corresponding  $x_j$  lie within that subinterval. For example, consider the function  $y = f(x)$  defined on the interval  $x \in [0, 3]$  such that  $f(1.5) = 1$  and  $f = 0$  elsewhere. Figure 1.1 shows the function that is produced by a cubic spline fit when  $h = .1$ . In this case 31 points are used to make the fit with  $x_0 = 0$ ,  $x_{30} = 3$ , and all  $y_j$  set to zero save for setting  $y_{15}=1$ . Also,  $f'(0)$  and  $f'(3)$  are set to zero. Evidently the spline fit falls rapidly to zero on either side of  $x = 1.5$ . In fact, it can be shown that the successive maxima *decay exponentially* as

$$y(x) \sim \exp[-\alpha(1/h)|x - 1.5|] \quad (18.1.25)$$

where

$$\alpha = \log(2 + \sqrt{3}) \simeq 1.317. \quad (18.1.26)$$

Similarly, Figure 1.2 shows the fit that is produced for the same setup when all  $y_j$  are set to zero,  $f'(0)$  is set to 1, and  $f'(3)$  is set to zero. Again the fit decays to zero exponentially with exponent  $-\alpha$ .

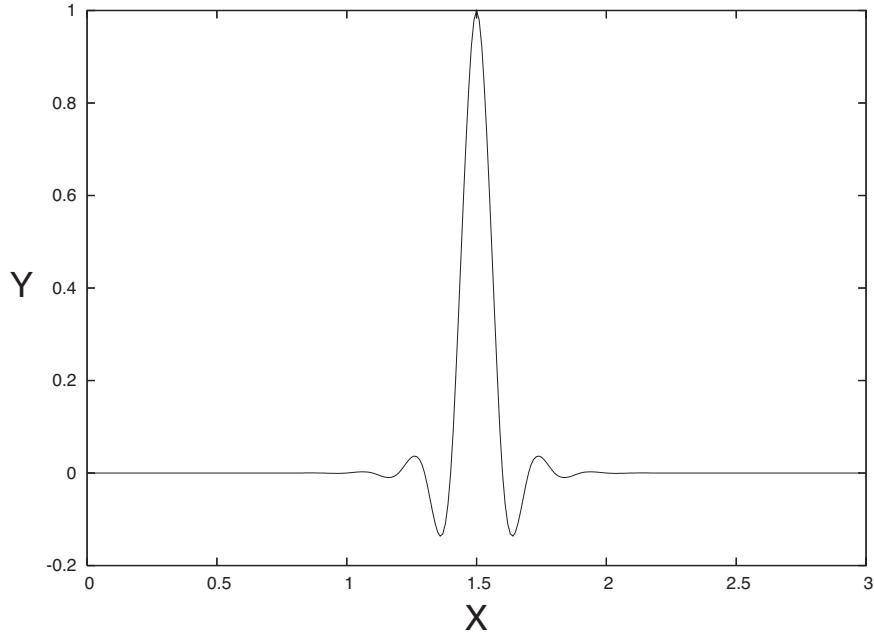


Figure 18.1.1: The 31-point spline fit associated with  $y_{15} = 1$  and all other  $y_j = 0$ . Also,  $f'(0)$  and  $f'(3)$  are set to zero. Note that the fit falls rapidly to zero on either side of  $x = 1.5$ .

### 18.1.2 Periodic Splines

The splines defined so far are useful for fitting a general function over an interval. Suppose we instead want to fit a function which is known to be periodic. Such functions will typically

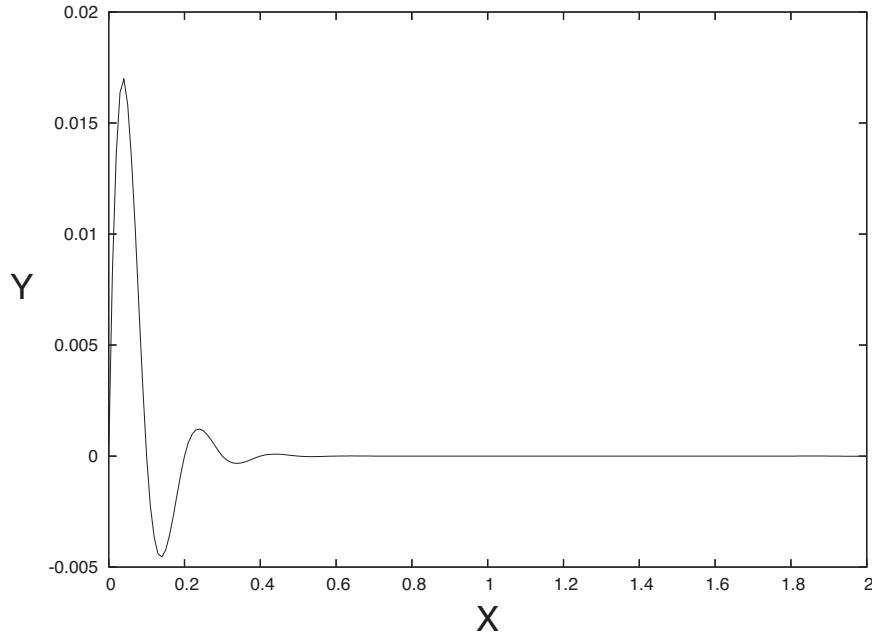


Figure 18.1.2: The spline fit associated with  $f'(x_0) = 1$ , all  $y_j = 0$ , and  $f'(x_{30}) = 0$ . Note that the fit falls rapidly to zero for  $x$  beyond  $x_0 = 0$ . Only the results over the interval  $x \in [0, 2]$  are displayed.

depend on angular variables.

As before, we imagine that there are known function values  $y_j$  at the points  $x_0$  to  $x_N$ , but that now  $y_0 = y_N$ . Begin the construction as before to find a unique set of cubic polynomials in terms of the  $y_j$  and  $\beta_0$  and  $\gamma_0$ . With this construction, the quantities  $f'_{N-1}(x_N)$  and  $f''_{N-1}(x_N)$  are specified. Indeed, there will be relations of the form

$$f'_{N-1}(x_N) = r(y_0, \dots, y_N) + s\beta_0 + t\gamma_0, \quad (18.1.27)$$

$$f''_{N-1}(x_N) = \rho(y_0, \dots, y_N) + \sigma\beta_0 + \tau\gamma_0 \quad (18.1.28)$$

where  $r$  and  $\rho$  are linear functions of the  $y_j$ ; and  $s, t, \sigma$ , and  $\tau$  are proportionality constants. Now adjust both  $\beta_0$  and  $\gamma_0$  such that there are the relations

$$f'_{N-1}(x_N) = f'_0(x_0) \quad (18.1.29)$$

and

$$f''_{N-1}(x_N) = f''_0(x_0). \quad (18.1.30)$$

So doing will make the spline fit periodic in that not only will  $y_0 = y_N$ , also the first and second derivatives will match at the endpoints. In view of (1.4), (1.5), (1.27), and (1.28), these matching relations are equivalent to the conditions

$$\beta_0 = r(y_0, \dots, y_N) + s\beta_0 + t\gamma_0, \quad (18.1.31)$$

$$\gamma_0 = \rho(y_0, \dots, y_N) + \sigma\beta_0 + \tau\gamma_0. \quad (18.1.32)$$

These equations have a (unique) solution provided the matrix  $M$  defined by

$$M = \begin{pmatrix} s-1 & t \\ \sigma & \tau-1 \end{pmatrix}. \quad (18.1.33)$$

has a nonzero determinant, which can be shown to be always the case. We conclude that a periodic cubic spline is uniquely specified by the values  $y_0, \dots, y_N$  with  $y_0 = y_N$ .

Here again we must remark that the pedagogically instructive algorithm we have been describing for computing periodic cubic splines is not numerically stable against roundoff errors. A stable periodic spline routine is also given in Appendix L.

### 18.1.3 Error Estimate for Spline Approximation

There remains the question of accuracy for a cubic spline approximation. Suppose the function  $f$  that is being approximated is known to have a continuous fourth-order derivative. Then it can be shown that the error involved in using its spline approximation  $f_{\text{sa}}$  has the estimate

$$\text{error}(x) = f(x) - f_{\text{sa}}(x) = (h^4/24)\theta^2(1-\theta)^2f^{\text{iv}}(x) + O(h^5) \quad (18.1.34)$$

for  $x$  in the subinterval  $[x_j, x_{j+1}]$  and

$$\theta = (x - x_j)/h. \quad (18.1.35)$$

Note that, according to (1.35),  $\theta$  lies in the interval  $\theta \in [0, 1]$ . It is easy to check that in this interval the quantity  $\theta^2(1-\theta)^2$  does not exceed  $1/16$ .

As an example, suppose

$$f(x) = 1 - x^4 \quad (18.1.36)$$

and we wish to approximate  $f$  over the interval  $[-1, 1]$ . That is, we set  $x_0 = -1$  and  $x_N = 1$ . Figure 1.3 shows  $f$  and its spline fit  $f_{\text{sa}}$  for  $h = .1$  (which corresponds to  $N = 20$ ). They are indistinguishable on the scale shown. Figure 1.4 shows the error that occurs when  $h = .1$ . By construction the error vanishes at the grid points  $x_j$ , and the global error is consistent with the estimate (1.34).

In making the spline fits for Figures 1.3 and 1.4, we have used as input for the end-point derivatives the exact results  $f'(-1) = 4$  and  $f'(1) = -4$  based on (1.36). Suppose we instead use (1.24) to estimate the end-point derivatives. Figure 1.5 shows the error that then occurs. We see that there is some error at the endpoints and that, as expected from localization, this error soon damps away so that only the error already seen in Figure 1.4 remains. We remark that the use of (1.24) gives the results  $f'(-1) = 3.994$  and  $f'(1) = -3.994$ . It is pleasantly surprising that, at its worst, the error in the spline fit is considerably less than might have naively been expected based on the error in the estimated end-point derivatives. Finally we remark that, for our applications, the end points occur in fringe-field regions where both the function being approximated and its derivatives are very small. Thus, we expect that the error made in using (1.22) or (1.24) in this case will be negligible.

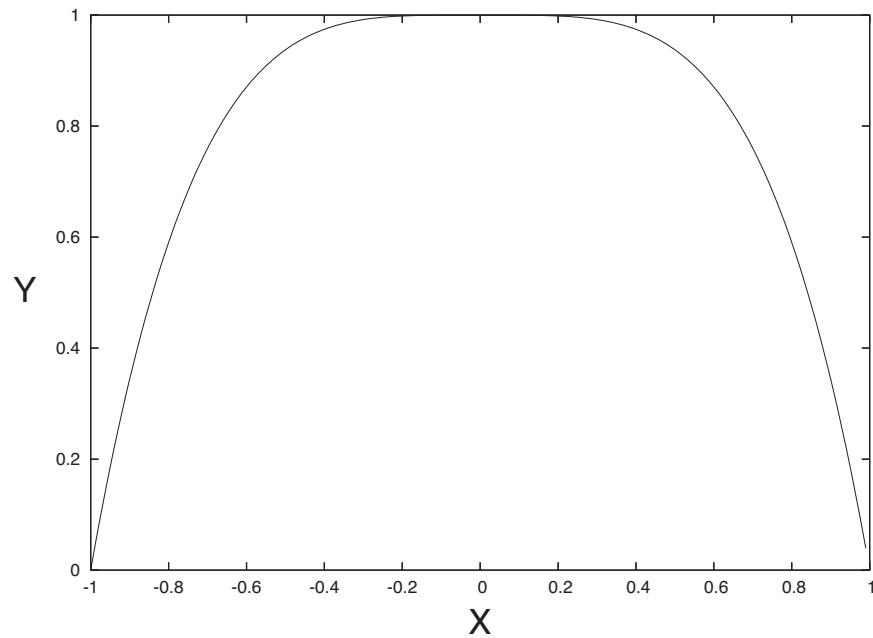


Figure 18.1.3: The function  $f$  and its spline fit  $f_{sa}$  for  $h = .1$ . They appear identical.

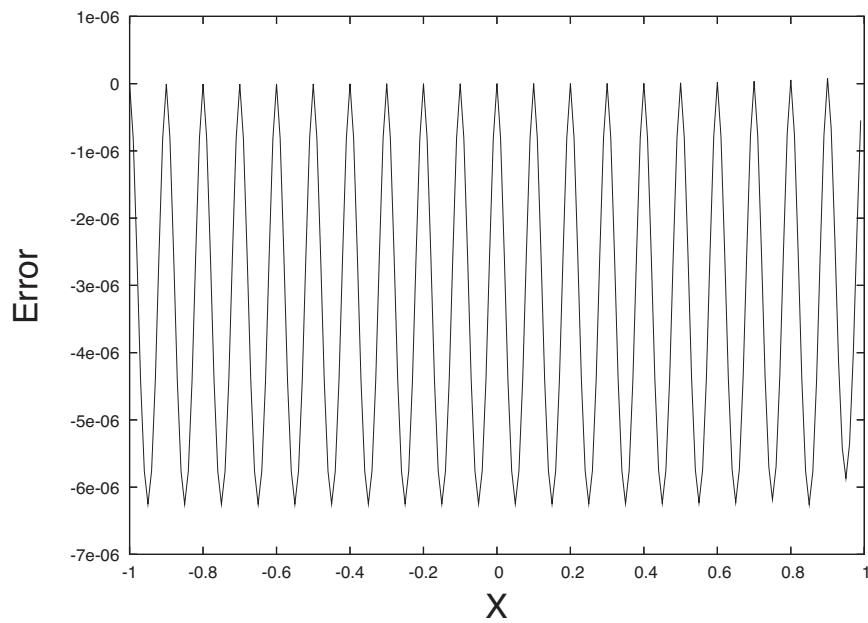


Figure 18.1.4: The difference between the function  $f$  and its spline fit  $f_{sa}$  for  $h = .1$ . Here  $\text{error} = f(x) - f_{sa}(x)$ . The spline  $f_{sa}$  is constructed using the exact values for the end-point derivatives.

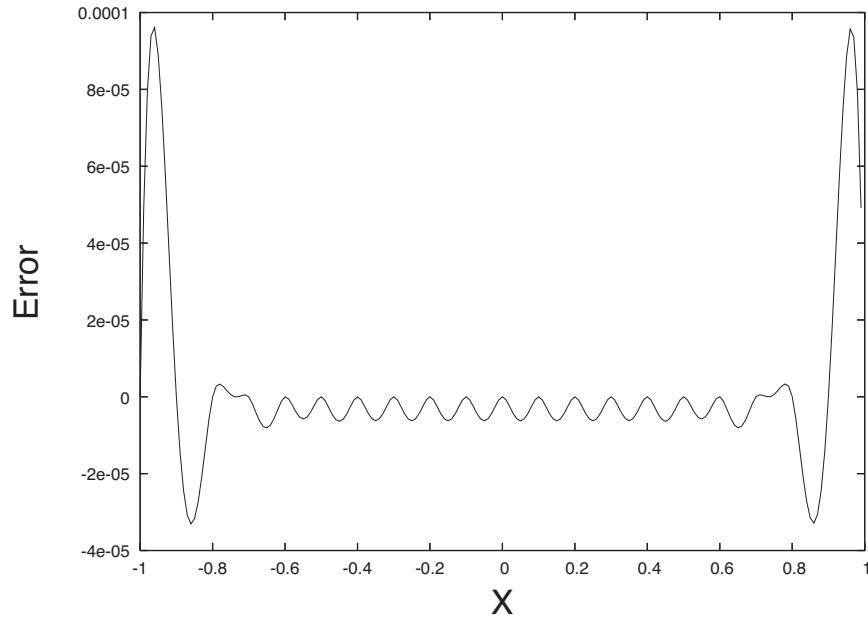


Figure 18.1.5: The difference between the function  $f$  and its spline fit  $f_{sa}$  for  $h = .1$ . Here  $\text{error} = f(x) - f_{sa}(x)$ . The spline  $f_{sa}$  is constructed using (1.24) to estimate the end-point derivatives.

## Exercises

**18.1.1.** Verify (1.22) through (1.24) using the finite difference calculus of Section 2.4.

## 18.2 Interpolation

The calculations of Chapter 14 begin with data provided on some regular Cartesian grid. With regard to locations in the  $z$  coordinate, we will use those provided, and call them  $Z_L$ . However, for the coordinates  $x$  and  $y$ , interpolation may be required.

In the case where a circular cylinder is employed, we need to interpolate to equi-angular locations given by the relations

$$\bar{x}_i = R \cos(\phi_i), \quad (18.2.1)$$

$$\bar{y}_i = R \sin(\phi_i) \quad (18.2.2)$$

where

$$\phi_0 = 0 \quad (18.2.3)$$

and

$$\phi_N = 2\pi. \quad (18.2.4)$$

See the second frame of Figure 14.1.1. Typically, in the circular cylinder case we take  $N$  to have the value  $N \approx 50$ .

In the case of an elliptical cylinder we write

$$\bar{x}_i = f \cosh(U) \cos(v_i), \quad (18.2.5)$$

$$\bar{y}_i = f \sinh(U) \sin(v_i) \quad (18.2.6)$$

where the  $v_i$  are equally spaced with

$$v_0 = 0 \quad (18.2.7)$$

and

$$v_N = 2\pi. \quad (18.2.8)$$

See Figure 14.4.3. Typically, in the elliptic cylinder case, we take  $N$  to have the value  $N \approx 120$ .

### 18.2.1 Bicubic Interpolation

For each  $\bar{x}_i, \bar{y}_i$  pair, find the closest 16 points in the regular grid in the  $x, y$  plane. See Figure 2.1. Note that the regular grid may be rectangular rather than square. Let  $X_J$  and  $Y_K$  be the coordinates of the grid point in the lower left corner. That is, we have the following inequalities:

$$X_{J+1} \leq \bar{x}_i \leq X_{J+2}, \quad (18.2.9)$$

$$Y_{K+1} \leq \bar{y}_i \leq Y_{K+2}. \quad (18.2.10)$$

Introduce local expansion variables  $\xi$  and  $\eta$  by the relations

$$x = X_J + \xi, \quad (18.2.11)$$

$$y = Y_K + \eta; \quad (18.2.12)$$

and also write

$$\bar{x}_i = X_J + \bar{\xi}_i, \quad (18.2.13)$$

$$\bar{y}_i = Y_K + \bar{\eta}_i. \quad (18.2.14)$$

We then interpolate the quantity of interest, be it a potential value or some transverse field component, from the regular grid to the point  $\bar{\xi}_i, \bar{\eta}_i$  with the aid of a *bicubic* polynomial  $P$  in the variables  $\xi$  and  $\eta$ .<sup>3</sup> This is a polynomial of the form

$$P(\xi, \eta) = \sum_{m,n=1}^4 c_{mn} \xi^{m-1} \eta^{n-1} = \sum_{m=1}^4 \left( \sum_{n=1}^4 c_{mn} \eta^{n-1} \right) \xi^{m-1} = \sum_{n=1}^4 \left( \sum_{m=1}^4 c_{mn} \xi^{m-1} \right) \eta^{n-1} \quad (18.2.15)$$

where the coefficients  $c_{mn}$  are to be determined. Note that  $P$  is cubic in each of the variables  $\xi$  and  $\eta$  separately. Hence the name bicubic. Also, note that  $P$  is *not* a homogeneous polynomial. For example, it contains the term  $\xi^3 \eta^3$ , but it does not contain the terms  $\xi^6$  or  $\eta^6$ . Finally note that, because  $c$  is  $4 \times 4$ , it requires 16 numbers to specify the  $c_{mn}$ . This is encouraging, because we have assumed that we have data on the 16 nearest-neighbor grid points.<sup>4</sup>

---

<sup>3</sup>If the point  $\bar{\xi}_i, \bar{\eta}_i$  happens to fall on a grid line, then only one-dimensional cubic interpolation is required. If it falls on a grid point, no interpolation is required at all.

<sup>4</sup>This count would not work out so neatly had we attempted what might appear to be more desirable, namely an expansion in homogenous polynomials.

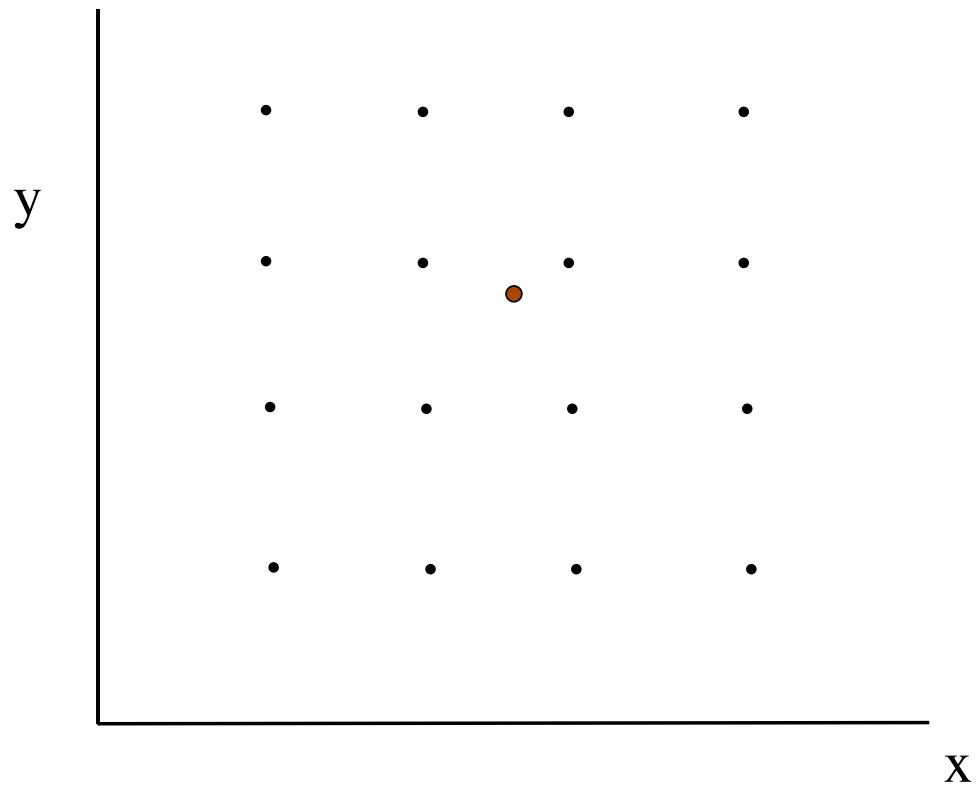


Figure 18.2.1: The point  $\bar{x}_i, \bar{y}_i$  and its 16 nearest-neighbor grid points. The coordinates of the grid point at the lower left corner are  $X_J$  and  $Y_K$ .

Let us first verify that  $P$  is uniquely defined in terms of values at the 16 nearest-neighbor grid points. For example, suppose we wish to interpolate a potential function  $\psi$ . Let  $h_x$  and  $h_y$  be the grid spacings in the  $x$  and  $y$  directions, respectively. Then we know the values  $\Psi_{jk}$  given by the relations

$$\Psi_{jk} = \psi[X_J + (j-1)h_x, Y_K + (k-1)h_y, Z_L] \text{ for } j, k \in [1, 4]. \quad (18.2.16)$$

In the spirit of (2.11) and (2.12), write

$$\xi_j = (j-1)h_x \quad (18.2.17)$$

and

$$\eta_k = (k-1)h_y. \quad (18.2.18)$$

Then we wish to have the relations

$$\Psi_{jk} = P(\xi_j, \eta_k) = \sum_{m,n=1}^4 c_{mn} \xi_j^{m-1} \eta_k^{n-1} \text{ for } j, k \in [1, 4], \quad (18.2.19)$$

and hope that these 16 desiderata will determine the 16  $c_{mn}$ .

To explore these relations, define vectors  $\hat{\xi}_j$  and  $\hat{\eta}_k$  by the rules

$$\hat{\xi}_j = (\xi_j^0, \xi_j^1, \xi_j^2, \xi_j^3)^T = (1, \xi_j, \xi_j^2, \xi_j^3)^T, \quad (18.2.20)$$

and

$$\hat{\eta}_k = (\eta_k^0, \eta_k^1, \eta_k^2, \eta_k^3)^T = (1, \eta_k, \eta_k^2, \eta_k^3)^T. \quad (18.2.21)$$

We will call  $\hat{\xi}_j$  and  $\hat{\eta}_k$  the *cubic vectors* associated with  $\xi_j$  and  $\eta_k$  since they are formed out of the cubic and lower powers of  $\xi_j$  and  $\eta_k$ , respectively. With this notation, (2.19) is equivalent to the inner product relations

$$\Psi_{jk} = (\hat{\xi}_j, c \hat{\eta}_k) \text{ for } j, k \in [1, 4], \quad (18.2.22)$$

where  $c$  is the matrix with entries  $c_{mn}$ .

Let  $\mathcal{X}$  be the matrix whose columns are the vectors  $\hat{\xi}_j$ ,

$$\mathcal{X} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ \xi_1 & \xi_2 & \xi_3 & \xi_4 \\ \xi_1^2 & \xi_2^2 & \xi_3^2 & \xi_4^2 \\ \xi_1^3 & \xi_2^3 & \xi_3^3 & \xi_4^3 \end{pmatrix}. \quad (18.2.23)$$

Then, by construction, we have the relations

$$\mathcal{X} e^j = \hat{\xi}_j \quad (18.2.24)$$

where the  $e^j$  are the standard orthonormal vectors (3.6.4). Similarly, if  $\mathcal{Y}$  is the matrix whose columns are the vectors  $\hat{\eta}_k$ , we have the relations

$$\mathcal{Y} e^k = \hat{\eta}_k. \quad (18.2.25)$$

Insert these relations into (2.22). So doing gives the result

$$\Psi_{jk} = (\mathcal{X}e^j, c \mathcal{Y}e^k) = (e^j, \mathcal{X}^T c \mathcal{Y}e^k), \quad (18.2.26)$$

which is equivalent to the matrix relation

$$\Psi = \mathcal{X}^T c \mathcal{Y}. \quad (18.2.27)$$

We will see shortly that the matrices  $\mathcal{X}$  and  $\mathcal{Y}$  are invertible. Assuming this to be the case, we may solve (2.27) for  $c$  to find the result

$$c = (\mathcal{X}^T)^{-1} \Psi \mathcal{Y}^{-1}. \quad (18.2.28)$$

We have found the  $c_{mn}$  in terms of the  $\Psi_{jk}$ , and therefore  $P$  is uniquely specified by the values  $\Psi_{jk}$ .

To see that  $\mathcal{X}$  is invertible, we examine  $\det \mathcal{X}$ , which (happily) is a *Vandermonde* determinant. It has the known value

$$\det \mathcal{X} = \prod_{j>k} (\xi_j - \xi_k). \quad (18.2.29)$$

Since the  $\xi_j$  are assumed to be *distinct* by construction,  $\det \mathcal{X}$  can never vanish. Thus  $\mathcal{X}$ , and similarly  $\mathcal{Y}$ , are invertible.

Now that  $P$  has been constructed, we find the desired interpolated result  $\psi(\bar{x}_i, \bar{y}_i, Z_L)$  by writing

$$\psi(\bar{x}_i, \bar{y}_i, Z_L) \approx P(\bar{\xi}_i, \bar{\eta}_i). \quad (18.2.30)$$

Note that the right side of (2.30) can be written an inner product form involving the two cubic vectors  $\hat{\bar{\xi}}_i$  and  $\hat{\bar{\eta}}_i$ ,

$$P(\bar{\xi}_i, \bar{\eta}_i) = (\hat{\bar{\xi}}_i, c \hat{\bar{\eta}}_i). \quad (18.2.31)$$

Hence the term *bicubic* interpolation again seems particularly appropriate.

In actual practice, since typically for our purposes any given  $P$  would be used only once, it is convenient to proceed somewhat differently. We will interpolate the quantity of interest, be it a potential value or some transverse field component, from the regular grid to the point  $\bar{x}_i, \bar{y}_i$  with the aid of four plus one cubic polynomials. This approach gives the same result as that obtained by first constructing  $P$  and then using (2.30), but employs somewhat more standard tools.

Suppose again that we wish to interpolate a potential function  $\psi$ . Using the four values  $\psi(X_J, Y_K, Z_L)$ ,  $\psi(X_J, Y_{K+1}, Z_L)$ ,  $\psi(X_J, Y_{K+2}, Z_L)$ ,  $\psi(X_J, Y_{K+3}, Z_L)$ , construct the cubic polynomial  $g_1(\sigma)$  given by

$$g_1(\sigma) = a_1 + b_1\sigma + c_1\sigma^2 + d_1\sigma^3 \quad (18.2.32)$$

such that

$$g_1[(m-1)h_y] = \psi(X_J, Y_{K+m-1}, Z_L) \text{ for } m = 1, 2, 3, 4. \quad (18.2.33)$$

(This is a standard construction in uniform-spacing *Lagrangian* interpolation.) Also form three more cubic polynomials  $g_2$  through  $g_4$  such that

$$g_k[(m-1)h_y] = \psi(X_{J+k-1}, Y_{K+m-1}, Z_L) \text{ for } m = 1, 2, 3, 4. \quad (18.2.34)$$

Together the four polynomials  $g_1$  through  $g_4$  allow us to interpolate in  $y$  for each of the four different  $x$  values  $X_K, X_{K+1}, X_{K+2}, X_{K+3}$ . (Pictorially, we are interpolating along the four different columns in Figure 2.1.) In particular, using (2.14), we have the four interpolated values

$$\psi_y(X_J, \bar{y}_i, Z_L) = g_1(\bar{\eta}_i), \quad (18.2.35)$$

$$\psi_y(X_{J+1}, \bar{y}_i, Z_L) = g_2(\bar{\eta}_i), \quad (18.2.36)$$

$$\psi_y(X_{J+2}, \bar{y}_i, Z_L) = g_3(\bar{\eta}_i), \quad (18.2.37)$$

$$\psi_y(X_{J+3}, \bar{y}_i, Z_L) = g_4(\bar{\eta}_i). \quad (18.2.38)$$

Next, using the four interpolated potential values  $\psi_y(X_{J+k-1}, \bar{y}_i, Z_L)$  with  $k \in [1, 4]$ , we will interpolate in  $x$ . Construct a fifth cubic polynomial  $f(\tau)$  of the form

$$f(\tau) = \alpha + \beta\tau + \gamma\tau^2 + \delta\tau^3 \quad (18.2.39)$$

such that

$$f[(k-1)h_x] = \psi_y(X_{J+k-1}, \bar{y}_i, Z_L) \text{ for } k = 1, 2, 3, 4. \quad (18.2.40)$$

Then, interpolating in  $x$  using  $f$  and (2.13) gives the final desired result

$$\psi(\bar{x}_i, \bar{y}_i, Z_L) \approx \psi_{xy}(\bar{x}_i, \bar{y}_i, Z_L) = f(\bar{\xi}_i). \quad (18.2.41)$$

At this point the observant reader has no doubt noticed that it also possible to interpolate in  $x$  first, and then in  $y$ , to get a result that we will call  $\psi_{yx}(\bar{x}_i, \bar{y}_i, Z_L)$ . How are  $\psi_{xy}(\bar{x}_i, \bar{y}_i, Z_L)$  and  $\psi_{yx}(\bar{x}_i, \bar{y}_i, Z_L)$  related? They are equal. In fact, there are the relations.

$$\psi_{xy}(\bar{x}_i, \bar{y}_i, Z_L) = \psi_{yx}(\bar{x}_i, \bar{y}_i, Z_L) = P(\bar{\xi}_i, \bar{\eta}_i). \quad (18.2.42)$$

That is, all three interpolation results agree. See Exercise 2.1

### 18.2.2 Bicubic Spline Interpolation

In the discussion so far, we have used cubic polynomials to interpolate in both  $x$  and  $y$ . An alternate approach is to use the cubic splines of Subsection 15.1.1 to perform interpolations. As before, one could interpolate first in  $y$  and then in  $x$ , or vice versa. It can be shown that these two results will again be the same. Look again at Figure 2.1, and consider the *central* square/rectangle that contains the point  $\bar{x}_i, \bar{y}_i$ . The coordinates of the grid point at the lower left corner of the central square/rectangle are  $X_{J+1}$  and  $Y_{K+1}$ . Now introduce *central* expansion variables, again call them  $\xi$  and  $\eta$ , by writing

$$\begin{aligned} x &= X_{J+1} + \xi, \\ y &= Y_{K+1} + \eta. \end{aligned} \quad (18.2.43)$$

Then it can be shown that this interpolation procedure is equivalent to using a bicubic polynomial again of the form (2.15), but now in the central expansion variables with the coefficients  $c_{mn}$  now determined from the  $\Psi_{jk}$  with the aid of cubic splines.

## Exercises

**18.2.1.** Verify that ....

## 18.3 Fourier Transforms

The work of Sections 14.2 through 14.5 required the computation of Fourier transforms. In this section we will describe numerical methods for this task. We will first define the Fourier transform and find its large  $|k|$  behavior. We will then define *discrete* Fourier transforms, and explore their large  $|k|$  behavior. Finally, we will define *spline-based* Fourier transforms that have, for our purposes, superior properties.

### 18.3.1 Exact Fourier Transform and Its Large $|k|$ Behavior

Suppose  $f(z)$  is a function that is nonzero (has *support*) only within the interval  $[a, b]$ , and that we wish to find its linear Fourier transform

$$\tilde{f}(k) = [1/(2\pi)] \int_{-\infty}^{\infty} dz \exp(-ikz) f(z) = [1/(2\pi)] \int_a^b dz \exp(-ikz) f(z). \quad (18.3.1)$$

Let us examine the behavior of  $\tilde{f}$  under the further supposition that  $f$  is differentiable and perhaps also has specific properties at the endpoints  $a$  and  $b$ . Then (3.1) may be integrated by parts to give the relation

$$\tilde{f}(k) = -[1/(2\pi)][1/(ik)] \exp(-ikz) f(z)|_a^b + [1/(2\pi)][1/(ik)] \int_a^b dz \exp(-ikz) f'(z). \quad (18.3.2)$$

The second term on the right in (3.2) may again be integrated by parts to give the result

$$\begin{aligned} & [1/(2\pi)][1/(ik)] \int_a^b dz \exp(-ikz) f'(z) \\ &= -[1/(2\pi)][1/(ik)^2] \exp(-ikz) f'(z)|_a^b + [1/(2\pi)][1/(ik)^2] \int_a^b dz \exp(-ikz) f''(z). \end{aligned} \quad (18.3.3)$$

Evidently, this process of integration by parts may be repeated at will as long as the required higher derivatives of  $f$  exist, and each such repetition produces one more power of  $1/k$ . For future use, we will repeat the process two more times to arrive at the result

$$\begin{aligned} \tilde{f}(k) &= -[1/(2\pi)][1/(ik)] \exp(-ikz) f(z)|_a^b - [1/(2\pi)][1/(ik)^2] \exp(-ikz) f'(z)|_a^b \\ &\quad - [1/(2\pi)][1/(ik)^3] \exp(-ikz) f''(z)|_a^b - [1/(2\pi)][1/(ik)^4] \exp(-ikz) f'''(z)|_a^b \\ &\quad + [1/(2\pi)][1/(ik)^4] \int_a^b dz \exp(-ikz) f^{iv}(z). \end{aligned} \quad (18.3.4)$$

We see that in general

$$|\tilde{f}(k)| \sim 1/|k| \text{ as } k \rightarrow \infty, \quad (18.3.5)$$

and that if  $f$  vanishes at the endpoints,  $f(a) = f(b) = 0$ , as will often be the case, then the first term on the right side of (3.4) will vanish so that

$$|\tilde{f}(k)| \sim 1/|k|^2 \text{ as } k \rightarrow \infty, \text{ etc.} \quad (18.3.6)$$

Thus,  $\tilde{f}(k)$  must vanish at least as fast as  $1/|k|$  for large  $|k|$ , and often vanishes as  $1/|k|^2$ .

### 18.3.2 Inverse Fourier Transform

One of the key features of the linear Fourier transform is the inverse Fourier transform relation

$$f(z) = \int_{-\infty}^{\infty} dk \exp(ikz) \tilde{f}(k). \quad (18.3.7)$$

That is, a function can be *reconstructed* from its linear Fourier transform by using an inverse Fourier transform. Let us further assume that the integral (3.7) can be *cut off* for  $|k| > K_c$  where  $K_c$  is some suitably large value, say a value where and beyond which the asymptotic behavior (3.5) or (3.6) has effectively driven  $\tilde{f}$  to zero. Thus, we write

$$f(z) \approx \int_{-K_c}^{K_c} dk \exp(ikz) \tilde{f}(k). \quad (18.3.8)$$

Eventually we will need to evaluate integrals of the form (3.8) numerically. Therefore, we would like to know something about the properties of  $\tilde{f}(k)$  in the interval  $[-K_c, K_c]$ . In particular, we would like to know how much  $\tilde{f}(k)$  oscillates. Suppose that  $f(z)$  has support only in the interval  $[a, b]$ . Then  $\tilde{f}(k)$  must encode two pieces of information: it must encode that  $f(z)$  is zero outside  $[a, b]$  and it must encode the behavior of  $f(z)$  within the interval  $[a, b]$ . We see from (3.1) that  $\tilde{f}(k)$  is a generalized sum (integral) of terms of the form  $\exp(-i\omega k)$  where  $\omega \in [a, b]$ . That is,  $\tilde{f}(k)$  contains all frequencies  $\omega \in [a, b]$  with weights  $f(\omega)$ . This is a potential disaster, because  $|a|$  and/or  $|b|$  could be quite large.

For example consider the functions  $f_{-1,1}(z)$  and  $f_{0,2}(z)$  defined by the relations

$$\begin{aligned} f_{-1,1}(z) &= 1 - z^4 \text{ for } z \in [-1, 1], \\ &= 0 \text{ for } z \text{ outside } [-1, 1]; \end{aligned} \quad (18.3.9)$$

$$\begin{aligned} f_{0,2}(z) &= 1 - (z - 1)^4 \text{ for } z \in [0, 2], \\ &= 0 \text{ for } z \text{ outside } [0, 2]. \end{aligned} \quad (18.3.10)$$

Figures 3.1 and 3.2 show their graphs which, evidently and by construction, are simply translations of each other. [See also (1.36) and Figure 1.3.] Calculation shows that their Fourier transforms are given by

$$\Re \tilde{f}_{-1,1}(k) = -[4/(\pi k^5)][k(k^2 - 6) \cos k - 3(k^2 - 2) \sin k], \quad (18.3.11)$$

$$\Im \tilde{f}_{-1,1}(k) = 0; \quad (18.3.12)$$

$$\Re \tilde{f}_{0,2}(k) = [1/(\pi k^5)][-4k(k^2 - 6)\cos^2 k + 6(k^2 - 2)\sin 2k], \quad (18.3.13)$$

$$\Im \tilde{f}_{0,2}(k) = [1/(\pi k^5)][12(k^2 - 2)\sin^2 k - 2k(k^2 - 6)\sin 2k]. \quad (18.3.14)$$

As expected,  $\tilde{f}_{-1,1}(k)$  contains frequency  $|\omega| = 1$  from the  $\sin k$  and  $\cos k$  terms. By contrast,  $\tilde{f}_{0,2}(k)$  contains frequency  $|\omega| = 2$  from the  $\sin 2k$ ,  $\cos^2 k$ , and  $\sin^2 k$  terms. These different behaviors are also evident in the graphs of these Fourier transforms as shown in Figures 3.3 through 3.5. Clearly the integral (3.8) is more difficult to evaluate for  $\tilde{f}_{0,2}(k)$  than for  $\tilde{f}_{-1,1}(k)$  because  $\tilde{f}_{0,2}(k)$  oscillates twice as often as  $\tilde{f}_{-1,1}(k)$ . There is an oscillation penalty to be paid for encoding the fact that some  $f$  has support in the interval  $[0, 2]$  rather than the interval  $[-1, 1]$ .

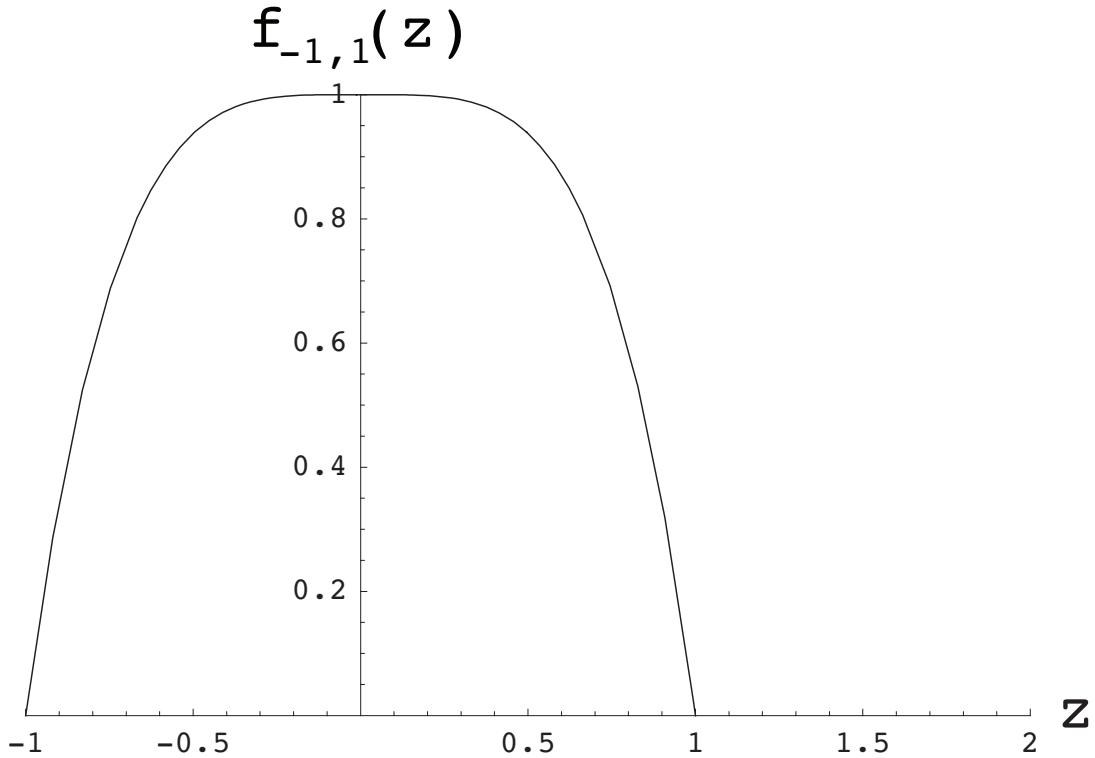
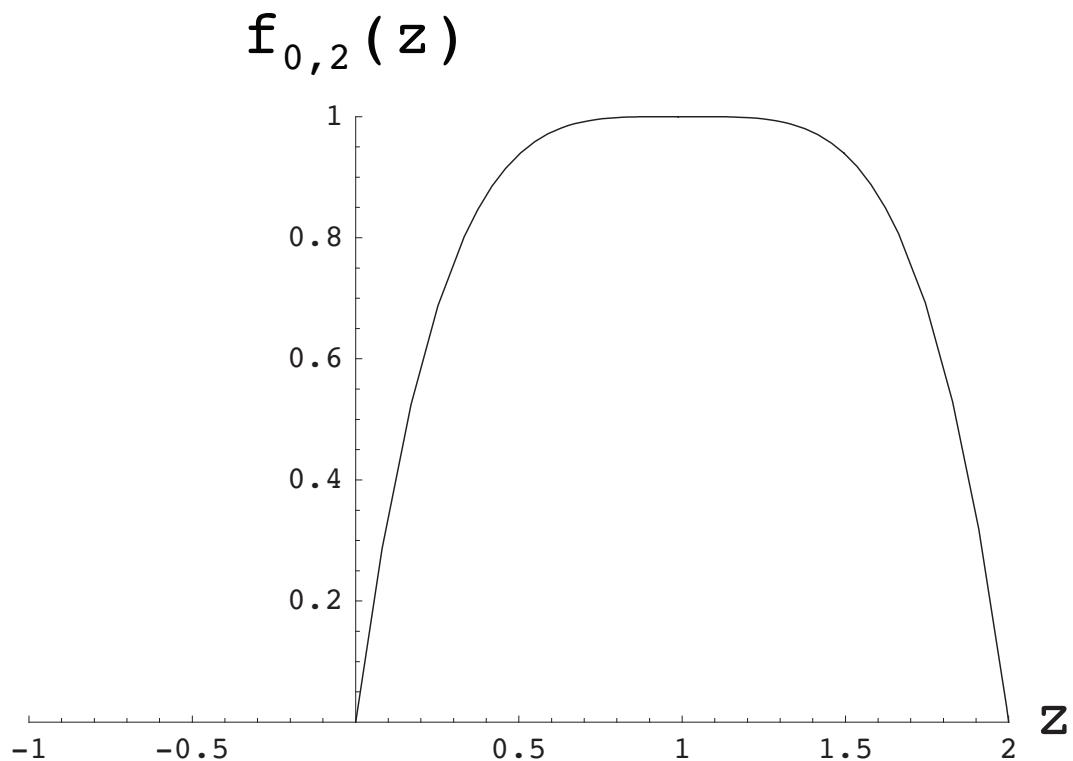
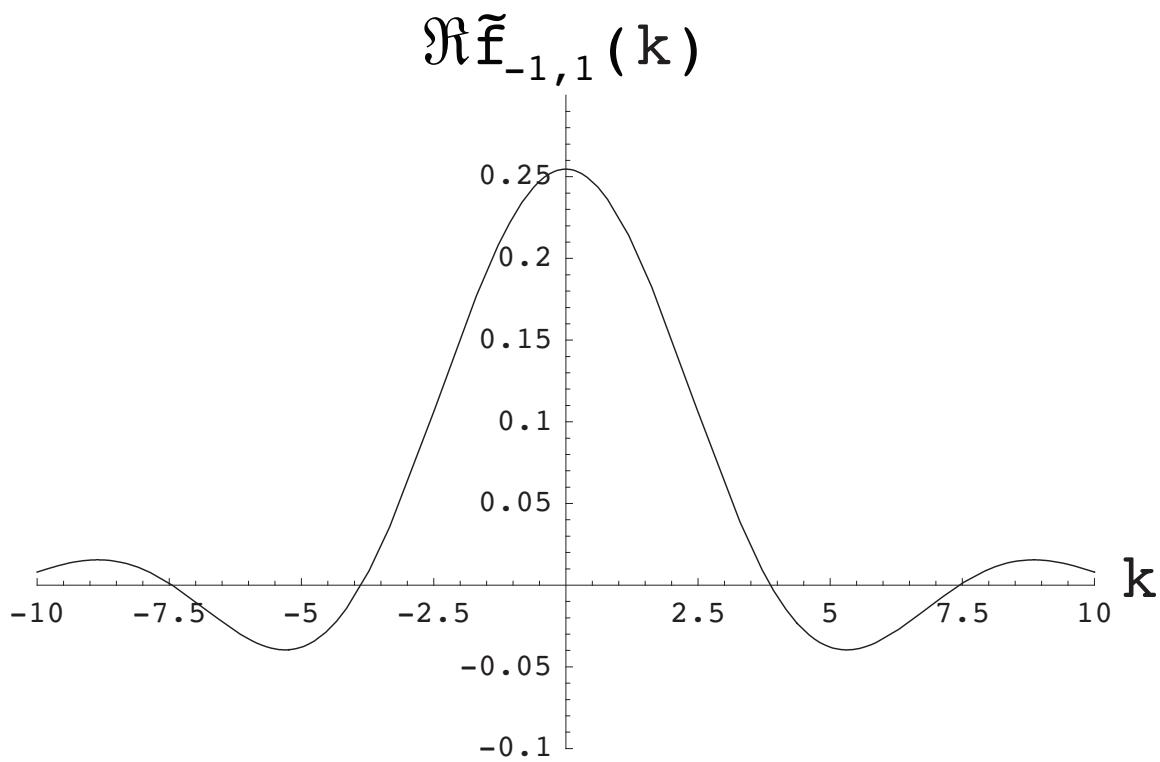


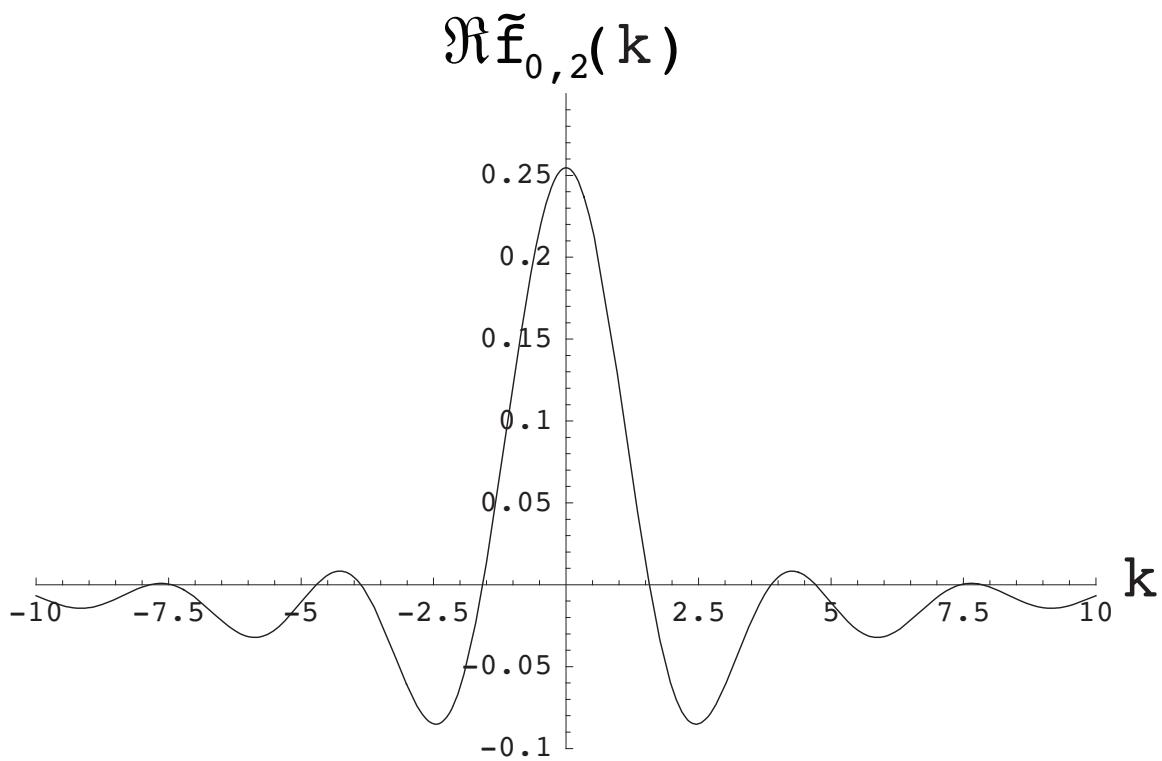
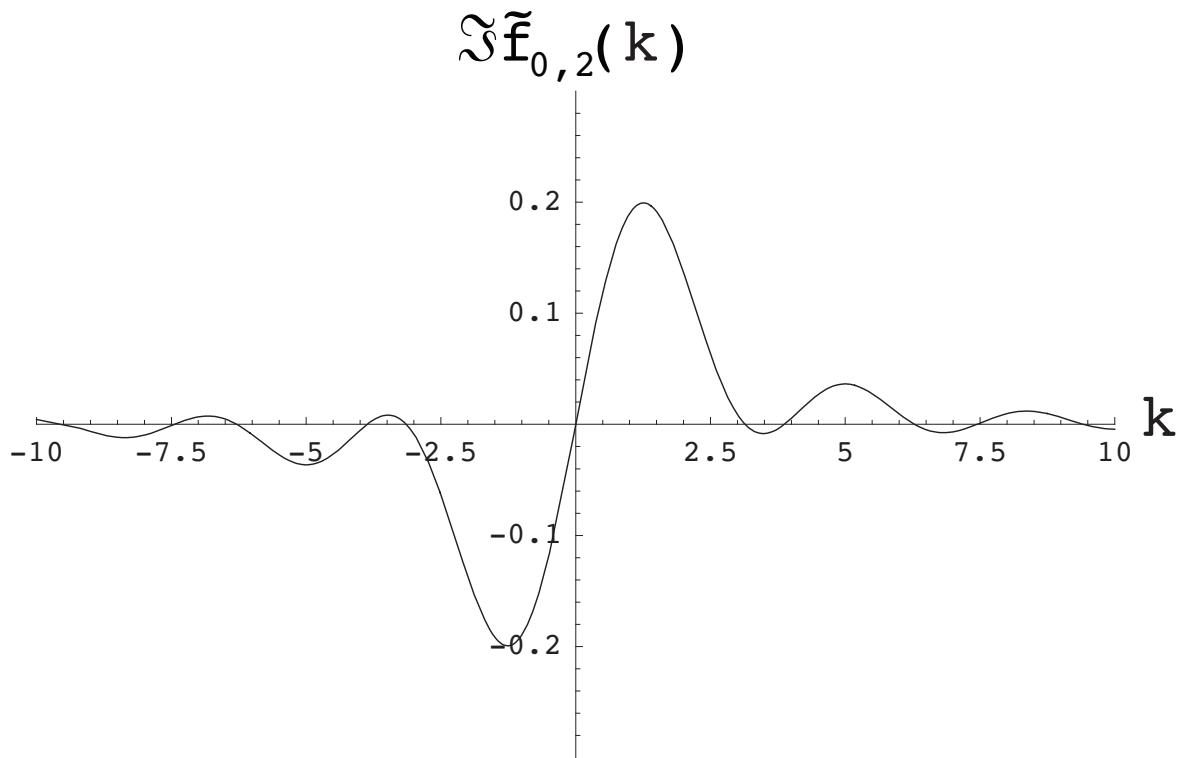
Figure 18.3.1: The function  $f_{-1,1}(z)$ .

What to do? By simple translation it is always possible to send the interval  $[a, b]$  to the interval  $[-Z_c, Z_c]$ . Therefore, without loss of generality, we may restrict our attention to integrals of the form

$$\tilde{f}(k) = [1/(2\pi)] \int_{-Z_c}^{Z_c} dz \exp(-ikz) f(z). \quad (18.3.15)$$

Then, after all operations have been carried out, we may undo, if we wish, the translation to obtain results in terms of the original coordinates. In this way, we only have to deal with  $\tilde{f}(k)$  that contain frequencies satisfying  $|\omega| \leq Z_c$ .

Figure 18.3.2: The function  $f_{0,2}(z)$ .Figure 18.3.3: The function  $\Re \tilde{f}_{-1,1}(k)$ .

Figure 18.3.4: The function  $\Re \tilde{f}_{0,2}(k)$ .Figure 18.3.5: The function  $\Im \tilde{f}_{0,2}(k)$ .

### 18.3.3 Discrete Fourier Transform

Let us now turn to the task of evaluating  $\tilde{f}$  numerically. Suppose the interval  $[-Z_c, Z_c]$  is subdivided into  $N$  subintervals, each of length  $h$ , by writing

$$z_n = -Z_c + nh \text{ with } n = 0, 1, \dots, N \quad (18.3.16)$$

where

$$h = 2Z_c/N \quad (18.3.17)$$

so that

$$z_0 = -Z_c \text{ and } z_N = Z_c. \quad (18.3.18)$$

See Figure 2.1.1 for an earlier analogous construction. Use this subdivision to approximate the integral (3.15) by the Riemann sum

$$\tilde{f}(k) \approx [1/(2\pi)]h \sum_{n=0}^{N-1} \exp(-ikz_n)f(z_n). \quad (18.3.19)$$

The quantity on the right side of (3.19) is called the *discrete Fourier transform* of  $f$ .

How accurate is the discrete Fourier transform? Let  $g(z)$  be any twice differentiable function. Then, according to the *trapezoidal rule*, there is the result

$$\begin{aligned} \int_{-Z_c}^{Z_c} dz g(z) &= h[(1/2)g(z_0) + g(z_1) + g(z_2) + \dots + g(z_{N-2}) + g(z_{N-1}) + (1/2)g(z_N)] \\ &\quad - (1/6)h^2g''(\zeta), \end{aligned} \quad (18.3.20)$$

which can be rewritten in the form

$$\begin{aligned} \int_{-Z_c}^{Z_c} dz g(z) &= h[g(z_0) + g(z_1) + g(z_2) + \dots + g(z_{N-2}) + g(z_{N-1})] \\ &\quad - (1/2)hg(z_0) + (1/2)hg(z_N) - (1/6)h^2g''(\zeta). \end{aligned} \quad (18.3.21)$$

Here  $\zeta$  is some point in the interval  $[-Z_c, Z_c]$ . In our case we have

$$g(z) = [1/(2\pi)] \exp(-ikz)f(z) \quad (18.3.22)$$

so that

$$g''(\zeta) = [1/(2\pi)][-k^2 f(\zeta) - 2ikf'(\zeta) + f''(\zeta)] \exp(-ik\zeta). \quad (18.3.23)$$

Thus, we have the result

$$\begin{aligned} \tilde{f}(k) &= [1/(2\pi)]h \sum_{n=0}^{N-1} \exp(-ikz_n)f(z_n) \\ &\quad - (1/2)[1/(2\pi)]hf(-Z_c) \exp(ikZ_c) + (1/2)[1/(2\pi)]hf(Z_c) \exp(-ikZ_c) \\ &\quad - (1/6)h^2[1/(2\pi)][-k^2 f(\zeta) - 2ikf'(\zeta) + f''(\zeta)] \exp(-ik\zeta). \end{aligned} \quad (18.3.24)$$

Upon comparing (3.19) and (3.24), we see that the discrete Fourier transform generally makes errors of order  $h$ . However, if  $f(-Z_c) = 0$  and  $f(Z_c) = 0$ , which is often the case, then the discrete transform makes errors of order  $h^2$ . But, the errors of order  $h^2$  can be very large if  $|k|$  is large.

Let us explore the large  $|k|$  behavior of the discrete Fourier transform. If we make use of (3.16), we see that (3.19) can also be written in the form

$$\tilde{f}(k) \approx [1/(2\pi)]h \exp(ikZ_c) \sum_{n=0}^{N-1} \exp(-iknh)f(z_n). \quad (18.3.25)$$

Observe that the function  $\exp(-ikh)$  is *periodic* in  $k$  with a period  $K$  given by

$$K = 2\pi/h = N\pi/Z_c. \quad (18.3.26)$$

Consequently, if we define a function  $F(k)$  by writing

$$F(k) = [1/(2\pi)]h \sum_{n=0}^{N-1} \exp(-iknh)f(z_n), \quad (18.3.27)$$

we have the relation

$$F(k + K) = F(k). \quad (18.3.28)$$

But (3.19) can be rewritten in terms of  $F$ . We find, using (3.25) and (3.27), the result

$$\tilde{f}(k) \approx \exp(ikZ_c)F(k). \quad (18.3.29)$$

We see that the discrete Fourier transform  $\tilde{f}$  is *quasi-periodic* in  $k$ . It is a product of the function  $\exp(ikZ_c)$ , which has the period  $2\pi/Z_c$  (so that  $|\omega| = Z_c$ ), and the *envelope* function  $F$  which has period  $K$ . Thus, the discrete Fourier transform can never satisfy (3.5) or (3.6).<sup>5</sup> In general, the discrete Fourier transform is reliable only in the interval  $k \in [-K/2, K/2]$ . The quantity  $K_{Ny} = K/2$  is called the *Nyquist critical frequency*.

As an example of the behavior of the discrete Fourier transform, consider again the function  $f_{-1,1}(z)$  given by (3.9) and shown in Figure 3.1. We have already seen that it has the exact Fourier transform given by (3.11) and (3.12) and shown in Figure 3.3. Figure 3.6 shows both the exact Fourier transform, and the discrete and spline-based Fourier transforms for the case  $h = .10$ .<sup>6</sup> (The spline-based Fourier transform is discussed in a subsequent subsection.) In this case  $K_{Ny} = \pi/h \simeq 31.4$ . We observe that, as warned, the discrete Fourier transform results are not reliable for  $|k| > K_{Ny}$ . Figure 3.7 shows the difference between the exact and discrete Fourier transforms within and somewhat beyond the *Nyquist band*  $|k| < K_{Ny}$ . Note that, as expected from the quasi-periodicity of the discrete Fourier transform, see Figure 3.6, the error grows as  $k$  leaves the Nyquist band.

---

<sup>5</sup>Observe that the discrete Fourier transform (3.19) can be viewed as the *exact* Fourier transform of the function  $h \sum_{n=0}^{N-1} f(z_n)\delta(z - z_n)$ , and that this finite sum of *delta* functions must have all frequency components present.

<sup>6</sup>All these Fourier transforms are real because  $f_{-1,1}(z)$  as given by (3.9) is an even function.

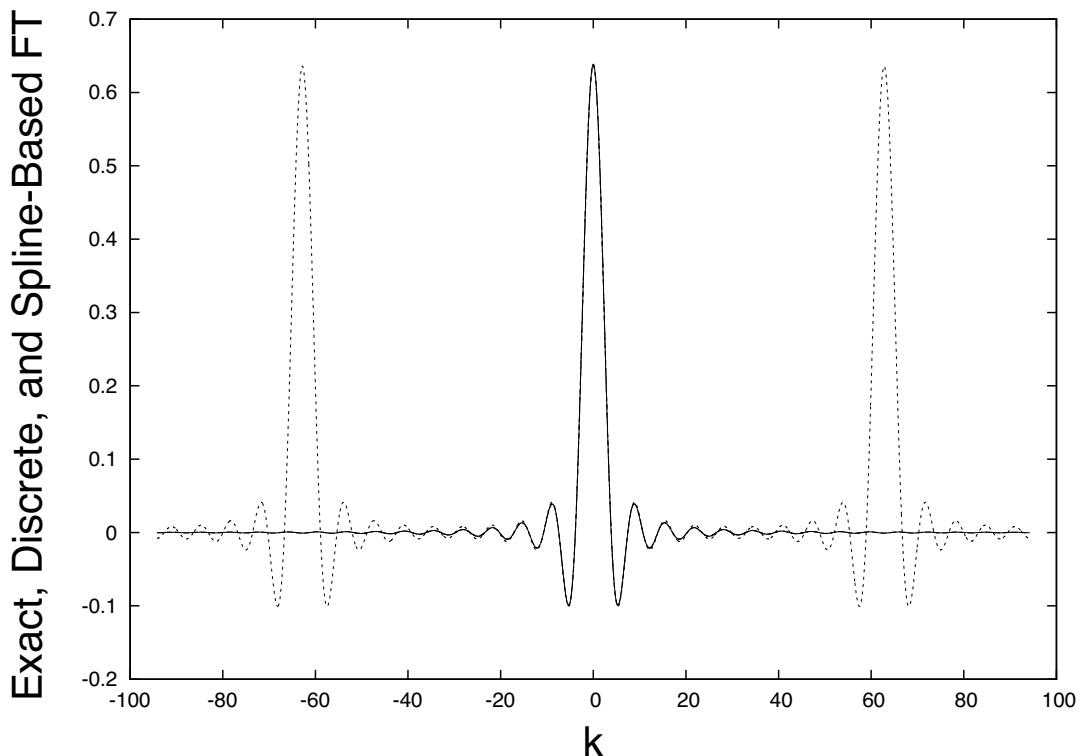


Figure 18.3.6: The exact, discrete, and spline-based Fourier transforms of  $f_{-1,1}(z)$  for  $h = .10$ . On the scale of this figure the exact and spline-based Fourier transforms are indistinguishable. They are both shown as a solid line. The discrete Fourier transform is shown as a dashed line. Note that it is quasi-periodic while the exact and spline-based Fourier transforms fall to zero for large  $|k|$ .

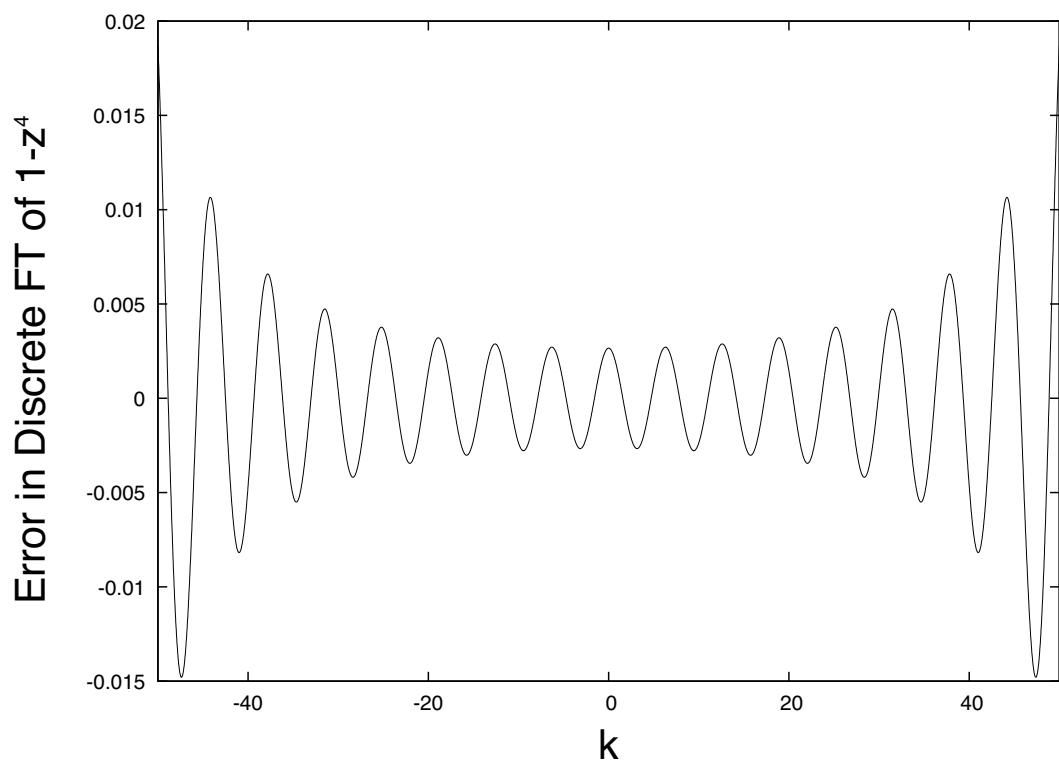


Figure 18.3.7: Difference between the exact and discrete Fourier transforms of  $f_{-1,1}(z)$  for  $h = .10$ .

### 18.3.4 Discrete Inverse Fourier Transform

Consider again the inverse Fourier transform (3.8). Let us explore to what extent this inverse transform can also be made discrete and to what extent the discrete forward and inverse Fourier transforms are related. Discretize what so far has been the continuous variable  $k$  by writing

$$k_m = -K_c + mH \text{ for } m = 0, 1, \dots, M \quad (18.3.30)$$

with

$$H = 2K_c/M \quad (18.3.31)$$

so that

$$k_0 = -K_c \text{ and } k_M = +K_c. \quad (18.3.32)$$

When this is done, the relation (3.8) can be approximated by the Riemann sum

$$f(z) \approx H \sum_{m=0}^{M-1} \exp(ik_m z) \tilde{f}(k_m). \quad (18.3.33)$$

Evidently both (3.19) and (3.33) are approximate, and presumably they become ever more accurate as  $M \rightarrow \infty$  and  $N \rightarrow \infty$ . However, if  $M = N$  and  $K_c = K_{Ny}$ , there are the *exact* relations

$$\tilde{f}(k_m) = [1/(2\pi)]h \sum_{n=0}^{N-1} \exp(-ik_m z_n) f(z_n), \quad (18.3.34)$$

and

$$f(z_n) = H \sum_{m=0}^{M-1} \exp(ik_m z_n) \tilde{f}(k_m). \quad (18.3.35)$$

Here (3.34) is taken to be the *definition* of the quantities  $\tilde{f}(k_m)$ . [They are *not* the exact values that would be found by doing the integral (3.15) exactly for the values  $k = k_m$ .] And when these approximate values  $\tilde{f}(k_m)$  are employed in the (approximate) formula (3.35), the values  $f(z_n)$  are recovered *exactly*. This a case where two wrongs do make a right!<sup>7</sup> See Exercise 3.1.

### 18.3.5 Spline-Based Fourier Transforms

Let  $f_{\text{sa}}$  be a cubic spline approximation to the function  $f$  appearing in (3.15). Then we may make the definition

$$\tilde{f}_{\text{sa}}(k) = [1/(2\pi)] \int_{-Z_c}^{Z_c} dz \exp(-ikz) f_{\text{sa}}(z). \quad (18.3.36)$$

As described in Section 15.1.1,  $f_{\text{sa}}(z)$  can be constructed from the values  $f(z_n)$  with  $n \in [0, N]$ . Moreover, observe that the definition (3.36) can be evaluated exactly (numerically to machine precision) for any value of  $k$  since the Fourier transforms of cubic polynomials

---

<sup>7</sup>The relation (3.35) does not give the exact values of  $f(z)$  when  $z \neq z_n$ . Also note that (3.35) produces a function of  $z$  that is *quasi-periodic* with quasi-period  $Z = 2\pi/H = 2$ . By contrast,  $f(z)$  is supposed to be zero for  $z$  outside the interval  $[-1, 1]$ .

can be found analytically and then evaluated numerically to machine precision. Thus, we have a relation of the form

$$\tilde{f}_{\text{sa}}(k) = \sum_{n=0}^N g_n^-(k) f(z_n) \quad (18.3.37)$$

where the  $g_n^-(k)$  are known functions of  $k$  that can be evaluated to machine precision.<sup>8</sup> Suppose we assume that  $f(z)$  vanishes at the endpoints  $\pm Z_c$ . Then  $f_{\text{sa}}$  will be a differentiable function of  $z$  that vanishes at the endpoints. Therefore we expect that  $\tilde{f}_{\text{sa}}(k)$ , unlike the  $\tilde{f}$  given by (3.19), will have more nearly appropriate large  $|k|$  behavior.<sup>9</sup> See (3.6). The accuracy of  $\tilde{f}_{\text{sa}}(k)$ , namely the difference between  $\tilde{f}_{\text{sa}}(k)$  and the exact  $\tilde{f}(k)$  given by (3.15), depends only on the quality of the spline fit  $f_{\text{sa}}(z)$ , and not on any approximation to the Fourier integral.

As an example of the behavior of  $\tilde{f}_{\text{sa}}(k)$ , consider again the  $f_{-1,1}(z)$  given by (3.9). As already described earlier, Figure 3.6 shows the exact Fourier transform, the discrete Fourier transform, and  $\tilde{f}_{\text{sa}}$  for the case  $h = .10$ . [Here we have used (1.24) to estimate the end-point derivatives required to construct  $f_{\text{sa}}(z)$ .] We have already noted that the exact and spline-based Fourier transforms appear identical on the scale shown. Figure 3.8 shows their difference. Note that their difference is small as expected from the error estimate (1.34) and the fact that the Fourier transform (3.36) of the spline approximation is evaluated exactly. In summary, as comparison of Figures 3.7 and 3.8 illustrates, for smooth functions the spline-based Fourier transform is much more accurate than the discrete Fourier transform.

What can be said about the inverse Fourier transform? Begin with (3.8). It is approximate because of the cutoff, but this cutoff  $K_c$  can in principle be made quite large to assure good accuracy. Next replace  $\tilde{f}(k)$  by  $\tilde{f}_{\text{sa}}(k)$  using (3.36) to get the approximation

$$f(z) \approx \int_{-K_c}^{K_c} dk \exp(ikz) \tilde{f}_{\text{sa}}(k). \quad (18.3.38)$$

The quality of this approximation depends on the quality of the spline fit  $f_{\text{sa}}(z)$ . Also suppose we carry out the operation (3.36) for  $M + 1$  discrete values of  $k$  using the  $k_m$  values given by (3.30). That is, we compute the quantities  $\tilde{f}_{\text{sa}}(k_m)$  by the rule

$$\tilde{f}_{\text{sa}}(k_m) = \sum_{n=0}^N g_n^-(k_m) f(z_n). \quad (18.3.39)$$

We may then use these quantities to try to reconstruct  $f(z)$ .

In particular, use the  $M + 1$  values  $\tilde{f}_{\text{sa}}(k_m)$  to construct a cubic spline approximation to  $\tilde{f}_{\text{sa}}(k)$  which we will call  $\tilde{f}_{\text{sasa}}(k)$ . [Again we use (1.24), this time applied to the values  $\tilde{f}_{\text{sa}}(k_m)$ , to estimate the required end-point derivatives.] Using this approximation in (3.38) gives the representation

$$f(z) \approx \int_{-K_c}^{K_c} dk \exp(ikz) \tilde{f}_{\text{sasa}}(k). \quad (18.3.40)$$

---

<sup>8</sup> The superscript “ $-$ ” indicates that  $\exp(-ikz)$  appears in (3.36).

<sup>9</sup>The function  $\tilde{f}_{\text{sa}}(k)$  given by (3.36) cannot fall off any faster than  $1/|k|^4$  at infinity because a cubic-spline has discontinuous third derivatives. See Exercise 3.2.

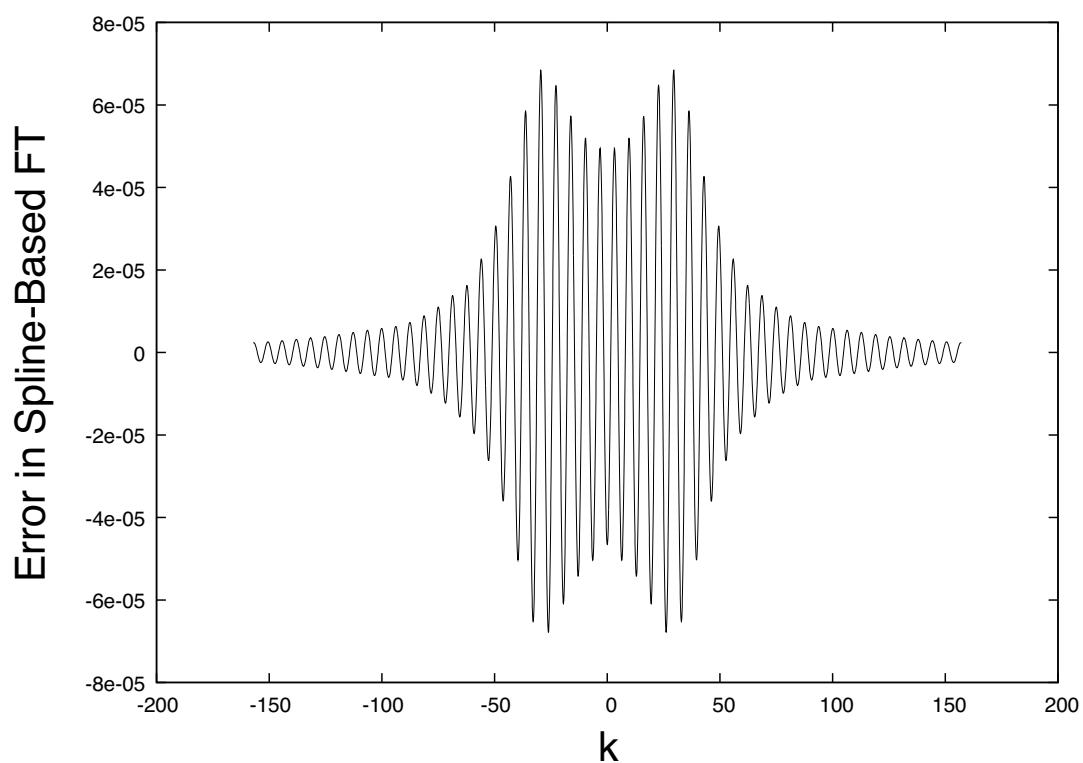


Figure 18.3.8: Difference between the exact and spline-based Fourier transforms of  $f_{-1,1}(z)$  for  $h = .10$ .

Because  $\tilde{f}_{\text{sasa}}(k)$  is a cubic spline approximation, the integral (3.40) can again be done exactly. That is, there are known functions  $g_m^+(z)$  such that<sup>10</sup>

$$f(z) \approx \int_{-K_c}^{K_c} dk \exp(ikz) \tilde{f}_{\text{sasa}}(k) = \sum_{m=0}^M g_m^+(z) \tilde{f}_{\text{sa}}(k_m). \quad (18.3.41)$$

The accuracy of this representation of  $f$  depends on how well  $\tilde{f}_{\text{sasa}}(k)$  approximates  $\tilde{f}_{\text{sa}}(k)$ . In general it will not yield an  $f$  that vanishes outside  $[-1, 1]$ . However, unlike the  $f$  produced by (3.33), the  $f$  produced by (3.41) will be very small for  $z$  outside the interval  $[-1, 1]$ .

In summary, the accuracy of direct and inverse spline-based Fourier transforms is governed primarily by the quality of spline approximations, and not by how well various Fourier integrals are approximated. Put another way, the use of the discrete Fourier transform does not make any optimistic assumptions about the smoothness properties of  $f(z)$ . (Indeed, it assumes the worst, a sum of delta functions approximation.) By contrast, the use of spline-based Fourier transforms capitalizes on the assumption that  $f(z)$  is not too badly behaved between sampling points  $z_n$ .

How well do the discrete and spline-based Fourier transforms work in reconstructing a function? Let us first consider the discrete case. As an example, we will again consider the  $f_{-1,1}$  given by (3.9). Figure 3.9 shows the reconstructed  $f_{-1,1}(z)$  produced by (3.33) with the  $\tilde{f}(k_m)$  given by (3.19) or, equivalently, (3.34). Figure 3.10 shows the difference between the exact  $f_{-1,1}$  and the reconstructed  $f_{-1,1}$ . Here again we have used  $h = .10$  so that  $N = 20$ ; and we have set  $K_c = K_{Ny}$  and  $M = N$ . We see from Figure 3.9 that the reconstructed  $f_{-1,1}$  is quasi-periodic as expected. We see from Figure 3.10 that the error is zero at the sampling points as expected, but rises to as high as 2% elsewhere.

What happens if we instead use spline-based Fourier transforms? We have already seen that, for this example, the *forward* spline-based Fourier transform  $\tilde{f}_{\text{sa}}(k)$  is more accurate than the discrete Fourier transform. This is because of the high accuracy of the spline approximation  $f_{\text{sa}}(z)$  to  $f(z)$ , and the fact that the Fourier transform of the spline approximation is performed exactly. We expect to be able to carry out the *inverse* spline-based Fourier transform with good accuracy provided the spline approximation  $\tilde{f}_{\text{sasa}}(k)$  to  $\tilde{f}_{\text{sa}}(k)$  has good accuracy. But now there is a possible problem. Figure 3.11 shows the 21-point spline approximation  $\tilde{f}_{\text{sasa}}(k)$  over the interval  $[-K_{Ny}, K_{Ny}]$  as well as  $\tilde{f}_{\text{sa}}(k)$  itself. We see that the 21-point spline approximation  $\tilde{f}_{\text{sasa}}(k)$  is not particularly good because of the oscillatory nature of  $\tilde{f}_{\text{sa}}(k)$ . Figure 3.12 shows the exact  $f_{-1,1}(z)$  and the reconstructed  $f_{-1,1}(z)$  based on using  $\tilde{f}_{\text{sasa}}(k)$  in (3.41) with  $K_c = K_{Ny}$ . Evidently the agreement is not particularly good, reflecting the poor quality of the 21-point spline approximation  $\tilde{f}_{\text{sasa}}(k)$ .

Suppose we instead make  $K_c$  somewhat larger than  $K_{Ny}$  by setting  $K_c = 50$  and also use a 51-point spline approximation to  $\tilde{f}_{\text{sasa}}(k)$  over this interval  $[-K_c, K_c]$ . When this is done, it is found that the difference between  $\tilde{f}_{\text{sa}}(k)$  and its spline fit  $\tilde{f}_{\text{sasa}}(k)$  is less than  $6 \times 10^{-5}$ . Correspondingly we expect the reconstruction of  $f_{-1,1}(z)$  to be much improved. This is indeed the case. Figure 3.13 shows the function  $f_{-1,1}(z)$  and its reconstruction using, in (3.41), the 51-point spline approximation  $\tilde{f}_{\text{sasa}}(k)$  over the interval  $k \in [-50, 50]$ . The agreement is much improved, and is even good outside the interval  $[-1, 1]$  where the discrete

<sup>10</sup>Here the superscript “+” indicates that  $\exp(+ikz)$  appears in (3.40) and (3.41).

### Reconstruction of $1 - z^4$ Using DFT

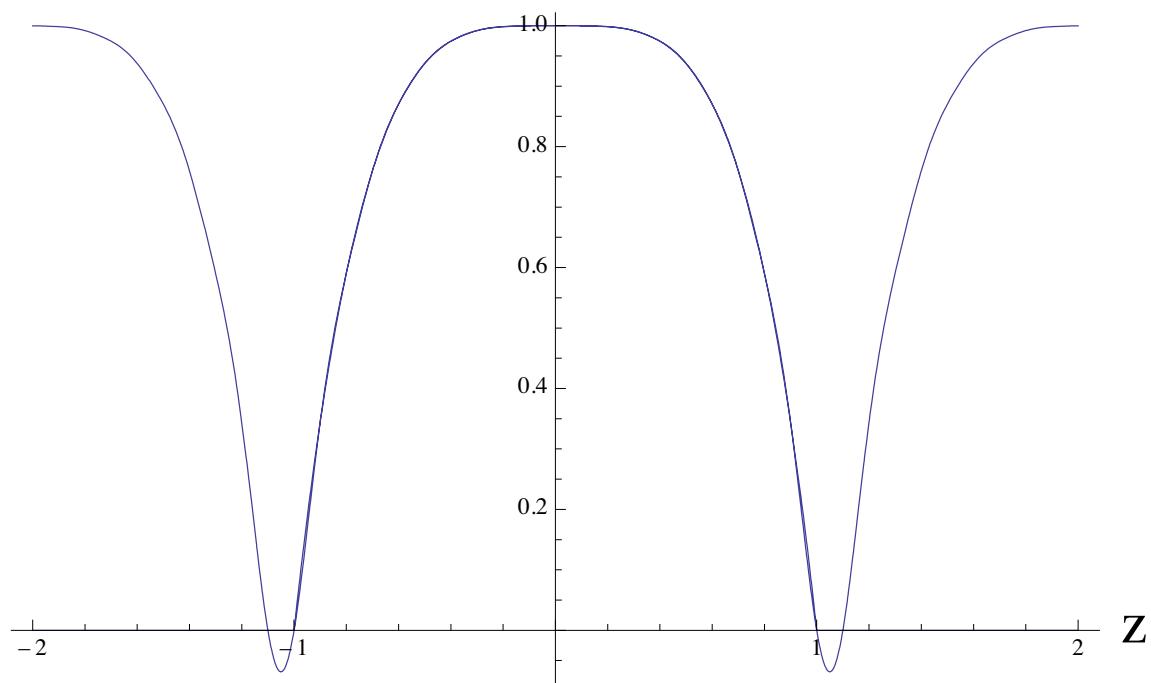


Figure 18.3.9: Reconstruction of  $f_{-1,1} = 1 - z^4$  using forward and inverse discrete Fourier transforms.

### Error in Reconstruction of $1-z^4$ Using DFT

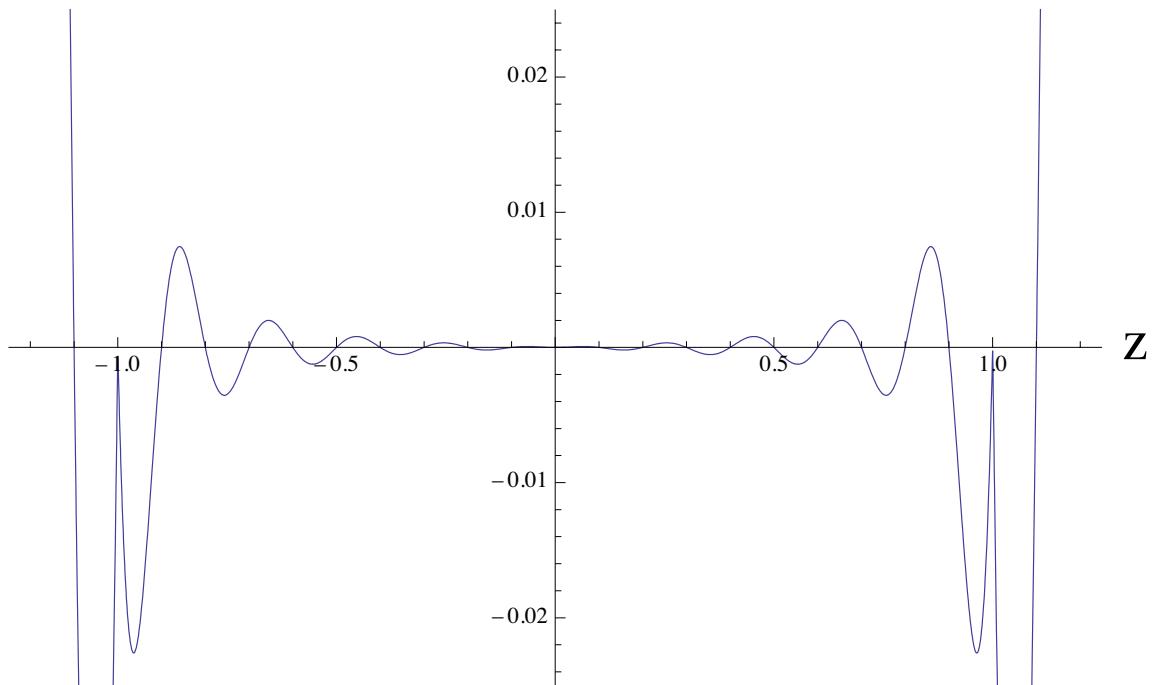


Figure 18.3.10: Error in reconstruction of  $f_{-1,1} = 1 - z^4$  using forward and inverse discrete Fourier transforms.

Fourier reconstruction fails. To provide further insight into the error, Figure 3.14 displays the difference between the exact  $f_{-1,1}(z)$  and the reconstructed  $f_{-1,1}(z)$ . Now the error is comparable to that for the discrete case, and, unlike the discrete case, is even small outside the interval  $[-1, 1]$ . Compare Figures 3.9, 3.10, 3.13, and 3.14. Of course, with more points the discrete-case error also decreases. But it decreases as a smaller power of  $h$  than in the spline-based case so that eventually the spline-based method wins.

Moreover, the apparent good performance of the discrete method is misleading. We already know from our previous discussion that its error when performing reconstructions must be zero at the sampling points due to the magic cancellation of errors in the forward and inverse discrete Fourier transformations at these points. But we are ultimately not interested in reconstruction. Rather, we are interested in forward Fourier transformation followed by inverse Fourier transformation with some  $k$ -dependent kernel. See (14.2.2), (14.2.6) and (14.3.1), (14.3.6) and (14.4.73), (14.4.74), (14.4.85), (14.4.86). In this context there is no reason to expect cancellation of errors when discrete Fourier transformations are employed. And, when spline-based Fourier transformations are employed, we may expect to see errors that are no worse than those encountered in the case of reconstruction. We conclude that as long as the spline approximations are done with care, the spline-based Fourier transforms should be superior to discrete Fourier transforms.

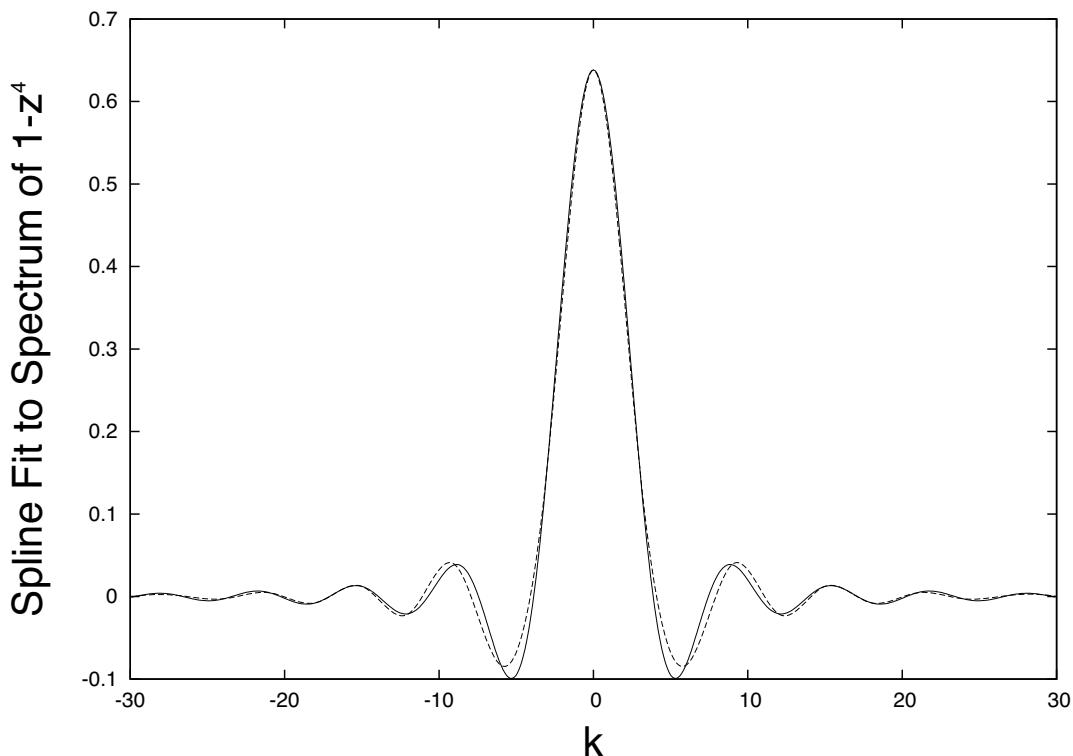


Figure 18.3.11: The function  $\tilde{f}_{\text{sa}}(k)$  (solid line) and its 21-point spline approximation  $\tilde{f}_{\text{sasa}}(k)$  (dashed line) over the Nyquist band  $k \in [-K_{Ny}, K_{Ny}]$ .

## Reconstruction of $1 - z^4$ Using Spline-Based FT

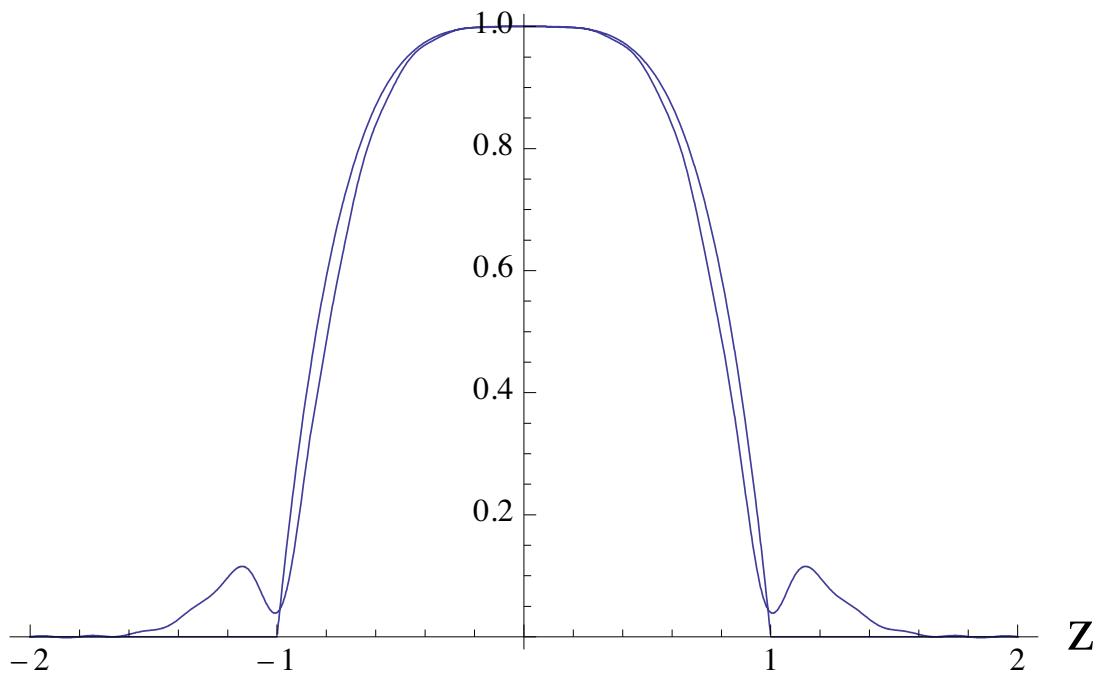


Figure 18.3.12: The function  $f_{-1,1}(z) = 1 - z^4$  and its reconstruction using the 21-point spline approximation  $\tilde{f}_{\text{sasa}}(k)$  in (3.41).

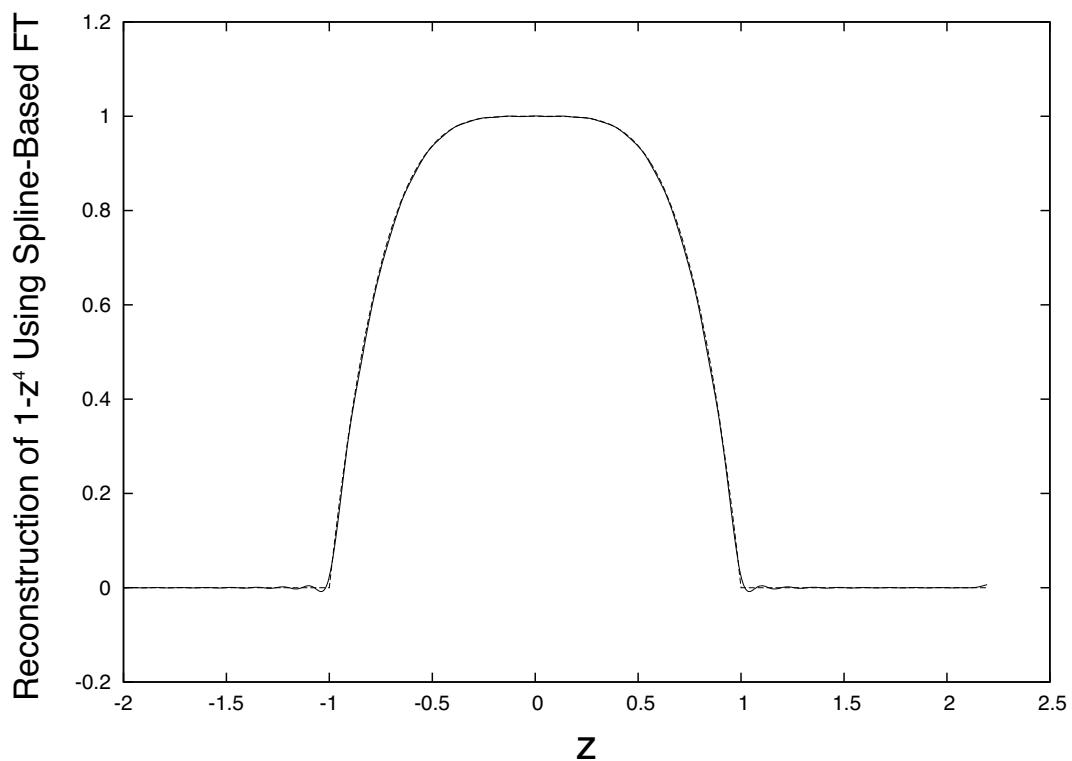


Figure 18.3.13: The function  $f_{-1,1}(z) = 1 - z^4$  and its reconstruction using, in (3.41), the 51-point spline approximation  $\tilde{f}_{\text{sasa}}(k)$  over the interval  $k \in [-50, 50]$ .

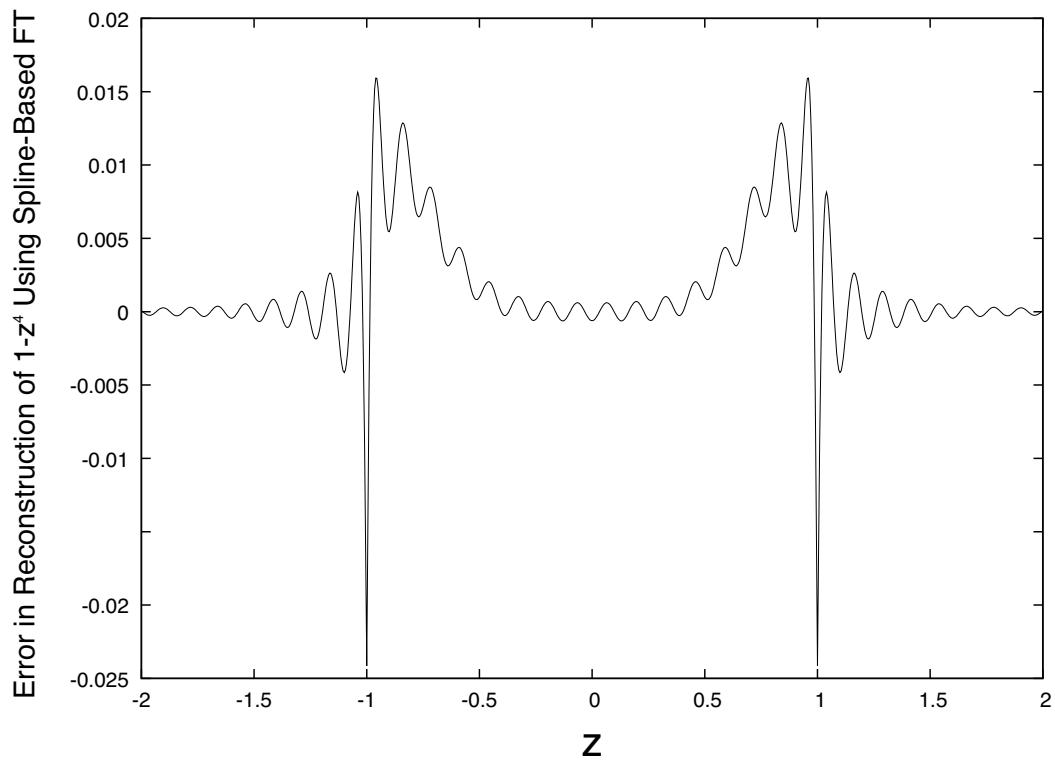


Figure 18.3.14: The difference between the exact function  $f_{-1,1}(z) = 1 - z^4$  and its reconstruction using, in (3.41), the 51-point spline approximation  $\tilde{f}_{\text{sasa}}(k)$  over the interval  $k \in [-50, 50]$ .

### 18.3.6 Fast Spline-Based Fourier Transforms

How much work is involved in computing Fourier transforms? Let  $M_{mn}$  be the matrix with entries

$$M_{mn} = [1/(2\pi)]h \exp(-ik_m z_n). \quad (18.3.42)$$

With this notation, the discrete Fourier transformation relation (3.34) can be written in the vector/matrix form

$$\tilde{f}(k_m) = \sum_{n=0}^{N-1} M_{mn} f(z_n). \quad (18.3.43)$$

Suppose we view the matrix  $M$  as being *precomputed*, so that we do not count its evaluation toward the work involved, but we do imagine carrying out (3.43) for a collection of  $k$  values. Then the major work involved will consist of  $N^2$  multiplications because there are  $N$  multiplications in (3.43) for each value of  $m$  and there are  $N$  such values. Of course, additions are also involved, but they are much less expensive in machine time than multiplications, and so they will be ignored. Thus, it would seem that the work involved in computing discrete Fourier transforms scales as  $N^2$ .

One of the celebrated realizations of 20th century computational science is that there are certain *favored* values of  $N$  for which there is a *fast* Fourier transform (FFT) algorithm such that the work scales only as  $N \log_2 N$  rather than  $N^2$ .<sup>11</sup> Although there are other possibilities, these favored values are most commonly taken to be integer powers of 2,  $N = 2^\ell$  for some integer  $\ell$ . The use of such  $N$  values does not cause any great complication because additional equally-spaced points  $z_n$  can be added at both ends of the general interval  $[a, b]$ , and  $f$  can be assigned the value 0 at these additional points. (This is called *padding with zeros*.) So doing does not affect the values of the discrete Fourier transform, nor does it affect  $K_{Ny}$ . It does affect, however, the location of the sampling points  $k_m$  in  $k$  space. [See (3.30) with  $M = N$  and  $K_c = K_{Ny}$ .] In fact, it makes the sampling points more finely spaced, which can be viewed as a virtue.

What can be said about the work involved in computing spline-based Fourier transforms? Examination of the logic involved in the construction of discrete FFT algorithms shows that exactly the same considerations apply to spline-based Fourier transforms. Therefore, for every favored value of  $N$ , there is also a fast spline-based Fourier transform algorithm for which the work also scales as  $N \log_2 N$ . Consequently, there is no computational penalty involved in the use of spline-based Fourier transforms.

## Exercises

**18.3.1.** Verify ....

**18.3.2.** Verify ....

---

<sup>11</sup>We say *realization* rather than *discovery* because, after the extensive FFT work of Danielson, Lanczos, Cooley, Tukey, and others in the mid 20th century, it became clear in retrospect that their remarkable accomplishments had been anticipated earlier by others including Gauss in 1805.

## 18.4 Bessel Functions

Look again at Section 14.2. The computation of on-axis gradients in terms of potential data on the surface of a circular cylinder required the calculation of Bessel functions  $I_m$ , and the equivalent computation in Section 14.3 based on field data required a knowledge of  $I'_m$ . For small values of the argument, say  $w \leq 1$ , these calculations can be done using the series (13.2.15). For larger argument values, and in view of the fact that the Bessel functions are needed for many equally-spaced argument values, it is convenient to compute Bessel functions by integrating the differential equation for  $I_m$  numerically (using the methods of Chapter 2) in the form

$$I''_m(w) + (1/w)I'_m(w) - (m^2/w^2)I_m(w) - I_m(w) = 0. \quad (18.4.1)$$

Here, as initial conditions, we use the series (13.2.15) to evaluate  $I_m$  and  $I'_m$  for  $w = 1$ .

We remark that many of the standard familiar transcendental functions satisfy or are defined by differential equations with good analytic properties. If their values are required at many equally spaced points, these values can often be conveniently and reliably obtained by numerical integration of these differential equations. We will use this method in the next section for Mathieu functions.

## 18.5 Mathieu Functions

This subsection describes briefly tools needed for the computation of Mathieu functions. In 1914 Whittaker remarked about Mathieu functions that

their actual analytical determination presents great difficulties.

He could have said the same thing about their numerical computation. Nearly 100 years later there still do not seem to be any open-source algorithms that are fully robust over the required range of the parameter  $q$ .

### 18.5.1 Calculation of Separation Constants $a_n(q)$ and $b_n(q)$

There is a fast algorithm, based on the use of continued fractions, for the computation of the separation constants  $a_n(q)$  and  $b_n(q)$ . There are also algorithms based on matrix diagonalization. Unfortunately, no routines have yet been found that are completely robust. Samples of the existing routines are listed in Appendix M along with a description of their performance. Although much work has been done on this subject by many authors, there is yet more to be done.

### 18.5.2 Calculation of Mathieu Functions

The calculation of the Mathieu functions themselves is also a delicate matter. As an indication of the difficulty of computing Mathieu functions reliably, we have found that *Mathematica*, useful as it is, also does not compute them accurately for some values of  $q$ . However, for our purposes, since we need them only for a relatively few equally-spaced values of the

arguments  $u$  and  $v$ , they can be obtained (with some care and recognition of their symmetry properties) by direct numerical integration (using the methods of Chapter 2) of their defining differential equations. Thus, for us, the major problem is accurate computation of the separation constants.

Specifically, for the  $\text{ce}_n(v, q)$ , we integrate the equation (14.4.22) over the interval  $[0, 2\pi]$  with the initial conditions

$$Q_n^c(0, q) = 1, \quad (18.5.1)$$

$$Q_n^c'(0, q) = 0. \quad (18.5.2)$$

Here, and in what follows, a  $'$  denotes differentiation with respect to  $v$ . Moreover, the notation  $Q_n^c(v, q)$  indicates that  $q$  is to be computed using (14.4.23), that this resulting  $q$  value is next used to compute  $a = a_n(q)$ , and that these values of  $q$  and  $a$  are then used in (14.4.22). Simultaneously, we integrate the first-order differential equation

$$N'_c(v) = [Q_n^c(v, q)]^2, \quad (18.5.3)$$

again over the interval  $[0, 2\pi]$ , with the initial condition

$$N_c(0) = 0. \quad (18.5.4)$$

Finally, we find  $\text{ce}_n(v, q)$  from the relation

$$\text{ce}_n(v, q) = [\sqrt{\pi/N_c(2\pi)}]Q_n^c(v, q). \quad (18.5.5)$$

In this way we generate a solution of (14.4.22) that is even in  $v$ , and also satisfies the normalization requirement (14.4.38) for  $m = n$ . Finally, we check numerically the periodicity requirements

$$\begin{aligned} Q_n^c(2\pi, q) &= 1, \\ Q_n^c'(2\pi, q) &= 0. \end{aligned} \quad (18.5.6)$$

We require that the relations (5.6) are always satisfied to high precision. So doing provides a check on both the accuracy of the  $a_n(q)$  and the numerical integration procedure. We remark that because of the symmetry conditions described at the end of Section 14.4.4, it is really only necessary to integrate over the interval  $[0, \pi/2]$  and then verify that (4.54) or (4.57) are satisfied. Moreover, if the  $a_n(q)$  are known to be accurate and there are strongly forbidden regions in  $v$ , it is only necessary to integrate over the still smaller interval  $[0, v_{\text{deep}}]$ .

The calculation of the  $\text{se}_n(v, q)$  is done in a similar way. Now we set  $a = b_n(q)$  and integrate (14.4.22) with the initial conditions

$$Q_n^s(0, q) = 0, \quad (18.5.7)$$

$$Q_n^s'(0, q) = 1. \quad (18.5.8)$$

At the same time we again integrate the differential equation

$$N'_s(v) = [Q_n^s(v, q)]^2 \quad (18.5.9)$$

with the initial condition

$$N_s(0) = 0. \quad (18.5.10)$$

Then we define  $\text{se}_n(v, q)$  by the relation

$$\text{se}_n(v, q) = [\sqrt{\pi/N_s(2\pi)}]Q_n^s(v, q). \quad (18.5.11)$$

Finally, we check numerically the periodicity requirements that now

$$\begin{aligned} Q_n^s(2\pi, q) &= 0, \\ Q_n^{s'}(2\pi, q) &= 1. \end{aligned} \quad (18.5.12)$$

We require that (5.12) be satisfied to high precision thereby providing a check on the accuracy of both the  $b_n(q)$  and, as before, the numerical integrator. Again, using symmetry it is really only necessary to integrate over the interval  $[0, \pi/2]$ . And if the  $b_n(q)$  are known to be accurate and there are strongly forbidden regions in  $v$ , it is only necessary to integrate over the still smaller interval  $[0, v_{\text{deep}}]$ .

We still have to describe the computation of  $\text{Ce}_n(u, q)$  and  $\text{Se}_n(u, q)$ . Now we will integrate (14.4.21) numerically. For the case of  $\text{Ce}_n(u, q)$  we find from (14.4.56) the initial condition

$$\text{Ce}_n(0, q) = \text{ce}_n(0, q) = \sqrt{\pi/N_c(2\pi)}. \quad (18.5.13)$$

Here we have also used (5.1) and (5.5). And, since  $\text{Ce}_n(u, q)$  is even in  $u$ , we have the second initial condition

$$\text{Ce}'_n(0, q) = 0. \quad (18.5.14)$$

Thus, given  $k$ , we find  $q$  and  $a = a_n(q)$ . Then, having selected  $U$ , we integrate (14.4.21) over the interval  $u \in [0, U]$  with the initial conditions

$$P_n^c(0, q) = \sqrt{\pi/N_c(2\pi)} \quad (18.5.15)$$

and

$$P_n^{c'}(0, q) = 0. \quad (18.5.16)$$

The result of this process is the value  $\text{Ce}_n(U, q) = P_n^c(U, q)$ .

The computation of  $\text{Se}_n(u, q)$  proceeds similarly. From (14.4.57) we see that there are the initial conditions

$$\text{Se}_n(0, q) = -i\text{se}_n(0, q) = 0, \quad (18.5.17)$$

$$\text{Se}'_n(u, q)|_{u=0} = -i\text{se}'_n(iu, q)|_{u=0}(i) = \text{se}'_n(0, q) = \sqrt{\pi/N_s(2\pi)}. \quad (18.5.18)$$

Here we have used (5.8) and (5.11). Thus, given  $k$ , we find  $q$  and  $a = b_n(q)$ . Then, having selected  $U$ , we integrate (14.4.21) over the interval  $u \in [0, U]$  with the initial conditions

$$P_n^s(0, q) = 0 \quad (18.5.19)$$

and

$$P_n^{s'}(0, q) = \sqrt{\pi/N_s(2\pi)}. \quad (18.5.20)$$

The result of this process is the value  $\text{Se}_n(U, q) = P_n^s(U, q)$ .

### 18.5.3 Calculation of Fourier and Mathieu-Bessel Connection Coefficients

The functions  $\text{ce}_r(v, q)$  and  $\text{se}_r(v, q)$  are periodic with period  $2\pi$  and therefore have Fourier expansions in terms of the functions  $\cos(mv)$  and  $\sin(mv)$ . See (14.4.52) through (14.4.55). As shown in Appendix N, the Fourier coefficients that appear in these expansions, which depend on  $q$ , are key to computing the Mathieu-Bessel connection coefficients  $\alpha_m^r(k)$  and  $\beta_m^r(k)$ . See (14.4.78) and (14.4.79). In this subsection we describe how these Fourier coefficients can be computed numerically.

Let us begin with the functions  $\text{ce}_r(v, q)$ . Since they are periodic and even, they have Fourier expansions of the form

$$\text{ce}_r(v, q) = \sum_{m=0}^{\infty} A_m^r(q) \cos(mv). \quad (18.5.21)$$

There are known algorithms for the computation of the Fourier coefficients  $A_m^r(q)$  but, as was the case with those for the computation of the  $\text{ce}_r(v, q)$ , we have found that they are not robust. However, by the orthogonality property of the trigonometric functions, it follows that

$$\begin{aligned} A_0^r(q) &= [1/(2\pi)] \int_0^{2\pi} dv \text{ce}_r(v, q), \\ A_m^r(q) &= (1/\pi) \int_0^{2\pi} dv \text{ce}_r(v, q) \cos(mv) \text{ for } m \geq 1. \end{aligned} \quad (18.5.22)$$

By (5.5) we may also write

$$\begin{aligned} A_0^r(q) &= [1/(2\pi)][\sqrt{\pi/N_c(2\pi)}] \int_0^{2\pi} dv Q_r^c(v, q), \\ A_m^r(q) &= (1/\pi)[\sqrt{\pi/N_c(2\pi)}] \int_0^{2\pi} dv Q_r^c(v, q) \cos(mv) \text{ for } m \geq 1. \end{aligned} \quad (18.5.23)$$

Let  $\hat{A}_m^r(v, q)$  be the functions defined for various values of  $m$  and  $r$  by the differential equations

$$\begin{aligned} \hat{A}_0^r'(v, q) &= (1/2)Q_r^c(v, q), \\ \hat{A}_m^r'(v, q) &= Q_r^c(v, q) \cos(mv) \text{ for } m \geq 1 \end{aligned} \quad (18.5.24)$$

with the common initial conditions

$$\hat{A}_m^r(0, q) = 0. \quad (18.5.25)$$

The differential equations (5.24) can be integrated numerically over the interval  $[0, 2\pi]$  simultaneously with those for the  $Q_r^c(v, q)$  and (5.3). Then we find that the  $A_m^r(q)$  are given by the relations

$$A_m^r(q) = (1/\pi)[\sqrt{\pi/N_c(2\pi)}]\hat{A}_m^r(2\pi, q). \quad (18.5.26)$$

Similarly, for the functions  $\text{se}_r(v, q)$ , there are Fourier expansions of the form

$$\text{se}_r(v, q) = \sum_{m=1}^{\infty} B_m^r(q) \sin(mv). \quad (18.5.27)$$

And, again by the orthogonality property of the trigonometric functions, it follows that

$$B_m^r(q) = (1/\pi) \int_0^{2\pi} dv \text{se}_r(v, q) \sin(mv). \quad (18.5.28)$$

With the aid of (5.11) this relation can also be written in the form

$$B_m^r(q) = (1/\pi)[\sqrt{\pi/N_s(2\pi)}] \int_0^{2\pi} dv Q_r^s(v, q) \sin(mv). \quad (18.5.29)$$

Now let  $\hat{B}_m^r(v, q)$  be the functions defined for various values of  $m$  and  $r$  by the differential equations

$$\hat{B}_m^r'(v, q) = Q_r^s(v, q) \sin(mv) \quad (18.5.30)$$

with the common initial conditions

$$\hat{B}_m^r(0, q) = 0. \quad (18.5.31)$$

Now we find that the  $B_m^r(q)$  are given by the relations

$$B_m^r(q) = (1/\pi)[\sqrt{\pi/N_s(2\pi)}]\hat{B}_m^r(2\pi, q). \quad (18.5.32)$$

At this point we remark that the functions  $\cos(mv)$  and  $\sin(mv)$  required to integrate the differential equations (5.24) and (5.30), as well as the equations of the form (14.4.22) for the  $Q_r^c(v, q)$  and the  $Q_r^s(v, q)$ , can also be computed on the fly by simultaneously integrating numerically the differential equations for the trigonometric functions. The needed hyperbolic functions in (14.4.21) can be calculated analogously. So doing is faster than using the built-in Fortran or *C* functions for the trigonometric functions.

# Bibliography

## Splines

- [1] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, Third Edition, Springer (2002).
- [2] F.B. Hildebrand, *Introduction to Numerical Analysis*, Second Edition, Dover (1987).
- [3] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in Fortran 77*, Second Edition, Cambridge (2003).

## Discrete Fourier Transforms

- [4] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in Fortran 77*, Second Edition, Cambridge (2003).
- [5] C.F. Van Loan, *Computational Frameworks for the Fast Fourier Transform*, SIAM (1992).
- [6] A. Zonst, *Understanding the FFT*, Citrus Press (1995).
- [7] R.E. Crandall, *Projects in Scientific Computation*, Springer-Verlag (2000).
- [8] R.E. Crandall, *Topics in Advanced Scientific Computation*, Springer-Telos (1996).
- [9] H. Joseph Weaver, *Theory of Discrete and Continuous Fourier Analysis*, John Wiley (1989).
- [10] E. Stein and R. Shakarchi, *Fourier Analysis: An Introduction*, Princeton University Press (2003).
- [11] B. P. Lathi, *Signal Processing & Linear Systems*, Berkeley-Cambridge Press (1998).  
See also [http://web.itu.edu.tr/hulyayalcin/Signal\\_Processing\\_Books/Lathi\\_Signal\\_Processing\\_and\\_Linear\\_Systems.pdf](http://web.itu.edu.tr/hulyayalcin/Signal_Processing_Books/Lathi_Signal_Processing_and_Linear_Systems.pdf).



# Chapter 19

## Numerical Benchmarks

How well do the surface methods of Chapter 14 work? To test them we need a problem that is sufficiently complex and challenging to fully exercise the numerical algorithms while at the same time being exactly soluble. In that way we will be able to gauge the accuracy of the numerical methods by comparing numerical results with exact analytic results. Such a test problem is that of the magnetic monopole doublet described in Section 13.7. For this problem we will assume that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm)<sup>2</sup>. See Section 13.7.1.

### 19.1 Circular Cylinder Numerical Results for Monopole Doublet

In this section we will apply the circular cylinder numerical method of Section 14.3 to the monopole doublet problem to investigate how accurately this method is able to reproduce the exact analytic results for the on-axis gradients found in Section 13.7.2.<sup>1</sup> For our benchmark calculation we will employ a cylinder with radius  $R = 2$  cm. See Figure 13.7.1. We will work up to the desired numerical comparison by stages. In this way we will be able to judge the accuracy of various intermediate steps.

Observe that the integrands of (14.3.8) and (14.3.23), apart from multiplicative constants, consist of the *product* of a *kernel* [ $k^{n+|m|-1}/I'_m(kR)$ ] and the *Fourier coefficients* [ $\tilde{\tilde{B}}_\rho(R, m', k')$ ] or [ $\tilde{\tilde{B}}_\rho^\alpha(R, m', k')$ ]. The kernels are *universal* (the same for all problems) and the Fourier coefficients are specific to each problem. In what follows we will be examining both. For convenience, we will use the Fourier coefficients [ $\tilde{\tilde{B}}_\rho^\alpha(R, m', k')$ ].

#### 19.1.1 Testing the Spline-Based Inverse ( $k \rightarrow z$ ) Fourier Transform

Suppose we know exactly the Fourier coefficients  $\tilde{\tilde{B}}_\rho^\alpha(R, m', k')$  as given by (14.3.14). We will see in the next two paragraphs that, for the case of the monopole doublet, the  $\tilde{\tilde{B}}_\rho^\alpha(R, m', k')$  can indeed be found exactly. We can insert these exact quantities into (14.3.23), and then

---

<sup>1</sup>A similar study could be made of the accuracy of the related method of Section 14.2, with analogous results.

perform the required integration numerically using the methods described in Section 15.3.5. In this way we will be able to test the accuracy of the numerical routines for  $I'_m(w)$  and for spline-based inverse Fourier transforms. This test will be performed in later paragraphs.

### Exact Fourier coefficients

To find the  $\tilde{B}_\rho^\alpha(R, m', k')$  exactly, suppose the  $C_{m,\alpha}^{[0]}(z)$  have the Fourier representation

$$C_{m,\alpha}^{[0]}(z) = \int_{-\infty}^{\infty} dk \tilde{C}_{m,\alpha}^{[0]}(k) \exp(ikz). \quad (19.1.1)$$

From (14.3.23) evaluated with  $n = 0$  we have the result

$$C_{m,\alpha}^{[0]}(z) = (1/2)^m (1/m!) \int_{-\infty}^{\infty} dk [k^{m-1}/I'_m(kR)] \tilde{B}_\rho^\alpha(R, m, k) \exp(ikz). \quad (19.1.2)$$

It follows from the uniqueness of the Fourier representation that there are the relations

$$\tilde{C}_{m,\alpha}^{[0]}(k) = (1/2)^m (1/m!) [k^{m-1}/I'_m(kR)] \tilde{B}_\rho^\alpha(R, m, k), \quad (19.1.3)$$

which can be solved for the  $\tilde{B}_\rho^\alpha(R, m, k)$  to give the relations

$$\tilde{B}_\rho^\alpha(R, m, k) = 2^m (m!) [I'_m(kR)/(k)^{m-1}] \tilde{C}_{m,\alpha}^{[0]}(k). \quad (19.1.4)$$

Of course, in view of (1.1), the  $\tilde{C}_{m,\alpha}^{[0]}(k)$  are also given by the inverse Fourier transform

$$\tilde{C}_{m,\alpha}^{[0]}(k) = [1/(2\pi)] \int_{-\infty}^{\infty} dz C_{m,\alpha}^{[0]}(z) \exp(-ikz). \quad (19.1.5)$$

Therefore, from (13.7.9) and (13.7.33), we conclude that for the monopole doublet there are the results

$$\tilde{C}_{m=0}^{[0]}(k) = 0, \quad (19.1.6)$$

$$\tilde{C}_{m,c}^{[0]}(k) = 0, \quad (19.1.7)$$

$$\tilde{C}_{m,s}^{[0]}(k) = 0 \text{ for } m \text{ even}, \quad (19.1.8)$$

and the only nonzero integrals on the right side of (1.5) are of the form

$$\int_{-\infty}^{\infty} dz C_{m,s}^{[0]}(z) \exp(-ikz) \text{ with } m \text{ odd}. \quad (19.1.9)$$

From (13.7.33) we see that we need the integrals

$$\int_{-\infty}^{\infty} dz \exp(-ikz) \beta^{2m+1}(z) \text{ for } m \text{ odd}. \quad (19.1.10)$$

These integrals can be done analytically, and have the values

$$\int_{-\infty}^{\infty} dz \exp(-ikz) \beta^{2m+1}(z) = \{2/[1 \cdot 3 \cdot 5 \cdot 7 \cdots (2m-1)]\} a^{m+1} |k|^m K_m(a|k|). \quad (19.1.11)$$

Putting everything together gives the final result that the only nonzero  $\tilde{\tilde{B}}_\rho^\alpha(R, m, k)$  are those given by the relations

$$\tilde{\tilde{B}}_\rho^s(R, m, k) = (-4g/\pi)(-1)^{(m-1)/2}|k|I'_m(kR)K_m(a|k|) \text{ for } m \text{ odd.} \quad (19.1.12)$$

We observe that the  $\tilde{\tilde{B}}_\rho^s(R, m, k)$  for the monopole doublet are pure real and are even functions of  $k$ . The imaginary part vanishes because, for the monopole doublet,  $B_\rho(R, \phi, z)$  is an even function of  $z$ . See (13.7.8) and (14.3.14) through (14.3.16).

As examples of the behavior of the  $\tilde{\tilde{B}}_\rho^s(R, m, k)$ , Figures 1.1 and 1.2 show the functions  $\Re\tilde{\tilde{B}}_\rho^s(R, 1, k)$  and  $\Im\tilde{\tilde{B}}_\rho^s(R, 7, k)$ . As already described,  $I'_m(kR)$  grows exponentially at infinity. See (14.3.7). By contrast,  $K_m(a|k|)$  decays exponentially to zero at infinity,

$$|K_m(a|k|)| \sim \exp(-|k|a)(\pi)^{1/2}/\sqrt{2|k|a} \text{ as } |k| \rightarrow \infty. \quad (19.1.13)$$

Since  $a > R$ , it follows that the  $\tilde{\tilde{B}}_\rho^s(R, m, k)$  are exponentially damped at infinity,

$$|\tilde{\tilde{B}}_\rho^s(R, m, k)| \sim \exp[-|k|(a - R)] \text{ as } |k| \rightarrow \infty. \quad (19.1.14)$$

The function  $I'_m(kR)$  is an entire function of  $k$ . The function  $K_m(a|k|)$  is singular at the origin. Analysis reveals that the product  $|k|I'_m(kR)K_m(a|k|)$  is finite at the origin and has additionally a logarithmic singularity at the origin of the form  $|k|^{2m} \log |k|$ .<sup>2</sup> Thus the  $\tilde{\tilde{B}}_\rho^s(R, m, k)$ , with  $m \geq 1$ , are finite for all  $k$  and vanish exponentially at infinity.

## Kernels

Since, as emphasized earlier, the integrand in (14.3.23) is the product of a kernel and a Fourier coefficient, we should also examine the kernels. Figures 1.3 through 1.5 display the kernels  $[k^{n+m-1}/I'_m(kR)]$  for the representative cases  $(m, n)=(1,0)$ ,  $(1,6)$ , and  $(7,0)$  when  $R = 2$  cm. We see, as expected, that they fall off rapidly for large  $|k|$ . The intermediate cases give analogous results.

## Exact Integrands

We are now ready to evaluate the integrals (14.3.23) to find the  $C_{m,s}^{[n]}(z)$ . Upon inserting (1.12) into (14.3.23), we find the results

$$C_{m,s}^{[n]}(z) = \left\{ \frac{-4g(-1)^{(m-1)/2}}{\pi 2^m(m!)} \right\} \int_{-\infty}^{\infty} dk \exp(ikz)(ik)^n |k|^m K_m(|k|a). \quad (19.1.15)$$

Or, more directly and as an algebraic check, we may use (1.5) to yield the result

$$\begin{aligned} \tilde{C}_{m,s}^{[0]}(k) &= [1/2\pi] \int_{-\infty}^{\infty} dz C_{m,s}^{[0]}(z) \exp(-ikz) \\ &= \{-4g/[\pi 2^m(m!)]\}(-1)^{(m-1)/2} |k|^m K_m(|k|a). \end{aligned} \quad (19.1.16)$$

---

<sup>2</sup>It can be shown that this singularity in  $\tilde{\tilde{B}}_\rho^s(R, m, k)$  is related to the  $\sim 1/|z|^3$  behavior of  $B_\rho(R, \phi, z)$  and the  $\sim 1/|z|^{2m+1}$  behavior of the  $C_{m,s}^{[0]}(z)$  for large  $|z|$ . See (13.7.8) and (13.7.34). In general, the faster the falloff at  $|z| = \infty$ , the milder the singularity at  $k = 0$ .

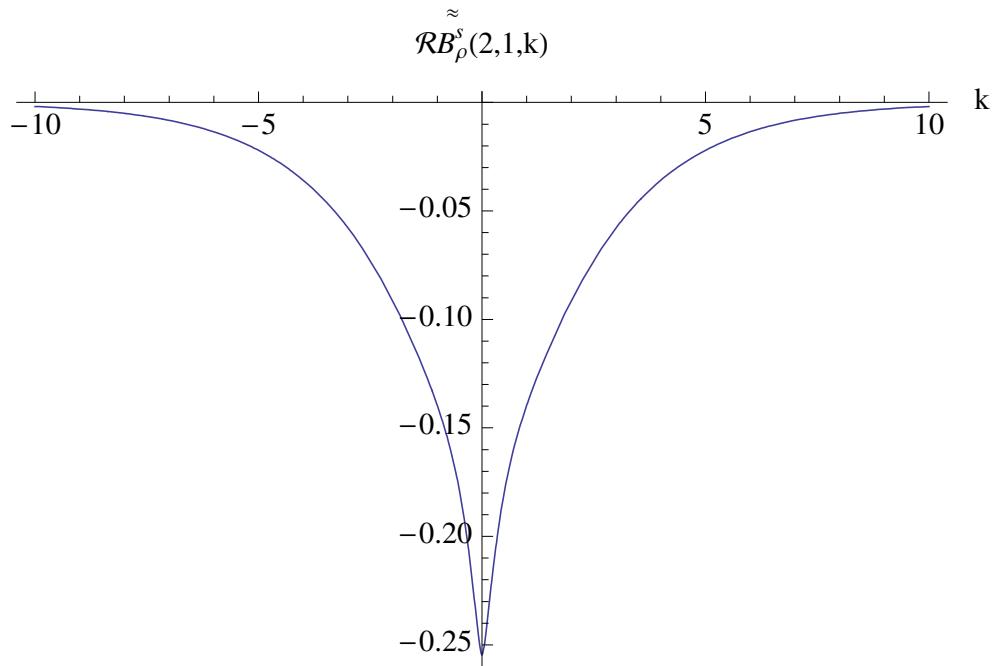


Figure 19.1.1: The real part of  $\tilde{\mathcal{R}}B_\rho^s(R, 1, k)$  as a function of  $k$  for the monopole doublet in the case that  $R = 2$  cm and  $a = 2.5$  cm. The imaginary part vanishes.

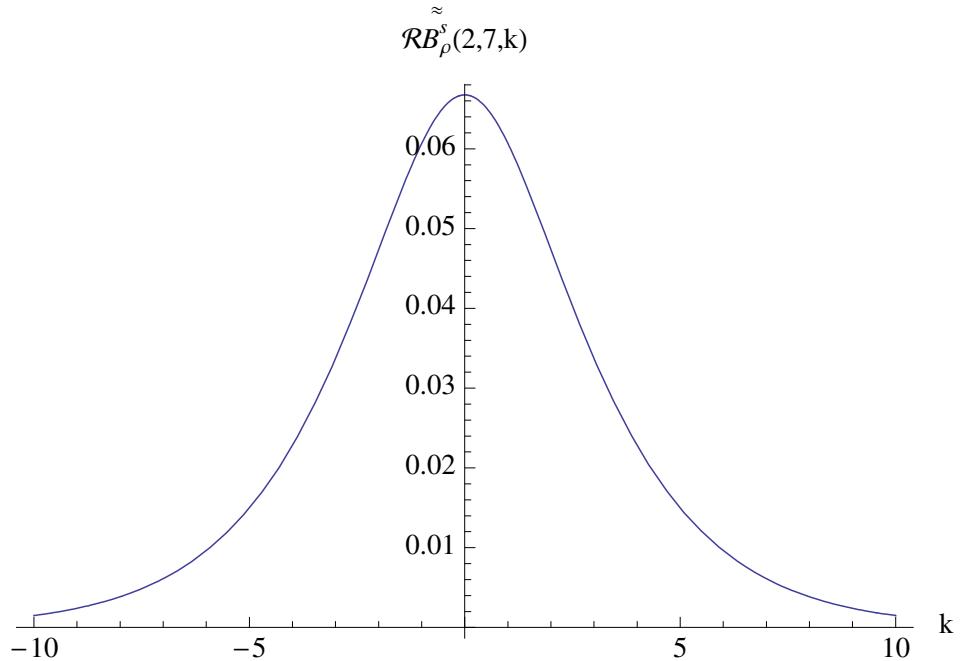


Figure 19.1.2: The real part of  $\tilde{\mathcal{R}}B_\rho^s(R, 7, k)$  as a function of  $k$  for the monopole doublet in the case that  $R = 2$  cm and  $a = 2.5$  cm. The imaginary part vanishes.

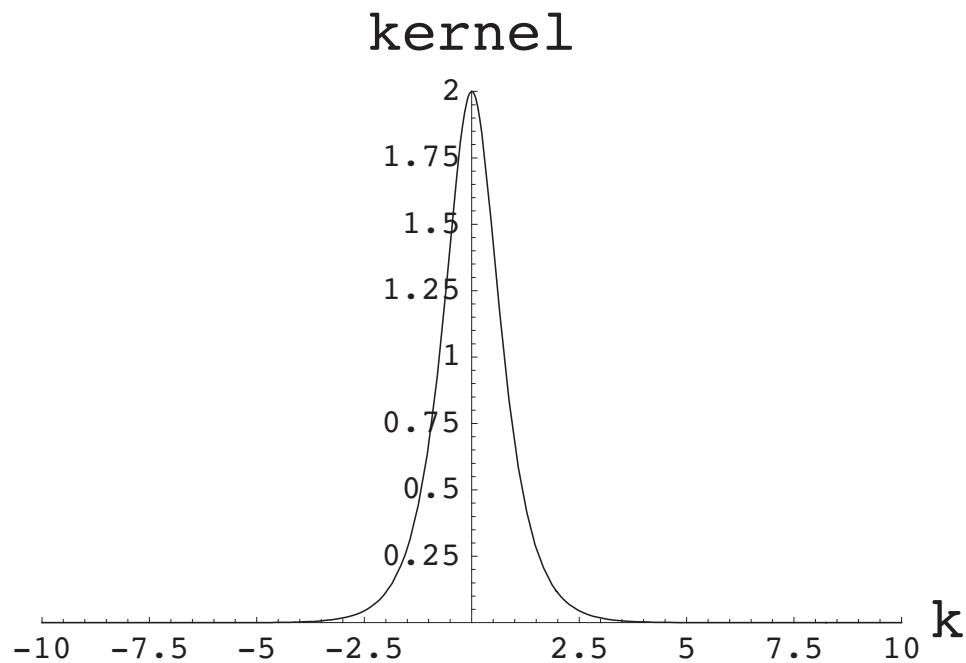


Figure 19.1.3: The kernel  $[k^{n+m-1}/I'_m(kR)]$  as a function of  $k$  in the case that  $m = 1$ ,  $n = 0$ , and  $R = 2$  cm.

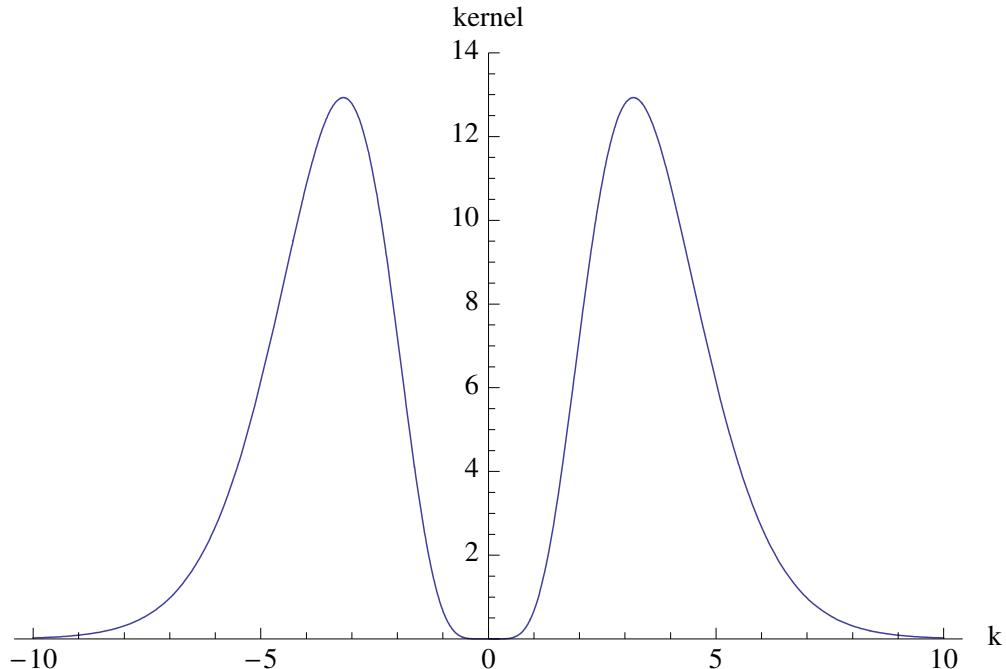


Figure 19.1.4: The kernel  $[k^{n+m-1}/I'_m(kR)]$  as a function of  $k$  in the case that  $m = 1$ ,  $n = 6$ , and  $R = 2$  cm.

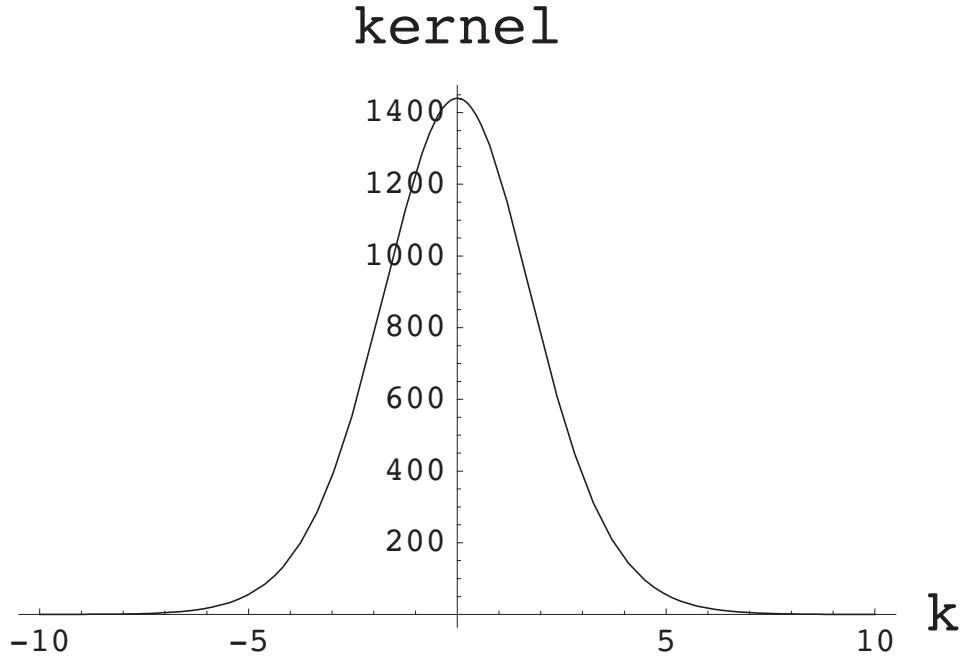


Figure 19.1.5: The kernel  $[k^{n+m-1}/I'_m(kR)]$  as a function of  $k$  in the case that  $m = 7$ ,  $n = 0$ , and  $R = 2$  cm.

Note that the nonzero  $\tilde{C}_{m,s}^{[0]}(k)$  for the monopole doublet are pure real because  $C_{m,\alpha}^{[0]}(z)$  is real and, for the monopole doublet case, is an even function of  $z$ . Correspondingly, the  $\tilde{C}_{m,s}^{[0]}(k)$  are even functions of  $k$ . Now, more compactly, we may write

$$C_{m,s}^{[n]}(z) = \int_{-\infty}^{\infty} dk (ik)^n \tilde{C}_{m,s}^{[0]}(k) \exp(ikz). \quad (19.1.17)$$

It is these integrals that we want to evaluate using spline-based inverse Fourier transforms in order to illustrate and verify the accuracy of this numerical method.

The integrands that are expected to be the hardest to integrate accurately are those for the extreme case  $\{m = 1, n = 0\}$  because of its  $|k|^2 \log |k|$  singularity at the origin and the extreme case  $\{m = 1, n = 6\}$  because of its oscillatory behavior. These integrands, which are required to compute  $C_{1,s}^{[0]}(z)$  and  $C_{1,s}^{[6]}(z)$ , respectively, are shown in Figures 1.6 and 1.7.

As expected, the integrands fall off rapidly for large  $|k|$  because, as seen earlier, both the Fourier coefficients and the kernels fall off rapidly for large  $|k|$ . Indeed, inspection of Figures 1.6 and 1.7 shows that the integrands have effectively fallen to zero when  $|k| > 10$ . Therefore, and in order to be conservative, we will evaluate the integrals (14.3.23) using the somewhat larger cutoff  $K_c = 20$ . Moreover, we will use 401-point spline fits to the integrands over the interval  $k \in [-K_c, K_c]$  so that  $H = .1$ . See (15.3.30).

### Spline-Based Inverse Fourier Transform Results

Figures 1.8 and 1.9 show the functions  $C_{1,s}^{[0]}(z)$  and  $C_{1,s}^{[6]}(z)$  obtained in this way as well as the exact results given by (13.7.33) and its derivatives. Figures 1.10 and 1.11 show for these

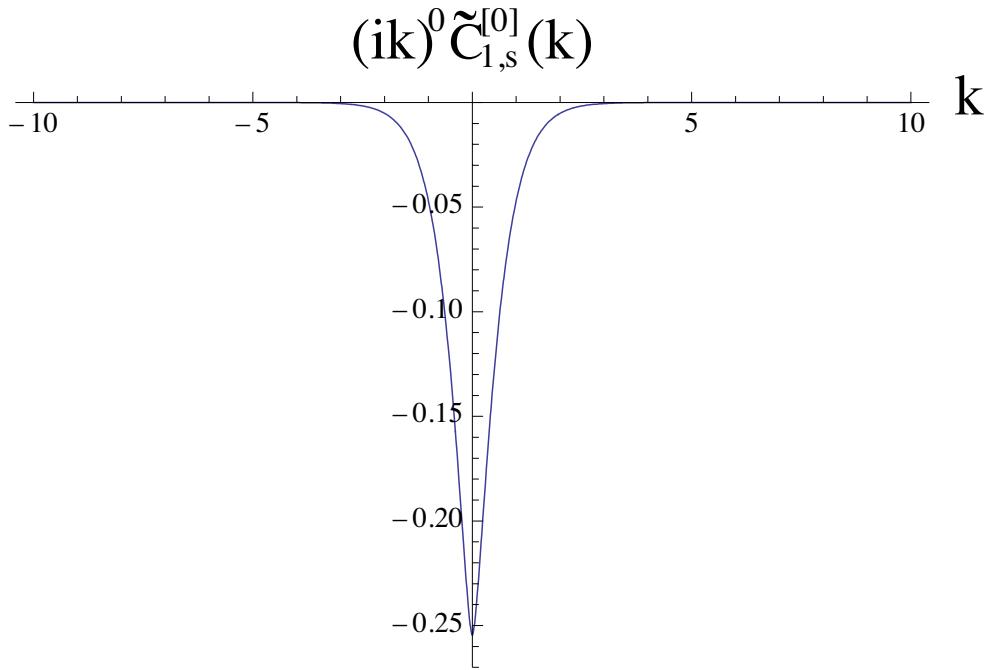


Figure 19.1.6: The integrand  $(ik)^n \tilde{C}_{m,s}^{[0]}(k)$  for  $m = 1$  and  $n = 0$  as a function of  $k$  in the case that  $R = 2$  cm. It is required to compute  $C_{1,s}^{[0]}(z)$ .

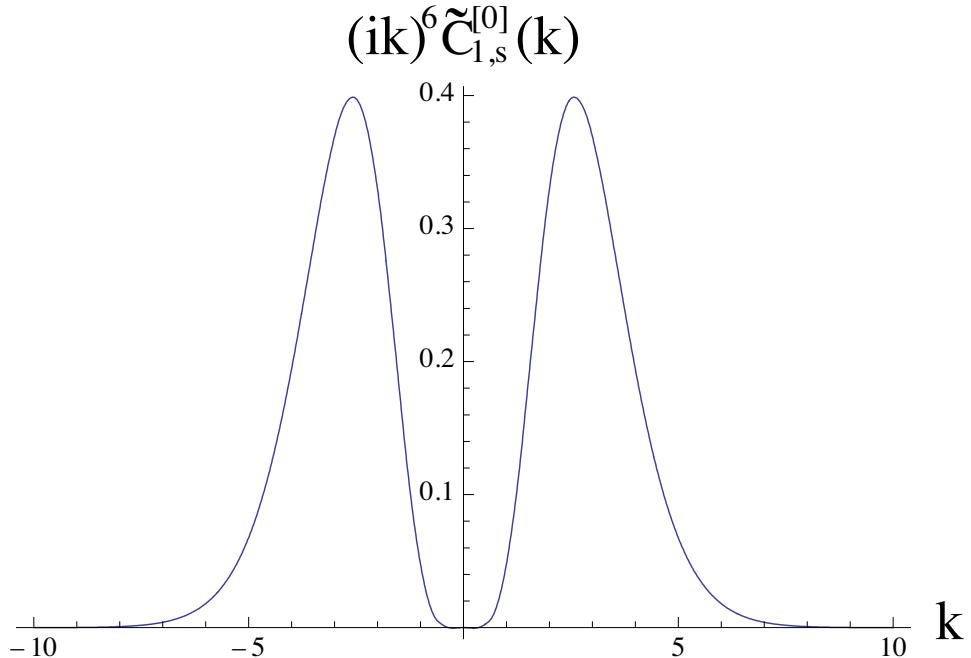


Figure 19.1.7: The integrand  $(ik)^n \tilde{C}_{m,s}^{[0]}(k)$  for  $m = 1$  and  $n = 6$  as a function of  $k$  in the case that  $R = 2$  cm. It is required to compute  $C_{1,s}^{[6]}(z)$ .

cases the differences between the exact and numerical results. Evidently the worst case is that of  $C_{1,s}^{[0]}(z)$ , and the error in this case is approximately 1.6 parts in  $10^4$ . Further numerical study shows that the error is smaller for the other  $C_{m,s}^{[n]}(z)$  listed in (13.7.35). Finally, further numerical study shows that the error can be made even smaller by increasing the number of points in the spline fit.<sup>3</sup> In this context we remark that the error in  $C_{1,s}^{[0]}(z)$  decreases only as  $H^3$  while the error in the other  $C_{m,s}^{[n]}(z)$  decreases as  $H^4$ . This behavior is due to the  $|k|^2 \log |k|$  singularity in  $\tilde{C}_{1,s}^{[0]}(k)$  so that  $\tilde{C}_{1,s}^{[0]}(k)$  does not have a finite fourth derivative at the origin. [See (15.1.34).] The integrands  $(ik)^n \tilde{C}_{m,s}^{[0]}(k)$  for other values of  $m, n$  are less singular at the origin.

We conclude that if the integrand is computed to high precision (exactly, in this case), the spline-based inverse Fourier transform gives accurate results.

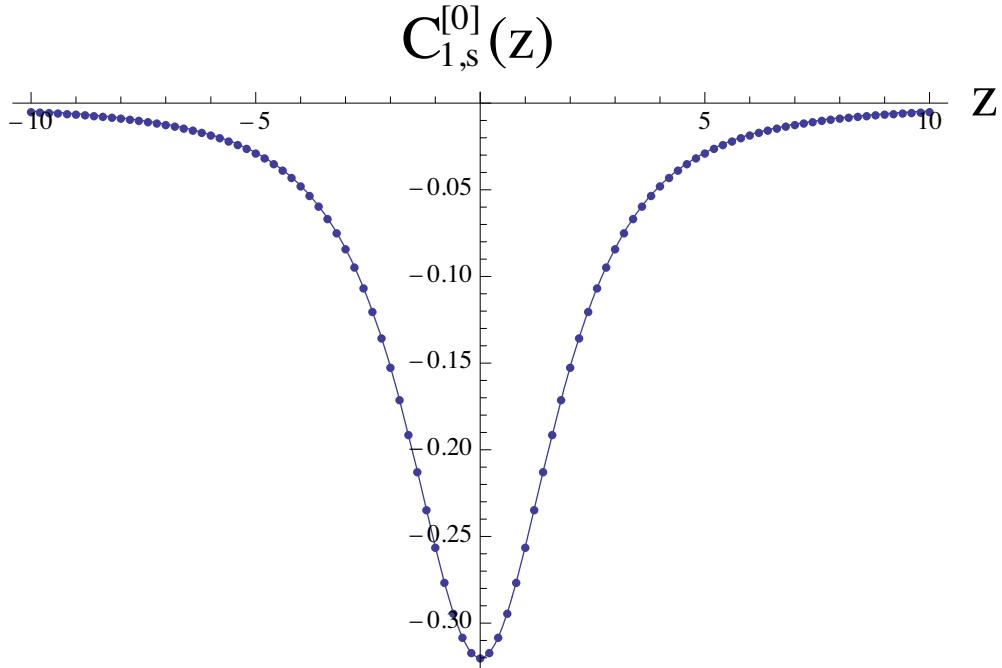


Figure 19.1.8: Exact and numerical results for  $C_{1,s}^{[0]}(z)$ . Exact results are shown as a solid line (see Figure 13.7.8), and numerical results are shown as dots.

### 19.1.2 Testing the Forward ( $z \rightarrow k$ ) and ( $\phi \rightarrow m$ ) Fourier Transforms

Having verified the accuracy of the spline-based inverse Fourier transform for carrying out the integration (14.3.23), let us also use splines to find  $\tilde{\tilde{B}}_\rho(R, m', k')$ . The radial component  $B_\rho(\rho, \phi, z)$  of the magnetic field  $\mathbf{B}$  is given by (13.7.8) and its behavior is displayed in Figures 13.7.6 and 13.7.7. According to (14.3.1), calculating  $\tilde{\tilde{B}}_\rho(R, m', k')$  requires both a Fourier

<sup>3</sup>The value of  $K_c$  can also be increased. However, this does not seem to be necessary unless much higher accuracy is required.

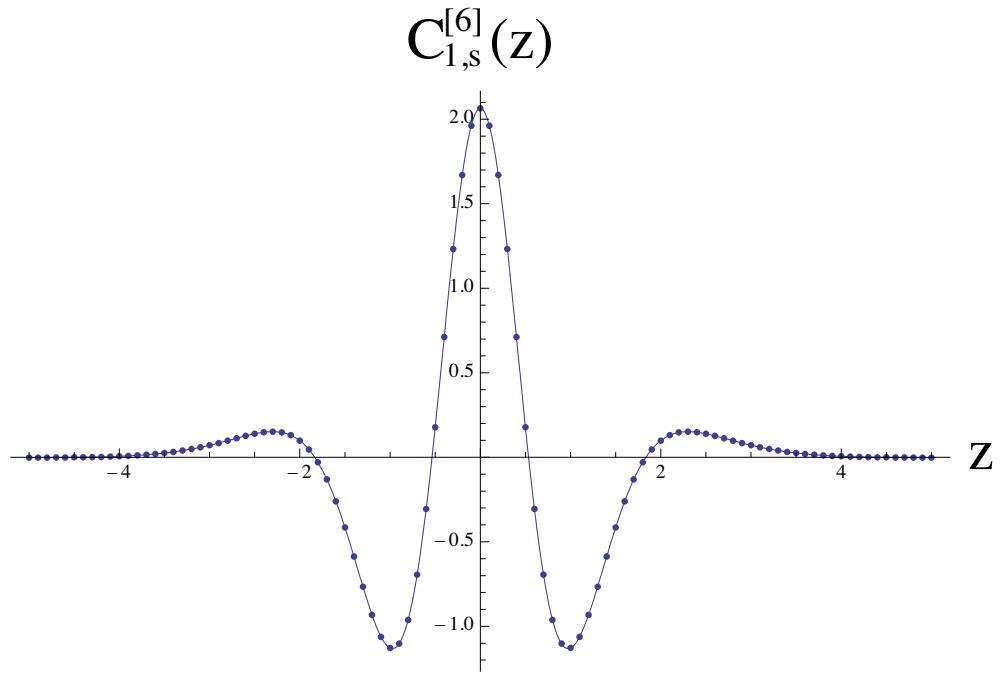


Figure 19.1.9: Exact and numerical results for  $C_{1,s}^{[6]}(z)$ . Exact results are shown as a solid line (see Figure 13.7.9), and numerical results are shown as dots.

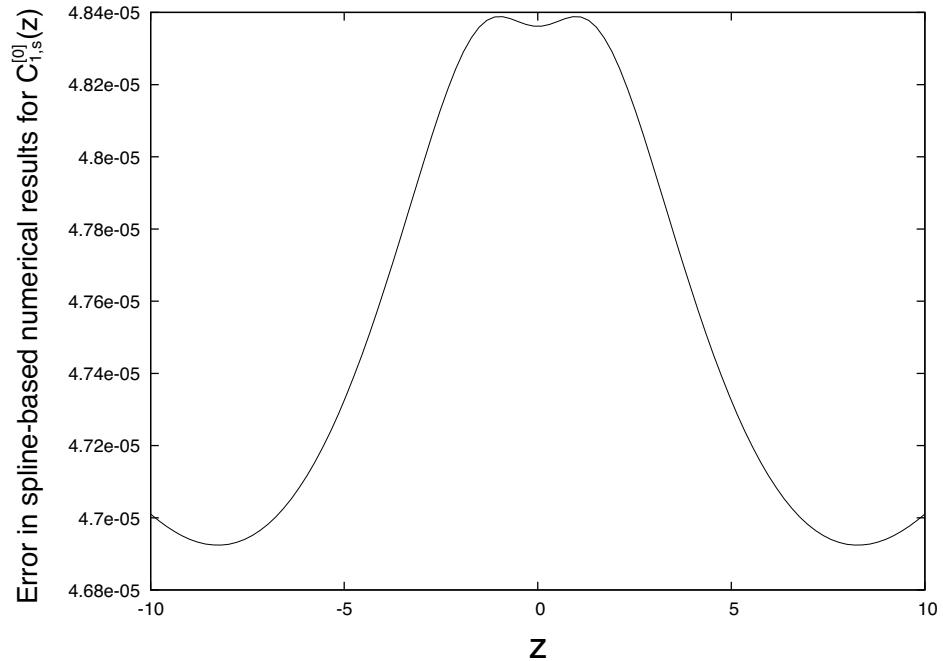


Figure 19.1.10: Difference between exact and spline-based numerical results for  $C_{1,s}^{[0]}(z)$  using an exact integrand in (14.3.23).

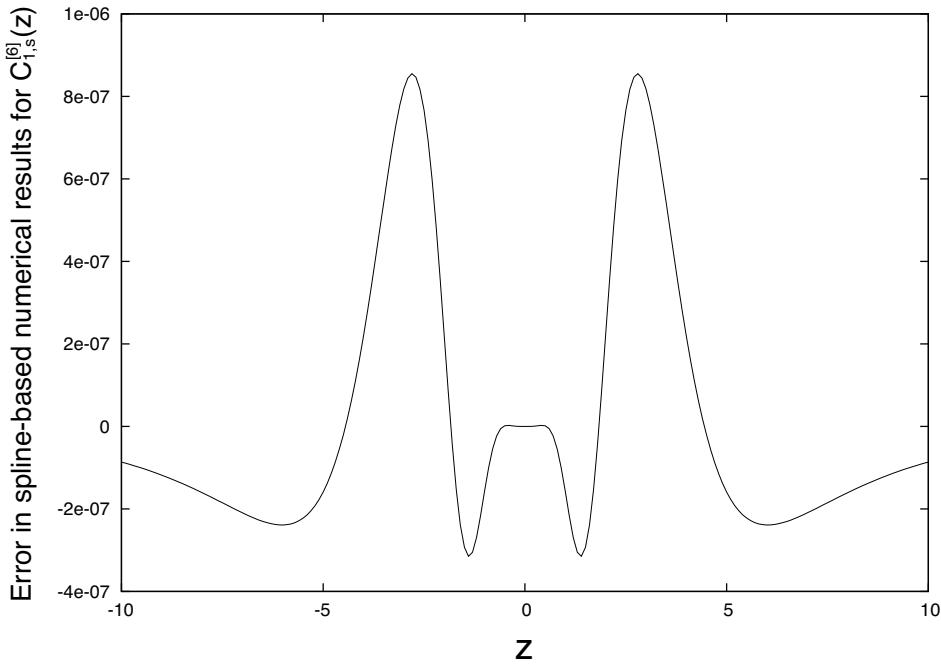


Figure 19.1.11: Difference between exact and spline-based numerical results for  $C_{1,s}^{[6]}(z)$  using an exact integrand in (14.3.23).

transform involving integration over  $z$ , and the computation of Fourier coefficients involving integration over  $\phi$ . The integrations may be performed in either order. Thus, we may define  $\tilde{B}_\rho(R, \phi, k')$  by writing

$$\tilde{B}_\rho(R, \phi, k') = [1/(2\pi)] \int_{-\infty}^{\infty} dz \exp(-ik'z) B_\rho(R, \phi, z), \quad (19.1.18)$$

and then obtain  $\tilde{\tilde{B}}_\rho(R, m', k')$  from the relation

$$\tilde{\tilde{B}}_\rho(R, m', k') = [1/(2\pi)] \int_0^{2\pi} d\phi \exp(-im'\phi) \tilde{B}_\rho(R, \phi, k'). \quad (19.1.19)$$

### Performing the Forward ( $z \rightarrow k$ ) Fourier Transform

Let us first attack the problem of evaluating (1.18). Based on inspection of Figure 13.7.7, we might imagine that we could safely cut off the Fourier transform integral (1.18) by setting the integrand to zero for  $|z| > Z_c$  with  $Z_c = 10$ . However, in this case looks are deceiving. Examine (13.7.8) for the case  $\phi = \pi/2$ . The two terms within the first set of curly brackets tend to cancel for large  $|z|$ . The two terms within the second set of curly brackets do not. The term within the second set of curly brackets that falls off most slowly for large  $|z|$  is given by the relation

$$\text{term with slowest falloff} = -ga[z^2 + (R + a)^2]^{-3/2} = -ga[z^2 + b^2]^{-3/2} \quad (19.1.20)$$

where

$$b = R + a. \quad (19.1.21)$$

If we are interested in evaluating  $\tilde{B}_\rho(R, \phi, k)$  accurately for  $\phi = \pi/2$  and  $k = 0$ , then we must make the comparison

$$\int_{-\infty}^{\infty} dz [z^2 + b^2]^{-3/2} \text{ versus } \int_{-Z_c}^{Z_c} dz [z^2 + b^2]^{-3/2}. \quad (19.1.22)$$

These two integrals have the values

$$\int_{-\infty}^{\infty} dz [z^2 + b^2]^{-3/2} = 2/b^2, \quad (19.1.23)$$

$$\int_{-Z_c}^{Z_c} dz [z^2 + b^2]^{-3/2} = (2/b^2) \{Z_c/[Z_c^2 + b^2]^{-1/2}\}. \quad (19.1.24)$$

Therefore the *fractional error* involved in imposing a cutoff is given by the relation

$$\text{fractional error} = 1 - Z_c/[Z_c^2 + b^2]^{-1/2} \approx (1/2)(b/Z_c)^2. \quad (19.1.25)$$

Suppose we are willing to accept a fractional error on the order of  $10^{-4}$ . Then we must have

$$Z_c \approx 100b/\sqrt{2}. \quad (19.1.26)$$

For  $a = 2.5$  and  $R = 2$ , which yields  $b = 4.5$ , this means that in reality we must have

$$Z_c \approx 300. \quad (19.1.27)$$

We remark that this is a worst case, where fringe fields fall off only as  $1/|z|^3$ . For cases where the fringe fields fall off more rapidly (e.g. quadrupoles, higher-order multipoles, dipoles with field clamps, etc.) the cutoff in  $z$  can be smaller.

### Performing the Forward ( $\phi \rightarrow m$ ) Fourier Transform

We are ready to carry out the spline-based calculation of  $\tilde{B}_\rho(R, m', k')$ . Let us select 4801 equally-spaced points  $z_j$  in the interval  $z \in [-Z_c, Z_c]$  with  $Z_c = 300$ , and let us select 49 equally-spaced values  $\phi_\ell$  in the interval  $\phi \in [0, 2\pi]$ . Then, for each  $\phi_\ell$ , we carry out a 4801-point (in  $z$ ) spline-based Fourier transform to find (for  $R = 2$ ) the quantities  $\tilde{B}_\rho(R, \phi_\ell, k')$ . Next we evaluate (1.19) using a 49-point (in  $\phi$ ) Riemann sum discrete angular Fourier transform to obtain  $\tilde{B}_\rho(R, m', k')$ . See Exercise 1.2 for an explanation of why 49 points should be adequate and, indeed, give good accuracy.

### Spline-Based Forward and Inverse Fourier Transform Results

As a last step in this part of our exercise, let us use the integrands based on this  $\tilde{B}_\rho(R, m', k')$  to carry out the same spline-based inverse Fourier transform described earlier. That is, we no longer work with exact integrands in (14.3.6), but rather use approximate integrands based on spline-based and discrete integrations over  $z$  and  $\phi$ . However, we still do use exact values of  $B_\rho(R, \phi, z)$  on the cylinder. The result of this process is an *almost completely* numerically calculated set of functions  $C_{m,\alpha}^{[n]}(z)$ . Examination of these numerically calculated functions

shows that they also well approximate the exact functions  $C_{m,\alpha}^{[n]}(z)$ . For example, Figures 1.12 and 1.13 show the differences between the exact and numerically calculated  $C_{1,s}^{[0]}(z)$  and  $C_{1,s}^{[6]}(z)$ . We see that Figure 1.12 resembles Figure 1.10. Surprisingly, the error in  $C_{1,s}^{[0]}(z)$  is now slightly less than before, but remains approximately 1.6 parts in  $10^4$ . Apparently in this case the errors involved in the approximate computation of  $\tilde{B}_\rho(R, m', k')$  cancel to some extent the errors involved in the spline-based inverse Fourier transform. We also see that Figure 1.13 somewhat resembles Figure 1.11, but now the error in  $C_{1,s}^{[6]}(z)$  is approximately 7 parts on  $10^5$  whereas it was 5 parts in  $10^7$  in Figure 1.11. In this case the errors involved in the approximate computation of  $\tilde{B}_\rho(R, m', k')$  add to the overall error. (However the overall error is still acceptably small.) Indeed, we find that if we compute  $\tilde{B}_\rho(R, m', k')$  more exactly by increasing the number of points in  $\phi$  beyond 49, increasing  $Z_c$  beyond 300, and increasing the number of points in  $z$  beyond 4801, then Figure 1.12 morphs into Figure 1.10, and Figure 1.13 morphs into Figure 1.11. With regard to the errors for the other nonzero  $C_{m,s}^{[n]}(z)$ , we find that they are also acceptably small. Finally, we find that the  $C_{m,\alpha}^{[0]}(z)$  that should vanish are, in fact, numerically very small.

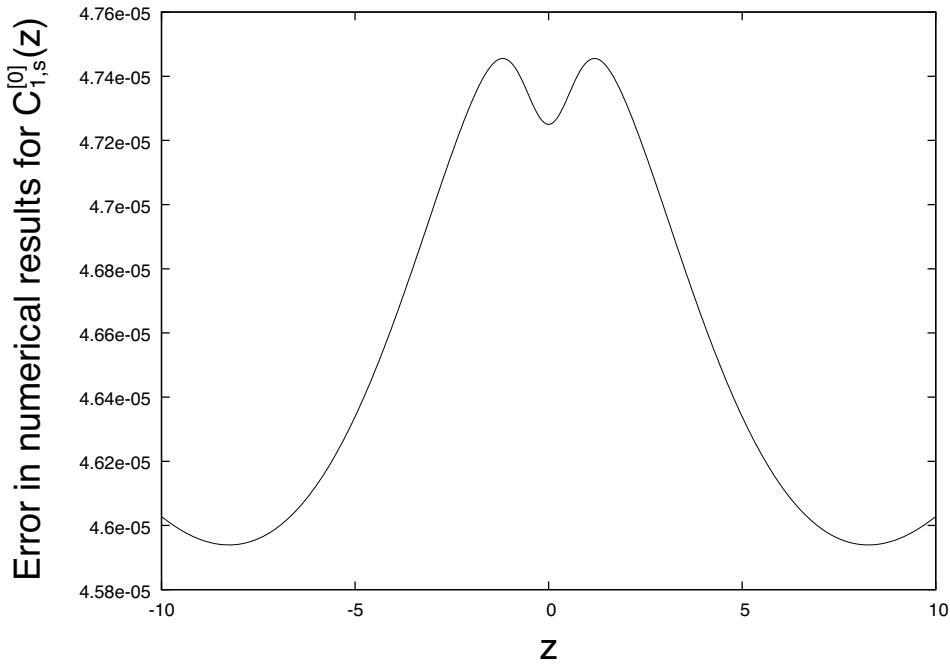


Figure 19.1.12: Difference between exact and numerical results for  $C_{1,s}^{[0]}(z)$  using a spline-based integrand in (14.3.23) and exact values of  $B_\rho(R, \phi, z)$  on the cylinder.

### 19.1.3 Test of Interpolation off a Grid

To complete our test, let us no longer use exact values of  $B_\rho(R, \phi, z)$  on the cylinder. Rather, suppose we set up a regular grid in  $x, y, z$  space centered on the origin  $(0, 0, 0)$ . Let  $x$  and  $y$  range over the values  $x \in [-2.4, 2.4]$  and  $y \in [-2.4, 2.4]$ , and (as before) let  $z$  range over the values  $z \in [-300, 300]$ . Use 49 grid points each in  $x$  and  $y$  so that  $h_x = h_y = .1$ ,

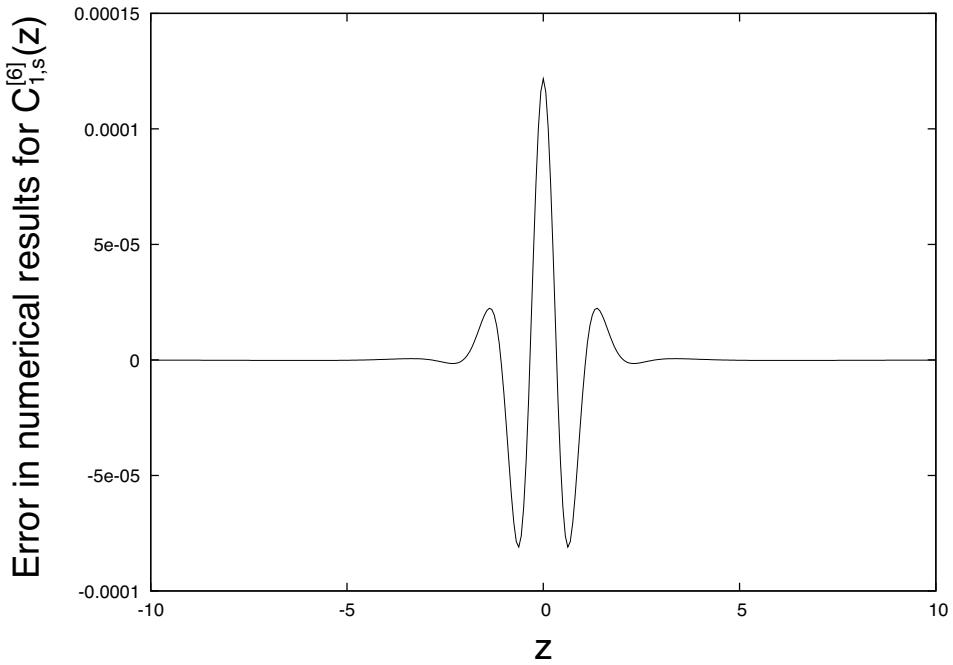


Figure 19.1.13: Difference between exact and numerical results for  $C_{1,s}^{[6]}(z)$  using a spline-based integrand in (14.3.23) and exact values of  $B_\rho(R, \phi, z)$  on the cylinder.

and (again as before) use 4801 grid points in  $z$  so that  $h_z = .125$ . Thus, use a total of  $49 \times 49 \times 4801 = 11,527,201$  grid points. For each grid point specify the three components  $B_x$ ,  $B_y$ , and  $B_z$  using (13.7.4) through (13.7.6) evaluated at these grid points. Employ bicubic interpolation (see Section 15.2.1) to interpolate  $\mathbf{B}$  at these grid points onto the selected angular points on the cylinder  $R = 2$ , and then compute  $B_\rho(R, \phi, z)$  at these angular points.<sup>4</sup> Finally, proceed as before using these approximate values of  $B_\rho(R, \phi, z)$  on the cylinder. In particular, evaluate the angular Fourier transforms with a Riemann sum using 49 angular points and evaluate the forward linear transforms for 401  $k$  values in the range  $k \in [-K_c, K_c]$  with  $K_c = 20$ . Use these same points in  $k$  space to evaluate the inverse Fourier transforms. The result of this process is a *completely numerically* calculated set of functions  $C_m^{[n]}(z)$ .

Examination of these completely numerically calculated functions shows that they also well approximate the exact functions  $C_{m,\alpha}^{[n]}(z)$ . For example, Figures 1.14 and 1.15 show the differences between the exact and the completely numerically calculated  $C_{1,s}^{[0]}(z)$  and  $C_{1,s}^{[6]}(z)$ . We see that Figure 1.14 is very similar to Figure 1.12, and Figure 1.15 is very similar to Figure 1.13. Consequently the error is little changed, and we conclude that interpolation from the grid onto the cylinder introduces little additional error. The errors for the other nonzero  $C_{m,s}^{[n]}(z)$  listed in (13.7.35) are comparable. For example, Figure 1.16 shows the

<sup>4</sup>Alternatively, one could use bicubic spline interpolation. See Section 15.2.2. Note that the  $B_z$  component does not, in fact, contribute to  $B_\rho(R, \phi, z)$ . Note also that only a relatively small number of the 11,527,201 points are actually used because only values at those points relatively near the circular cylinder are needed to interpolate onto the cylinder. And, after interpolation onto the cylinder, only  $49 \times 4801 = 235,249$  surface values of  $B_\rho$  are used in the remainder of the calculation.

exact and completely numerical results for  $C_{7,s}^{[0]}(z)$ , and Figure 1.17 shows the difference between the exact and completely numerical results for  $C_{7,s}^{[0]}(z)$ . We see that the error is approximately 4 parts in  $10^4$ . Finally, the  $C_{m,\alpha}^{[0]}(z)$  that should vanish are, in fact, again numerically very small.

We have demonstrated, for the monopole-doublet problem, that the steps in the first three boxes shown in Figure 14.1.1 can be carried out to yield results having good numerical accuracy. As remarked earlier, again see Figures 13.7.6 and 13.7.7, the surface field we have been working with is quite singular, more singular than fields likely to be encountered in practice. Thus the fact that the circular cylinder surface method has succeeded in this rather extreme case indicates that it is likely to work even better in actual physical applications.

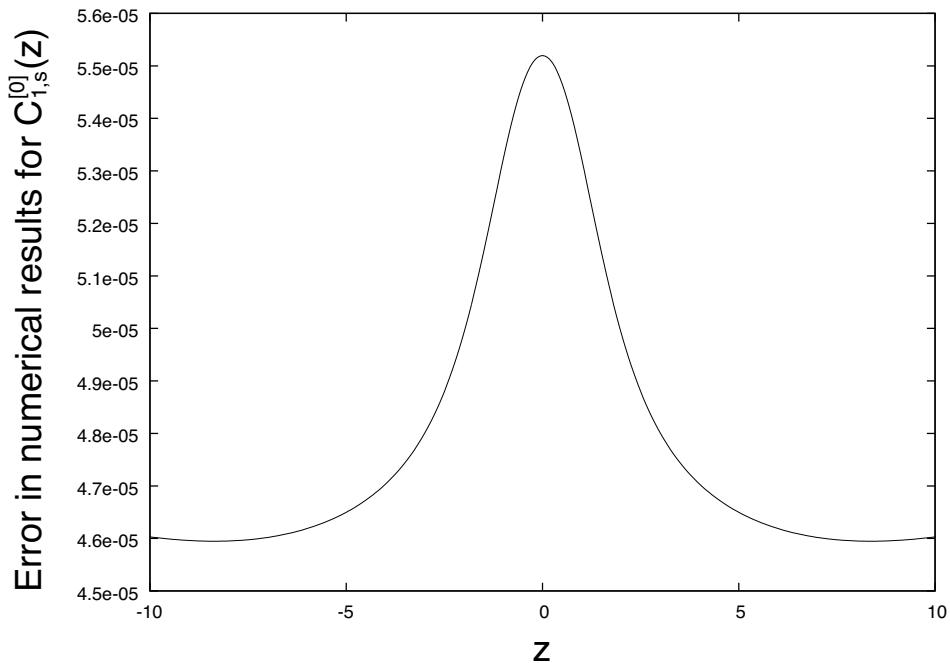


Figure 19.1.14: Difference between exact and completely numerical results for  $C_{1,s}^{[0]}(z)$  using a spline-based integrand in (14.3.23) and interpolated values of  $B_\rho(R, \phi, z)$  on the cylinder based on field data provided on a grid.

#### 19.1.4 Reproduction of Interior Field Values

Another, but less stringent, test of accuracy is to use the completely numerically obtained on-axis gradients to compute  $\mathbf{B}$  at the *interior* grid points with the aid of (13.2.69) through (13.2.71). These computed values can be compared with the known values of  $\mathbf{B}$  at the interior grid points.<sup>5</sup> Before making such a comparison, some discussion is required.

---

<sup>5</sup>This test is less stringent because it does not compare derivatives.

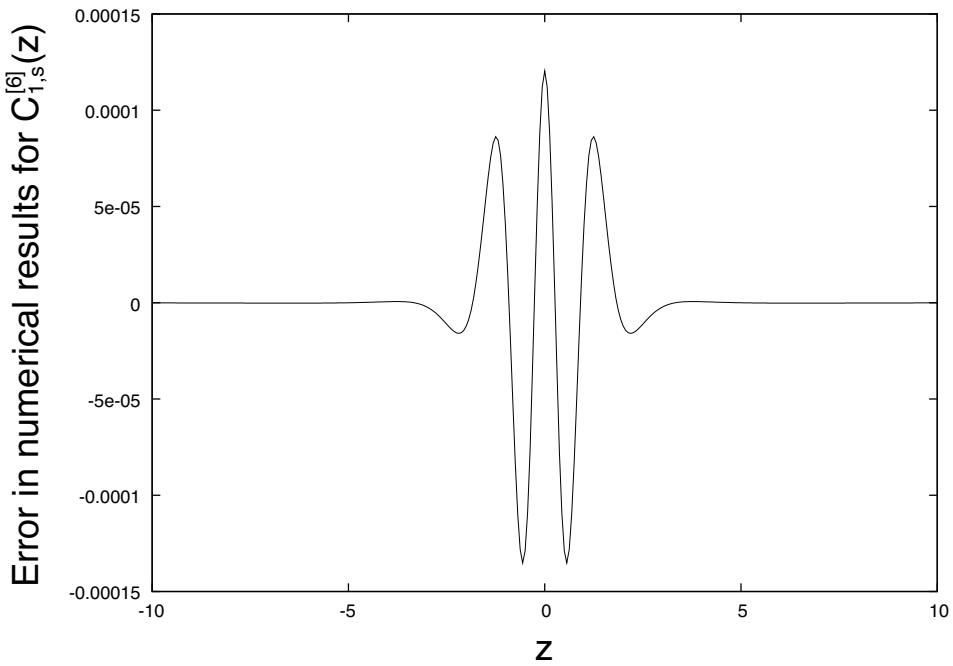


Figure 19.1.15: Difference between exact and completely numerical results for  $C_{1,s}^{[6]}(z)$  using a spline-based integrand in (14.3.23) and interpolated values of  $B_\rho(R, \phi, z)$  on the cylinder based on field data provided on a grid.

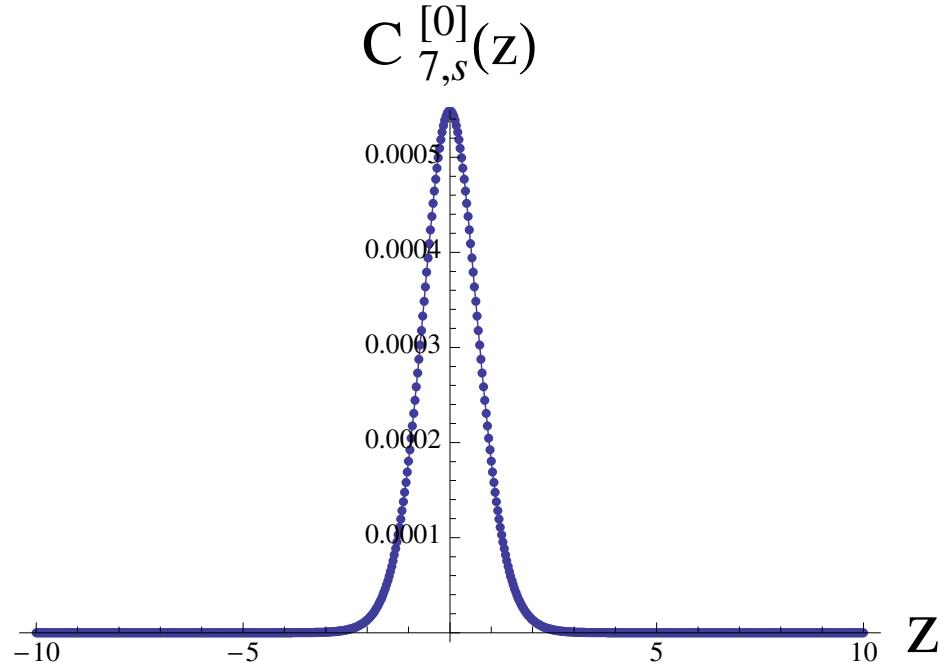


Figure 19.1.16: Exact and completely numerical results for  $C_{7,s}^{[0]}(z)$ . Exact results are shown as a solid line (see Figure 13.7.15), and numerical results are shown as dots.

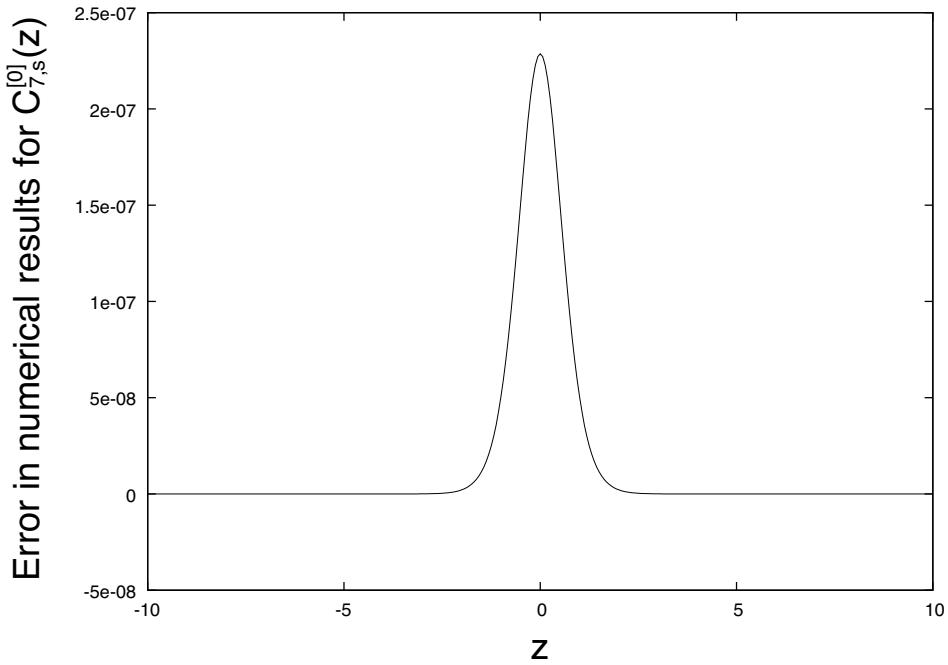


Figure 19.1.17: Difference between exact and completely numerical results for  $C_{7,s}^{[0]}(z)$  using a spline-based integrand in (14.3.23) and interpolated values of  $B_\rho(R, \phi, z)$  on the cylinder based on field data provided on a grid.

### What to Hope for

First, since the surface fields we have been working with are quite singular, let us investigate the region over which the  $C_{m,s}^{[n]}(z)$  that we have decided to employ, see (13.7.35), can be expected to give a good representation of  $\mathbf{B}(x, y, z)$ . As an initial exploration, let us consider the behavior of the Fourier series representation

$$B_\rho(R = 2, \phi, z = 0) = \sum_{m=-\infty}^{\infty} a_m \exp(im\phi) \quad (19.1.28)$$

for the function  $B_\rho(R = 2, \phi, z = 0)$  shown in Figure 13.7.6. Suppose this series is *truncated* so that only terms for which  $|m| \leq 7$  are retained. See Exercise 1.3. Call the resulting function  $B_\rho^{\text{Tr}}(R = 2, \phi, z = 0)$ . Figure 1.18 displays  $B_\rho^{\text{Tr}}(R = 2, \phi, z = 0)$  as a function of  $\phi$ , and Figure 1.19 shows the difference between  $B_\rho(R = 2, \phi, z = 0)$  and  $B_\rho^{\text{Tr}}(R = 2, \phi, z = 0)$ . Evidently terms well beyond  $|m| = 7$  must be retained in (1.28) to adequately represent the surface field. See also the table of Fourier coefficients in Exercise 1.3. It follows that many  $C_{m,s}^{[n]}(z)$  beyond those listed in (13.7.35) are required to represent the field near the surface  $R = 2$ .

As an illustration of this conclusion, let  $\mathbf{B}^{\text{TrA}}$  denote the field computed using the series (13.2.69) through (13.2.71) *truncated* so that only the  $C_{m,s}^{[n]}(z)$  listed in (13.7.35) are retained, and using the *analytic* expressions (13.7.33) and their  $z$  derivatives for these  $C_{m,s}^{[n]}(z)$ .<sup>6</sup> Also,

<sup>6</sup>Note that the use of only the  $C_{m,s}^{[n]}(z)$  listed in (13.7.35) to compute  $\mathbf{B}^{\text{TrA}}$  amounts to using 6<sup>th</sup>-order

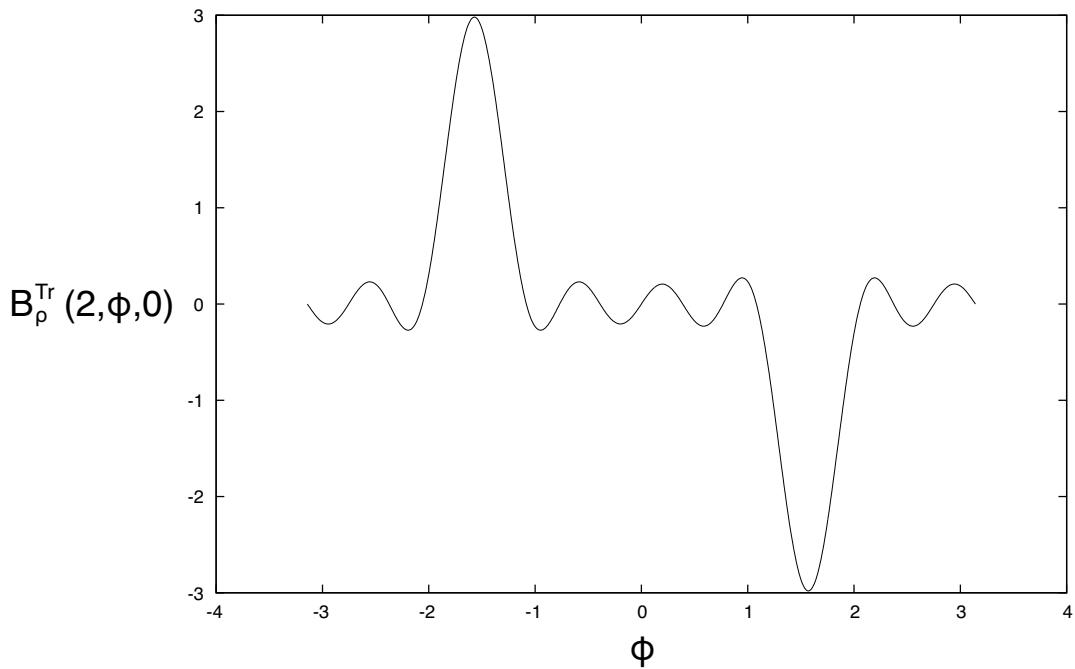


Figure 19.1.18: The quantity  $B_\rho^{\text{Tr}}(R = 2, \phi, z = 0)$  for the monopole doublet in the case that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm) $^2$ .

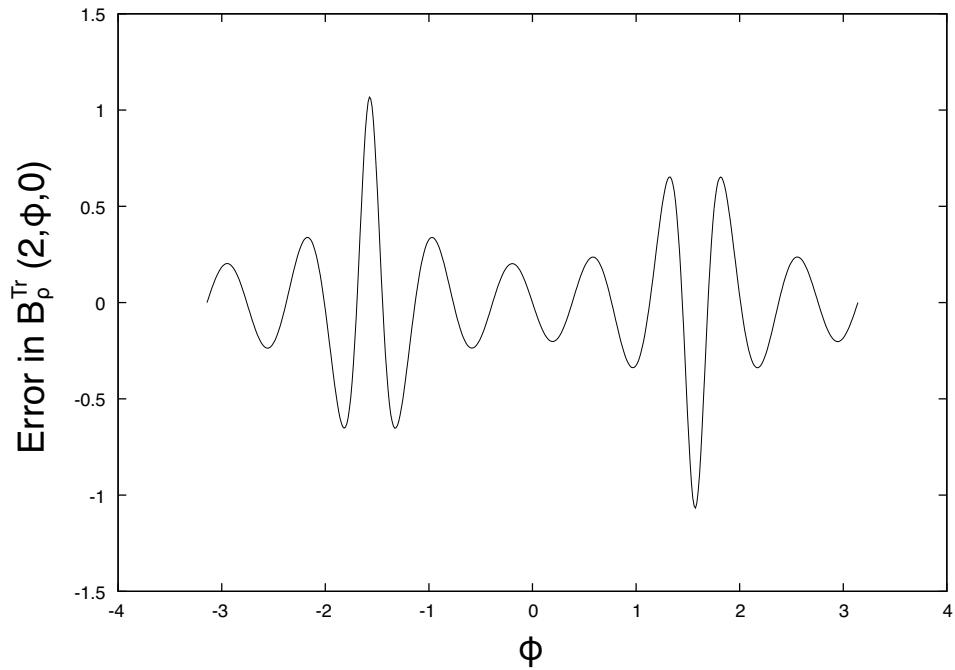


Figure 19.1.19: Difference between  $B_\rho(R = 2, \phi, z = 0)$  and  $B_\rho^{\text{Tr}}(R = 2, \phi, z = 0)$  for the monopole doublet in the case that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm) $^2$ .

let  $\mathbf{B}^{\text{Exact}}$  denote the known *exact* values of  $\mathbf{B}$  computed using (13.7.4) through (13.7.6). Figures 1.20 through 1.22 show, as a function of  $\phi$ , the quantity  $\|\mathbf{B}^{\text{TrA}} - \mathbf{B}^{\text{Exact}}\|/\|\mathbf{B}\|^{\text{Max}}$  for various values of  $\rho$  and  $z$ . Here  $\|\mathbf{B}\|^{\text{Max}}$  is the *maximum* value of  $\|\mathbf{B}\|$  within the cylinder with radius  $\rho$ . We conclude that within the cylinder  $\rho \leq 1/2$  the relative error in the field due to truncating the cylindrical multipole expansion is less than a few parts in  $10^5$ . Note also that the domain of good approximation opens up as  $z$  leaves the plane  $z = 0$ . This behavior is to be expected based on the analytic properties of  $\psi(x, y, z)$  when  $x, y, z$  are treated as three *complex* variables. See Examples 2.1 and 2.2 in Section 31.2; and Examples 3.1 and 3.2 in Section 31.3.

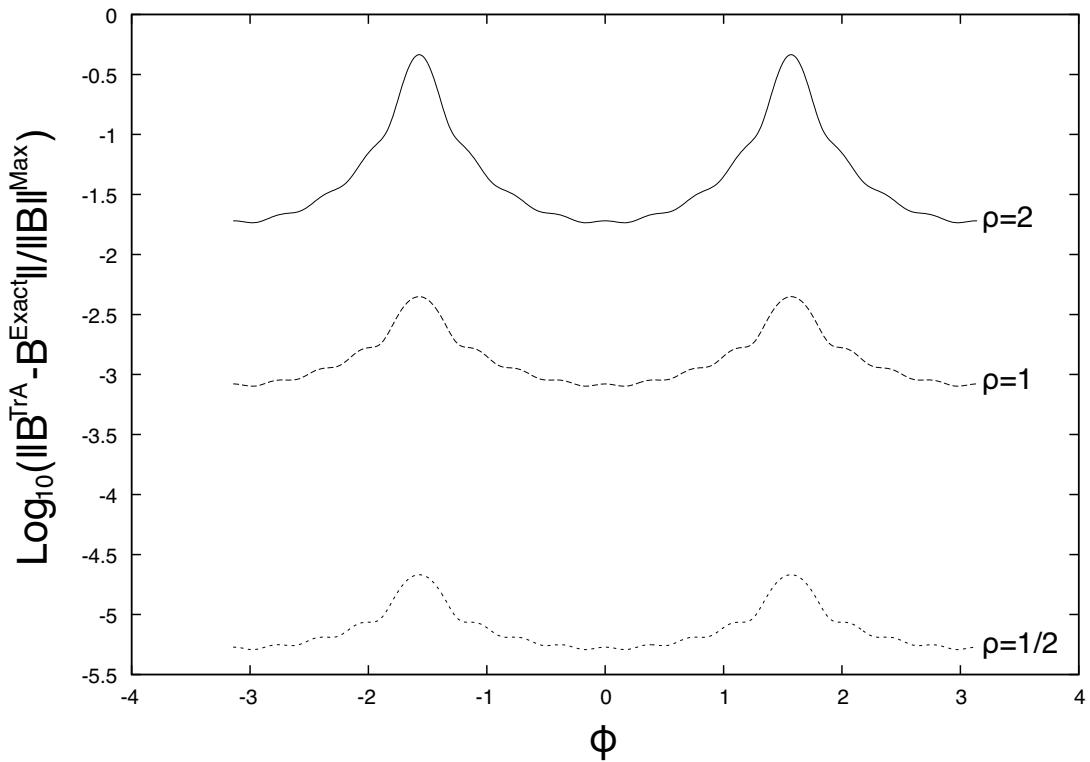


Figure 19.1.20: The logarithm base 10 of the quantity  $\|\mathbf{B}^{\text{TrA}} - \mathbf{B}^{\text{Exact}}\|/\|\mathbf{B}\|^{\text{Max}}$  as a function of  $\phi$  for three  $\rho$  values and  $z = 0$ , for the monopole doublet, in the case that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm) $^2$ . The solid line corresponds to  $\rho = 2$ , the dashed line to  $\rho = 1$ , and the dotted line to  $\rho = 1/2$ .

### Test of Field Reproduction within the Cylinder $\rho \leq 1/2$

The previous discussion has determined (for the case of a monopole doublet) the region where the truncated cylindrical multipole expansion, which retains only the terms listed in (13.7.35), is expected to be accurate assuming the listed  $C_{m,s}^{[n]}(z)$  are known exactly. For real

---

polynomials in the  $x, y$  variables (with  $z$ -dependent coefficients) for the components  $B_x$  and  $B_y$ , and a 5<sup>th</sup>-order polynomial in  $x, y$  for  $B_z$ .

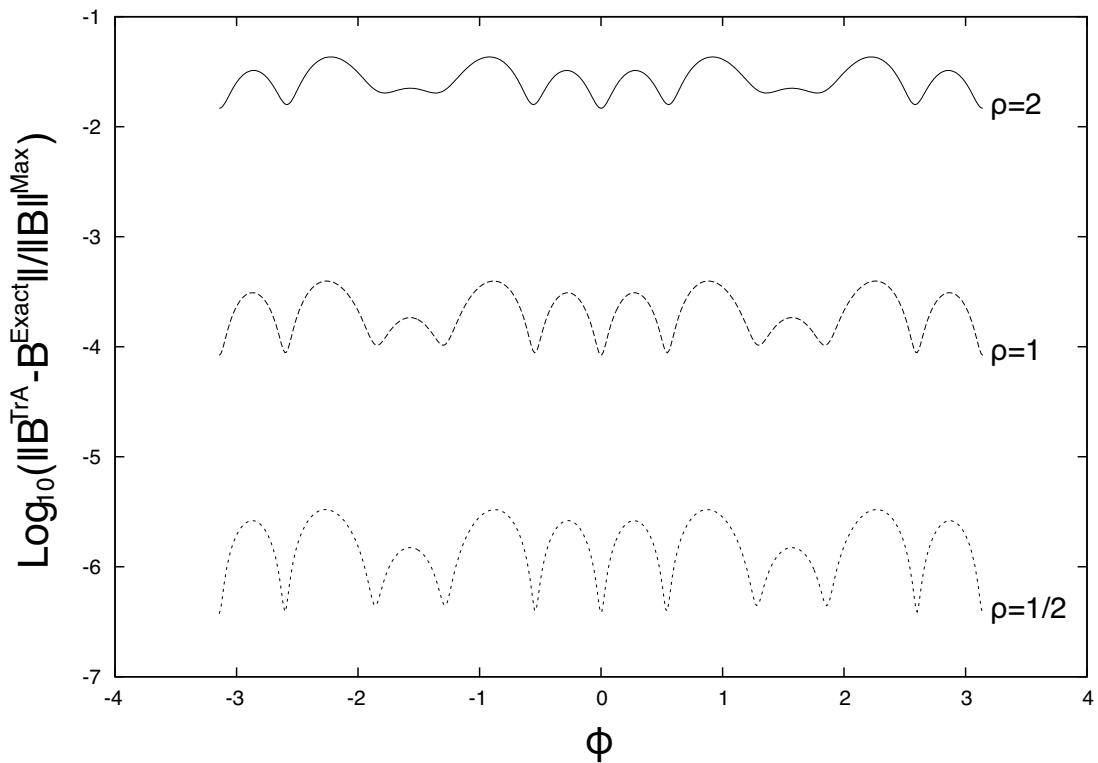


Figure 19.1.21: The logarithm base 10 of the quantity  $\|\mathbf{B}^{\text{TrA}} - \mathbf{B}^{\text{Exact}}\| / \|\mathbf{B}\|_{\text{Max}}$  as a function of  $\phi$  for three  $\rho$  values and  $z = 2.5$ , for the monopole doublet, in the case that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm) $^2$ . The solid line corresponds to  $\rho = 2$ , the dashed line to  $\rho = 1$ , and the dotted line to  $\rho = 1/2$ .

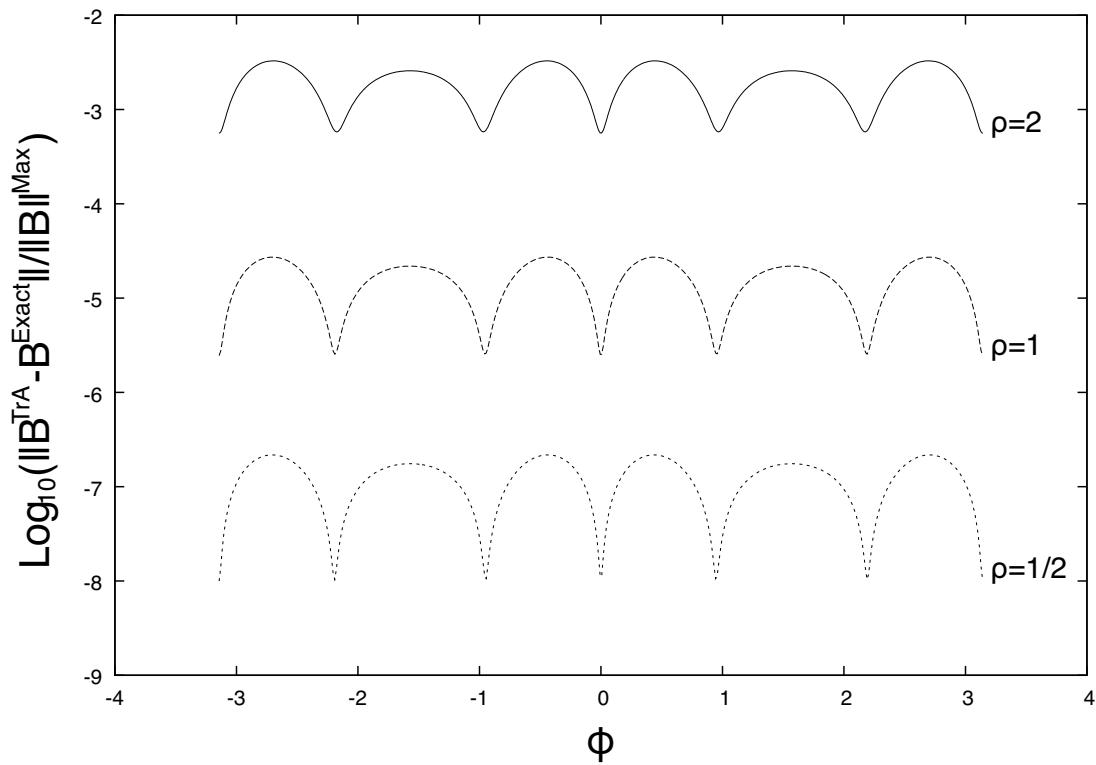


Figure 19.1.22: The logarithm base 10 of the quantity  $\|\mathbf{B}^{\text{TrA}} - \mathbf{B}^{\text{Exact}}\| / \|\mathbf{B}\|_{\text{Max}}$  as a function of  $\phi$  for three  $\rho$  values and  $z = 5$ , for the monopole doublet, in the case that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm) $^2$ . The solid line corresponds to  $\rho = 2$ , the dashed line to  $\rho = 1$ , and the dotted line to  $\rho = 1/2$ .

problems we only know the  $C_{m,\alpha}^{[n]}(z)$  as calculated numerically based on grid data interpolated onto a surface, and only have the values of  $\mathbf{B}$  at the grid points. Let  $\mathbf{B}^{\text{Num}}$  denote the completely *numerically* calculated values of  $\mathbf{B}$  based on the completely numerically calculated  $C_{m,\alpha}^{[n]}(z)$  with  $(m + n) \leq 7$ , and let  $\mathbf{B}^{\text{Grid}}$  now denote the values of  $\mathbf{B}$  at the *grid* points obtained using (13.7.4) through (13.7.6). Examination of the completely numerically calculated values and the corresponding grid values shows that

$$\|\mathbf{B}^{\text{Num}} - \mathbf{B}^{\text{Grid}}\| < 6 \times 10^{-5} \quad (19.1.29)$$

for all grid points within the cylinder  $\rho \leq (1/2)$ . Observe that, according to Figure 13.7.3, the magnitude of the maximum on-axis field is .32. Thus, the maximum error within the cylinder  $\rho = 1/2$  compared to the maximum on-axis field is approximately 2 parts in  $10^4$ . The smallness of this error again illustrates the accuracy of the method.<sup>7</sup> It also is a data-based indication that the cylindrical harmonic expansion is converging well for the interior  $\rho \leq 1/2$  and that the Maxwell equations are well satisfied by the interior values of  $\mathbf{B}^{\text{Grid}}$ .<sup>8</sup>

## Exercises

**19.1.1.** Verify (1.11), (1.12), (1.15), and (1.16). You may need the identity

$$(2m)! = 2^m m! [1 \cdot 3 \cdot 5 \cdot 7 \cdots (2m - 1)]. \quad (19.1.30)$$

Prove this identity, say, by induction.

**19.1.2.** This exercise studies the discrete Fourier transform of a periodic function. Suppose  $f(\phi)$  is a  $2\pi$  periodic function and therefore has a Fourier expansion of the form

$$f(\phi) = \sum_{m=-\infty}^{\infty} a_m \exp(im\phi). \quad (19.1.31)$$

We know that the coefficients  $a_j$  are given by the relation

$$a_j = [1/(2\pi)] \int_0^{2\pi} d\phi f(\phi) \exp(-ij\phi). \quad (19.1.32)$$

Select  $N$  discrete phi values  $\phi_n$  according to the rule

$$\phi_n = n(2\pi/N) \text{ for } n = 0, 1, \dots, N - 1. \quad (19.1.33)$$

---

<sup>7</sup>This small error means, among other things, that Figures 13.7.3 through 13.7.5 are well reproduced by  $\mathbf{B}^{\text{Num}}$ . Note that *only* the values of  $B_x^{\text{Grid}}$  and  $B_y^{\text{Grid}}$  on grid points near the surface of the cylinder  $R = 2$  were used to determine all three components of  $\mathbf{B}^{\text{Num}}$  at all grid points within the cylinder.

<sup>8</sup>By construction, the  $\mathbf{B}^{\text{Num}}$  satisfy the Maxwell equations to good precision. And, in this case, the  $\mathbf{B}^{\text{Grid}}$  also satisfy the Maxwell equations since they were obtained by evaluating the Maxwell solution (13.7.4) through (13.7.6) at the grid points. However, in the general case, the  $\mathbf{B}^{\text{Grid}}$  are supplied by some 3-D electromagnetic code, and the ‘Maxwellian goodness’ of these quantities depends on the quality of the 3-D electromagnetic code.

Next define quantities  $A_j$  by the rule

$$A_j = (1/N) \sum_{n=0}^{N-1} f(\phi_n) \exp(-ij\phi_n). \quad (19.1.34)$$

Verify that (1.34) is the discrete version of (1.32).

Show that combining (1.31), (1.33), and (1.34) gives the result

$$\begin{aligned} A_j &= \sum_{m=-\infty}^{\infty} a_m \left\{ (1/N) \sum_{n=0}^{N-1} \exp[inm(2\pi/N)] \exp[-ijn(2\pi/N)] \right\} \\ &= \sum_{m=-\infty}^{\infty} a_m \left\{ (1/N) \sum_{n=0}^{N-1} \exp[in(m-j)(2\pi/N)] \right\}. \end{aligned} \quad (19.1.35)$$

However, there is the identity

$$\sum_{n=0}^{N-1} x^n = (1 - x^N)/(1 - x). \quad (19.1.36)$$

Using this identity, show that

$$(1/N) \sum_{n=0}^{N-1} \exp[in(m-j)(2\pi/N)] = (1/N) \{1 - \exp[i(m-j)2\pi]\} / \{1 - \exp[i(m-j)(2\pi/N)]\}. \quad (19.1.37)$$

From (1.37), show that

$$(1/N) \sum_{n=0}^{N-1} \exp[in(m-j)(2\pi/N)] = 0 \text{ if } (m-j)/N \text{ is not an integer}, \quad (19.1.38)$$

and

$$(1/N) \sum_{n=0}^{N-1} \exp[in(m-j)(2\pi/N)] = 1 \text{ if } (m-j)/N \text{ is an integer}. \quad (19.1.39)$$

[Note that  $(m-j)$  is always an integer.] Verify that these results can be written more compactly in the form

$$(1/N) \sum_{n=0}^{N-1} \exp[in(m-j)(2\pi/N)] = \sum_{\ell=-\infty}^{\infty} \delta_{m,j+\ell N}. \quad (19.1.40)$$

Employ (1.35) and (1.40) to get the final result

$$A_j = \sum_{\ell=-\infty}^{\infty} a_{j+\ell N}. \quad (19.1.41)$$

Verify that (1.41) implies the periodicity relation

$$A_{j+N} = A_j. \quad (19.1.42)$$

The relation (1.41) can be rewritten in the form

$$A_j = a_j + \sum_{\ell \neq 0}^{\infty} a_{j+\ell N}. \quad (19.1.43)$$

Thus,  $A_j$  is a good approximation to  $a_j$  provided  $N$  is large enough that the  $a_{j+\ell N}$  are small for  $\ell = \pm 1, \pm 2, \dots$  in such a way that the sum in (1.43) is also small.

Note that under the assumptions made, namely periodicity and rapid falloff of the Fourier coefficients, the discrete angular Fourier transform is much more accurate than the estimate (15.3.24) would promise.

**19.1.3.** This exercise is a continuation of Exercise 1.2. It shows that good performance of the discrete Fourier transform for periodic functions can be assured under the assumption of suitable *analyticity*. As an application, it justifies the use of the  $N = 49$  discrete Fourier transform to obtain  $\tilde{B}_\rho(R, m', k')$  in the case of the monopole doublet.

As in (1.28), let  $f(\phi)$  be the function  $B_\rho(R = 2, \phi, z = 0)$  shown in Figure 13.7.6. Its first few nonzero Fourier coefficients, obtained by accurate numerical integration of (1.32) using *Mathematica*, are listed below in Table 1.1. As shown in Figure 1.23, they fall off exponentially with increasing  $j$  in the fashion

$$|a_j| \sim (.8)^j. \quad (19.1.44)$$

Consequently, the contribution of the sum in (1.43) will be small provided  $N$  is reasonably large and  $j$  is considerably smaller than  $N$ . Table 1.1 also lists the first few nonzero values of  $A_j$  obtained from (1.34) with  $N = 49$ . We see from the table that (when  $N = 49$ ) the  $A_j$  are good approximations to the  $a_j$  for  $j \leq 7$ , as advertised.

Table 19.1.1: The exact and discrete (with  $N = 49$ ) Fourier coefficients of  $f(\phi)$ .

$j$	$2\Im a_j$	$2\Im A_j$
1	0.986833050662540	0.9868330457500823
3	-0.859714503908633	-0.8597145139462157
5	0.656703778334560	0.6567037866425125
7	-0.477523703835796	-0.47752370010754797
9	0.337961137907473	0.33796112868607286
11	-0.235072232641152	-0.235072209887016
13	0.161532491767950	0.16153245497610078
15	-0.110003634817578	-0.11000357184645376
17	0.074393215827366	0.0743931186747128
19	-0.050032679038838	-0.05003252910493426
21	0.033497429399684	0.0334971980691082
23	-0.022342757613886	-0.022342400791792692
25	0.0148553386831777	0.01485478841656739

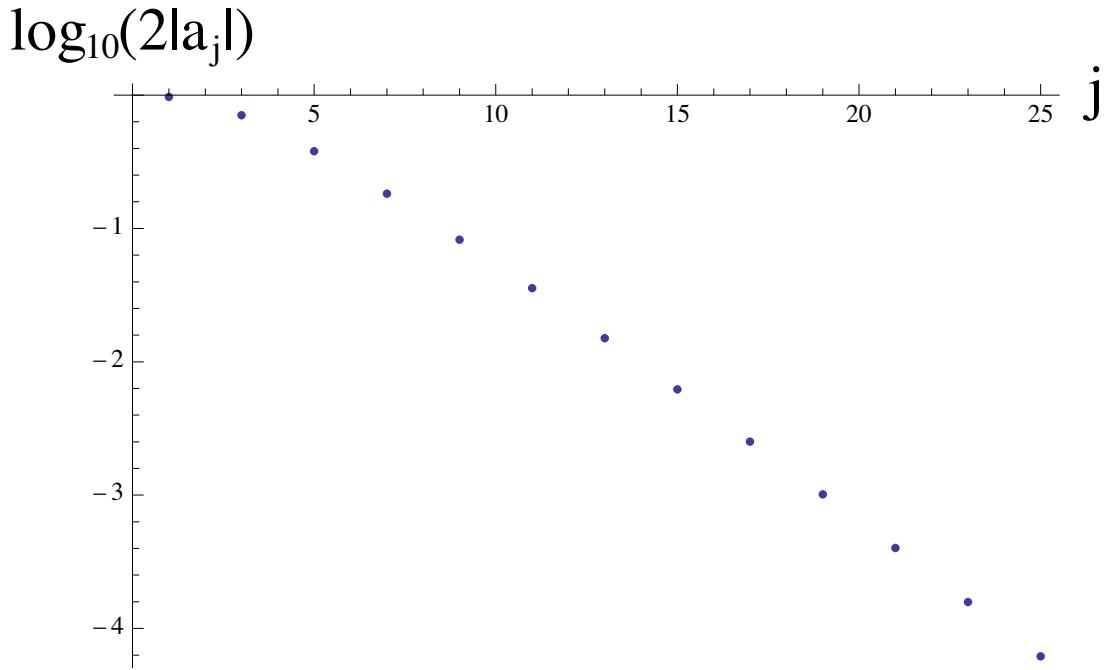


Figure 19.1.23: The quantities  $\log_{10}(2|a_j|)$  as a function of  $j$ . For large  $j$  the points fall on a straight line having slope  $\log_{10}(.8)$ .

Similar considerations apply to the function  $\tilde{B}_\rho(R, \phi, k)$ . Figures 1.24 and 1.25 illustrate the cases  $\tilde{B}_\rho(R = 2, \phi, k = 0)$  and  $\tilde{B}_\rho(R = 2, \phi, k = 20)$ , and Figures 1.26 and 1.27 display their Fourier coefficients. Again the Fourier coefficients fall off as  $(.8)^j$ . This behavior can be understood as follows: It can be shown that the Fourier coefficients of  $B_\rho(R, \phi, z)$  for  $z \neq 0$  fall off even more rapidly than (1.44). We conclude that since the Fourier coefficients of  $B_\rho(R, \phi, z)$  fall off at least as rapidly as (1.44) for all  $z$ , then the Fourier coefficients of  $\tilde{B}_\rho(R, \phi, k)$  must also fall off in this fashion, because  $\tilde{B}_\rho(R, \phi, k)$  may be viewed as a linear combination of the  $B_\rho(R, \phi, z)$ . See (1.22).

How might one anticipate the relation (1.44)? Introduce the complex variable  $\lambda$  by writing

$$\lambda = \exp i\phi. \quad (19.1.45)$$

With this change of variable the integral (1.32) becomes

$$a_j = [-i/(2\pi)] \oint_C d\lambda f(-i \log \lambda) \lambda^{-(j+1)} \quad (19.1.46)$$

where the contour  $C$  is the unit circle. Now deform the contour to make  $C$  as large a circle as possible without encountering singularities of  $f(-i \log \lambda)$ . Suppose this circle has radius  $\Lambda$ . Then the integral (1.46) has the asymptotic behavior

$$|a_j| \sim (1/\Lambda)^j. \quad (19.1.47)$$

Evidently (13.7.8), when evaluated at  $z = 0$ , is singular when

$$\sin \phi = \pm [(R^2 + a^2)/(2aR)]. \quad (19.1.48)$$

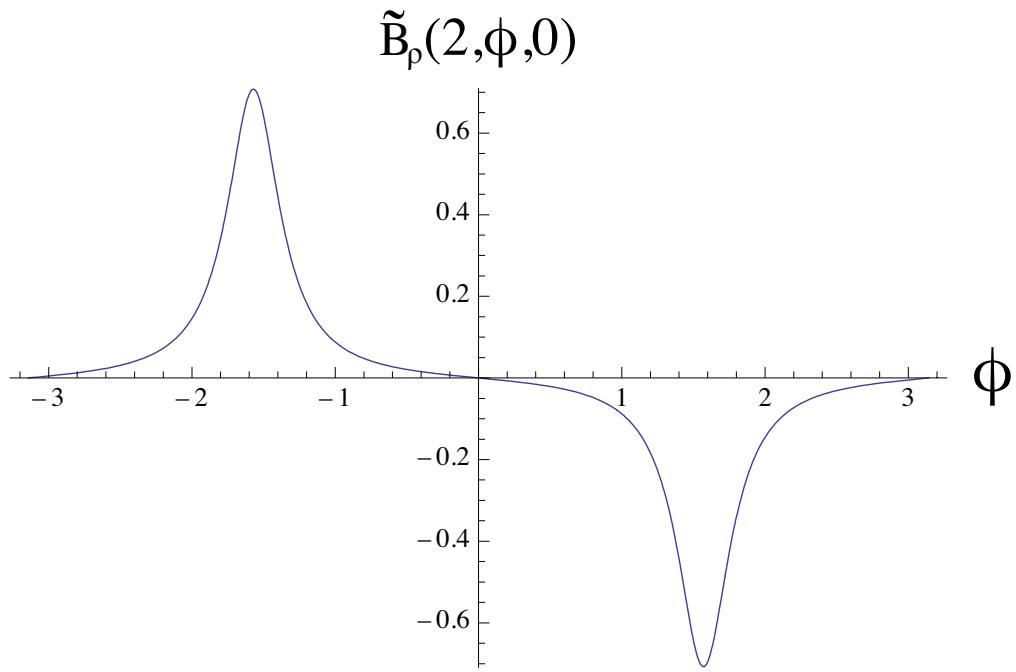


Figure 19.1.24: The real part of  $\tilde{B}_\rho(R = 2, \phi, k = 0)$  for the monopole doublet in the case that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm) $^2$ . The imaginary part vanishes.

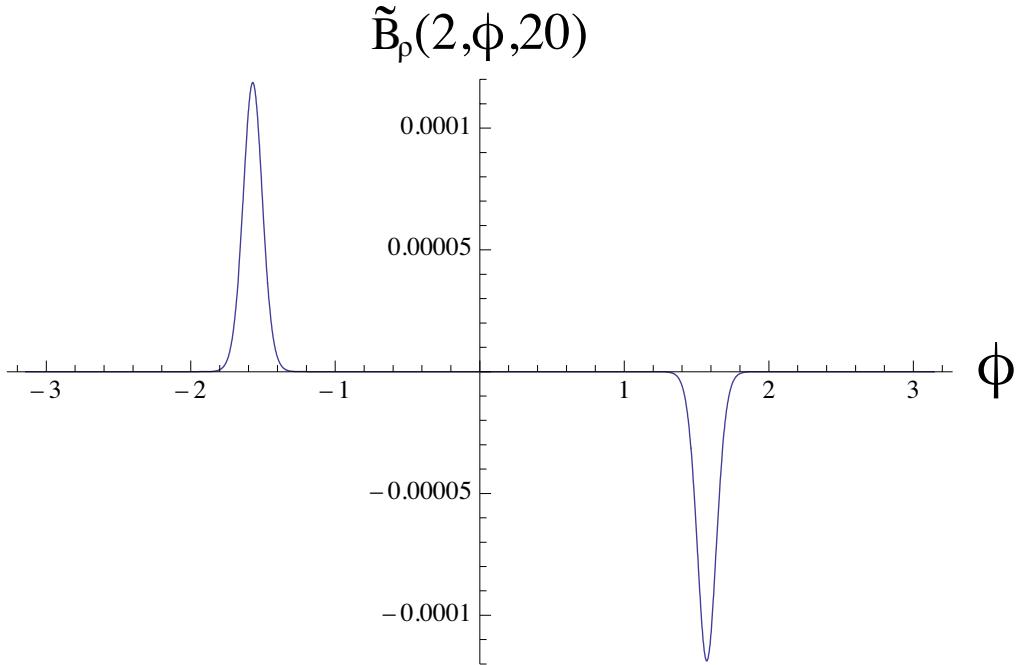


Figure 19.1.25: The real part of  $\tilde{B}_\rho(R = 2, \phi, k = 20)$  for the monopole doublet in the case that  $a = 2.5$  cm and  $g = 1$  Tesla-(cm) $^2$ . The imaginary part vanishes.

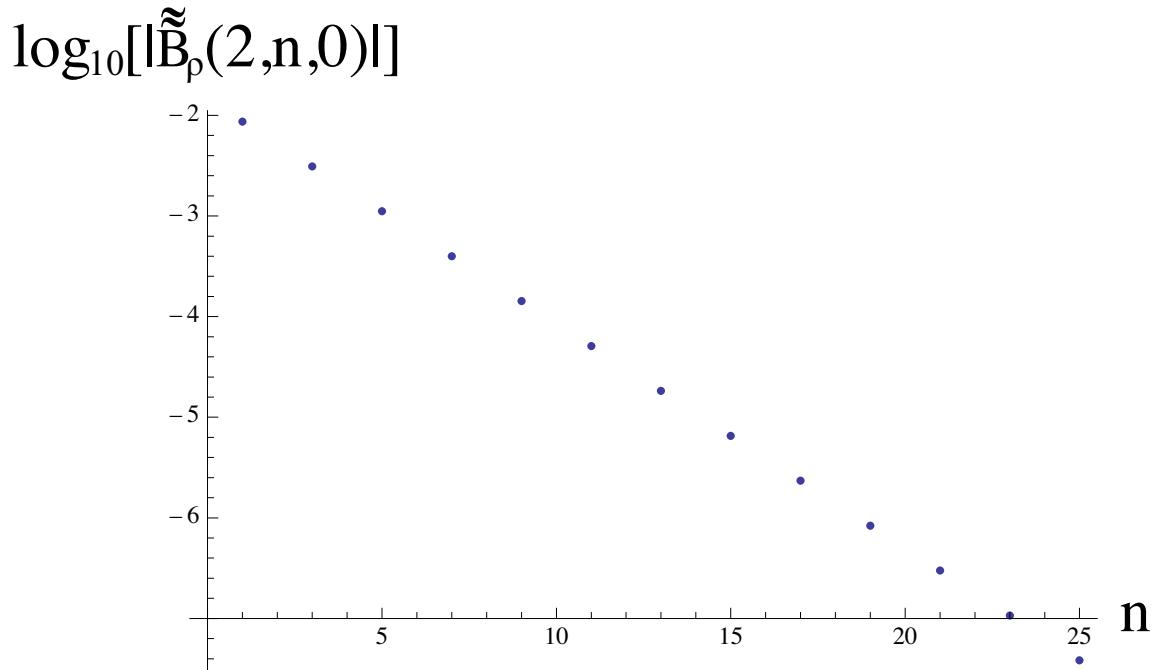


Figure 19.1.26: The quantities  $\log_{10}[|\tilde{\tilde{B}}_\rho(R = 2, n, k = 0)|]$  as a function of  $n$ . For large  $n$  the points fall on a straight line having slope  $\log_{10}(.8)$ .

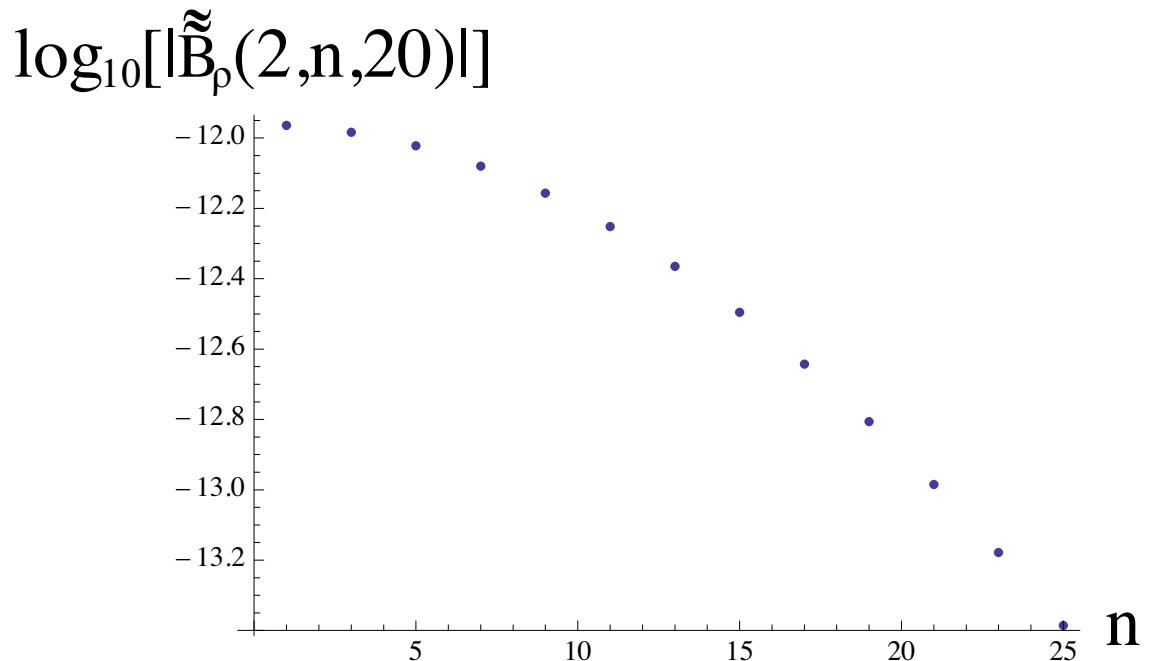


Figure 19.1.27: The quantities  $\log_{10}[|\tilde{\tilde{B}}_\rho(R = 2, n, k = 20)|]$  as a function of  $n$ . For large  $n$  the points fall on a straight line having slope  $\log_{10}(.8)$ .

Show that (1.48) yields the relation

$$1/\Lambda = R/a = 2/2.5 = .8. \quad (19.1.49)$$

Suppose  $z \neq 0$  in (13.7.8). Show that then  $\Lambda$  is larger than the value given in (1.49). Thus, we have analyzed the worst case, the case with the slowest falloff with increasing  $j$ .

**19.1.4.** Review Exercise 1.3. Look at the relation (14.3.2), the expansion (13.2.37), and the monopole doublet results (13.7.19), (13.7.29), and (13.7.33). Use these quantities to find explicit expressions for the Fourier Coefficients  $\tilde{B}_\rho(R, m, z)$ , and find their large  $m$  behavior.

## 19.2 Elliptical Cylinder Numerical Results for Monopole Doublet

In this subsection we will benchmark the numerical method of Section 14.4 to demonstrate the use of an elliptical cylinder. Here we have two goals: First, as a practical matter, the infinite sums over  $r$  that occur in (14.4.83) through (14.4.86) must be truncated, and we must establish that this can be done while still achieving a desired accuracy. Second, we must demonstrate that all our numerical machinery actually works.

As just done for the case of a circular cylinder, we will again try to reproduce the exact results for the on-axis gradients of the same monopole doublet. But now, as an example, we will use as our surface that of an elliptical cylinder for which the ellipse has a semi-major axis ( $x^{\max}$ ) of 4 cm in the  $x$  direction and a semi-minor axis ( $y^{\max}$ ) of 2 cm in the  $y$  direction. See Figure 14.4.2. This is achieved by setting

$$u = U = \tanh^{-1}(y^{\max}/x^{\max}) = \tanh^{-1}(2/4) = .549306144 \quad (19.2.1)$$

and

$$f = 4/\cosh(U) = \sqrt{12} = 3.464101615 \text{ cm} \quad (19.2.2)$$

in equations (14.4.1) and (14.4.2) so that we have the relations

$$x^{\max} = f \cosh U = 4 \text{ cm}, \quad (19.2.3)$$

$$y^{\max} = f \sinh U = 2 \text{ cm}. \quad (19.2.4)$$

### 19.2.1 Finding the Mathieu Coefficients

In the elliptical cylinder case there are fewer calculations we can carry out exactly compared to the circular cylinder case. However, some of the routines we will be using in the elliptic case will be the same as in the circular case, and we have already benchmarked them in the circular case.

### Exact Results for the Forward ( $z \rightarrow k$ ) Fourier Transform

There is a quantity we can still compute exactly for the elliptic cylinder when applied to the monopole doublet case, and that is the function  $\tilde{F}(v, k)$ . See (14.4.72).<sup>9</sup> Upon combining (14.4.67) and (14.4.72) we see that

$$\tilde{F}(v, k) = (f \sinh U \cos v) \tilde{B}_x(U, v, k) + (f \cosh U \sin v) \tilde{B}_y(U, v, k) \quad (19.2.5)$$

where

$$\tilde{B}_x(U, v, k) = [1/(2\pi)] \int_{-\infty}^{\infty} dz \exp(-ikz) B_x(U, v, z), \quad (19.2.6)$$

$$\tilde{B}_y(U, v, k) = [1/(2\pi)] \int_{-\infty}^{\infty} dz \exp(-ikz) B_y(U, v, z). \quad (19.2.7)$$

Examine  $B_x(U, v, z)$  and  $B_y(U, v, z)$  as given by (14.4.1), (14.4.2), (13.7.4), and (13.7.5). Define quantities  $b_{\pm}(v)$  by the rule

$$b_{\pm}(v) = [x^2 + (y \pm a)^2]^{1/2} = \{[f \cosh(U) \cos(v)]^2 + [f \sinh(U) \sin(v) \pm a]^2\}^{1/2}. \quad (19.2.8)$$

Using this definition we may write

$$[x^2 + (y \pm a)^2 + z^2] = [z^2 + b_{\pm}^2(v)], \quad (19.2.9)$$

and  $B_x(U, v, z)$  and  $B_y(U, v, z)$  take the form

$$\begin{aligned} B_x(U, v, z) &= gf \cosh(U) \cos(v) [z^2 + b_-^2(v)]^{-3/2} \\ &\quad - gf \cosh(U) \cos(v) [z^2 + b_+^2(v)]^{-3/2}, \end{aligned} \quad (19.2.10)$$

$$\begin{aligned} B_y(U, v, z) &= g[f \sinh(U) \sin(v) - a] [z^2 + b_-^2(v)]^{-3/2} \\ &\quad - g[f \sinh(U) \sin(v) + a] [z^2 + b_+^2(v)]^{-3/2}. \end{aligned} \quad (19.2.11)$$

Next recall the Fourier transform relation

$$[1/(2\pi)] \int_{-\infty}^{\infty} dz \exp(-ikz) [z^2 + b_{\pm}^2(v)]^{-3/2} = (1/\pi)[|k|/b_{\pm}(v)] K_1[|k|b_{\pm}(v)]. \quad (19.2.12)$$

For convenience, define the functions  $F_{\pm}(v, k)$  by the rule

$$F_{\pm}(v, k) = (1/\pi)[|k|/b_{\pm}(v)] K_1[|k|b_{\pm}(v)]. \quad (19.2.13)$$

Then, in terms of these functions, we have the relations

$$\begin{aligned} \tilde{B}_x(U, v, k) &= gf \cosh(U) \cos(v) F_{-}(v, k) \\ &\quad - gf \cosh(U) \cos(v) F_{+}(v, k), \end{aligned} \quad (19.2.14)$$

---

<sup>9</sup>Of course, we will eventually want to demonstrate that we can compute this function numerically with high accuracy using field data on grid points.

$$\begin{aligned}\tilde{B}_y(U, v, k) &= g[f \sinh(U) \sin(v) - a]F_-(v, k) \\ &\quad - g[f \sinh(U) \sin(v) + a]F_+(v, k).\end{aligned}\tag{19.2.15}$$

Combining (2.5), (2.8), and (2.13) through (2.15) gives a final expression for  $\tilde{F}(v, k)$ ,

$$\tilde{F}(v, k) = .\tag{19.2.16}$$

Note that for the magnetic monopole doublet the functions  $F_{\pm}(v, k)$  and  $\tilde{F}(v, k)$  are pure real and are even functions of  $k$ .

Let us pause to study the functions  $F_{\pm}(v, k)$ . The function  $K_1(w)$  has, at the origin, the behavior

$$K_1(w) \approx (1/w) + (w/2) \log(w/2);\tag{19.2.17}$$

and therefore

$$wK_1(w) \approx 1 + (w^2/2) \log(w/2)\tag{19.2.18}$$

at the origin. At  $w = +\infty$ ,  $K_1(w)$  has the behavior

$$K_1(w) \approx [\pi/(2w)]^{1/2} \exp(-w).\tag{19.2.19}$$

Consequently  $F_{\pm}(v, k)$ , and therefore also  $\tilde{F}(v, k)$ , are well behaved for all  $k$ , but are not analytic at the origin because of the log term.<sup>10</sup>

### Exact Results for the Mathieu Coefficients

According to (14.4.73) and (14.4.74), the next step is to compute the Mathieu coefficients by performing the angular integrals<sup>11</sup>

$$\tilde{\tilde{F}}_r^c(k) = (1/\pi) \int_0^{2\pi} dv \operatorname{ce}_r(v, q) \tilde{F}(v, k),\tag{19.2.20}$$

$$\tilde{\tilde{F}}_r^s(k) = (1/\pi) \int_0^{2\pi} dv \operatorname{se}_r(v, q) \tilde{F}(v, k).\tag{19.2.21}$$

For the monopole doublet, because the transverse field components are even in  $z$ , the functions  $\tilde{\tilde{F}}_r^{\alpha}(k)$  will be real and even in  $k$ . Unfortunately, unlike the circular cylinder case for which we were able to find the  $\tilde{B}_{\rho}(R, m, k)$  analytically, see (1.14) through (1.16), we do not have analytic results for the  $\tilde{\tilde{F}}_r^{\alpha}(k)$ . However, since we do know  $\tilde{F}(v, k)$  analytically, we can compute the  $\tilde{\tilde{F}}_r^{\alpha}(k)$  numerically.<sup>12</sup> With these functions in hand, we will be able to explore how many of them need to be retained in the sums (14.4.85) and (14.4.86).

---

<sup>10</sup>As emphasized earlier, this lack of analyticity at the origin is related to the fact that  $\mathbf{B}$  for the monopole doublet falls off only as  $|z|^{-3}$  at  $\pm\infty$ .

<sup>11</sup>Note that, unlike the circular cylinder case where the  $(z \rightarrow k)$  and  $(\phi \rightarrow m)$  Fourier transforms can be performed in either order, see Section 16.1.2, in the elliptic cylinder case we must first perform the  $(z \rightarrow k)$  transform and then the  $(v \rightarrow r)$  transform. This is because  $\operatorname{ce}_r(v, q)$  and  $\operatorname{se}_r(v, q)$  depend on  $k$  through (14.4.23).

<sup>12</sup>Of course, we will eventually want to demonstrate that we can also compute these functions to high accuracy using only data on grid points.

To see what we might expect for the Mathieu coefficients,  $\tilde{F}_r^\alpha(k)$ , it is useful to examine the angular dependence of  $\tilde{F}(v, k)$ . As examples, Figures 2.1 and 2.2 display the functions  $\Re \tilde{F}(v, k = 0)$  and  $\Re \tilde{F}(v, k = 20)$ . Also shown, in Figure 2.3, is the quantity  $\Re \tilde{F}(v = \pi/2, k)$ . Note that the maximum amplitude of  $\tilde{F}(v, k)$  decreases dramatically as  $k$  increases, in accord with (2.19), and the peaks become sharper. As is evident from Figures 2.1 and 2.2, and can also be seen from (2.16),  $\tilde{F}(v, k)$  is an odd function of  $v$ . Therefore we immediately know, in this monopole doublet case, that

$$\tilde{F}_r^c(k) = 0 \quad (19.2.22)$$

since the  $ce_r(v, q)$  are even functions of  $v$ . Moreover, these figures and (2.16) show that  $\tilde{F}(v, k)$  is symmetric about the values  $v = \pm\pi/2$ . Since the  $se_r(v, q)$  for even  $r$  are antisymmetric about  $v = \pm\pi/2$ , we conclude, again in this monopole doublet case, that the only nonvanishing angular integrals will be

$$\tilde{F}_r^s(k) = (1/\pi) \int_0^{2\pi} dv se_r(v, q) \tilde{F}(v, k) \quad (19.2.23)$$

with  $r = 1, 3, 5, 7 \dots$ .

Finally, what range of  $q$  values is of interest? According to (14.4.23) and (2.2),  $q$  and  $k$  are connected in this instance by the relation

$$q = -k^2 f^2 / 4 = -3k^2. \quad (19.2.24)$$

If we use  $k$  values in the range  $k \in [-K_c, K_c]$ ,  $q$  will lie in the range  $q \in [q_{\min}, 0]$  with

$$q_{\min} = -3K_c^2. \quad (19.2.25)$$

Since the fields on the surface of the elliptical cylinder are no more singular than those on the surface of the circular cylinder, we could set  $K_c = 20$  as done before for the circular cylinder case. Doing so yields  $q_{\min} = -1200$ . Or, being less conservative, we might use  $K_c = 10$ , in which case  $q_{\min} = -300$ . This is the extreme  $q$  value used in making Figures 14.4.13 through 14.4.21.

With this background in mind, examine Figures 2.4 and 2.5. They show the Mathieu coefficients  $\tilde{F}_r^s(k)$  as a function of  $k$  for the cases  $r = 1$  through  $r = 11$  and  $r = 17$  through  $r = 25$ . The curves for the intervening values  $r = 13$  and  $r = 15$  behave analogously. We see that all the  $\tilde{F}_r^s(k)$  tend to zero with increasing  $|k|$ .

This large  $|k|$  behavior can be understood as follows: Suppose  $f(v)$  and  $g(v)$  are any two  $2\pi$  periodic functions. Into the vector space of such functions introduce a scalar product by the usual rule

$$(f, g) = [1/(2\pi)] \int_0^{2\pi} dv \bar{f}(v)g(v) \quad (19.2.26)$$

where a bar denotes complex conjugation. Also define a norm by the usual rule

$$(\|f\|)^2 = (f, f). \quad (19.2.27)$$

Then, by the Schwarz inequality, there is the result

$$|(f, g)| \leq \|f\| \|g\|. \quad (19.2.28)$$

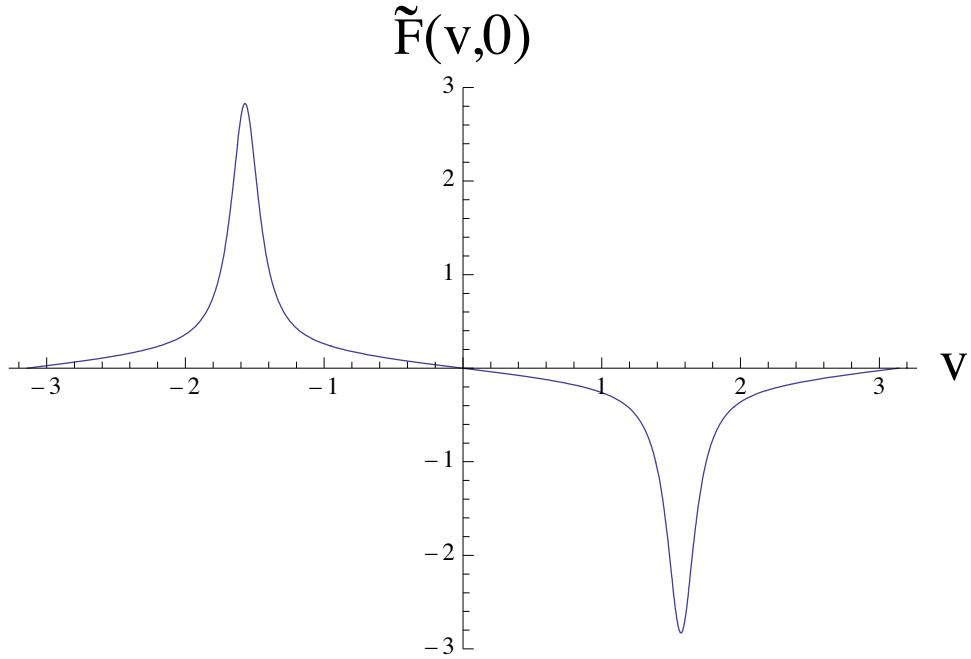


Figure 19.2.1: The real part of  $\tilde{F}(v, k = 0)$  for the monopole doublet in the case that  $x^{\max} = 4$  cm,  $y^{\max} = 2$  cm,  $a = 2.5$  cm, and  $g = 1$  Tesla-(cm) $^2$ . The imaginary part vanishes.

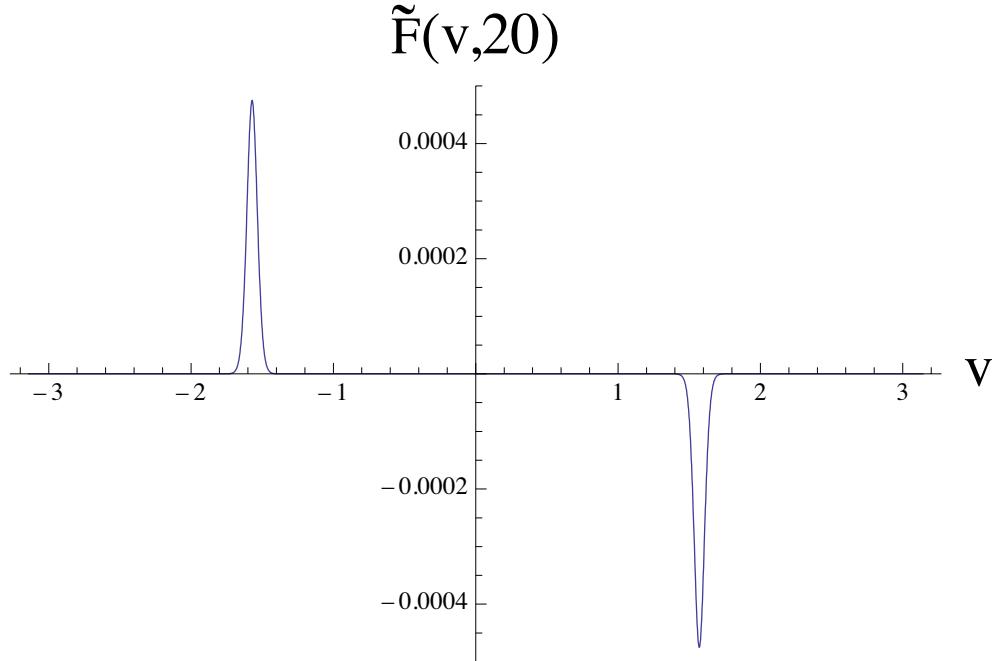


Figure 19.2.2: The real part of  $\tilde{F}(v, k = 20)$  for the monopole doublet in the case that  $x^{\max} = 4$  cm,  $y^{\max} = 2$  cm,  $a = 2.5$  cm, and  $g = 1$  Tesla-(cm) $^2$ . The imaginary part vanishes.

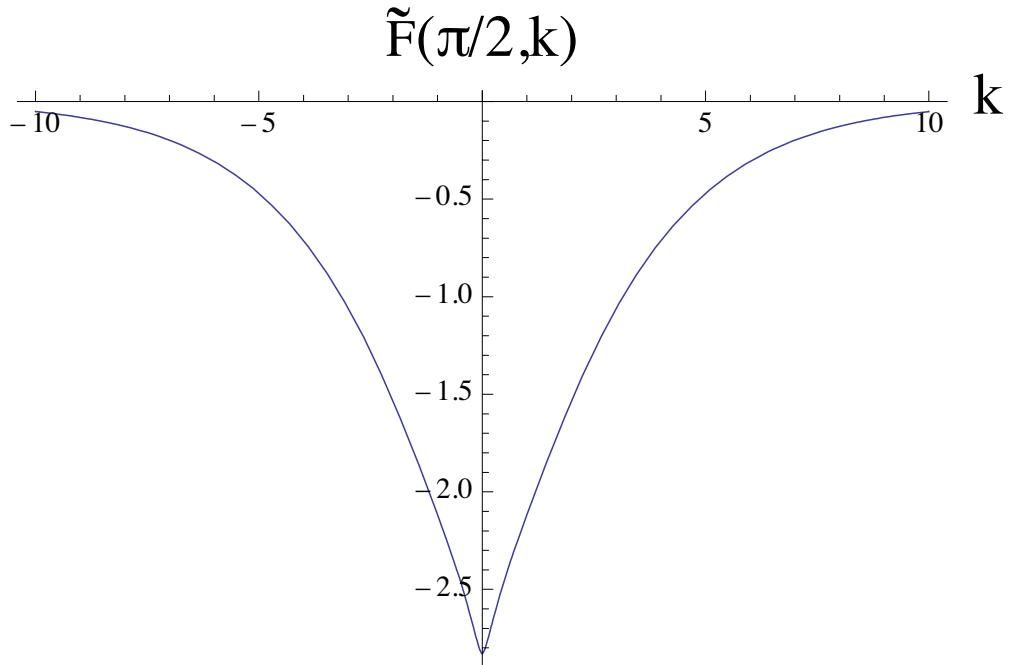


Figure 19.2.3: The real part of  $\tilde{F}(v = \pi/2, k)$  for the monopole doublet in the case that  $x^{\max} = 4$  cm,  $y^{\max} = 2$  cm,  $a = 2.5$  cm, and  $g = 1$  Tesla-(cm) $^2$ . The imaginary part vanishes.

With this notation in mind, we see that (2.23) can be rewritten in the form

$$\tilde{F}_r^s(k) = 2(\mathbf{s}\mathbf{e}_r, \tilde{F}). \quad (19.2.29)$$

Therefore, using (2.28), we have the inequality

$$|\tilde{F}_r^s(k)| \leq 2\|\mathbf{s}\mathbf{e}_r\| \|\tilde{F}\|. \quad (19.2.30)$$

Also, from the normalization (14.4.39), we see that

$$\|\mathbf{s}\mathbf{e}_r\| = 1/\sqrt{2}. \quad (19.2.31)$$

We conclude that there is the  $r$  independent bound

$$|\tilde{F}_r^s(k)| \leq \sqrt{2} \|\tilde{F}\| \quad (19.2.32)$$

where

$$(\|\tilde{F}\|)^2 = [1/(2\pi)] \int_0^{2\pi} dv [\tilde{F}(v, k)]^2. \quad (19.2.33)$$

(See also Exercise 2.1.) Figure 2.6 displays the quantity  $\sqrt{2} \|\tilde{F}\|$  as a function of  $k$ . Evidently it also decreases with increasing  $|k|$ . In fact, we can get a loose bound on  $\|\tilde{F}\|$  from its definition (2.33) by estimating the integral,

$$(\|\tilde{F}\|)^2 = [1/(2\pi)] \int_0^{2\pi} dv [\tilde{F}(v, k)]^2 \leq [\tilde{F}(\pi/2, k)]^2 \quad (19.2.34)$$

from which it follows that

$$||\tilde{F}|| \leq |\tilde{F}(\pi/2, k)|. \quad (19.2.35)$$

Here we have used the fact that  $[\tilde{F}(v, k)]^2$  takes its maxima at  $v = \pm\pi/2$ . See Figures 2.1 and 2.2. We already know that  $|\tilde{F}(\pi/2, k)|$  decreases exponentially with increasing  $|k|$ . Again see Figure 2.3 and (2.16) through (2.19). We conclude that the  $|\tilde{F}_r^s(k)|$  must all decrease at least this rapidly as well.

In fact there are two reasons why, for fixed  $r$ , the  $|\tilde{F}_r^s(k)|$  must eventually decrease even more rapidly as  $|k| \rightarrow \infty$ . First, as comparison of Figures 2.1 and 2.2 indicates, the peaks in  $\tilde{F}(v, k)$  at  $v = \pm\pi/2$  become more narrow with increasing  $|k|$ . Second, look at (2.21). We expect that the greatest contribution to this integral will come from  $v = \pm\pi/2$  because  $\tilde{F}(v, k)$  is peaked there. But it is precisely at these  $v$  values that the  $se_r(v, q)$  vanish rapidly as  $q \rightarrow -\infty$  since these  $v$  values are in the middle of the forbidden region. See, for example, Figures 14.4.19 and 14.4.20. To observe these considerations in action for the case of  $|\tilde{F}_r^s(k)|$  with  $r = 29$ , see Figure 2.30 in Exercise 2.4.

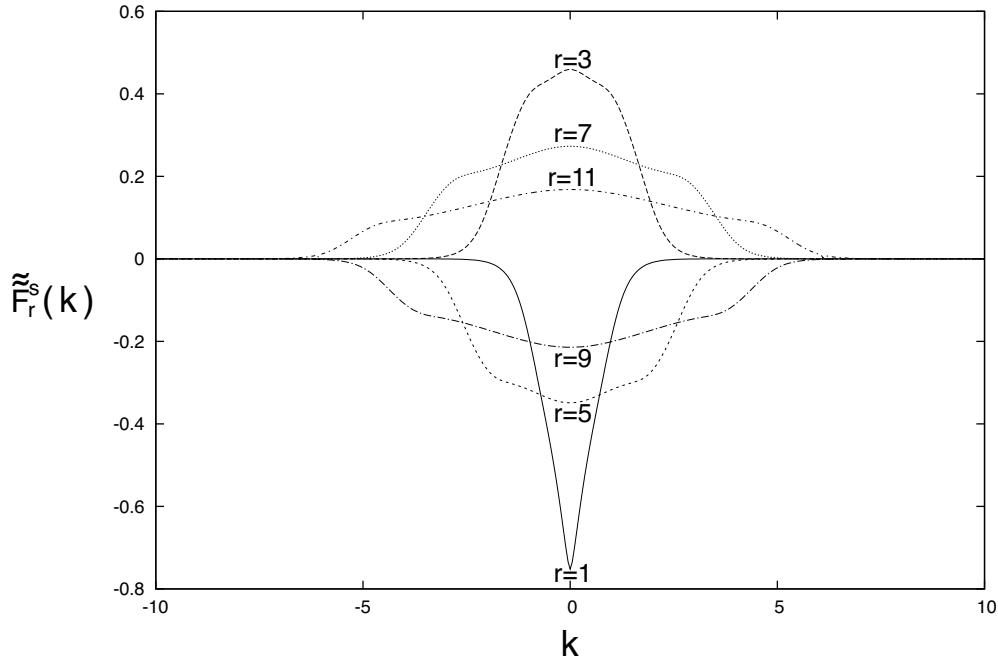


Figure 19.2.4: The real parts of the Mathieu coefficients  $\tilde{F}_r^s(k)$  as a function of  $k$ , with  $r = 1, 3, 5, 7, 9, 11$ , for the monopole doublet in the case that  $x^{\max} = 4$  cm,  $y^{\max} = 2$  cm,  $a = 2.5$  cm, and  $g = 1$  Tesla-(cm) $^2$ . The imaginary parts vanish. The solid curve, the one with the largest negative excursion at  $k = 0$ , is that for  $r = 1$ . The curves alternate in sign, and the magnitudes of their values at  $k = 0$  decrease, for each successive value of  $r$ . For example, the curve with the largest positive excursion at  $k = 0$  is that for  $r = 3$ .

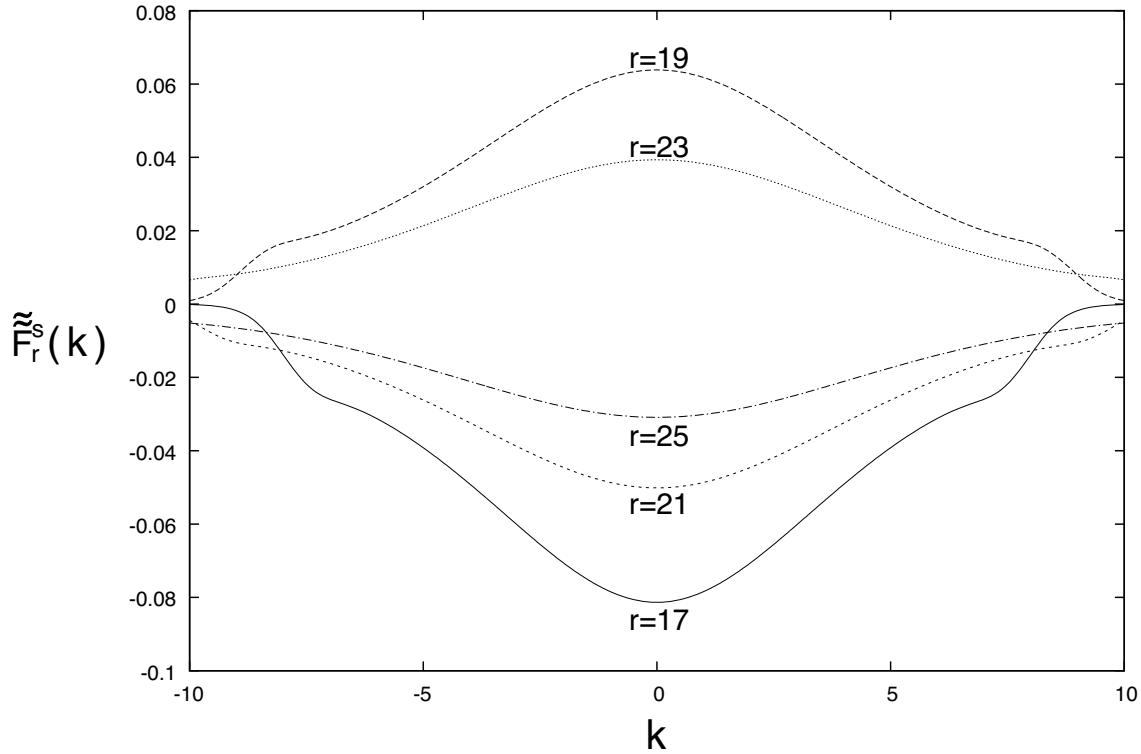


Figure 19.2.5: The real parts of the Mathieu coefficients  $\tilde{F}_r^s(k)$  as a function of  $k$ , with  $r = 17, 19, 21, 23, 25$ , for the monopole doublet in the case that  $x^{\max} = 4$  cm,  $y^{\max} = 2$  cm,  $a = 2.5$  cm, and  $g = 1$  Tesla-(cm) $^2$ . The imaginary parts vanish. The solid curve, the one with the largest negative excursion at  $k = 0$ , is that for  $r = 17$ . The curves alternate in sign, and the magnitudes of their values at  $k = 0$  decrease for each successive value of  $r$ . For example, the curve with the largest positive excursion at  $k = 0$  is that for  $r = 19$ .

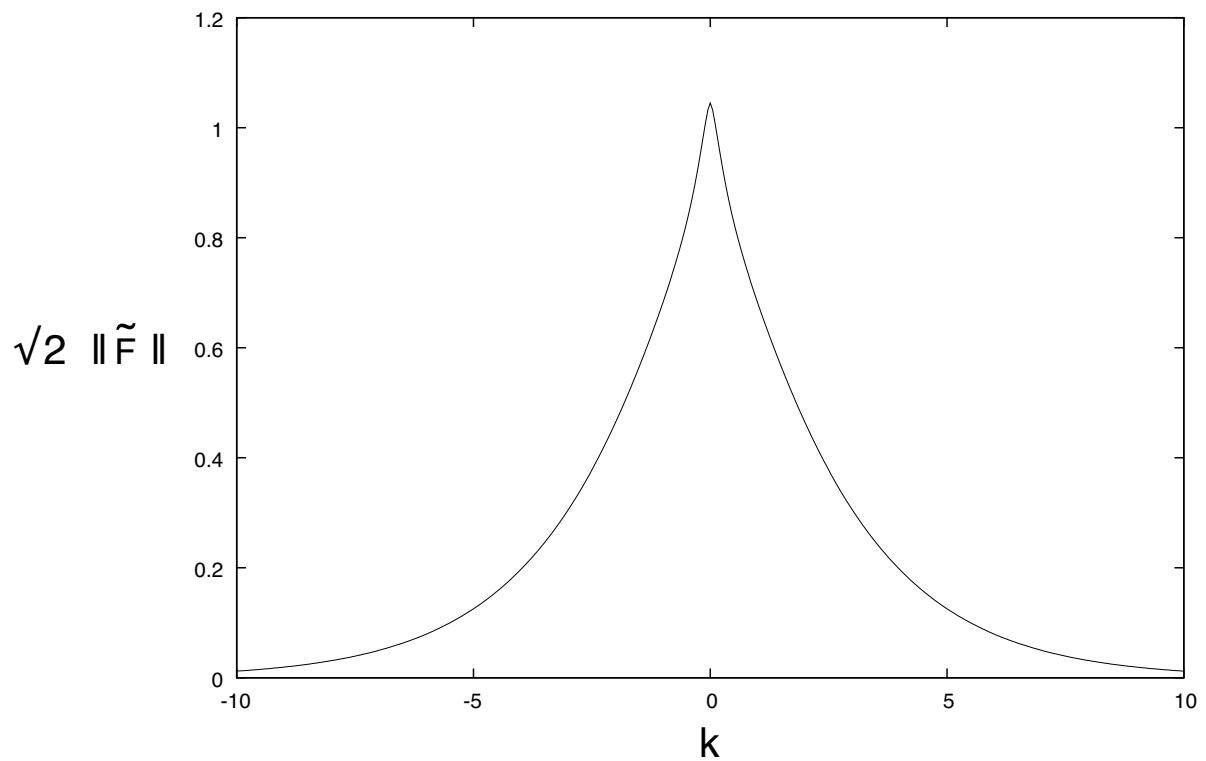


Figure 19.2.6: The quantity  $\sqrt{2} \|\tilde{F}\|$  as a function of  $k$  for the monopole doublet in the case that  $x^{\max} = 4$  cm,  $y^{\max} = 2$  cm,  $a = 2.5$  cm, and  $g = 1$  Tesla-(cm) $^2$ .

### 19.2.2 Behavior of Kernels

We have determined the behavior of the Mathieu coefficients  $\tilde{F}_r^s(k)$  for the case of the monopole doublet. In analogy with our previous discussion of the circular cylinder case in Section 16.1.1, let us next examine the kernels  $k^m \beta_m^r(k)/\text{Se}'_r(U, q)$  that appear in (14.4.86). Figure 2.7 shows the kernels  $k^m \beta_m^r(k)/\text{Se}'_r(U, q)$  for the case  $m = 1$  and  $r = 1, 3, 5, 7, 9, 11$ . We see that each kernel has constant sign and the absolute value of each goes monotonically to 0 as  $|k| \rightarrow \infty$ . Also, as Figure 2.8 shows, their absolute values at  $k = 0$  go monotonically to zero as  $r$  increases.

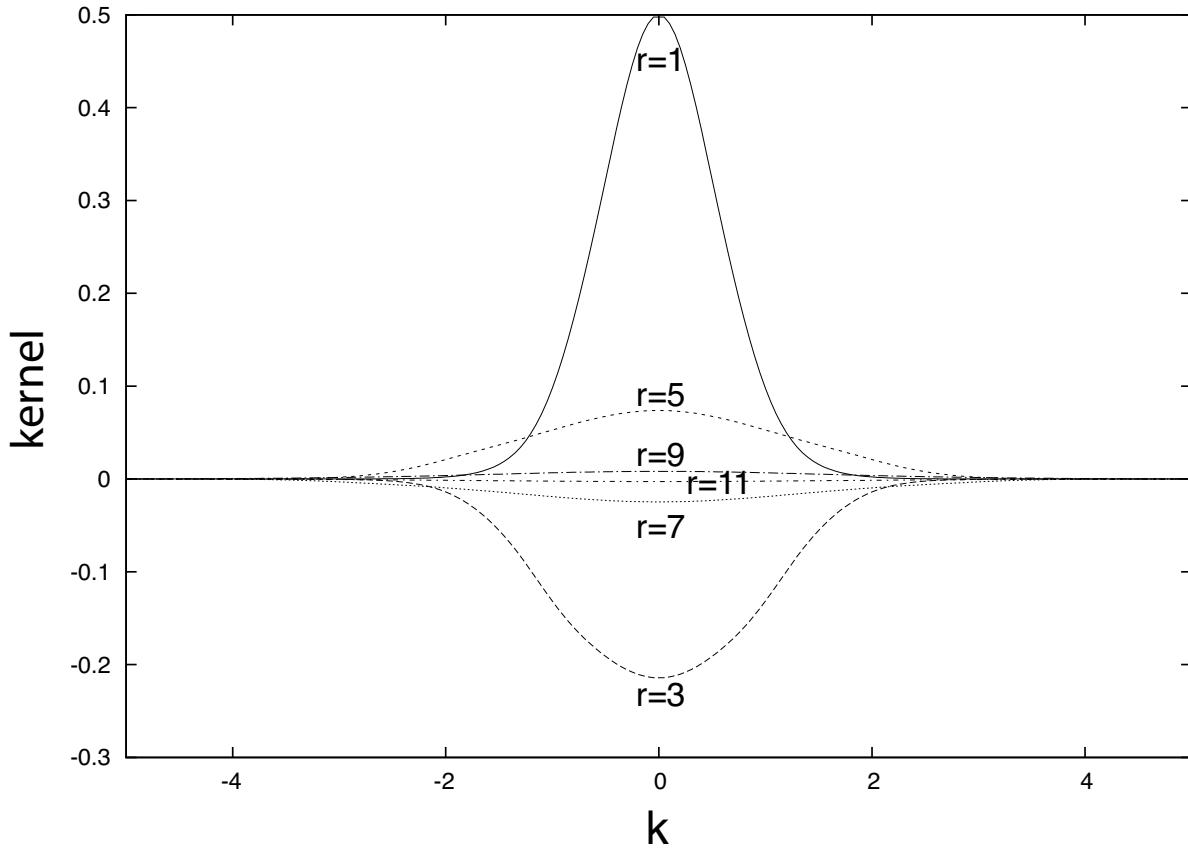


Figure 19.2.7: The kernels  $k^m \beta_m^r(k)/\text{Se}'_r(U, q)$  for the case  $m = 1$  and  $r = 1, 3, 5, 7, 9, 11$ , as a function of  $k$ , with  $q$  and  $k$  related by (2.24) and  $U$  given by (2.1). The kernel for  $r = 1$  is the one with the largest positive value at  $k = 0$ . Kernels for successive values of  $r$  alternate in sign. Their absolute values at  $k = 0$  decrease monotonically with increasing  $r$ .

Figure 2.9 shows the kernels  $k^m \beta_m^r(k)/\text{Se}'_r(U, q)$  for the case  $m = 7$  and  $r = 1, 3, 5$ . Figure 2.10 shows the kernels  $k^m \beta_m^r(k)/\text{Se}'_r(U, q)$  for the case  $m = 7$  and  $r = 7, 9, 11$ . Figure 2.11 shows the kernels  $k^m \beta_m^r(k)/\text{Se}'_r(U, q)$  for the case  $m = 7$  and  $r = 13, 15, 17, 21, 23$ . We see that the kernels for  $r < m$  vanish at  $k = 0$ . Those with  $r \geq m$  are finite at  $k = 0$ , but ultimately go to 0 as  $r \rightarrow \infty$ . See Figure 2.12. Finally, all kernels go rapidly to zero as  $|k| \rightarrow \infty$ . We have documented the extreme cases  $m = 1$  and  $m = 7$ . The intermediate cases  $m = 3$  and  $m = 5$  are similar to the  $m = 7$  case: the kernels for  $r < m$  vanish at

## | kernel |

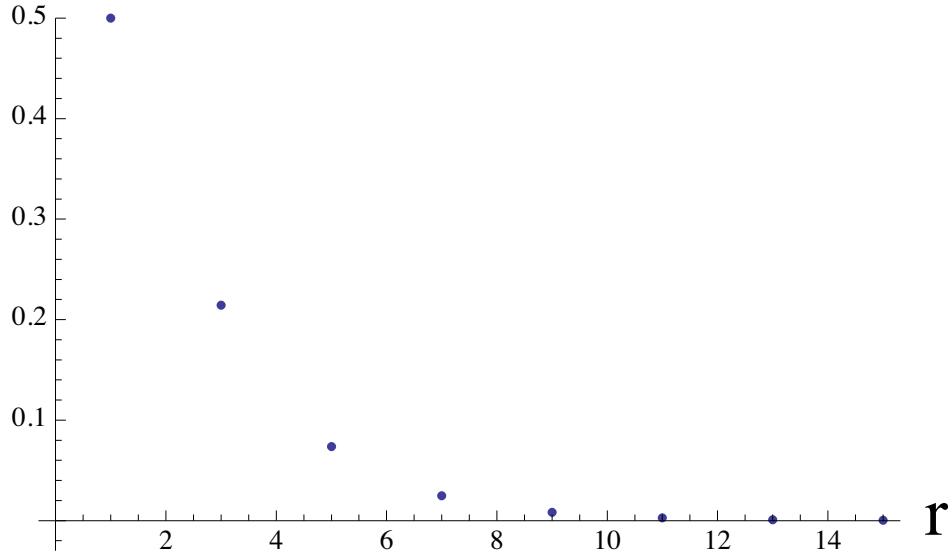


Figure 19.2.8: Absolute values of the kernels  $k^m \beta_m^r(k)/\text{Se}_r'(U, q)$  evaluated at  $k = 0$  for the case  $m = 1$  and  $r \in [1, 15]$  with  $U$  given by (2.1).

$k = 0$ ; those with  $r \geq m$  are finite at  $k = 0$ , but ultimately go to 0 as  $r \rightarrow \infty$ ; all kernels go rapidly to zero as  $|k| \rightarrow \infty$ .

### 19.2.3 Truncation of Series

We have studied the Mathieu coefficients  $\tilde{F}_r^s(k)$  and the kernels  $k^m \beta_m^r(k)/\text{Se}_r'(U, q)$ . Next, we need to study their combinations as they occur in (14.4.84) and (14.4.86). In particular, let us look at the quantities  $(1/2)^m (1/m!) k^m G_{m,s}(k)$ . Of course, in the monopole doublet case, we also know exactly what the result should be. Comparison of (14.4.86) and (1.1) gives the relation

$$(1/2)^m (1/m!) k^m G_{m,s}(k) = \tilde{C}_{m,s}^{[0]}(k), \quad (19.2.36)$$

and we know the right side of (2.36) from (1.20).

Figure 2.13 shows their values for the cases  $m = 1, 3, 5, 7$ . Figures 2.14 through 2.16 show their values for the cases  $m = 3, 5, 7$  separately. These are the cases that we need for our magnet monopole doublet example.

As described in the beginning of this subsection, we must truncate the infinite sums over  $r$  that occur in (14.4.83) through (14.4.86) in order to obtain practical results. We will do this by assuming that the truncation error is comparable to the size of the last retained term.<sup>13</sup> By this criterion, we retained all terms with values of  $r$  through  $r = r_{\max}(m)$  with  $r_{\max}(m) = 11, 19, 25, 29$  for the cases  $m = 1, 3, 5, 7$ , respectively. Figures 2.17 through

---

<sup>13</sup>This assumption is justified because both the Mathieu coefficients  $\tilde{F}_r^s(k)$  and the kernels  $k^m \beta_m^r(k)/\text{Se}_r'(U, q)$  fall off exponentially in  $r$  for large  $r$ .

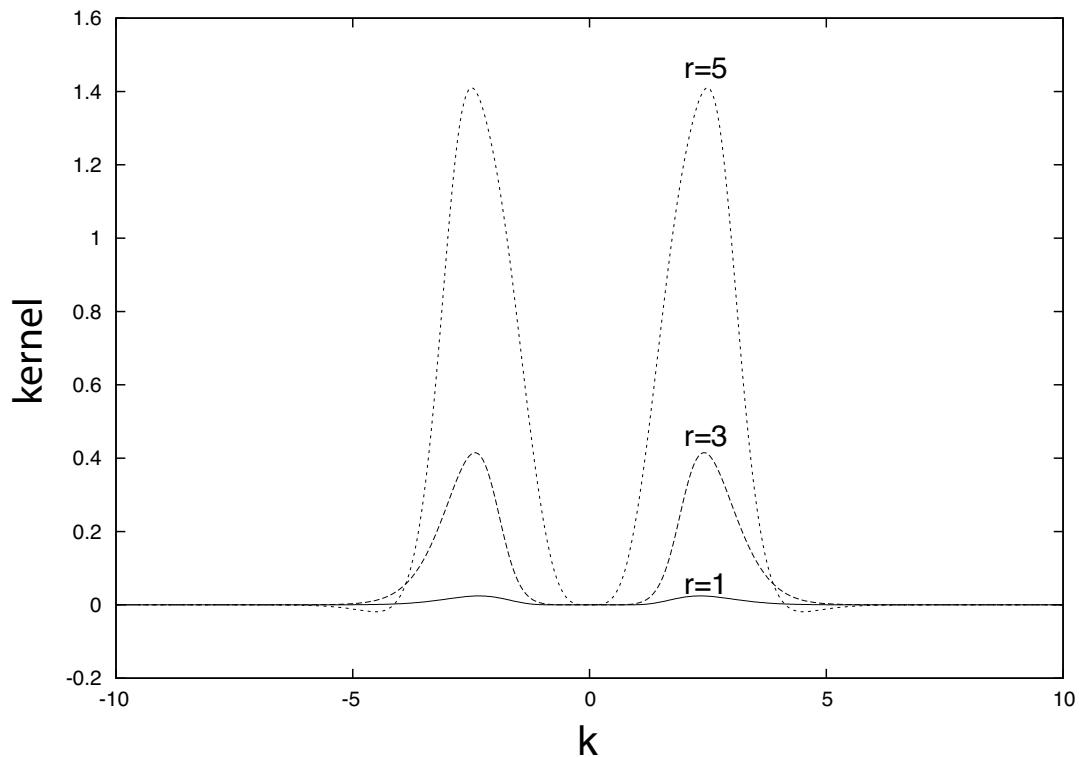


Figure 19.2.9: The kernels  $k^m \beta_m^r(k) / \text{Se}'_r(U, q)$  for the case  $m = 7$  and  $r = 1, 3, 5$ , as a function of  $k$ , with  $q$  and  $k$  related by (2.24) and  $U$  given by (2.1). The kernel that has the largest positive value is that for  $r = 5$ . The kernel with the next largest positive value is that for  $r = 3$ . The remaining kernel is that for  $r = 1$ .

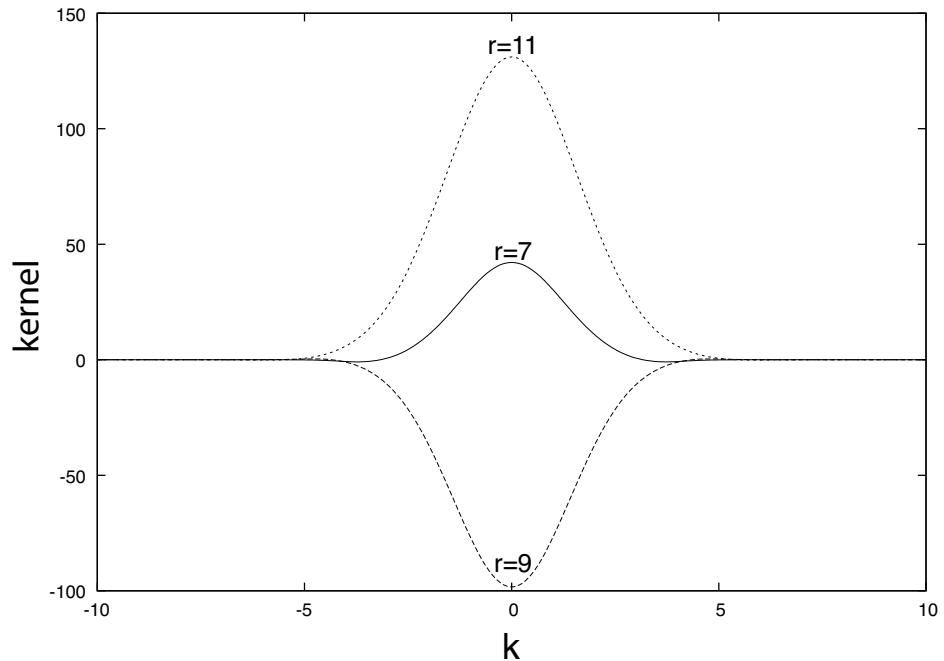


Figure 19.2.10: The kernels  $k^m \beta_m^r(k) / \text{Se}'_r(U, q)$  for the case  $m = 7$  and  $r = 7, 9, 11$ , as a function of  $k$ , with  $q$  and  $k$  related by (2.24) and  $U$  given by (2.1). The kernel for  $r = 7$  is the one with the smallest positive value at  $k = 0$ . Kernels for successive values of  $r$  alternate in sign. Their magnitudes at  $k = 0$  increase monotonically with increasing  $r$  in the range  $r \in [7, 11]$ .

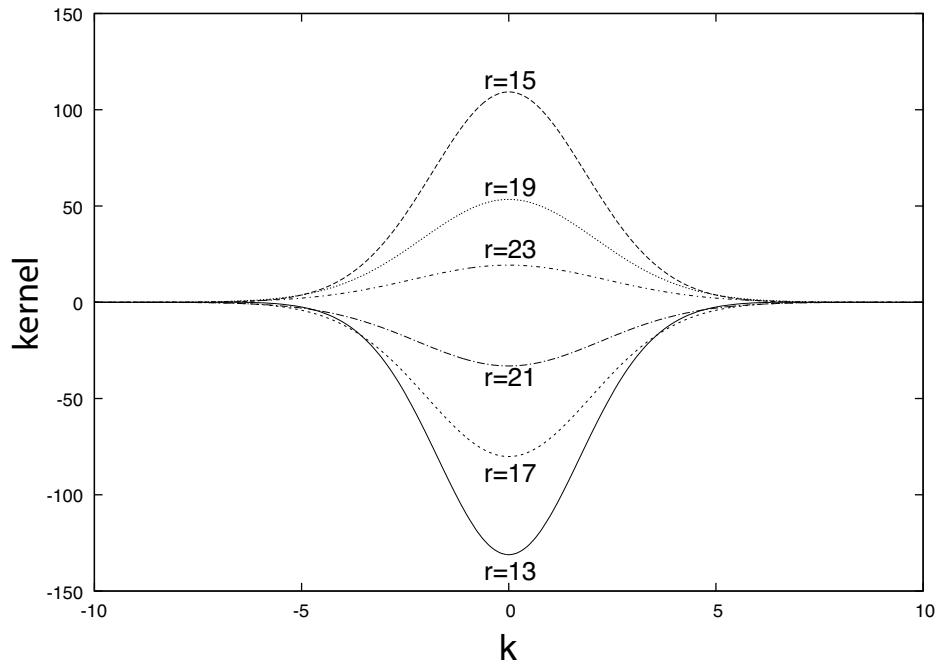


Figure 19.2.11: The kernels  $k^m \beta_m^r(k)/\text{Se}'_r(U, q)$  for the case  $m = 7$  and  $r = 13, 15, 17, 19, 21, 23$ , as a function of  $k$ , with  $q$  and  $k$  related by (2.24) and  $U$  given by (2.1). The kernel for  $r = 13$  is the one with the largest negative value at  $k = 0$ . Kernels for successive values of  $r$  alternate in sign. Their magnitudes at  $k = 0$  decrease monotonically with increasing  $r$  in the range  $r \in [13, 23]$ .

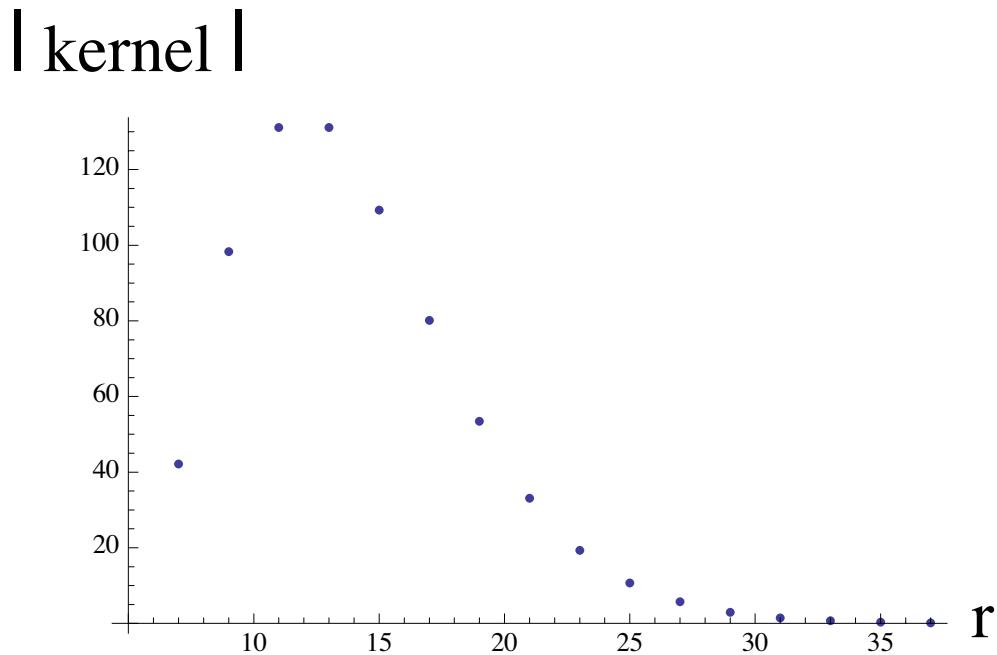


Figure 19.2.12: Absolute values of the kernels  $k^m \beta_m^r(k)/\text{Se}'_r(U, q)$  evaluated at  $k = 0$  for the case  $m = 7$  and  $r \in [7, 37]$  with  $U$  given by (2.1).

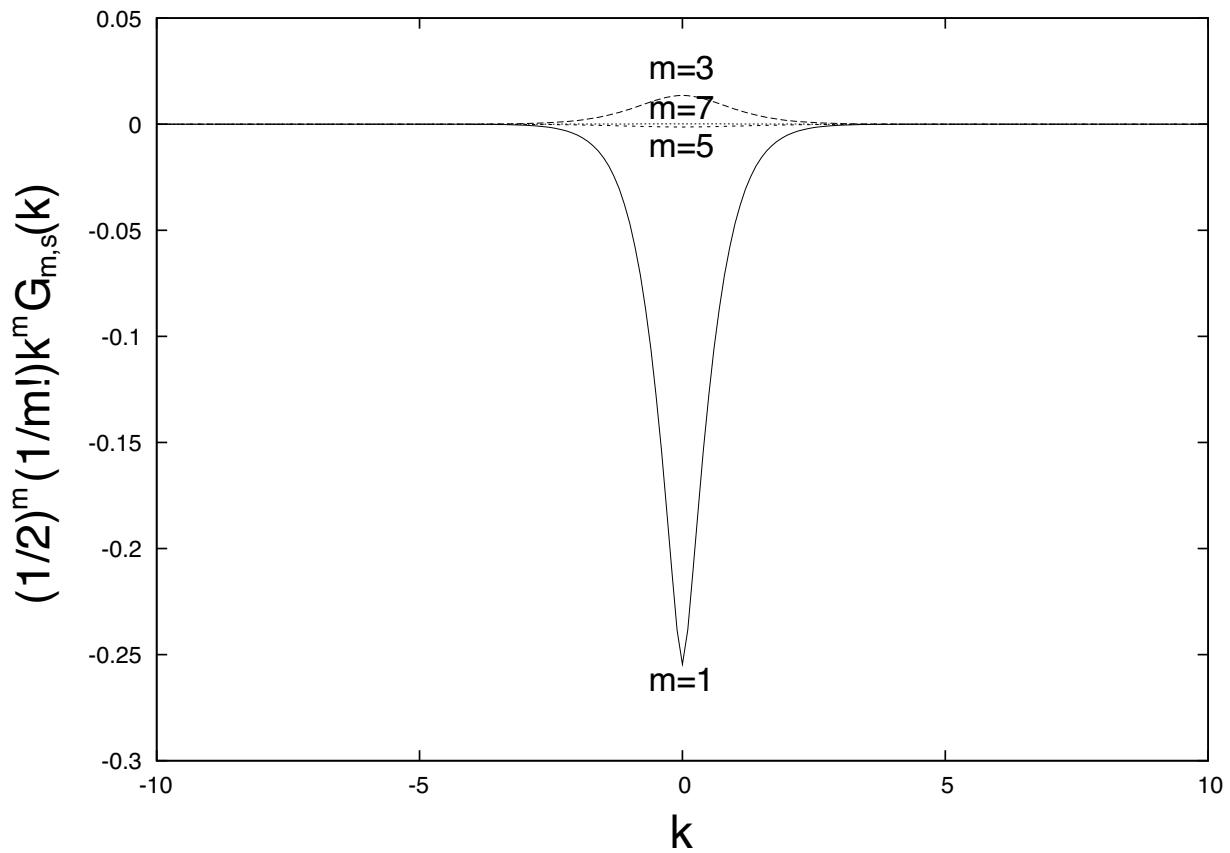


Figure 19.2.13: The real parts of  $(1/2)^m (1/m!) k^m G_{m,s}(k)$  for the monopole doublet when  $m = 1, 3, 5, 7$ . The imaginary parts vanish. The quantities decrease in magnitude with increasing  $m$ . For example, the curve with the largest negative value at  $k = 0$  is that for  $m = 1$ .

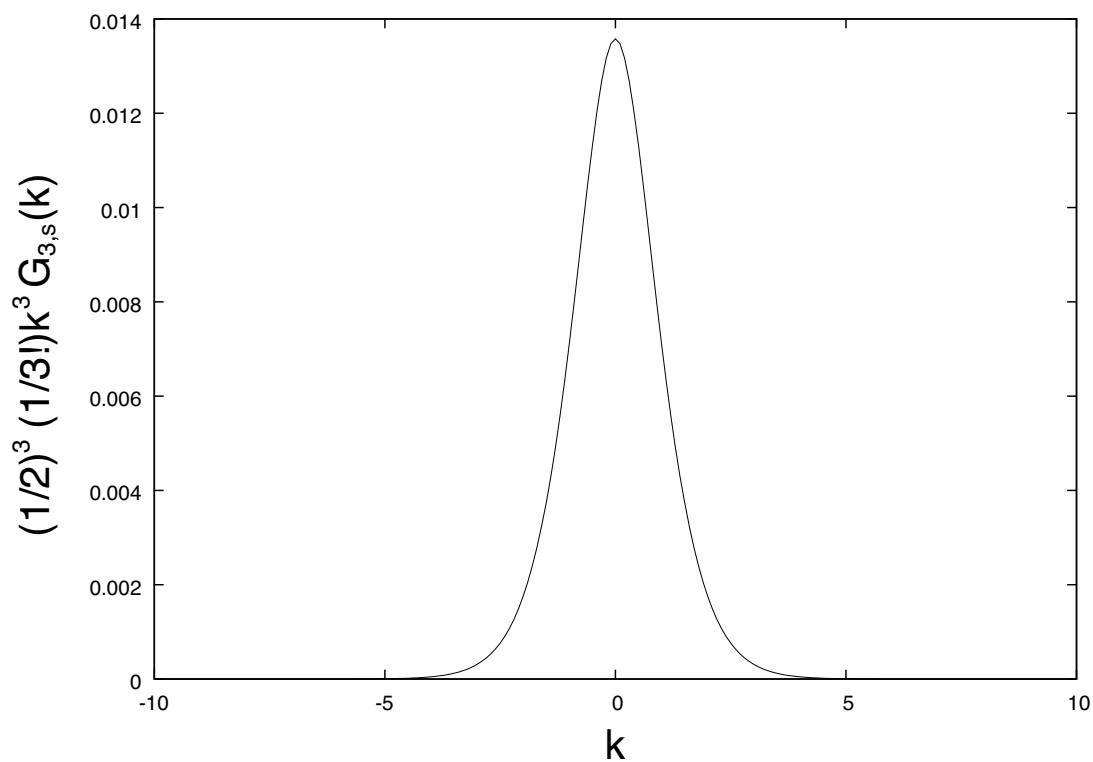


Figure 19.2.14: The real part of  $(1/2)^m (1/m!) k^m G_{m,s}(k)$  for  $m = 3$ .

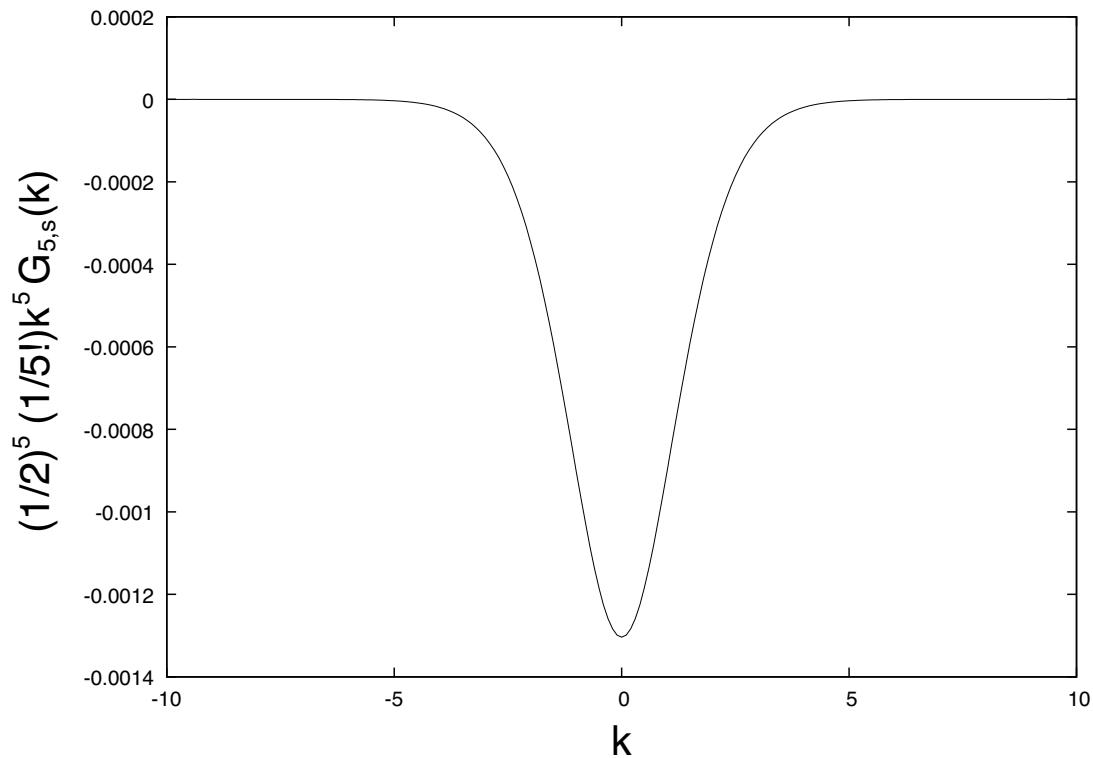


Figure 19.2.15: The real part of  $(1/2)^m (1/m!) k^m G_{m,s}(k)$  for  $m = 5$ .

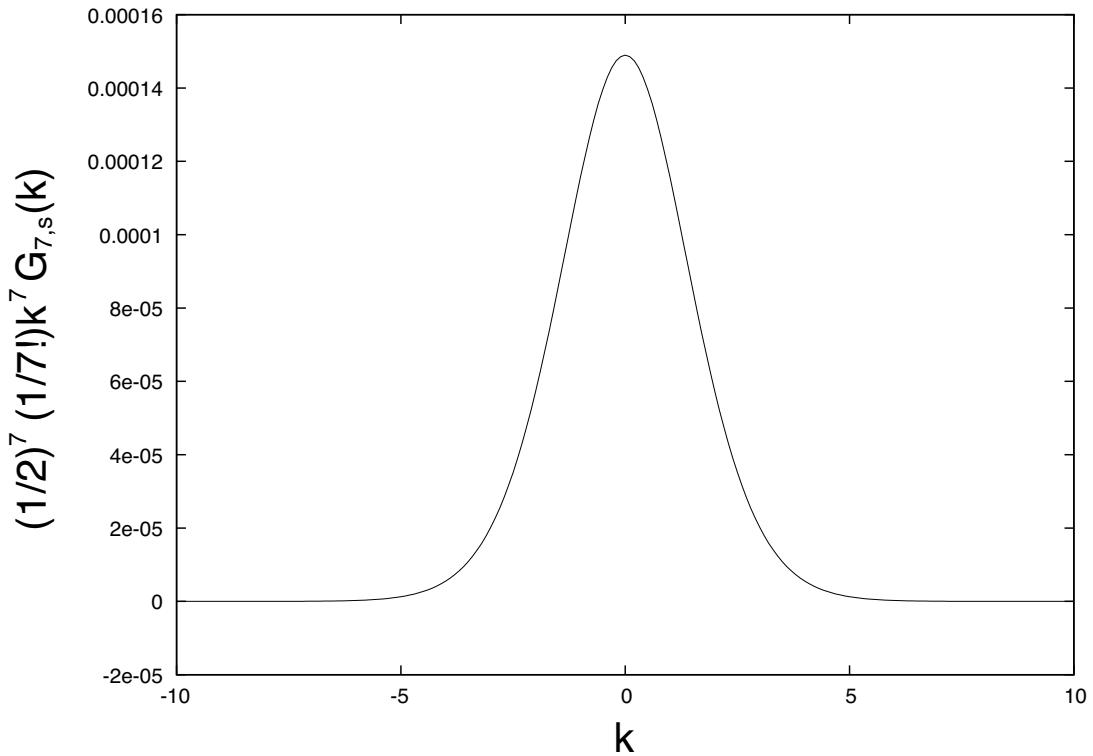


Figure 19.2.16: The real part of  $(1/2)^m (1/m!) k^m G_{m,s}(k)$  for  $m = 7$ .

2.20 show the last retained term in each case. Judging from Figures 2.13 through 2.16, we estimate the errors to be 10,4,3,6 parts in  $10^4$  for the cases  $m = 1, 3, 5, 7$  respectively.

Figures 2.21 through 2.24 show the actual truncation error, the difference between the truncated result and the exact result (2.36). Comparison of Figures 2.13 through 2.16 and Figures 2.21 through 2.24 shows that the actual truncation errors are 3,2,2,4 parts in  $10^4$  for  $m = 1, 3, 5, 7$ , respectively.

#### 19.2.4 Approximation of Angular Integrals by Riemann Sums

At this point we have verified that the truncation criterion is adequate and that the routines used to compute the Mathieu functions and the Mathieu-Bessel function connection coefficients are working properly. The next step is to explore the replacement of angular integrals by Riemann sums.

Select  $N$  discrete  $v$  values  $v_n$  according to the rule

$$v_n = n(2\pi/N) \text{ for } n = 0, 1, \dots, N-1. \quad (19.2.37)$$

What we want to do is to approximate the integrals (2.20) and (2.21) by the Riemann sums

$$\tilde{F}_r^c(k) \approx (2/N) \sum_{n=0}^{N-1} \text{ce}_r(v_n, q) \tilde{F}(v_n, k), \quad (19.2.38)$$

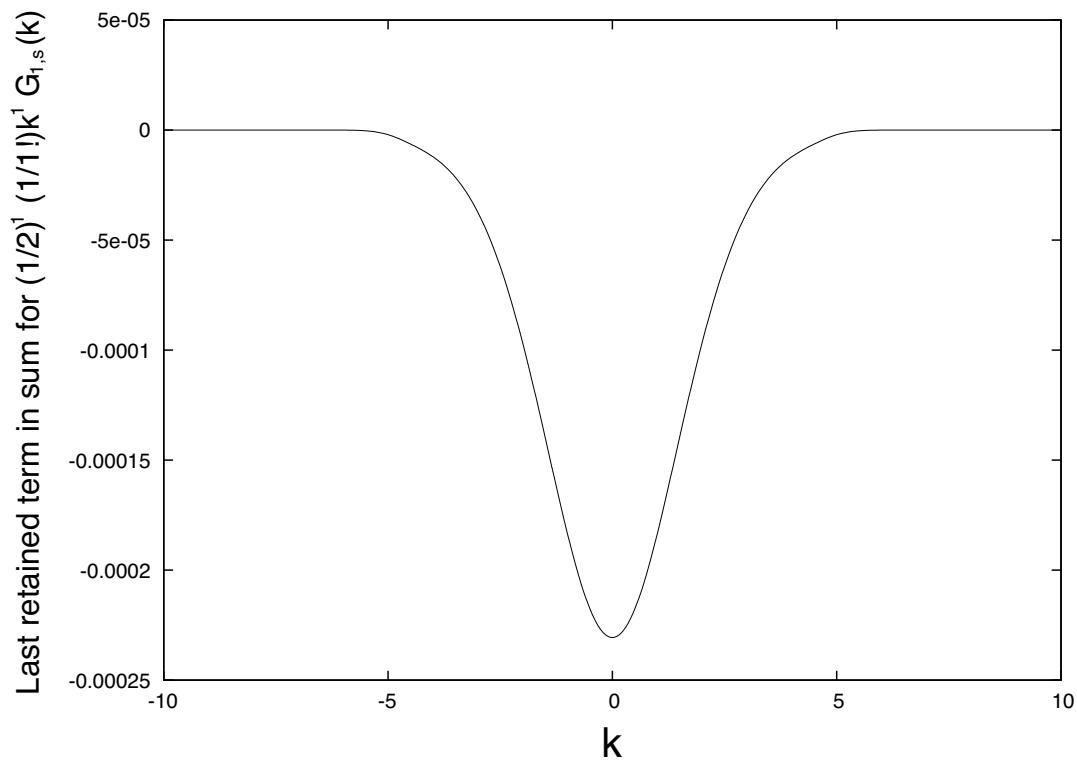


Figure 19.2.17: The real part of the last retained term in the sum for  $(1/2)^m (1/m!) k^m G_{m,s}(k)$  with  $m = 1$  based on truncating the series (14.4.84) beyond  $r = r_{\max}(1) = 11$ . The imaginary part vanishes.

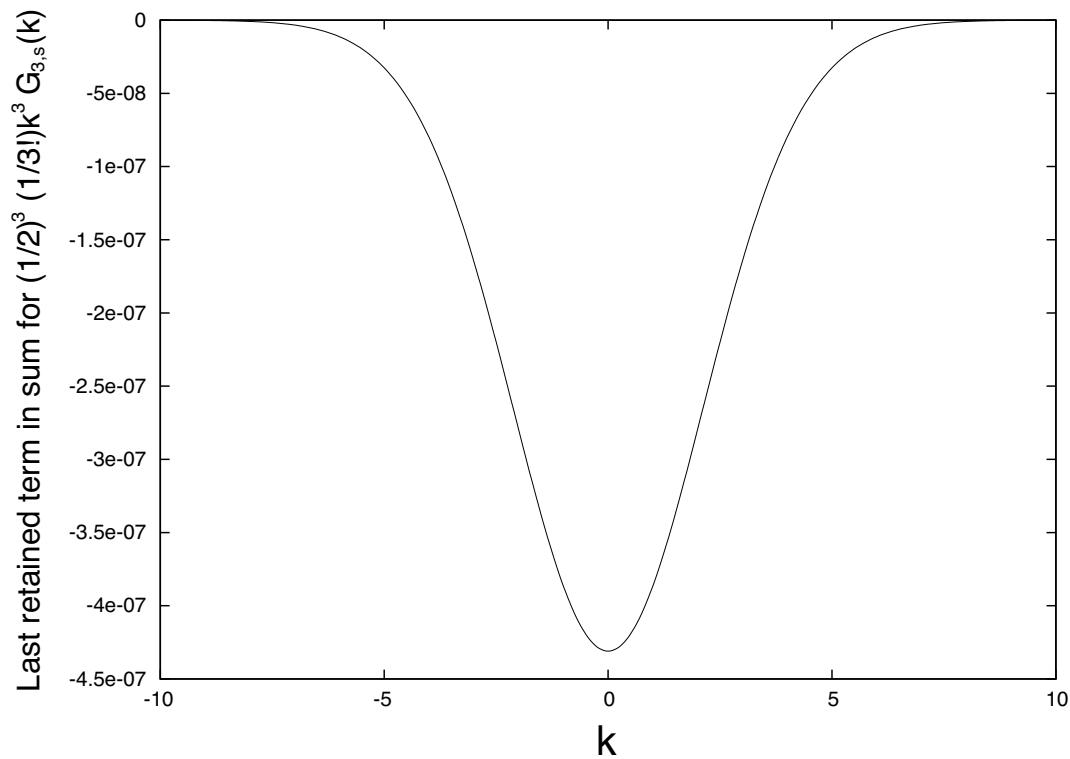


Figure 19.2.18: The real part of the last retained term in the sum for  $(1/2)^m (1/m!) k^m G_{m,s}(k)$  with  $m = 3$  based on truncating the series (14.4.84) beyond  $r = r_{\max}(3) = 19$ . The imaginary part vanishes.

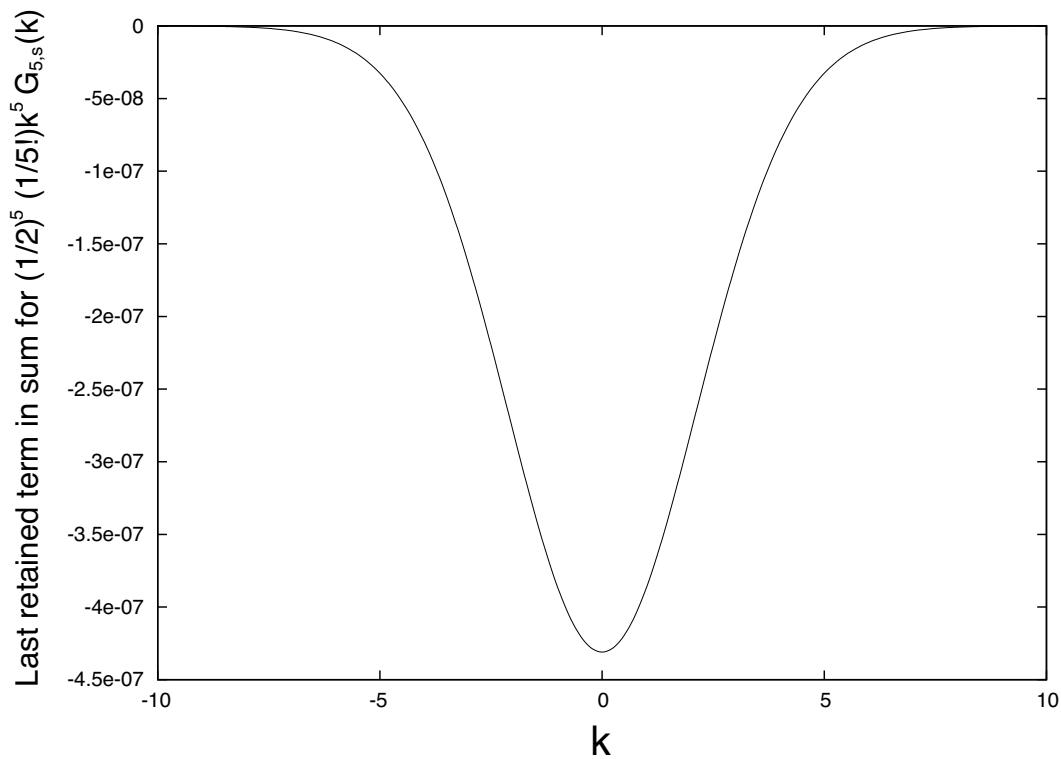


Figure 19.2.19: The real part of the last retained term in the sum for  $(1/2)^m (1/m!) k^m G_{m,s}(k)$  with  $m = 5$  based on truncating the series (14.4.84) beyond  $r = r_{\max}(5) = 25$ . The imaginary part vanishes.

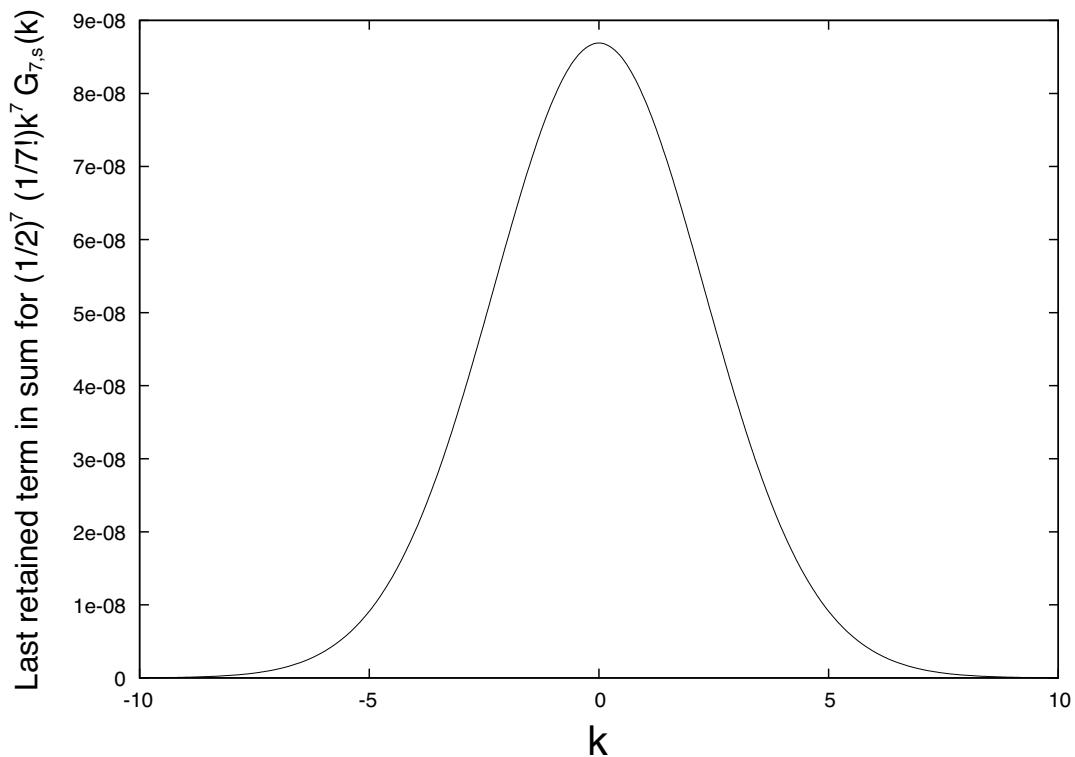


Figure 19.2.20: The real part of the last retained term in the sum for  $(1/2)^m(1/m!)k^mG_{m,s}(k)$  with  $m = 7$  based on truncating the series (14.4.84) beyond  $r = r_{\max}(7) = 29$ . The imaginary part vanishes.

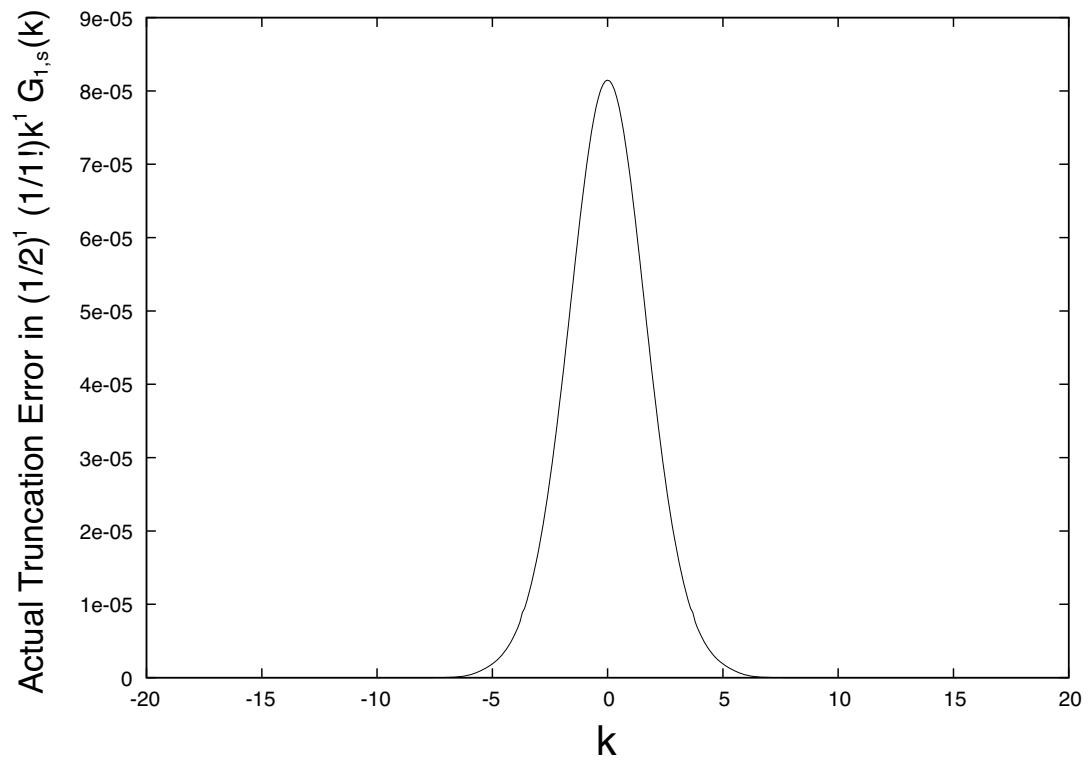


Figure 19.2.21: Real part of actual truncation error in  $(1/2)^m (1/m!) k^m G_{m,s}(k)$  for  $m = 1$  produced by truncating the series (14.4.84) beyond  $r = r_{\max}(1) = 11$ . The imaginary part vanishes.

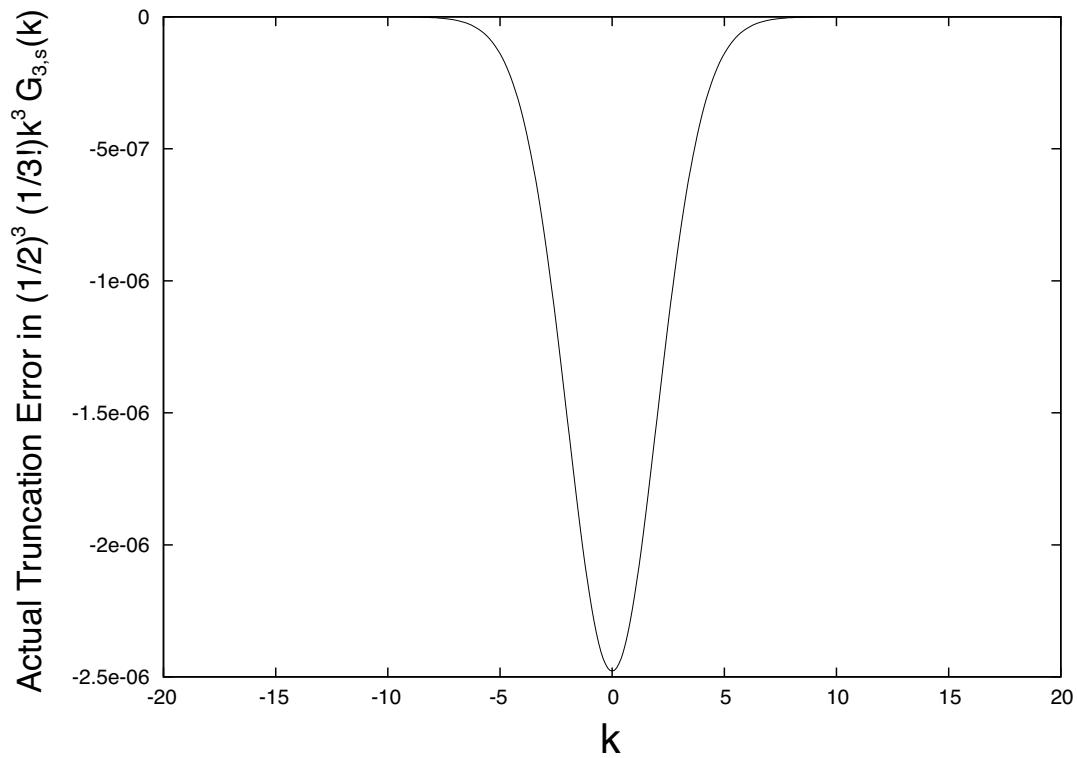


Figure 19.2.22: Real part of actual truncation error in  $(1/2)^m(1/m!)k^mG_{m,s}(k)$  for  $m = 3$  produced by truncating the series (14.4.84) beyond  $r = r_{\max}(3) = 19$ . The imaginary part vanishes.

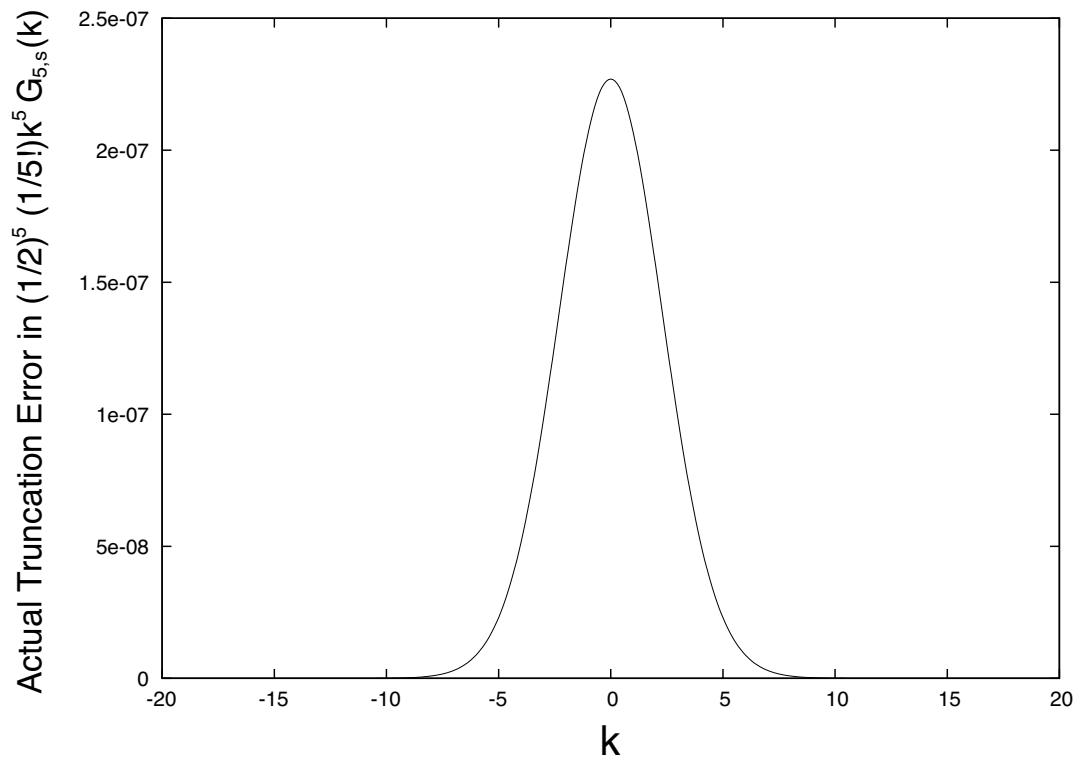


Figure 19.2.23: Real part of actual truncation error in  $(1/2)^m (1/m!) k^m G_{m,s}(k)$  for  $m = 5$  produced by truncating the series (14.4.84) beyond  $r = r_{\max}(5) = 25$ . The imaginary part vanishes.

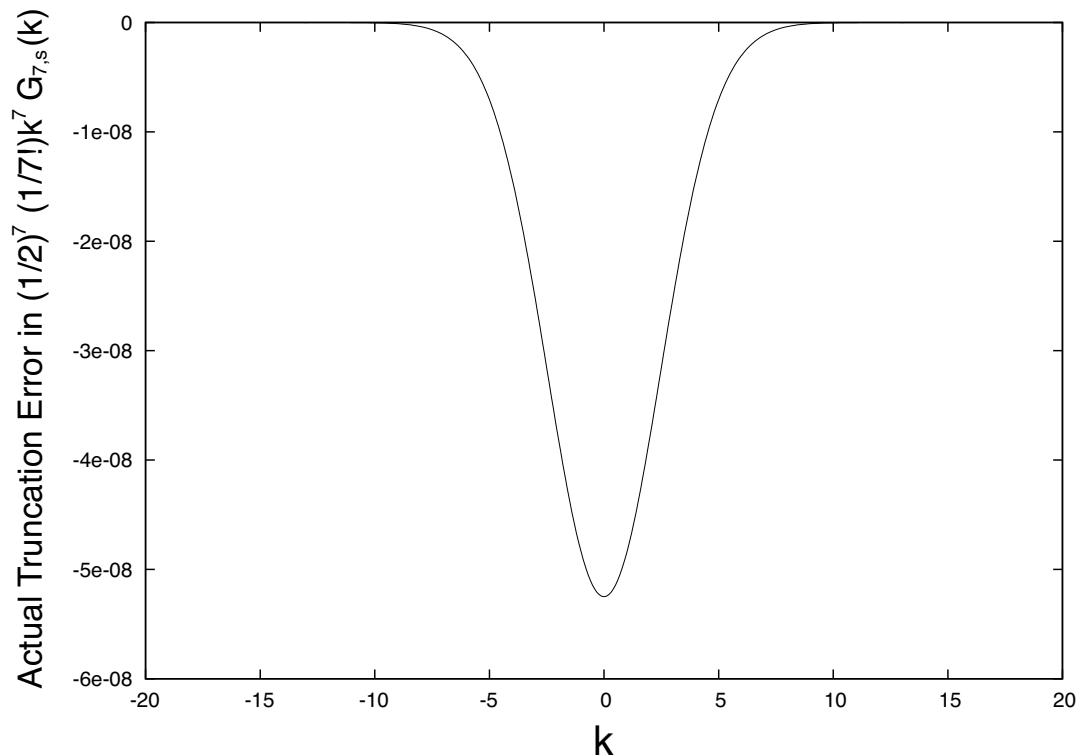


Figure 19.2.24: Real part of actual truncation error in  $(1/2)^m(1/m!)k^mG_{m,s}(k)$  for  $m = 7$  produced by truncating the series (14.4.84) beyond  $r = r_{\max}(7) = 29$ . The imaginary part vanishes.

$$\tilde{F}_r^s(k) \approx (2/N) \sum_{n=0}^{N-1} \text{se}_r(v_n, q) \tilde{F}(v_n, k). \quad (19.2.39)$$

The accuracy of these approximations depends on the large  $m$  behavior of the Fourier coefficients for  $\tilde{F}(v, k)$  and the Fourier coefficients for the Mathieu functions  $\text{ce}_r(v, q)$  and  $\text{se}_r(v, q)$ . See Exercise 2.2. There it is shown that (2.38) and (2.39) are good approximations providing the Fourier coefficients are sufficiently small for  $m \geq N$ . For the monopole doublet case this proves to be true for  $N \geq 120$ . See Exercises 2.3 and 2.4. In that case the use of (2.38) and (2.39) with  $N = 120$  is accurate to 5 parts in  $10^4$  for all  $k$  and  $r$  values of interest. Indeed, the  $\tilde{F}_r^s(k)$  used in making Figures 2.4 and 2.5 were obtained using (2.39) with  $N = 120$ .

### 19.2.5 Further Tests

Following the pattern established in the circular cylinder case, the next items we should verify are these:

- So far, we have been using exact results for  $\tilde{F}(v_n, k)$ . We should verify that sufficiently accurate results for  $\tilde{F}(v_n, k)$  can be obtained by evaluating (14.4.72) numerically using splines.
- With the use of this spline-computed  $\tilde{F}(v_n, k)$  and the related Riemann sum  $\tilde{F}_r^s(k)$  derived from it employing (2.39), we should verify that (2.36) is still well satisfied.
- Assuming that (2.36) is well satisfied, we should verify that inserting these results into (14.4.86) and then carrying out the indicated inverse Fourier transform, again numerically using splines, yields sufficiently accurate approximations to the functions  $C_{m,s}^{[n]}(z)$ .
- Finally, all these procedures should also yield satisfactory results when data is taken from a grid and interpolated onto the elliptic cylinder. See Figures 13.7.2 and 14.4.3.

The first three items above depend on the accuracy of the numerical spline-based Fourier transform routines. Since this accuracy has already been established in the circular cylinder case, and the demands made of the spline-based routines are no more stringent in the elliptic cylinder case, it seems sensible to proceed directly to verifying the last item. Moreover, its verification also provides added evidence of the veracity of the first three.

### 19.2.6 Completion of Test

Therefore, to complete our test, we again set up a regular grid in  $x, y, z$  space. We again let  $y$  and  $z$  range over the intervals  $y \in [-2.4, 2.4]$  with  $h_y = .1$  and  $z \in [-300, 300]$  with  $h_z = .125$ . However, for  $x$  we will use the interval  $x \in [-4.4, 4.4]$  with  $h_x = .1$  in order to ensure that all the ellipse given by (2.1) and (2.2) lies within the grid. Thus, we use 89 grid points in  $x$ , 49 grid points in  $y$ , and 4801 grid points in  $z$  for a total of  $89 \times 49 \times 4801 = 20,937,161$  grid

points.<sup>14</sup> As before, we evaluate the angular integrals using a Riemann sum with  $N = 120$ . Doing so requires interpolation off the grid onto the elliptical cylinder at 120 angular points for each of the 4801 selected values of  $z$ . And we evaluate the forward linear transforms for 401  $k$  values in the range  $k \in [-K_c, K_c]$  with  $K_c = 20$ . We also use these same points in  $k$  space to evaluate the inverse Fourier transforms. That is,  $H = .1$ . Finally we use the same values for  $r_{\max}(m)$  as before. The result of this process is a *completely numerically* calculated set of functions  $C_m^{[n]}(z)$ .

We find that the results so obtained for the  $C_m^{[n]}(z)$  are to the eye indistinguishable, for example, from Figures 1.7, 1.8, and 1.15. As more precise indicators of accuracy, Figures 2.25 through 2.27 show differences between exact and grid-value based completely numerically computed results for  $C_{1,s}^{[0]}(z)$ ,  $C_{1,s}^{[6]}(z)$ , and  $C_{7,s}^{[0]}(z)$ . We see that the error is between 4 and 5 parts in  $10^4$ , which is just slightly larger than the circular cylinder error result.

We have demonstrated, for the monopole-doublet problem, that the steps in the first three boxes shown in Figure 14.1.1 can also be carried out to yield results having good numerical accuracy for the case of a cylinder with elliptical cross section. As remarked earlier, again see Figure 13.7.7 and now also Figure 2.28, the surface field we have been working with is quite singular, more singular than fields likely to be encountered in practice. Thus the fact that the elliptical cylinder surface method has succeeded in this rather extreme case indicates that it is likely to work even better in actual physical applications.

## Exercises

**19.2.1.** From (2.20) through (2.22) we know that  $\tilde{F}(v, k)$  must have an expansion of the form

$$\tilde{F}(v, k) = \sum_r \tilde{\tilde{F}}_r^s(k) \text{se}_r(v, q). \quad (19.2.40)$$

Use this expansion to show that there is the relation

$$\|\tilde{F}\|^2 = (1/2) \sum_n |\tilde{\tilde{F}}_n^s(k)|^2, \quad (19.2.41)$$

which is a form of *Parseval's* theorem applied to Mathieu expansions. From (2.41) conclude that

$$|\tilde{\tilde{F}}_r^s(k)| = \sqrt{2}[\|\tilde{F}\|^2 - \sum_{n \neq r} |\tilde{\tilde{F}}_n^s(k)|^2]^{1/2} \leq \sqrt{2}\|\tilde{F}\|. \quad (19.2.42)$$

**19.2.2.** This exercise is a generalization of Exercise 1.2. Suppose  $f(\phi)$  and  $g(\phi)$  are  $2\pi$  periodic functions and therefore have Fourier expansions of the form

$$f(\phi) = \sum_{m=-\infty}^{\infty} f_m \exp(im\phi), \quad (19.2.43)$$

---

<sup>14</sup>Of course, as in the circular cylinder case, most of these grid points are actually unused since only relatively few are sufficiently near the surface of the elliptic cylinder to be employed in the interpolation procedure. After interpolation onto the surface of the elliptic cylinder,  $120 \times 4801 = 576,120$  surface values of  $F(U, v, z)$  are used in the remainder of the calculation.

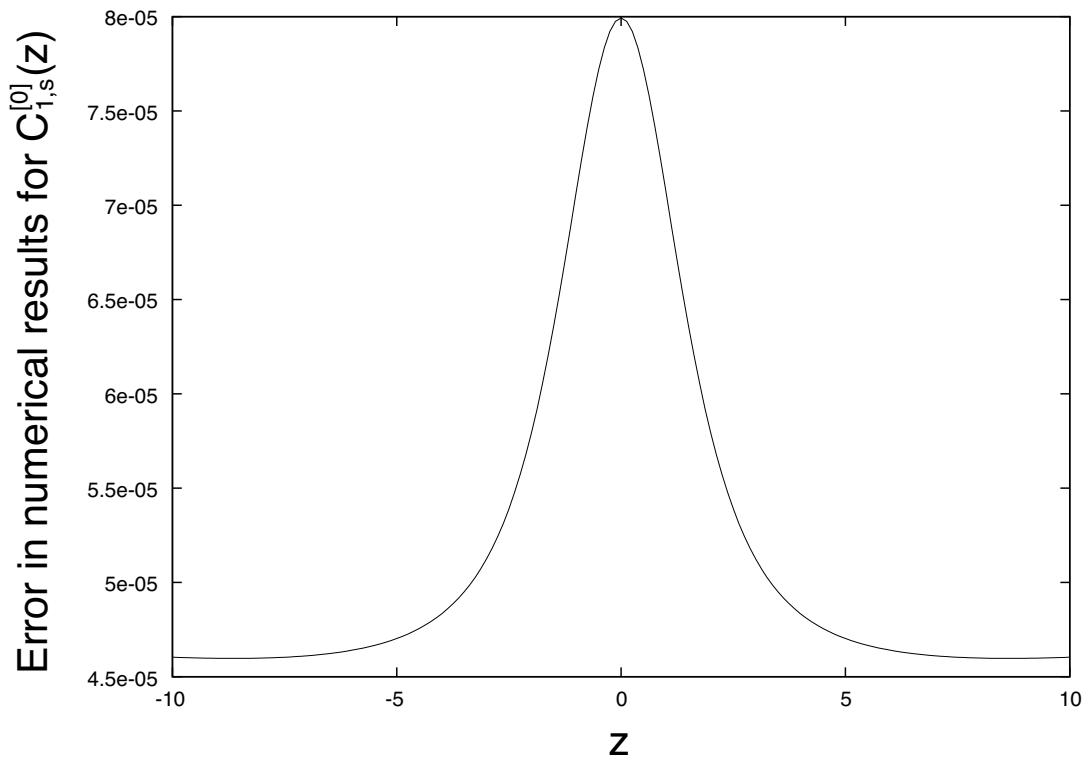


Figure 19.2.25: Difference between exact and completely numerically computed results for  $C_{1,s}^{[0]}(z)$  based on field data provided on a grid and interpolated onto an elliptic cylinder with  $x_{\max} = 4$  cm and  $y_{\max} = 2$  cm.

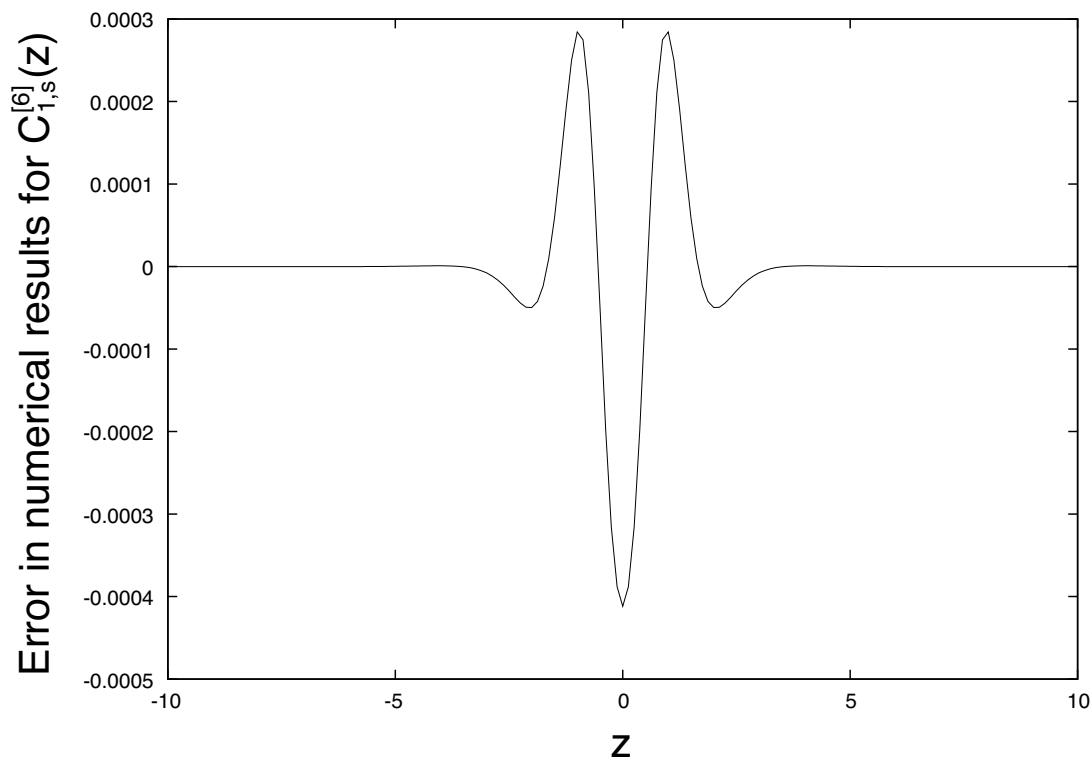


Figure 19.2.26: Difference between exact and completely numerically computed results for  $C_{1,s}^{[6]}(z)$  based on field data provided on a grid and interpolated onto an elliptic cylinder with  $x_{\max} = 4$  cm and  $y_{\max} = 2$  cm.

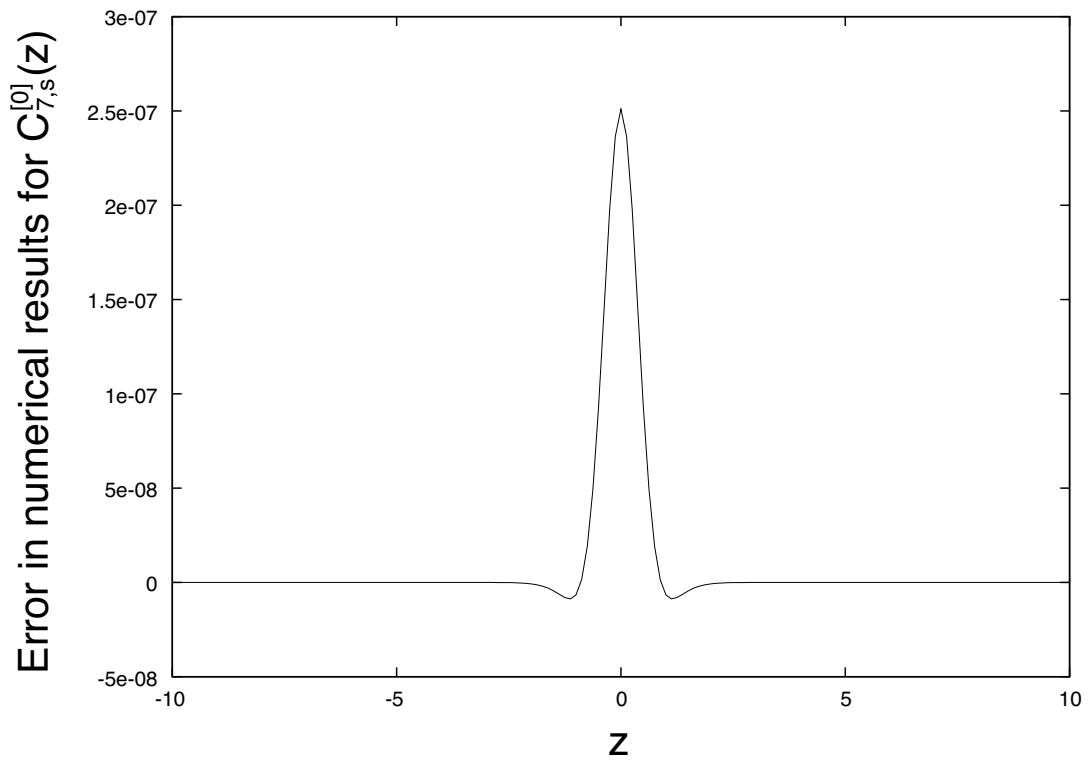


Figure 19.2.27: Difference between exact and completely numerically computed results for  $C_{7,s}^{[0]}(z)$  based on field data provided on a grid and interpolated onto an elliptic cylinder with  $x_{\max} = 4$  cm and  $y_{\max} = 2$  cm..

$$g(\phi) = \sum_{m=-\infty}^{\infty} g_m \exp(im\phi). \quad (19.2.44)$$

Into the vector space of such functions introduce a scalar product by the usual rule

$$(f, g) = [1/(2\pi)] \int_0^{2\pi} d\phi \bar{f}(\phi)g(\phi) \quad (19.2.45)$$

where a bar denotes complex conjugation. Show that

$$(f, g) = \sum_{m=-\infty}^{\infty} \bar{f}_m g_m, \quad (19.2.46)$$

which is *Parseval's* theorem for Fourier expansions. Again select  $N$  discrete phi values  $\phi_n$  according to the rule (6.1.37), and use these discrete values to define a *discrete* scalar product by the rule

$$(f, g)_d = (1/N) \sum_{n=0}^{N-1} \bar{f}(\phi_n)g(\phi_n). \quad (19.2.47)$$

Verify that (2.47) is the discrete version of (2.45). Show that

$$(f, g)_d = \sum_{m=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} \bar{f}_m g_{m+\ell N}, \quad (19.2.48)$$

which can be rewritten in the form

$$\begin{aligned} (f, g)_d &= \sum_{m=-\infty}^{\infty} \bar{f}_m g_m + \sum_{m=-\infty}^{\infty} \sum_{\ell \neq 0} \bar{f}_m g_{m+\ell N} \\ &= (f, g) + \sum_{m=-\infty}^{\infty} \sum_{\ell \neq 0} \bar{f}_m g_{m+\ell N}. \end{aligned} \quad (19.2.49)$$

Thus,  $(f, g)_d$  is a good approximation to  $(f, g)$  provided the  $f_m$  and  $g_m$  fall off sufficiently rapidly for large  $|m|$ , and  $N$  is sufficiently large, so that the sum in (2.49) can be neglected.

Let us apply these results to the computation of the  $\tilde{F}_r^s(k)$  as given by (2.21). Employing scalar product notation, we may write (2.21) in the form

$$\tilde{F}_r^s(k) = 2(\text{se}_r, \tilde{F}). \quad (19.2.50)$$

We know that the  $\text{se}_r$  have the Fourier expansion (15.5.27), which we rewrite in the complex form

$$\text{se}_r(v, q) = \sum_{m=-\infty}^{\infty} \hat{B}_m^r(q) \exp(imv). \quad (19.2.51)$$

Also, let  $\tilde{F}^F(m, k)$  denote the *Fourier* coefficients of  $\tilde{F}(v, k)$  so that we may write

$$\tilde{F}(v, k) = \sum_{m=-\infty}^{\infty} \tilde{F}^F(m, k) \exp(imv). \quad (19.2.52)$$

Then we have the result

$$(\text{se}_r, \tilde{F})_d = (\text{se}_r, \tilde{F}) + \sum_{m=-\infty}^{\infty} \sum_{\ell \neq 0} \bar{B}_m^r(q) \tilde{F}^F(m + \ell N, k). \quad (19.2.53)$$

In the next two exercises we will study the large  $|m|$  behavior of the Fourier coefficients  $\tilde{F}^F(m, k)$  and  $\bar{B}_m^r(q)$ .

**19.2.3.** This exercise is a continuation of Exercise 2.2. Here we will examine the falloff of the Fourier coefficients of  $F(U, v, z)$  and  $\tilde{F}(v, k)$  for the monopole doublet example. If you have not already done so, read Exercise 1.3. Here in analogy, we want to view  $F(U, v, z)$  as a function of  $\lambda$ , with

$$\lambda = \exp iv, \quad (19.2.54)$$

and locate its singularities in  $\lambda$ .

From (14.4.67) we see that  $F$  is analytic in  $v$  save for those points where  $B_x(U, v, z)$  and  $B_y(U, v, z)$  have singularities in  $v$ . And from (13.7.4) and (13.7.5) we see that these singularities occur where  $\psi(x, y, z)$  as given by (13.7.3) is singular, namely the points where

$$x^2 + (y \pm a)^2 + z^2 = 0. \quad (19.2.55)$$

Use (14.4.1), (14.4.2), and (2.1) through (2.4) to show that (2.25) is equivalent to the condition

$$\sin^2(v) \pm (2a/f) \sinh(U) \sin(v) - \{[f \cosh(U)]^2 + a^2 + z^2\}/f^2 = 0, \quad (19.2.56)$$

which can also be written in the form

$$\sin^2(v) \pm (2ay^{\max}/f^2) \sin(v) - [(x^{\max})^2 + a^2 + z^2]/f^2 = 0. \quad (19.2.57)$$

For the given values of  $f$ ,  $x^{\max}$ ,  $y^{\max}$ , and  $a = 2.5$ , and setting  $z = 0$  as before, verify that (2.57) has the solutions

$$\sin(v) = \pm 1.0073, \pm 1.84067 \quad (19.2.58)$$

Correspondingly, verify that  $\lambda$  has the values

$$\lambda = \pm 1.12863i, \pm .88603i, \pm .29533i, \pm 3.38601i, \quad (19.2.59)$$

and therefore  $\Lambda$  has the value

$$\Lambda = 1.12863 \quad (19.2.60)$$

and  $1/\Lambda$  has the value

$$1/\Lambda = 1/1.12863 = .88603. \quad (19.2.61)$$

Upon comparing (2.61) with (1.53), we see that the Fourier coefficients of  $F(U, v, z = 0)$  are expected to fall off less rapidly than those of  $B_\rho(R = 2, \phi, z = 0)$ . Figure 2.28 displays the function  $F(U, v, z = 0)$ , with  $U$  given by (2.1). Observe that the peaks at  $v = \pm\pi/2$  are sharper than those in Figure 13.7.6 for the corresponding case of a circular cylinder. Sharper peaks imply the existence of greater high ‘frequency’ content, which is consistent with the slower falloff of the Fourier coefficients.

Finally, it can be shown that, like the circular case,  $\Lambda$  is larger when  $z \neq 0$ . It follows, because the  $\tilde{F}(v, k)$  may be viewed as linear combinations of the  $F(U, v, z)$ , that the  $\tilde{F}^F(m, k)$  must also have the asymptotic behavior

$$|\tilde{F}^F(m, k)| \sim (.88603)^{|m|}. \quad (19.2.62)$$

This slower falloff is also consistent with the properties of  $\tilde{F}(v, k)$ . Figures 2.1 and 2.2 display the functions  $\tilde{F}(v, k = 0)$  and  $\tilde{F}(v, k = 20)$ . For comparison, Figures 1.23 and 1.24 show  $\tilde{B}_\rho(R, \phi, k = 0)$  and  $\tilde{B}_\rho(R, \phi, k = 20)$  for the case of the circular cylinder. We see that the angular behavior is similar, but more peaked in the case of the elliptic cylinder. Since  $\tilde{F}(v, k)$  and  $\tilde{B}_\rho(R, \phi, k)$  have similar angular behavior, we conclude that the large  $m$  behavior of  $\tilde{F}^F(m, k)$  and  $\tilde{B}_\rho(R, m, k)$  should be similar. However, since the elliptic case is more peaked, we expect the falloff with  $m$  will be slower, as observed.

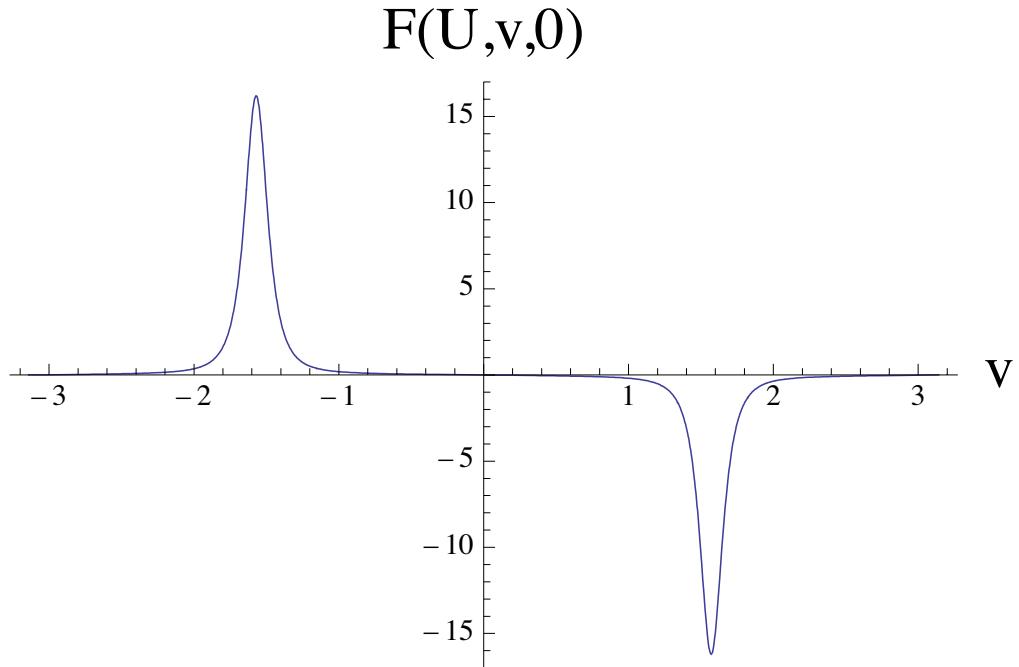


Figure 19.2.28: The quantity  $F(U, v, z = 0)$  for the monopole doublet in the case that  $x^{\max} = 4$  cm,  $y^{\max} = 2$  cm,  $a = 2.5$  cm, and  $g = 1$  Tesla-(cm) $^2$ .

**19.2.4.** This exercise is a continuation of Exercise 2.2. Here we will examine the Fourier coefficients for the Mathieu functions  $\text{ce}_r(v, q)$  and  $\text{se}_r(v, q)$ . That is, we want to study the functions  $A_m^r(q)$  and  $B_m^r(q)$  given in (15.5.21) and (15.5.27).

When  $q = 0$  the Mathieu functions become the trigonometric functions. See (14.4.35) through (14.4.37). Consequently, we have the results

$$A_m^r(0) = \delta_{mr} \text{ when } m \neq 0, \quad (19.2.63)$$

$$A_m^r(0) = \sqrt{2}\delta_{mr} \text{ when } m = 0, \quad (19.2.64)$$

$$B_m^r(0) = \delta_{mr} \text{ with } m, r \geq 1. \quad (19.2.65)$$

For  $q \neq 0$  the  $A_m^r(q)$  and  $B_m^r(q)$  need to be computed numerically, and the methods of Section 15.5.3 are convenient for doing so.

From (15.5.22) and (15.5.28), and applying the logic of Exercise 1.3, we see that as  $m \rightarrow \infty$  the  $A_m^r(q)$  and  $B_m^r(q)$ , for fixed  $q$  and  $r$ , must fall off faster than  $(1/\Lambda)^m$  for any  $\Lambda > 1$  because the Mathieu functions  $\text{ce}_r(v, q)$  and  $\text{se}_r(v, q)$  are entire functions of  $v$ . This is good news, but we still would like to know how large  $m$  must be for this asymptotic behavior to set in. We will see that the answer to this question depends on  $q$ .

For simplicity, let us study the behavior of  $B_m^r(q)$  as a function of  $m$  and  $q$  for the case  $r = 7$ , which is relevant to the case of the magnetic monopole doublet. Table 2.1 lists the values of  $B_m^7(q)$  for various values of  $m$  and  $q$ . We see that, as  $q$  becomes more negative, the  $m$  value for which  $|B_m^7(q)|$  peaks becomes ever larger. This is because, as  $q$  becomes ever more negative, more and more of the  $v$  axis is forbidden. See Section 14.4.4. However, we also know that  $\text{se}_7(v, q)$  must have 7 zeroes in the half-open interval  $v \in [0, \pi]$ . Thus, these oscillations are crowded into an ever smaller regions about 0 and  $\pm\pi$ , therefore leading to ever higher effective frequencies of oscillation.

Table 19.2.1: The coefficients  $B_m^7(q)$ .

$m \setminus q$	0	-50	-100	-150	-200	-250	-300
1	0	-.564487	-.375636	-.311321	-.274811	-.250197	-.232052
3	0	.293386	-.123551	-.237834	-.284830	-.307162	-.318251
5	0	.165556	.448665	.354776	.253034	.170433	.105346
7	1	-.465523	-.069863	.210855	.324419	.362328	.365811
9	0	.095015	-.435315	-.337074	-.163346	-.018371	.089069
11	0	.473149	.013745	-.311063	-.395713	-.372390	-.308973
13	0	.321478	.442586	.161649	-.088508	-.245162	-.327431
15	0	.118454	.432784	.453953	.318139	.151822	.003982
17	0	.029177	.236414	.409101	.456423	.415762	.331884
19	0	.005265	.089434	.234953	.355749	.421796	.437909
21	0	.000734	.025531	.099327	.196896	.287659	.356275
23	0	.000082	.005770	.032960	.084599	.149978	.216970
25	0	.000007	.001065	.008913	.029531	.063243	.106189
27	0	5.81E-7	.000164	.002014	.008615	.022275	.043350
29	0	3.81E-8	.000021	.000387	.002142	.006694	.015110

We saw on Exercise 2.2 that the discrete scalar product (discrete angular Riemann sum) is a good approximation to the true scalar product if  $N$  is sufficiently large such that

$$\tilde{F}^F(m, k) \approx 0 \text{ for } m \geq N \quad (19.2.66)$$

and

$$B_m^r(q) \approx 0 \text{ for } m \geq N \quad (19.2.67)$$

with  $k \in [-K_c, K_c]$  and  $r$  being within its required range. See (2.53). In our benchmarking we have set  $N = 120$ . Verify that  $(.88603)^{120} = 4.9 \times 10^{-7}$  so that, in view of (2.62), (2.66) is well satisfied. We next have to worry about the condition (2.67). From Table 2.1 we infer that  $B_m^7(q = -50)$  begins to fall off rapidly with increasing  $m$  for  $m$  somewhat greater than 17, and  $B_m^7(q = -300)$  begins to fall off rapidly with increasing  $m$  for  $m$  somewhat greater than 29. Assuming this trend continues as  $q$  becomes ever more negative, estimate by linear extrapolation that  $B_m^7(q = -1200)$  begins to fall off rapidly with increasing  $m$  for  $m$  somewhat greater than 72. Also, we know that the largest  $r$  value we are interested in is  $r_{\max}(7) = 29$ . Therefore we might infer that  $B_m^{29}(q = -1200)$  begins to fall off rapidly for  $m$  somewhat larger than  $72 + (29 - 7) = 94$ . Since 120 is significantly larger than 94, we infer that (2.67) should also be well satisfied for the choice  $N = 120$ .

These expectations are verified by the following calculations: Let  ${}^N\tilde{\tilde{F}}_r^s(k)$  denote the result of the Riemann sum (2.93), and let

$${}^\infty\tilde{\tilde{F}}_r^s(k) = \tilde{\tilde{F}}_r^s(k) \quad (19.2.68)$$

denote its  $N \rightarrow \infty$  limit given by the integral (2.21). Figure 2.29 shows the quantity  ${}^\infty\tilde{\tilde{F}}_{29}^s(k)$  obtained by careful numerical evaluation of the integral in (2.21) when  $r = 29$ . En passant, we also take this opportunity to show in Figure 2.30 the base 10 logarithm of the three quantities  $[-{}^\infty\tilde{\tilde{F}}_{29}^s(k)]$ ,  $[\sqrt{2}||\tilde{F}||]$ , and  $[\sqrt{2}|\tilde{F}(\pi/2, k)|]$  to illustrate the inequalities (2.32) and (2.35). More to the point of this exercise, Figures 2.31 through 2.33 show the error quantities

$${}^N\tilde{\tilde{F}}_{29}^s(k) - {}^\infty\tilde{\tilde{F}}_{29}^s(k)$$

for  $N = 40, 80$ , and  $120$ . We see that the error decreases rapidly with increasing  $N$ , and is less than 6 parts in  $10^4$  when  $N = 120$ . Indeed, further calculation shows that the error is approximately 4 parts in  $10^6$  when  $N = 160$ .

**19.2.5.** Show that the quantities  $B_m^n$  comprise the entries of an infinite orthogonal matrix.

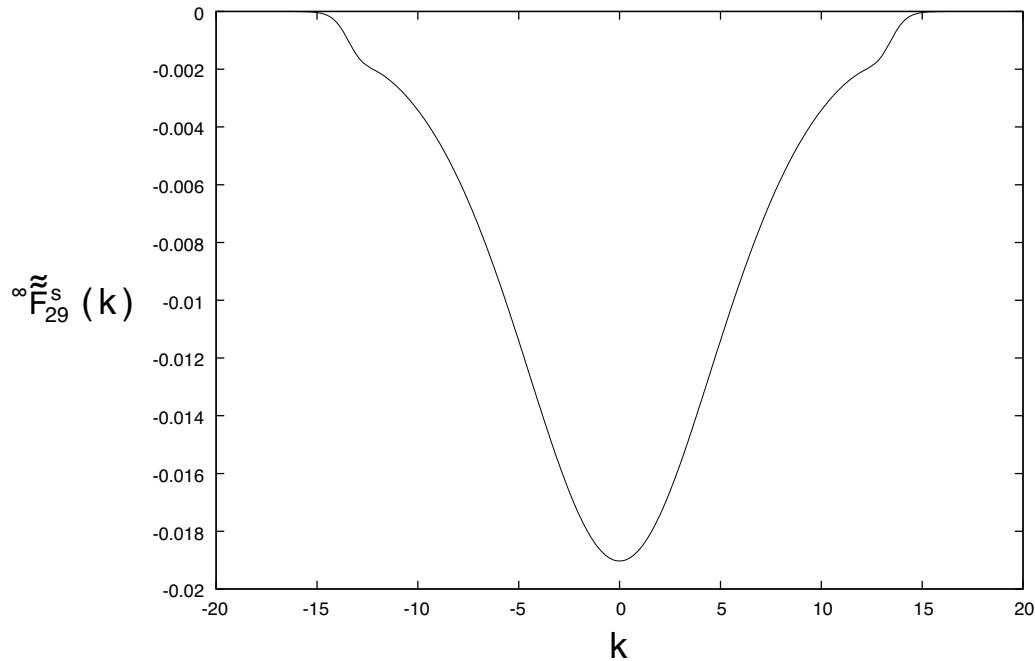


Figure 19.2.29: Real part of  ${}^\infty \tilde{F}_{29}^s(k)$ . The imaginary part vanishes.

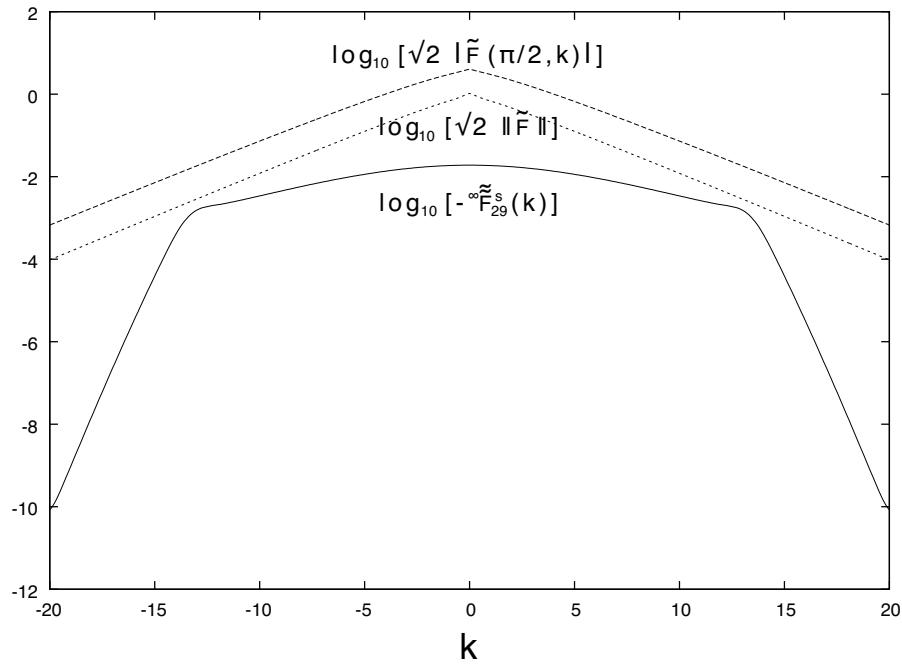


Figure 19.2.30: The base 10 logarithm of three quantities. The quantity  $\log_{10}[\sqrt{2}|\tilde{F}(\pi/2, k)|]$  is the top curve. The middle curve is the quantity  $\log_{10}[\sqrt{2}||\tilde{F}||]$ , and the bottom curve is  $\log_{10}[-{}^\infty \tilde{F}_{29}^s(k)]$ . Together they illustrate the inequalities (2.32) and (2.35).

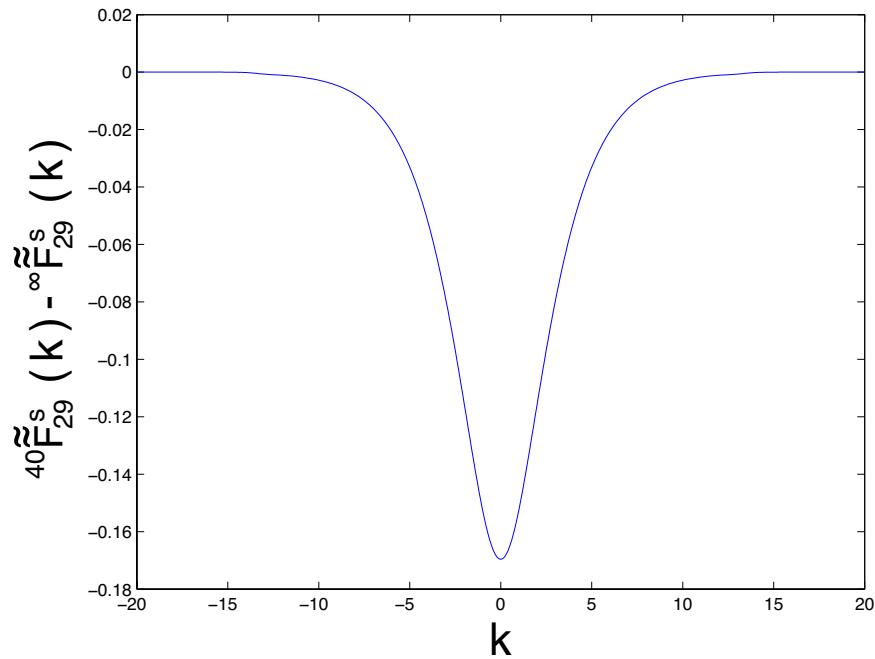


Figure 19.2.31: Real part of the error quantity  ${}^N\tilde{F}_{29}^s(k) - {}^\infty\tilde{F}_{29}^s(k)$  for  $N = 40$ . The imaginary part vanishes.

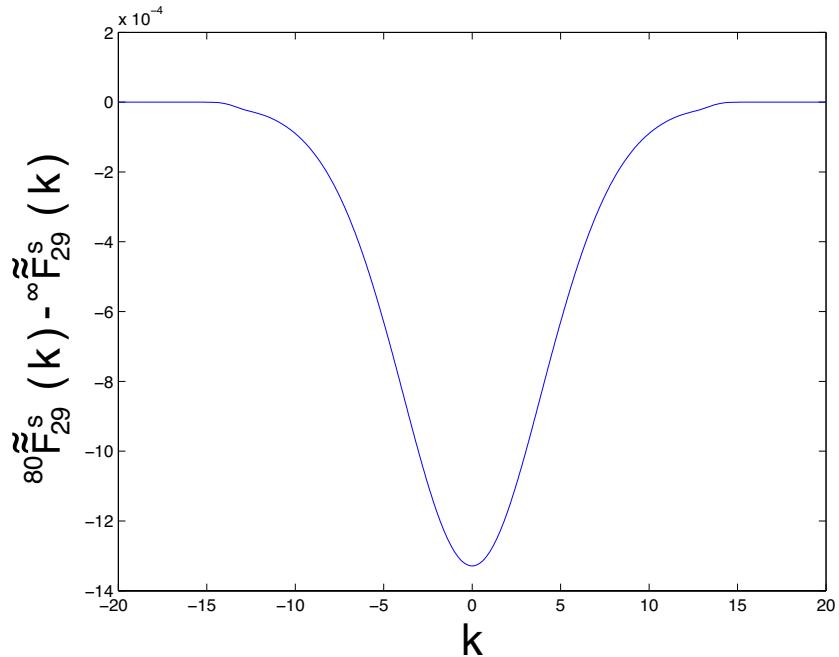


Figure 19.2.32: Real part of the error quantity  ${}^N\tilde{F}_{29}^s(k) - {}^\infty\tilde{F}_{29}^s(k)$  for  $N = 80$ . The imaginary part vanishes.

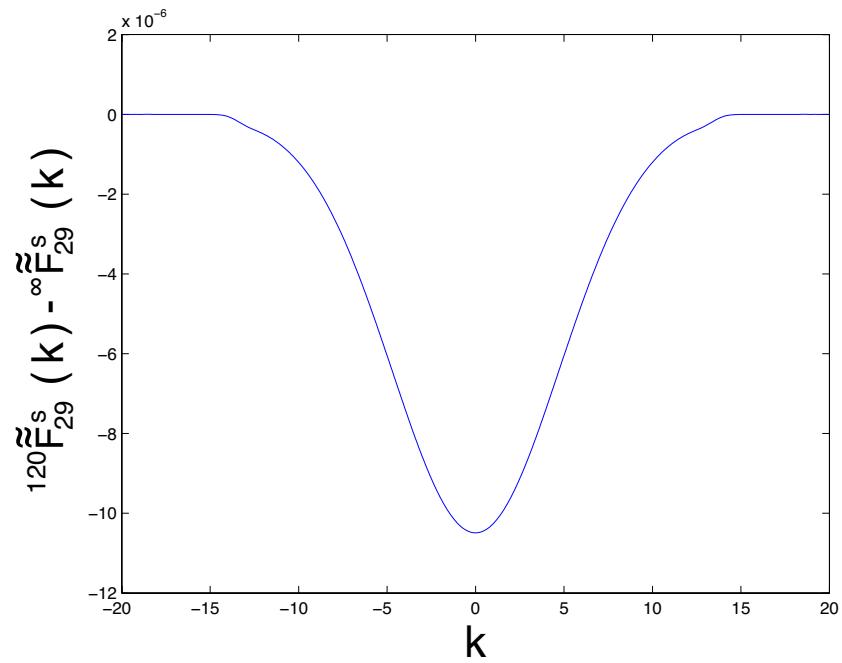


Figure 19.2.33: Real part of the error quantity  ${}^N\tilde{\mathcal{F}}_{29}^s(k) - {}^\infty\tilde{\mathcal{F}}_{29}^s(k)$  for  $N = 120$ . The imaginary part vanishes.

### 19.3 Rectangular Cylinder Numerical Results for Monopole Doublet

# Bibliography

## General References

- [1] C. Mitchell, “Calculation of Realistic Charged-Particle Transfer Maps”, University of Maryland Physics Department Ph.D. Thesis (2007).
- [2] C. Mitchell and A. Dragt, “Accurate transfer maps for realistic beam-line elements: Part I, straight elements”, *Physical Review Special Topics - Accelerators and Beams* **13** 064001 (2010).

## Bessel Functions

- [3] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, Chapter 9, Dover (1972). Also available on the Web by Googling “abramowitz and stegun 1972”.
- [4] 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/>

## Fourier Integrals

- [5] A. Erdélyi, Edit., *Tables of Integral Transforms*, McGraw-Hill (1954).

## Angular Integrals

- [6] M. Javed and L. Trefethen, “A trapezoidal rule error bound unifying the Euler-Maclaurin formula and geometric convergence for periodic functions”, *Proc. R. Soc. A Math. Phys. Eng. Sci.* **470**, 20130571 (2013). See also the Web site [http://people.maths.ox.ac.uk/trefethen/publication/PDF/2014\\_150.pdf](http://people.maths.ox.ac.uk/trefethen/publication/PDF/2014_150.pdf)
- [7] L. Trefethen and J. Weideman, “The exponentially convergent trapezoidal rule”, (2014). *SIAM Rev.* **56**, 385–458 (2014). See the Web site [http://people.maths.ox.ac.uk/trefethen/publication/PDF/2014\\_149.pdf](http://people.maths.ox.ac.uk/trefethen/publication/PDF/2014_149.pdf)



# Chapter 20

## Smoothing and Insensitivity to Errors

### 20.1 Introduction

In the previous Chapters 13 and 14 mention was made of the smoothing feature of surface methods. In this chapter we will explore the smoothing behavior of surface methods in more detail.

By way of introduction, imagine for simplicity that we initially use the surface of a circular cylinder. Suppose there are measurement or computational errors in the radial surface field values  $B_\rho(\rho = R, \phi, z)$ . What effect do these errors have on the determination of the on-axis gradients and their derivatives? We will see that due to smoothing the effect of these errors is rather mild.

#### 20.1.1 Preliminary Considerations

The relative insensitivity of surface methods to errors arises from a basic property of solutions to Laplace's equation: the value of  $\psi$  at some interior point is an appropriately weighted average of its values over any surrounding boundary. Consequently,  $\psi$  is smoother in the interior of a region than it may be on a boundary of this region. Correspondingly, errors in boundary values are averaged.

Something can also be said about surface methods and fitting errors. Suppose  $\psi$  is a *harmonic* function (satisfies  $\nabla^2\psi = 0$ ) in some domain  $\mathcal{D}$ . Then, it can be shown that  $\psi$  assumes its maxima and minima on the *boundary* of  $\mathcal{D}$ . Imagine that  $\psi^{\text{exact}}$  is the true scalar potential and  $\psi^{\text{approx}}$  is some approximation to it. We know that  $\psi^{\text{exact}}$  is harmonic, and suppose that  $\psi^{\text{approx}}$  has been constructed to be harmonic. Then we know that  $\psi^{\text{error}} = \psi^{\text{approx}} - \psi^{\text{exact}}$  is harmonic. Therefore the magnitude of the error must be largest on the boundary of  $\mathcal{D}$ . However, if we do a good job of fitting  $\psi^{\text{exact}}$  by  $\psi^{\text{approx}}$  on the boundary of  $\mathcal{D}$ , then the error on the boundary will be small. And, thanks to  $\psi^{\text{error}}$  being harmonic, the error in the interior of  $\mathcal{D}$  will be even smaller.

#### 20.1.2 Analyticity

For a preliminary exploration of smoothing, suppose, for example, that the magnetic field is produced by an iron-dominated magnet, and is therefore localized in space. In this case the

integrals (14.3.2) can be considered to have, in practice, finite limits of integration. With some care, an effective cut-off can also be found even if the fields extend to infinity since they fall off sufficiently rapidly at infinity. Also, since the generalized Bessel function  $I'_m$  increases exponentially as described by (14.3.9), there is also, in effect, a cut-off in  $k$  for the integrals (14.3.8) defining the on-axis gradients.

Next suppose that the  $\tilde{B}_\rho(R, m', z)$  are absolutely integrable,

$$\int_{-\infty}^{\infty} dz |\tilde{B}_\rho(R, m', z)| < \infty. \quad (20.1.1)$$

This will certainly be the case if  $B_\rho(R, \phi, z)$  and hence the  $\tilde{B}_\rho(R, m', z)$  are localized in  $z$  space. It follows from (14.3.2) that the Fourier transforms  $\tilde{\tilde{B}}_\rho(R, m', k')$  are then bounded,

$$|\tilde{\tilde{B}}_\rho(R, m', k')| < [1/(2\pi)] \int_{-\infty}^{\infty} dz |\tilde{B}_\rho(R, m', z)| < \infty. \quad (20.1.2)$$

Now look at the integral representations (14.3.8) for the on-axis gradients. We see that, due to the bounds (1.2) and the fall off in  $k$  at infinity produced by the  $I'_m(kR)$  denominators, the integrals (14.3.8) are absolutely convergent in the domain

$$\Re(z) \in (-\infty, \infty), \quad \Im(z) \in (-R, R). \quad (20.1.3)$$

Thus, under very mild assumptions about the surface data  $B_\rho(\rho = R, \phi, z)$ , including the possibility of errors, we conclude that the on-axis gradients are *analytic* in the strip (1.3). Note that commonly used fringe-field models, those that assume constant ( $z$  independent) fields for  $z$  within the body of a magnet, zero fields outside beyond the fringe-field regions, and interpolating linear ramps in the fringe-field regions, violate this analyticity property because of singularities in the first derivative at the beginnings and ends of the ramps. These models are therefore unphysical, and their use could lead to erroneous conclusions.

### 20.1.3 Equivalent Spatial Kernel

In Section 14.3 the on-axis gradients were related to fields on the surface of a circular cylinder by various Fourier transform operations. We will now see that, for the circular cylinder case, they can also be viewed as being related by integration against a spatial kernel. Later, analogous results will be found for the cases of elliptic and rectangular cylinders.

Begin by relabeling variables so that (14.3.14) through (14.3.16) take (for  $m > 0$ ) the form

$$\tilde{\tilde{B}}_\rho^\alpha(R, m, k) = [1/(2\pi)] \int_{-\infty}^{\infty} dz' \exp(-ikz') \tilde{B}_\rho^\alpha(R, m, z') \quad (20.1.4)$$

with

$$\tilde{B}_\rho^s(R, m, z') = (1/\pi) \int_0^{2\pi} d\phi \sin(m\phi) B_\rho(R, \phi, z'), \quad (20.1.5)$$

$$\tilde{B}_\rho^c(R, m, z') = (1/\pi) \int_0^{2\pi} d\phi \cos(m\phi) B_\rho(R, \phi, z'), \quad (20.1.6)$$

and (again for  $m > 0$ ) (14.3.23) takes the form

$$C_{m,\alpha}^{[n]}(z) = i^n (1/2)^m (1/m!) \int_{-\infty}^{\infty} dk [k^{n+m-1} / I'_m(kR)] \tilde{B}_{\rho}^{\alpha}(R, m, k) \exp(ikz). \quad (20.1.7)$$

Our goal is to re-express the relations (1.4) and (1.7) in terms of a spatial kernel.

To do this, insert (1.4) into (1.7) to find the relation

$$\begin{aligned} C_{m,\alpha}^{[n]}(z) &= [1/(2\pi)] i^n (1/2)^m (1/m!) \times \\ &\quad \int_{-\infty}^{\infty} dk [k^{n+m-1} / I'_m(kR)] \exp(ikz) \int_{-\infty}^{\infty} dz' \exp(-ikz') \tilde{B}_{\rho}^{\alpha}(R, m, z') \\ &= [1/(2\pi)] i^n (1/2)^m (1/m!) \times \\ &\quad \int_{-\infty}^{\infty} dz' \tilde{B}_{\rho}^{\alpha}(R, m, z') \int_{-\infty}^{\infty} dk [k^{n+m-1} / I'_m(kR)] \exp(ikz) \exp(-ikz') \\ &= [1/(2\pi)] i^n (1/2)^m (1/m!) \times \\ &\quad \int_{-\infty}^{\infty} dz' \tilde{B}_{\rho}^{\alpha}(R, m, z') \int_{-\infty}^{\infty} dk [k^{n+m-1} / I'_m(kR)] \exp[ik(z - z')]. \end{aligned} \quad (20.1.8)$$

Define the kernel  $K_m^{[n]}$  by the rule

$$K_m^{[n]}(z, z') = [1/(2\pi)] i^n (1/2)^m (1/m!) \int_{-\infty}^{\infty} dk [k^{n+m-1} / I'_m(kR)] \exp[ik(z - z')]. \quad (20.1.9)$$

With this definition, we may rewrite (1.8) in the form

$$C_{m,\alpha}^{[n]}(z) = \int_{-\infty}^{\infty} dz' K_m^{[n]}(z, z') \tilde{B}_{\rho}^{\alpha}(R, m, z'). \quad (20.1.10)$$

That is,  $C_{m,\alpha}^{[n]}(z)$  has been expressed as the result of integrating  $\tilde{B}_{\rho}^{\alpha}(R, m, z')$  against the spatial kernel  $K_m^{[n]}(z, z')$ .

Let us now explore the properties of  $K_m^{[n]}$ . In the definition (1.9) make the substitution

$$\lambda = kR \text{ or } k = \lambda/R. \quad (20.1.11)$$

So doing gives the result

$$K_m^{[n]}(z, z') = [1/(2\pi)] i^n (1/2)^m (1/m!) (1/R)^{n+m} \int_{-\infty}^{\infty} d\lambda [\lambda^{n+m-1} / I'_m(\lambda)] \exp[i\lambda(z - z')/R]. \quad (20.1.12)$$

Staring at (1.12) suggests writing

$$K_m^{[n]}(z, z') = (1/R)^{n+m} L_m^{[n]}(\Delta) \quad (20.1.13)$$

where

$$\Delta = (z - z')/R \quad (20.1.14)$$

and

$$L_m^{[n]}(\Delta) = [1/(2\pi)]i^n(1/2)^m(1/m!) \int_{-\infty}^{\infty} d\lambda [\lambda^{n+m-1}/I'_m(\lambda)] \exp(i\lambda\Delta). \quad (20.1.15)$$

We observe, consistent with the notation being employed, that there is the relation

$$L_m^{[n]}(\Delta) = (\partial_{\Delta})^n L_m^{[0]}(\Delta). \quad (20.1.16)$$

We also observe that the integrand factor  $[\lambda^{m-1}/I'_m(\lambda)]$  appearing in

$$L_m^{[0]}(\Delta) = [1/(2\pi)](1/2)^m(1/m!) \int_{-\infty}^{\infty} d\lambda [\lambda^{m-1}/I'_m(\lambda)] \exp(i\lambda\Delta) \quad (20.1.17)$$

is *even* in  $\lambda$ . It follows that the  $L_m^{[n]}(\Delta)$ , and hence also the  $K_m^{[n]}(z, z')$ , are purely real. Moreover, the  $L_m^{[0]}(\Delta)$  are even in  $\Delta$ .

Graphs of some of the functions  $L_m^{[n]}(\Delta)$  are shown in Figures 1.1 through 1.3 below. Before commenting on them, it is also useful to examine the integrands  $\tilde{L}_m^{[0]}(\lambda)$ , which are the Fourier transforms of the  $L_m^{[0]}(\Delta)$ , defined by the relations

$$\tilde{L}_m^{[0]}(\lambda) = [1/(2\pi)](1/2)^m(1/m!)[\lambda^{m-1}/I'_m(\lambda)]. \quad (20.1.18)$$

Some of them are displayed in Figures 1.4 and 1.5. Again we recall the asymptotic behavior

$$|I'_m(\lambda)| \sim \exp(|\lambda|)/\sqrt{2\pi|\lambda|} \text{ as } |\lambda| \rightarrow \infty, \quad (20.1.19)$$

from which we conclude that the  $\tilde{L}_m^{[0]}(\lambda)$  essentially vanish exponentially at infinity.

What insights can be gained from examining Figures 1.1 through 1.5? First, we observe from Figure 1.3 that the  $L_m^{[0]}$  become ever narrower with increasing  $m$ . Correspondingly, in accord with the uncertainty principle relating Fourier transform pairs, Figure 1.5 shows that the  $\tilde{L}_m^{[0]}$  become ever broader with increasing  $m$ .

Next, from (1.10), we see that it is desirable that the  $K_m^{[0]}(z, z')$  be slowly varying in  $z'$ , because then noise in  $\tilde{B}_{\rho}^{\alpha}(R, m, z')$  will be averaged over a large interval in  $z'$ . From (1.14) and Figure 1.3 we see that the  $K_m^{[0]}(z, z')$  will be more nearly constant in  $z'$  the larger the value of  $R$ , and therefore there is ever more smoothing in  $z'$  as  $R$  is increased. We have already observed that the  $L_m^{[0]}$  become somewhat narrower as  $m$  increases. Therefore there is somewhat less smoothing in  $z'$  for larger  $m$ . However, there is still ever more smoothing in  $z'$  as  $R$  is increased. Also, inspection of (1.13) shows that there is a  $(1/R)^{n+m}$  factor relating  $K_m^{[n]}$  and  $L_m^{[n]}$ . When the associated  $C_{m,\alpha}^{[n]}$  are used in (13.2.37) or analogous expansions for  $\mathbf{A}$ , it is evident that the effective dimensionless expansion factor is  $(\rho/R)^{n+m}$ . Thus, although there is somewhat less smoothing in  $z'$  for larger  $m$ , there is increased suppression of high angular harmonic noise, and moreover this suppression is enhanced as  $R$  is increased. We also observe that there is increased suppression of high angular harmonic noise as  $n$  is increased.

Finally, let us examine smoothing from the perspective of  $k$  space. In the integral (1.7) make the change of variables

$$\lambda = kR \quad (20.1.20)$$

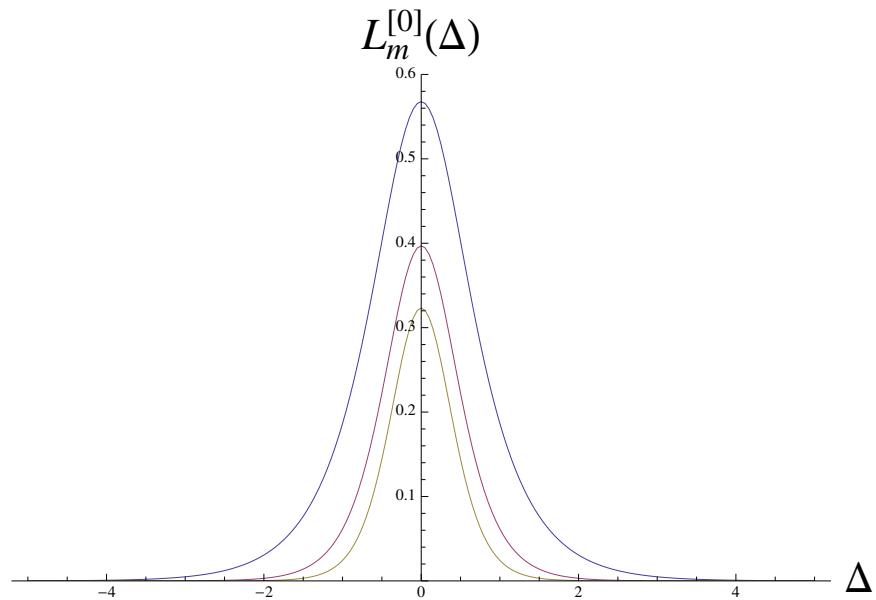


Figure 20.1.1: The spatial kernels  $L_1^{[0]}(\Delta)$  through  $L_3^{[0]}(\Delta)$ . For  $\Delta = 0$ , the kernels  $L_m^{[0]}(\Delta)$  decrease with increasing  $m$ .

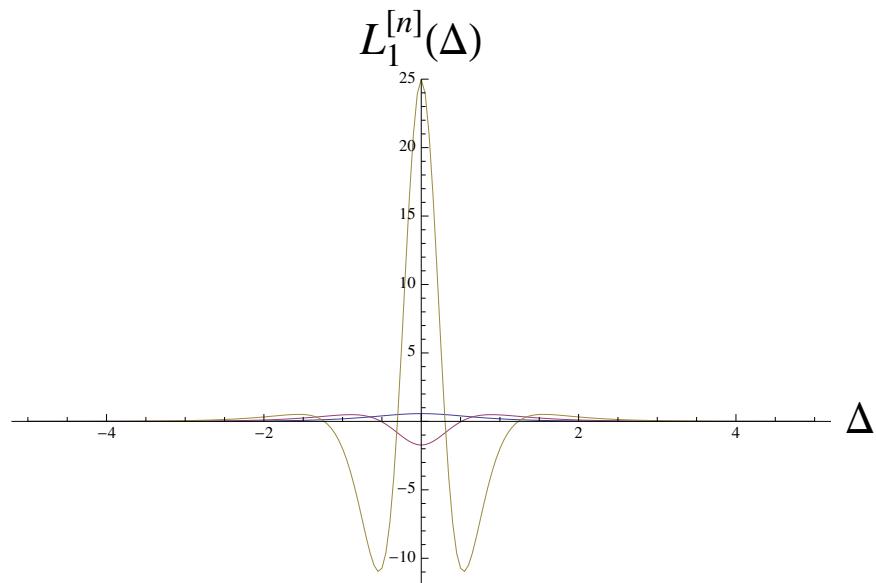


Figure 20.1.2: The spatial kernels  $L_1^{[n]}(\Delta)$  for  $n = 0, 2, 4$ . Note that they satisfy (1.16). In particular, they have  $n$  zeroes.

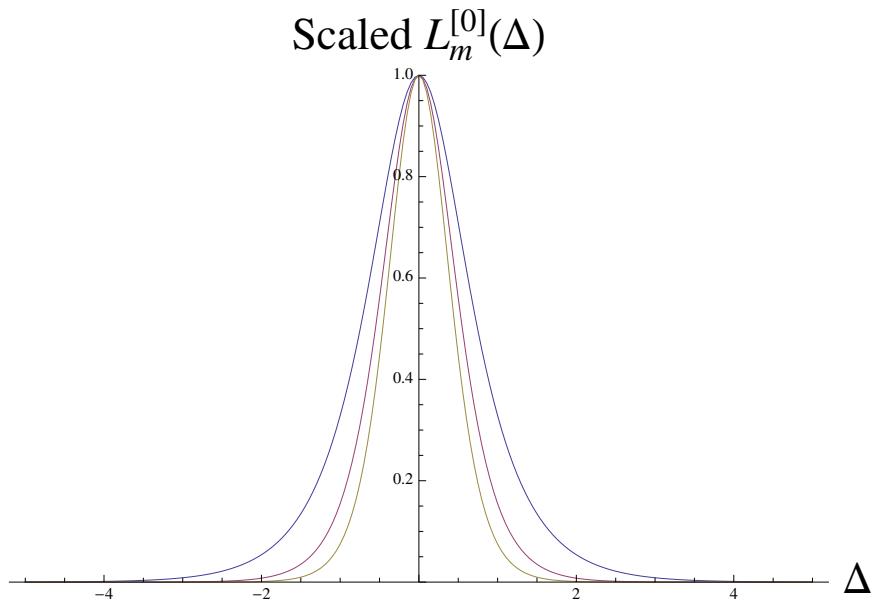


Figure 20.1.3: The *scaled* spatial kernels  $L_1^{[0]}(\Delta)$  through  $L_3^{[0]}(\Delta)$ , all normalized to 1 at  $\Delta = 0$ . The scaled  $L_m^{[0]}$  become ever narrower with increasing  $m$ .

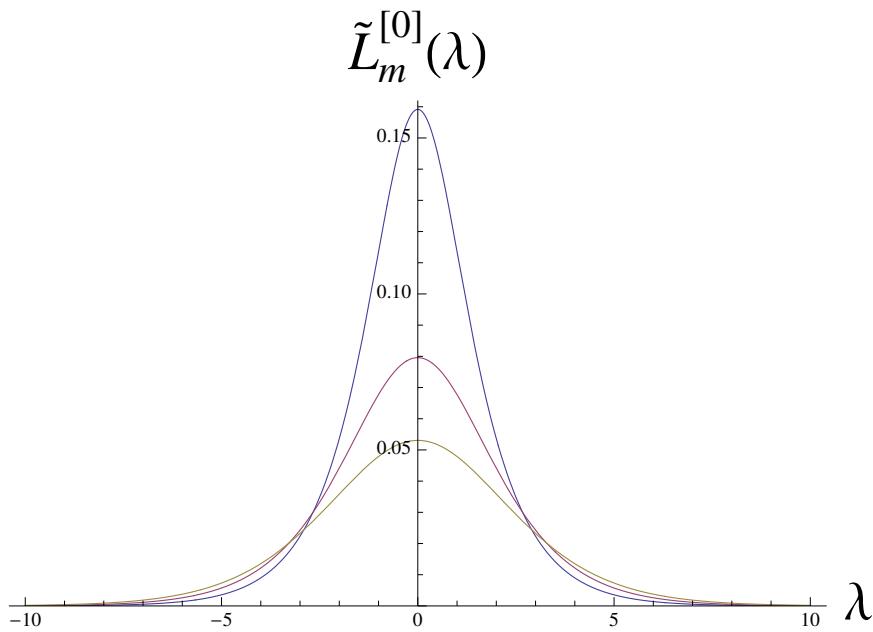


Figure 20.1.4: The integrands  $\tilde{L}_m^{[0]}(\lambda)$  for  $m = 1, 2, 3$ . For  $\lambda = 0$ , the integrands decrease with increasing  $m$ .

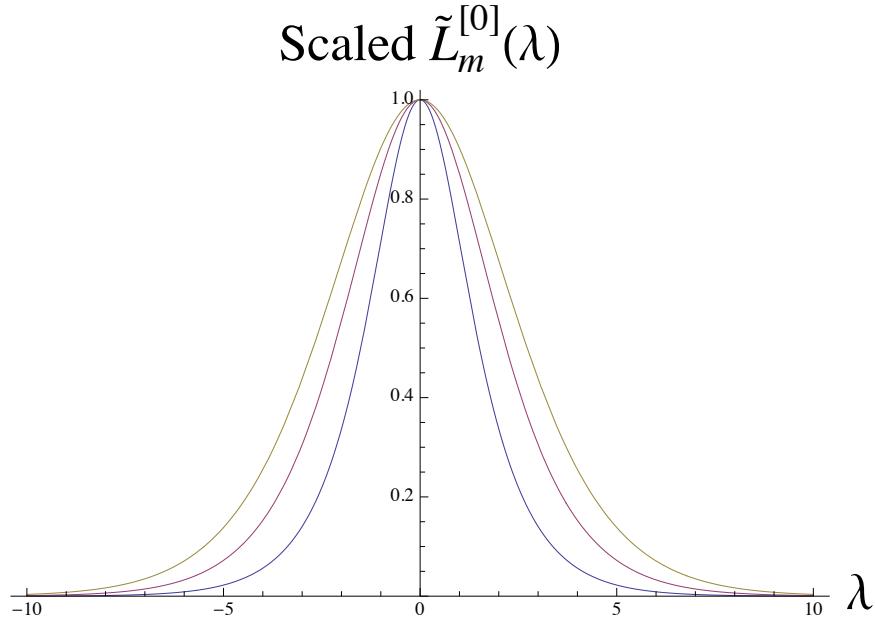


Figure 20.1.5: The *scaled* integrands  $\tilde{L}_m^{[0]}(\lambda)$  for  $m = 1, 2, 3$ , all normalized to 1 at  $\lambda = 0$ . The scaled integrands become ever broader with increasing  $m$ .

to find the result

$$C_{m,\alpha}^{[n]}(z) = (1/R)^{n+m} i^n (1/2)^m (1/m!) \int_{-\infty}^{\infty} d\lambda [\lambda^{n+m-1}/I'_m(\lambda)] \tilde{B}_{\rho}^{\alpha}(R, m, \lambda/R) \exp(i\lambda z/R). \quad (20.1.21)$$

With the aid of the definition (1.18) this result can be rewritten in the form

$$C_{m,\alpha}^{[n]}(z) = 2\pi i^n (1/R)^{n+m} \int_{-\infty}^{\infty} d\lambda \tilde{L}_m^{[n]}(\lambda) \tilde{B}_{\rho}^{\alpha}(R, m, \lambda/R) \exp(i\lambda z/R). \quad (20.1.22)$$

Consider, for example, the case  $n = 0$ . Since the  $\tilde{L}_m^{[0]}(\lambda)$  are peaked about  $\lambda = 0$ , we have the approximate result

$$\tilde{L}_m^{[0]}(\lambda) \tilde{B}_{\rho}^{\alpha}(R, m, \lambda/R) \approx \tilde{L}_m^{[0]}(\lambda) \tilde{B}_{\rho}^{\alpha}(R, m, 0), \quad (20.1.23)$$

and this result becomes ever more exact the larger the value of  $R$ . From (1.4) we see that evaluating  $\tilde{B}_{\rho}^{\alpha}(R, m, k)$  for  $k \approx 0$  essentially amounts to averaging  $\tilde{B}_{\rho}^{\alpha}(R, m, z')$  over  $z'$ , thereby suppressing the effect of noise. From Figure 1.5 we see that this smoothing in  $z'$  becomes somewhat less effective with increasing  $m$  because then the  $\tilde{L}_m^{[0]}(\lambda)$  are less peaked about  $\lambda = 0$ . But again there is a  $(1/R)^{n+m}$  factor that comes into play so that the effective dimensionless expansion factor is again  $(\rho/R)^{n+m}$ . Thus, although there is somewhat less smoothing in  $z'$  for larger  $m$ , there is increased suppression of high angular harmonic noise as  $n$  and  $m$  are increased, and again this suppression is further enhanced as  $R$  is increased.

### 20.1.4 What Work Lies Ahead

We will now study smoothing and insensitivity to errors in more detail depending on what surface is used. We will do so for the monopole-doublet examples treated in Sections 16.1 through 16.3, but our conclusions will be general. Section 17.2 treats the use of circular cylinders, and Sections 17.3 and 17.4 treat the use of elliptic and rectangular cylinders.

## Exercises

**20.1.1.** Show that the on-axis gradients associated with the magnetic monopole doublet and given by (13.7.33) are analytic in the domain (1.3), and have singularities on the boundary. Show that the same is true for the on-axis gradient associated with the air-core solenoid of Section 11.11.

**20.1.2.** Verify that that the integrand factor  $[\lambda^{m-1}/I'_m(\lambda)]$  is *even* in  $\lambda$ , and therefore the  $L_m^{[n]}(\Delta)$ , and also the  $K_m^{[n]}(z, z')$ , are purely real. Verify that the  $L_m^{[0]}(\Delta)$  are even in  $\Delta$ .

## 20.2 Circular Cylinders

Review Section 16.1.3 that described, for the monopole-doublet test case and the use of the surface of a circular cylinder, the calculation of on-axis gradients based on field data provided on a grid. Suppose we add to the field data at each grid point small random field components in the  $x$  and  $y$  directions.<sup>1</sup> What will be the effect of this noise on the on-axis gradients? We could use the noisy data to compute the on-axis gradients, and compare the results with those obtained in the absence of noise (and which we know agree very well with exact results). However, observe that the on-axis gradients are *linear* functions of the input field values on the grid points. Therefore, we can just as well compute the on-axis gradients that arise from *pure* noise without any background field. Doing so will give us better insight. If these purely noise-generated on-axis gradients are small compared to those for the noise-free data, then we will know that the effect of noise is small.

How shall we assign random field values to each grid point? Suppose the grid points are numbered from 1 to  $N$ . For example, in the calculation described in Section 16.1.3,  $N = 11,527,201$ . Let  $(x_j, y_j, z_j)$  be the coordinates of the  $j^{\text{th}}$  grid point. Let  $B_y(0, 0, z)$  be the vertical on-axis field arising from the monopole doublet and displayed in Figure 13.7.3. To model noise we make the Ansätze

$$B_x^{\text{noise}}(x_j, y_j, z_j) = \epsilon B_y(0, 0, z_j) \delta_x(j), \quad (20.2.1)$$

$$B_y^{\text{noise}}(x_j, y_j, z_j) = \epsilon B_y(0, 0, z_j) \delta_y(j). \quad (20.2.2)$$

Here the  $\delta_x(j)$  and  $\delta_y(j)$  are random numbers uniformly distributed in the interval  $[-1, 1]$ , and we set  $\epsilon = .01$ . By this prescription we produce a random field that is proportional, at

---

<sup>1</sup>Note, as observed earlier, that the  $z$  component of the field makes no contributions to the on-axis gradients.

the 1% level, to the strength of the monopole-doublet on-axis vertical field.<sup>2</sup>

The first step in the purely numerical calculation is to interpolate the field onto the surface of the cylinder and find its normal component at each of the cylinder sampling points. Figures 2.1 and 2.2 show the resulting  $B_\rho(R, \phi, z = 0)$  for two different random number seeds. Compare with Figure 13.7.6. Correspondingly, Figures 2.3 and 2.4 show  $B_\rho(R, \phi = \pi/2, z)$ . Compare with Figure 13.7.7. We see that the surface field is noisy as expected, and the noise field falls to zero as  $z \rightarrow \pm\infty$  because  $B_y(0, 0, z)$  does so.

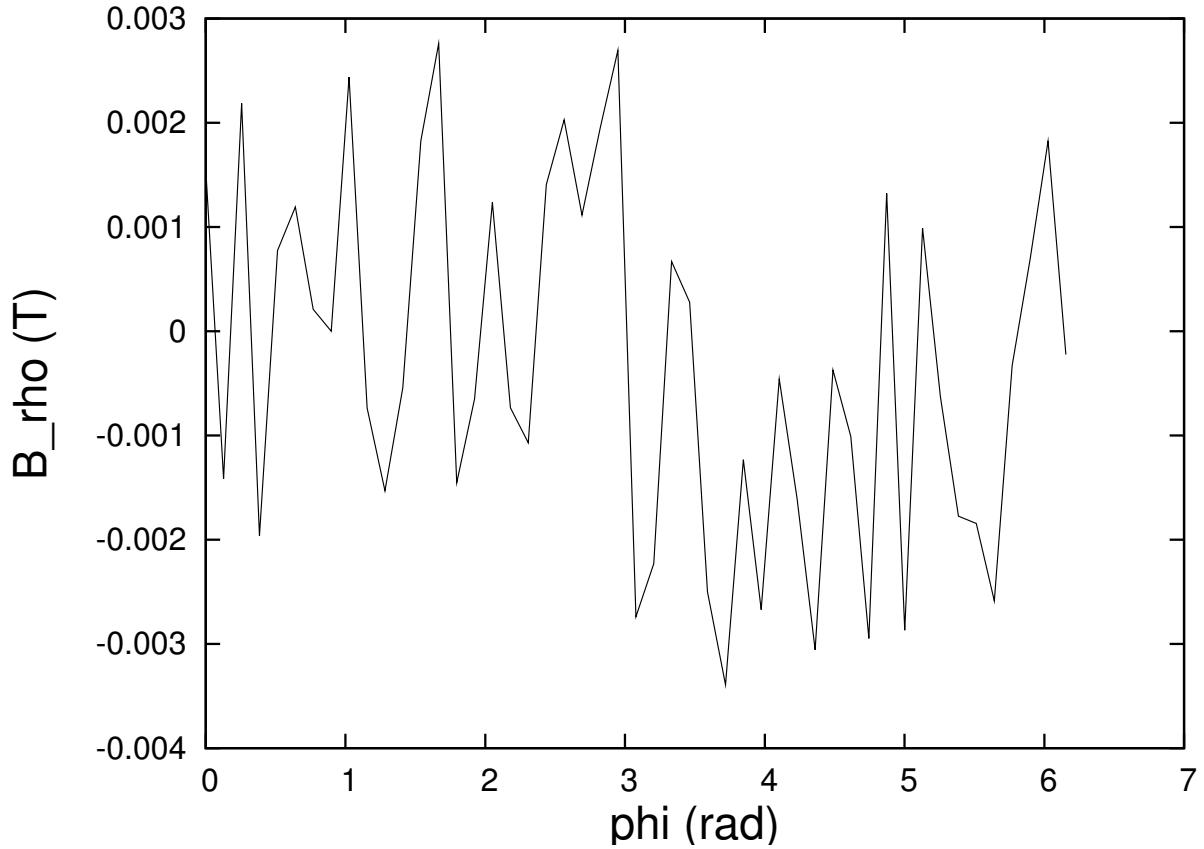


Figure 20.2.1: The function  $B_\rho(R, \phi, z = 0)$  produced by a pure noise field.

Suppose, for example, that we now wish to compute the  $C_{1,s}^{[0]}(z)$  produced by the pure noise field. Then, according to (14.3.23), we first need to compute  $\tilde{B}_\rho^s(R, m = 1, k)$ . And, because no particular symmetry is assumed for the noise,  $\tilde{B}_\rho^s(R, m = 1, k)$  will have both real and imaginary parts. Figures 2.5 and 2.6 display the real parts for the two different choices of random number seed. [The imaginary parts behave analogously. They no longer vanish because  $\tilde{B}_\rho^s(R, m = 1, z)$  for the noise is not assumed to be even in  $z$ .] In both cases

---

<sup>2</sup>This would be the general procedure. Actually, for this study, we generate a random field at all the  $N = 20,937,161$  points described in Section 16.2.6; but, of course, only those required for interpolation onto the surface of the circular cylinder are actually used. We do this because in Section 17.3 we want to compare the use of circular and elliptic cylinders, and for this purpose we want to have a common data base. Note that the grid points described in Section 16.1.3 are a subset of those described in Section 16.2.6.

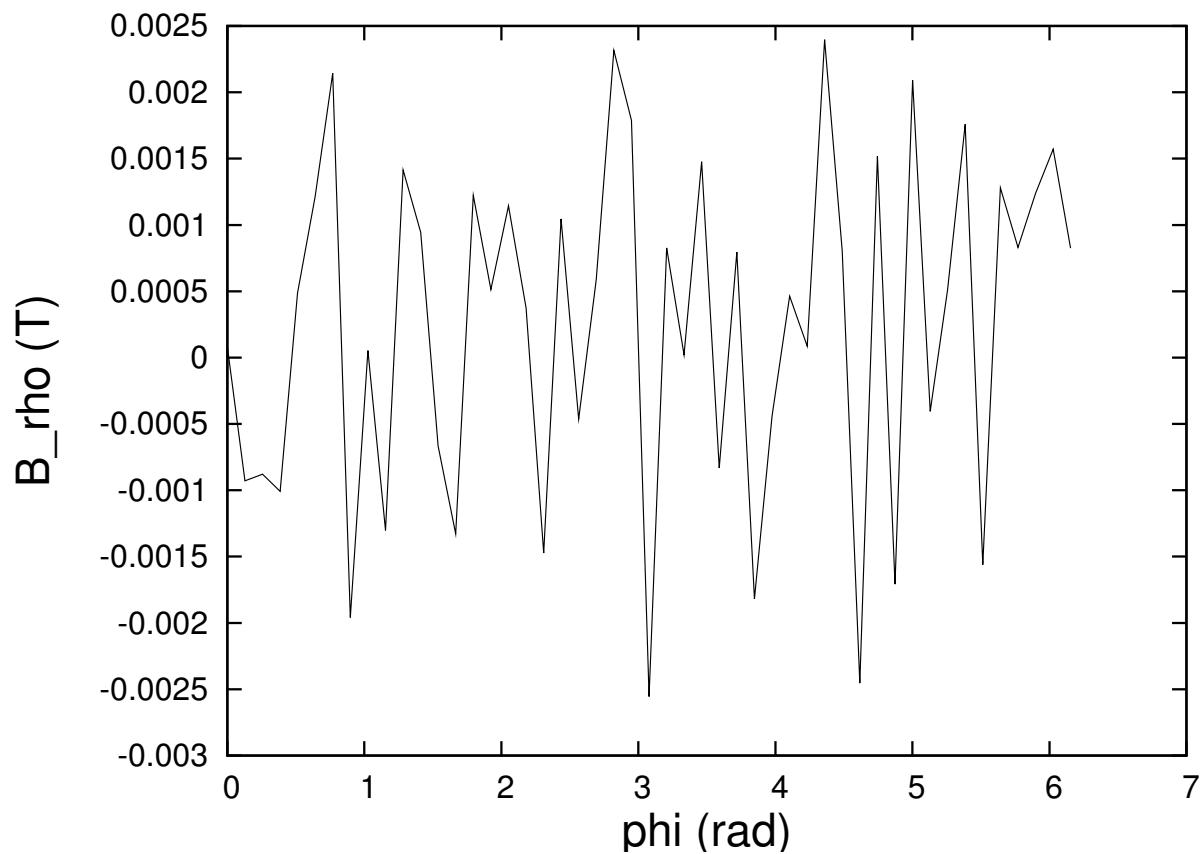


Figure 20.2.2: The function  $B_\rho(R, \phi, z = 0)$  produced by a pure noise field arising from a second different random number seed.

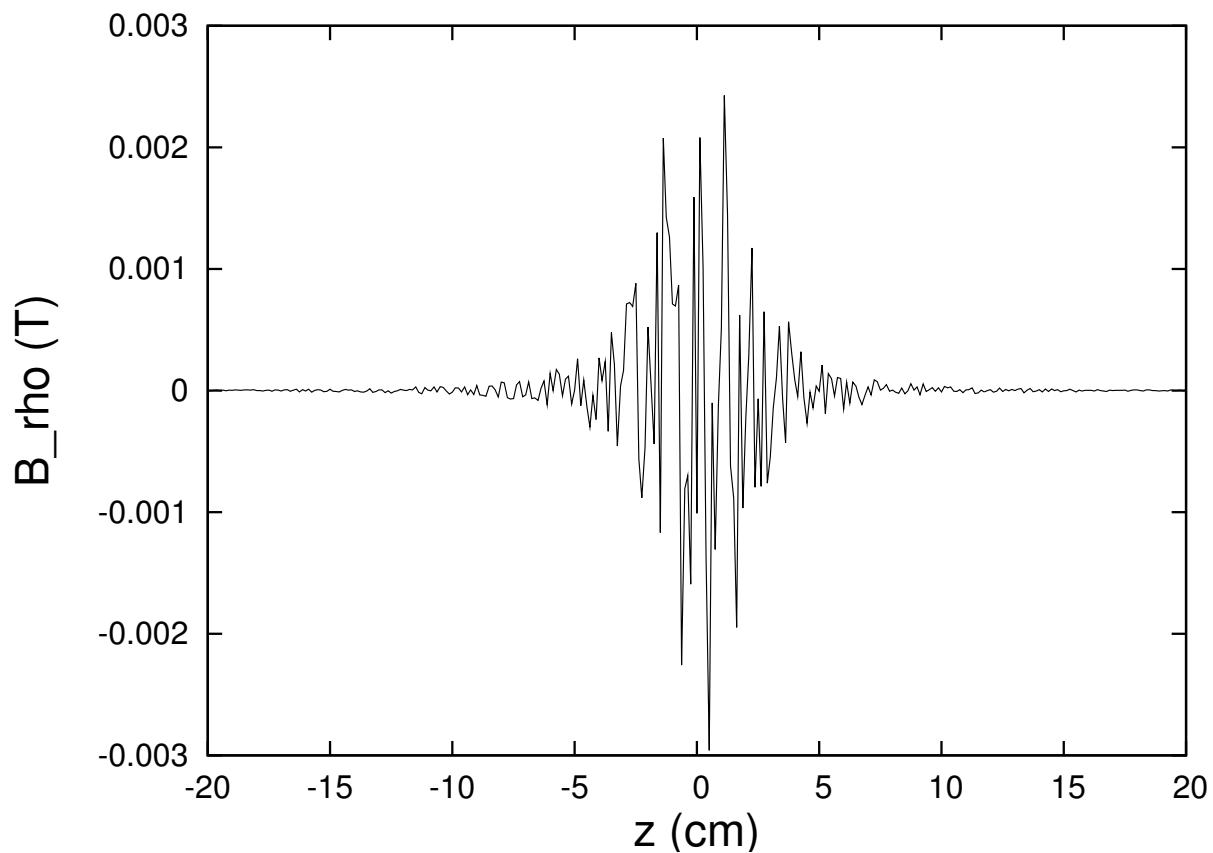


Figure 20.2.3: The function  $B_\rho(R, \phi = \pi/2, z)$  produced by a pure noise field.

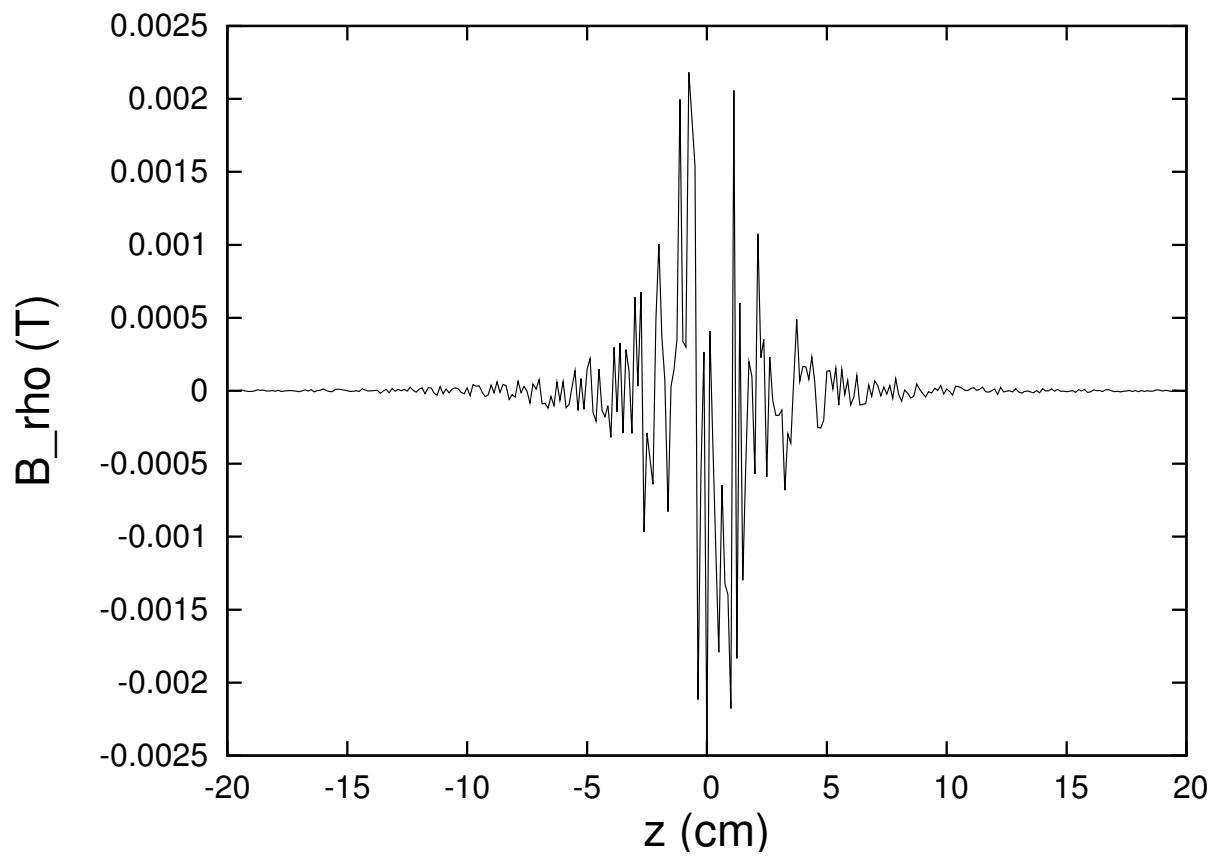


Figure 20.2.4: The function  $B_\rho(R, \phi = \pi/2, z)$  produced by a pure noise field arising from a second different random number seed.

they have support for large values of  $|k|$  as expected due to noise. Compare with Figure 16.1.1. In fact, because  $h_z = .125$ , we expect the noise to have Fourier contributions out to  $K_{Ny} = \pi/h_z = 8\pi$ , which is consistent with what is displayed.

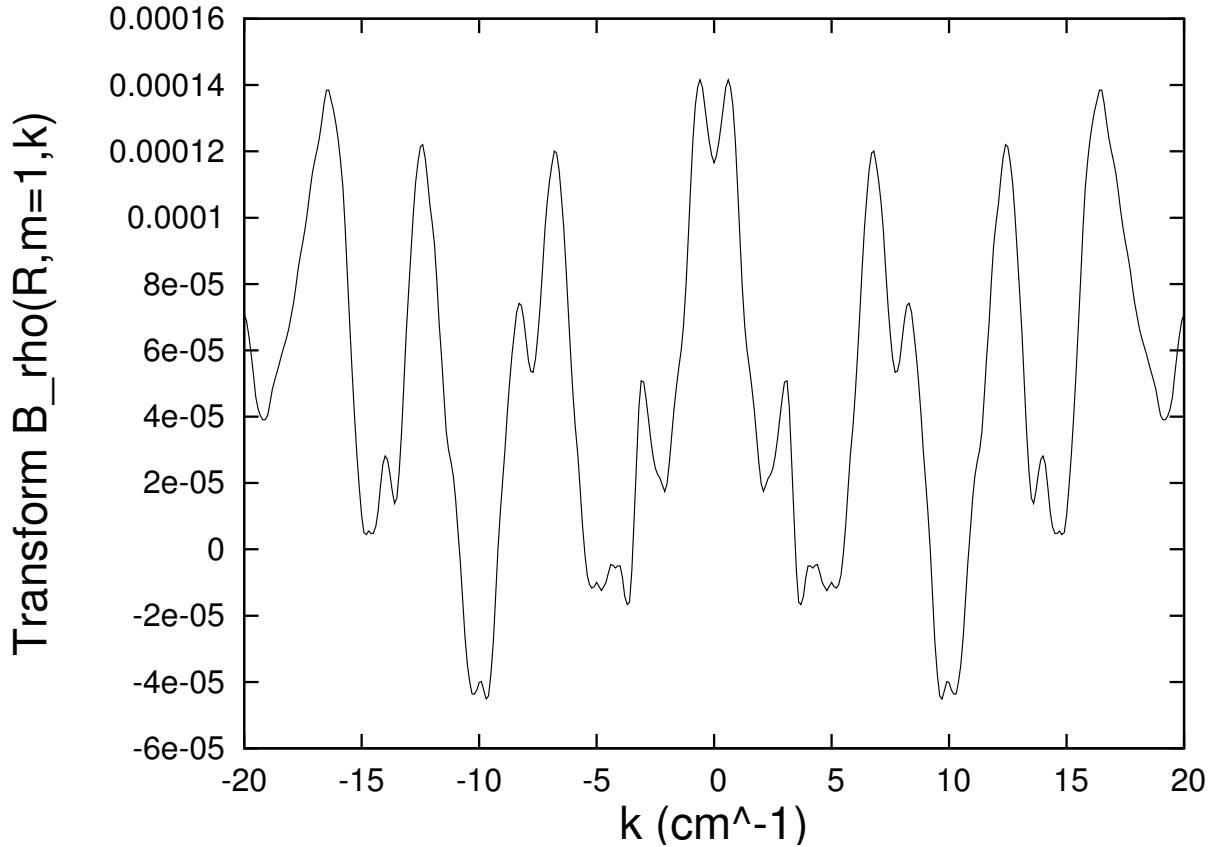


Figure 20.2.5: Real part of  $\tilde{\tilde{B}}_\rho^s(R, m = 1, k)$  produced by a pure noise field. The imaginary part is comparable.

Next, again according to (14.3.23), we need to multiply  $\tilde{\tilde{B}}_\rho^s(R, m = 1, k)$  by the kernel shown in Figure 16.1.3. Figures 2.7 and 2.8 show (for the real part) the results of this multiplication for the two different seed cases. We see, in both cases, that high spatial frequency noise is filtered out by the kernel.

Finally, we need to carry out the integration in (14.3.23). Figures 2.9 and 2.10 show the  $C_{1,s}^{[0]}(z)$  so obtained for each noise realization. Comparison of these figures with Figure 16.1.7 shows that in this study a 1% noise in field data produces at most a .03% error in  $C_{1,s}^{[0]}(z)$ . Note that, unlike the case of Figure 16.1.7,  $C_{1,s}^{[0]}(z)$  in Figures 2.9 and 2.10 is not symmetric about  $z = 0$ . There is no assumed symmetry for the noise.

What can be said about the other  $C_{m,s}^{[n]}(z)$ ? They too are small. For example, Figures 2.11 through 2.14 show the functions  $C_{1,s}^{[6]}(z)$  and  $C_{7,s}^{[0]}(z)$ . Comparison with Figures 16.1.8 and 16.1.15 shows that in this case a 1% noise in field data produces at most a .02% error in  $C_{1,s}^{[6]}(z)$  and a .08% error in  $C_{7,s}^{[0]}(z)$ . It is remarkable that the error in the on-axis gradients is considerably smaller than that in the field data. It seems particularly remarkable that

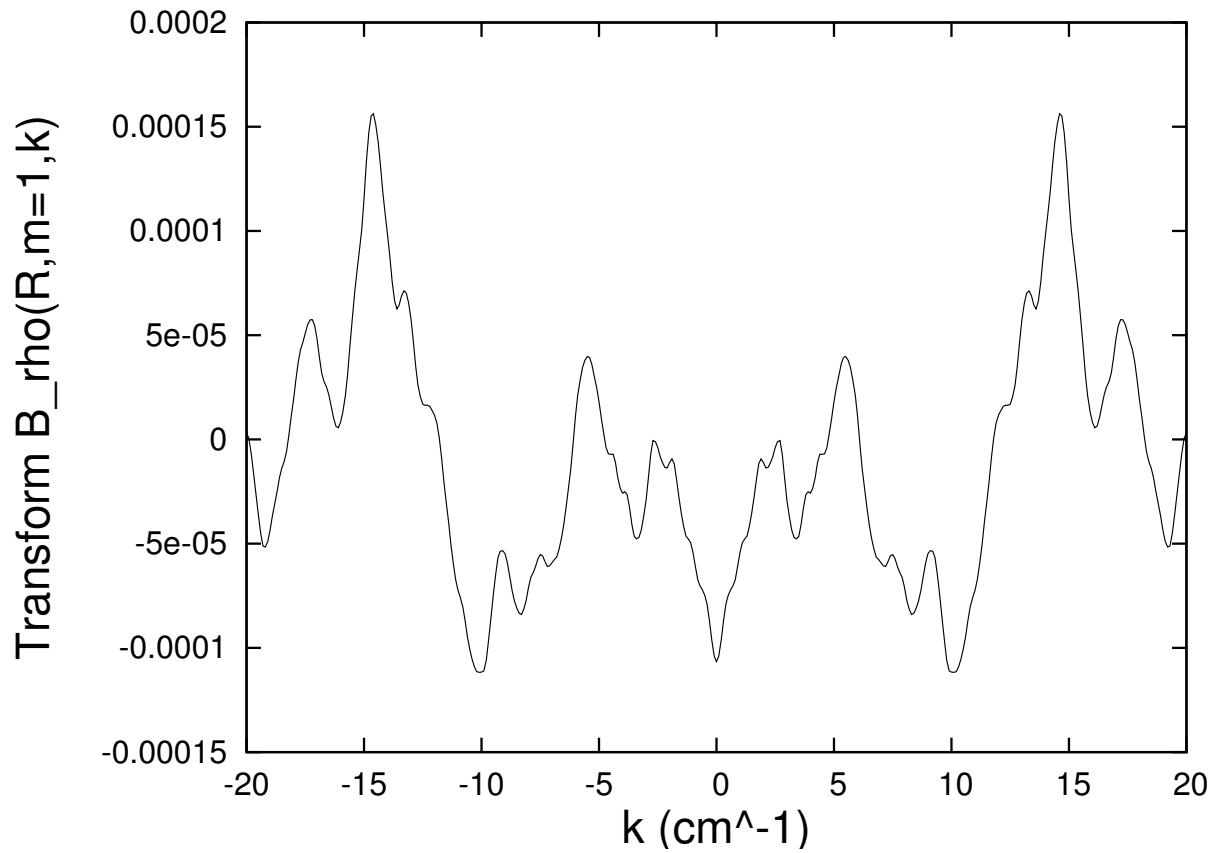


Figure 20.2.6: Real part of  $\tilde{B}_\rho(R, m = 1, k)$  produced by a pure noise field arising from a second different random number seed. The imaginary part is comparable.

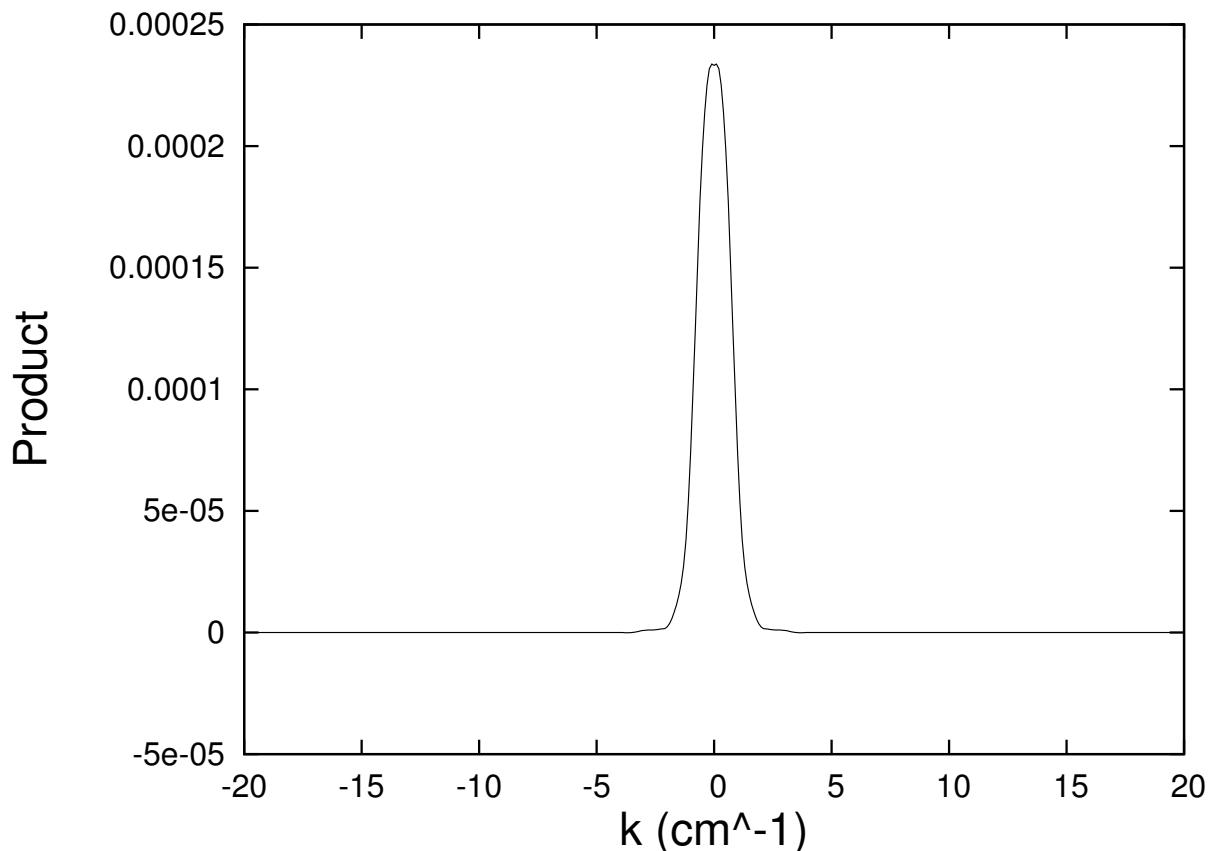


Figure 20.2.7: The product of  $\Re \tilde{B}_\rho(R, m = 1, k)$  for the first random number seed and the kernel of Figure 16.1.3.

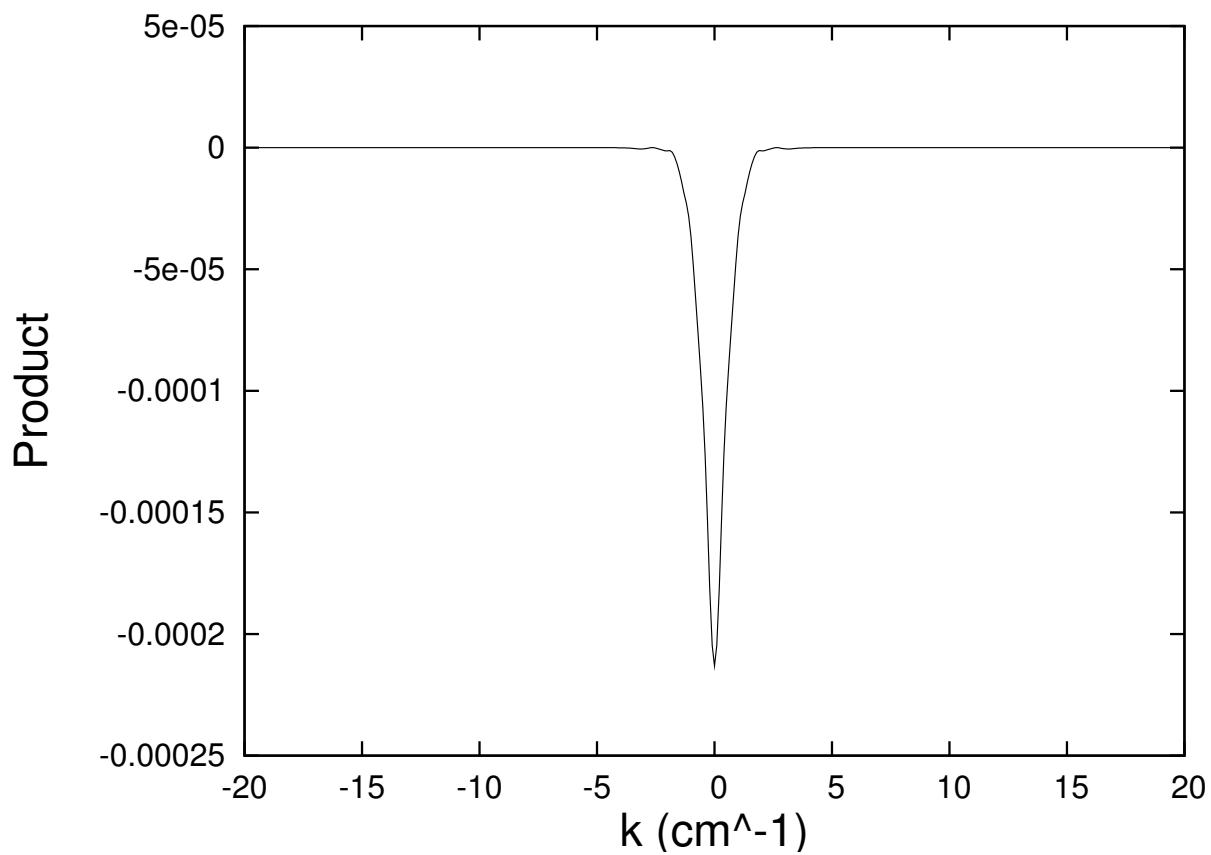


Figure 20.2.8: The product of  $\Re \tilde{B}_\rho(R, m = 1, k)$  for the second different random number seed and the kernel of Figure 16.1.3.

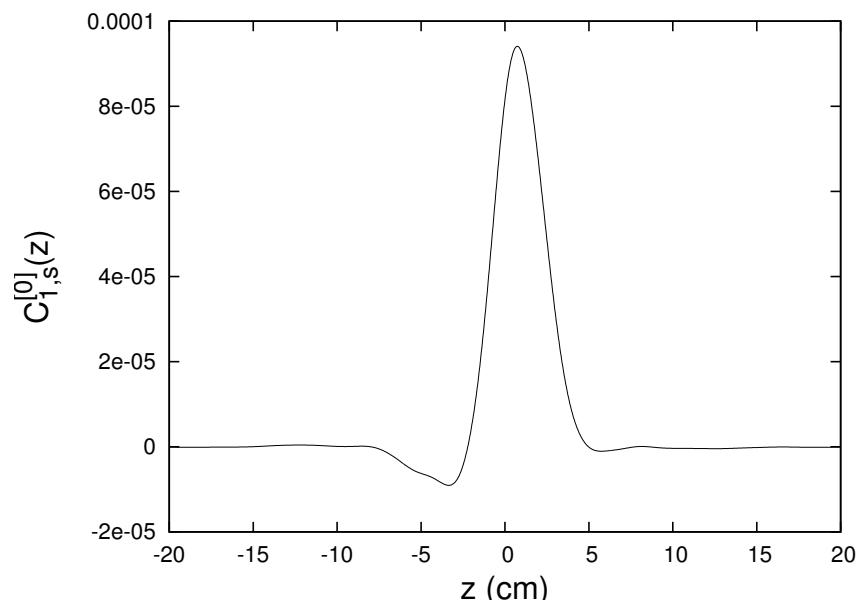


Figure 20.2.9: The function  $C_{1,s}^{[0]}(z)$  produced by a pure noise field.

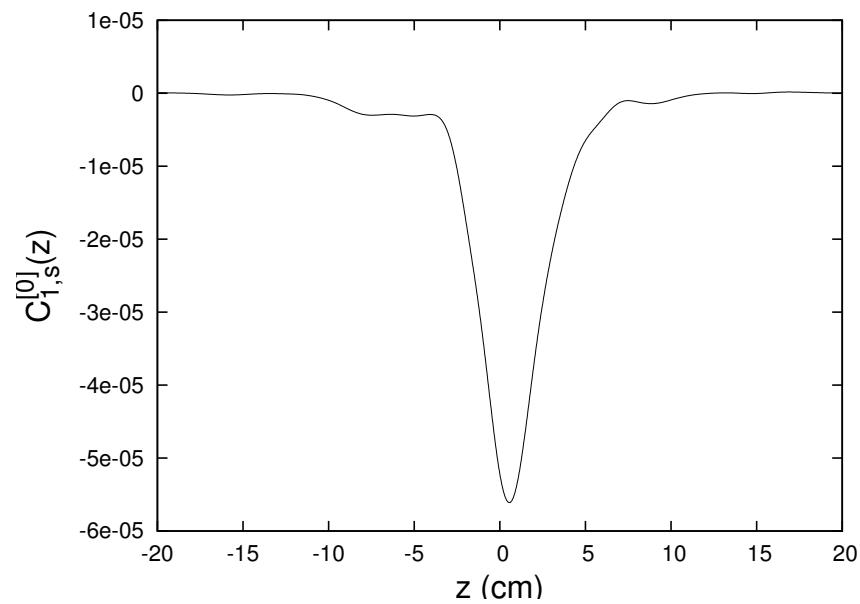


Figure 20.2.10: The function  $C_{1,s}^{[0]}(z)$  produced by a pure noise field arising from a second different random number seed.

the error in  $C_{1,s}^{[6]}(z)$  is so small because it involves 6 derivatives and the interpolated surface data, as evidenced by Figures 2.3 through 2.6, essentially has no derivatives.

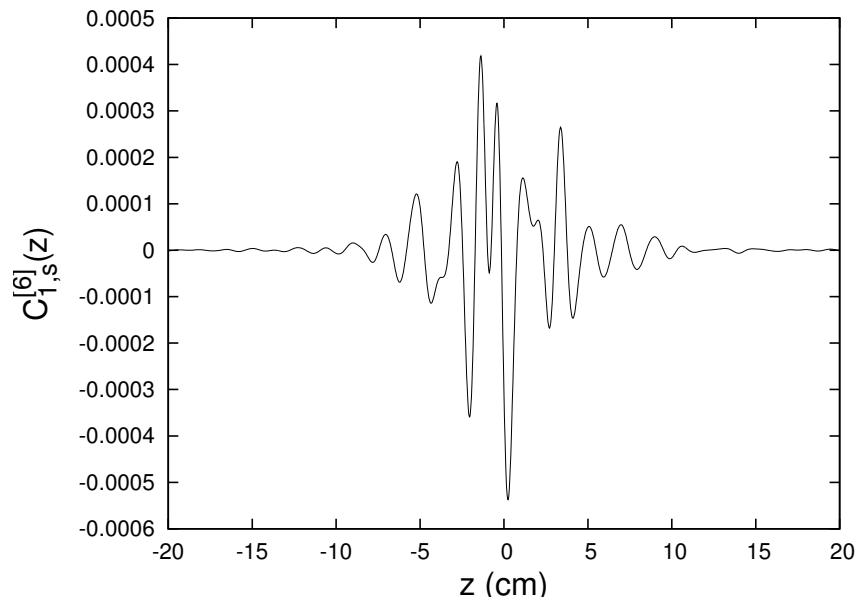


Figure 20.2.11: The function  $C_{1,s}^{[6]}(z)$  produced by a pure noise field.

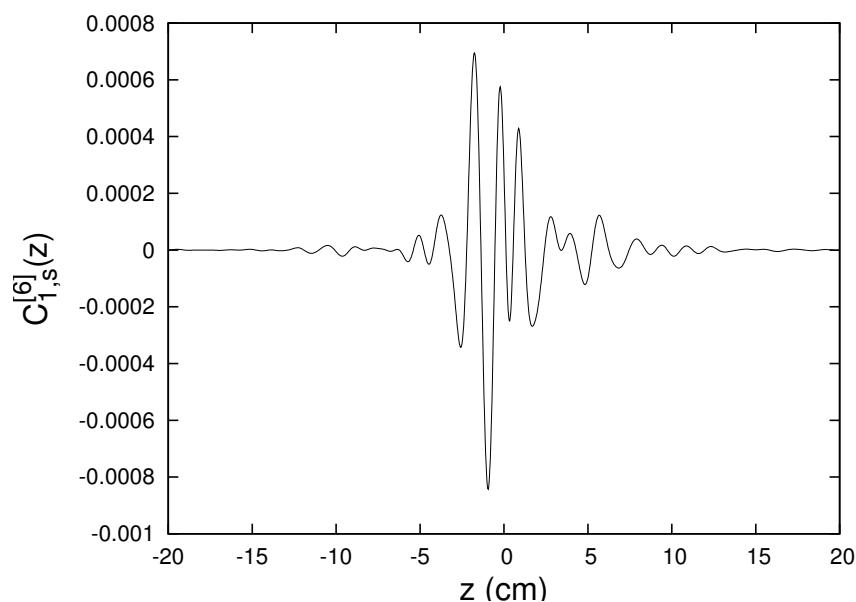


Figure 20.2.12: The function  $C_{1,s}^{[6]}(z)$  produced by a pure noise field arising from a second different random number seed.

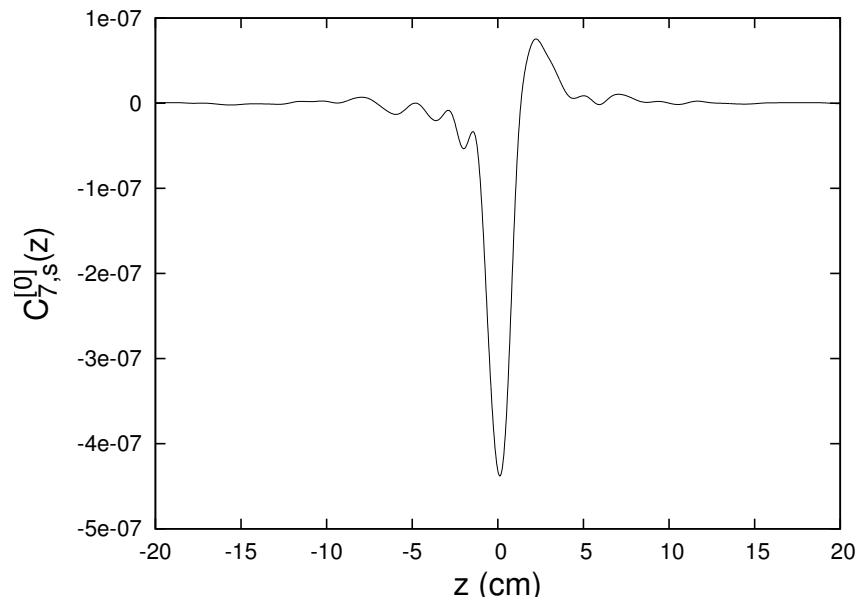


Figure 20.2.13: The function  $C_{7,s}^{[0]}(z)$  produced by a pure noise field.

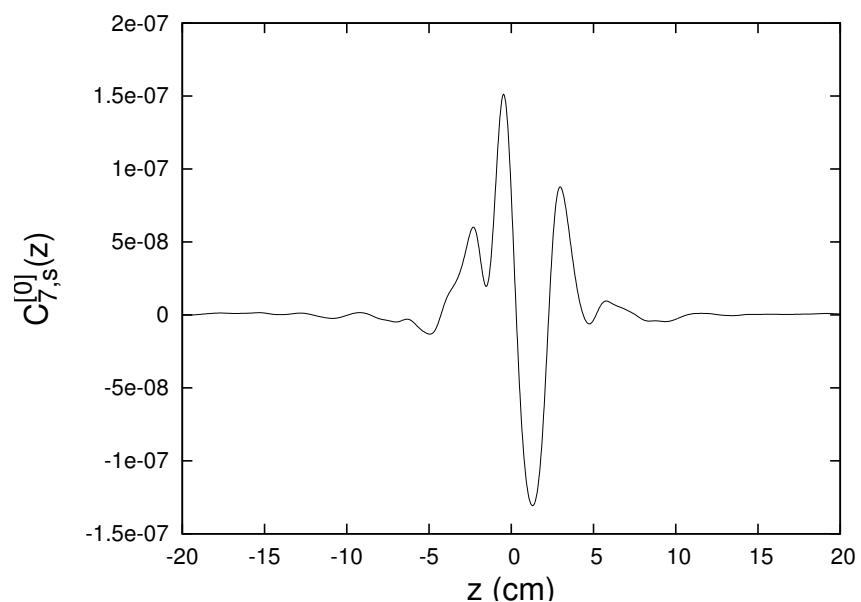


Figure 20.2.14: The function  $C_{7,s}^{[0]}(z)$  produced by a pure noise field arising from a second different random number seed.

We end this section by exploring how smoothing depends on the radius of the circular cylinder. Figures 2.9 through 2.14 presented results for the case of a circular cylinder having radius  $R = 2$ . What if we had instead used a circular cylinder with  $R = 1$ ? Presumably the effect of noise will be larger because there will then be less filtering. See (14.2.7).

Let us make a simple model of what to expect. Look at (14.3.23). Consistent with arising from a noise source, suppose the associated  $\tilde{B}_\rho(R, m, k)$  is essentially independent of  $k$ . Then we have the bound

$$|C_{m,\alpha}^{[n]}(z)| \leq |\tilde{B}_\rho^\alpha(R, m, k \approx 0)|[(1/2)^m(1/m!)] \int_{-\infty}^{\infty} dk [|k|^{n+m-1} / |I'_m(kR)|]. \quad (20.2.3)$$

By the change of variables (1.20), the integral appearing in (2.3) can be brought to the form

$$\int_{-\infty}^{\infty} dk [|k|^{n+m-1} / |I'_m(kR)|] = (1/R)^{n+m} \int_{-\infty}^{\infty} d\lambda [|\lambda|^{n+m-1} / |I'_m(\lambda)|]. \quad (20.2.4)$$

Correspondingly, the bound (2.3) now takes the form

$$|C_{m,\alpha}^{[n]}(z)| \leq |\tilde{B}_\rho^\alpha(R, m, k \approx 0)|[(1/2)^m(1/m!)](1/R)^{n+m} \int_{-\infty}^{\infty} d\lambda [|\lambda|^{n+m-1} / |I'_m(\lambda)|]. \quad (20.2.5)$$

On the assumption that the noise itself is independent of  $R$ , we see that because of smoothing the  $C_{m,\alpha}^{[n]}(z)$  due to noise may be expected to decrease with increasing  $R$  as  $(1/R)^{n+m}$ .

What actually happens? Figures 2.15 through 2.20 compare the results for  $R = 1$  and  $R = 2$ . We see that, as a general trend, noise indeed has a larger effect when the smaller cylinder is employed. This is particularly true, as expected, for large values of  $n + m$ . But, in the case of Figure 2.15, the  $C_{1,s}^{[0]}(z)$  computed from noise on the  $R = 1$  cylinder is smaller than that computed from noise on the  $R = 2$  cylinder. How can this be? As explained in the beginning of this section, see (2.1) and (2.2), in our model the noise values at the various grid points are independent. It can happen, through statistical fluctuations, that the net noise on the  $R = 1$  cylinder is considerably less than on the  $R = 2$  cylinder, so much so that this fluctuation effect more than compensates the poorer smoothing supplied by the smaller cylinder.

As a check on this explanation, suppose we attempt to make the noise on the  $R = 1$  cylinder nearly the same as that on the  $R = 2$  cylinder. One way to do this is the following: Suppose a noise value is required at some point on the  $R = 1$  cylinder having angle  $\phi$ . Instead of interpolating off nearby grid points, we may find the point on the  $R = 2$  cylinder having the same  $\phi$  value, and then interpolate off grid points near this  $R = 2$  point. Figure 2.21 shows what happens when this done. Evidently the effect of noise on the  $R = 1$  cylinder is now larger than the effect of essentially the same noise on the  $R = 2$  cylinder.

We conclude that, as hoped, expected, and advertised, the use of surface methods (in this case the surface of a circular cylinder) does indeed yield results that are relatively insensitive to noise, and that this insensitivity is improved by placing the surface farther from the axis.

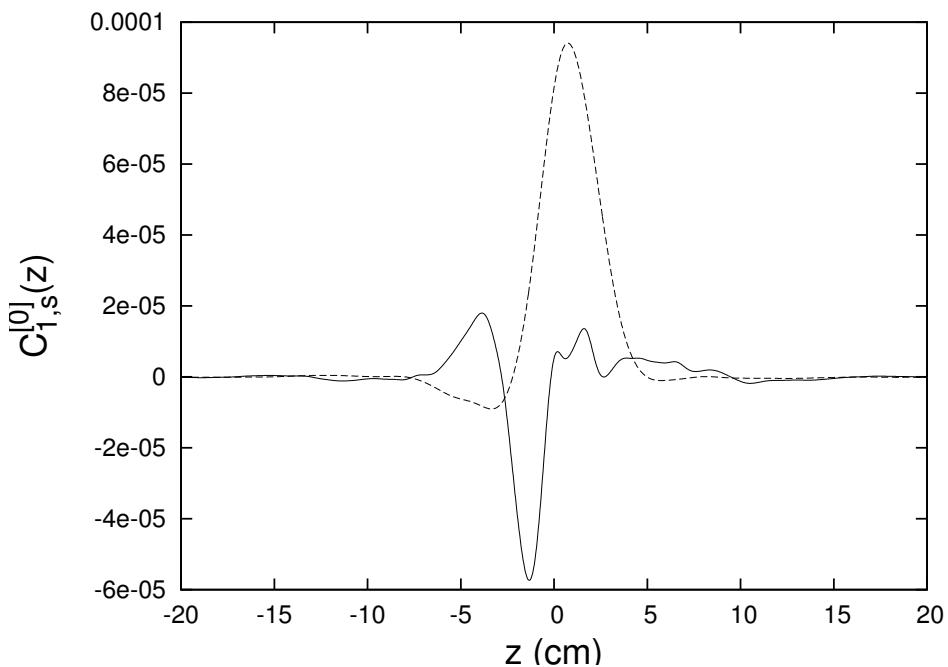


Figure 20.2.15: The functions  $C_{1,s}^{[0]}(z)$  produced by a pure noise field on circular cylinders having  $R = 1$  (solid line) and  $R = 2$  (dashed line).

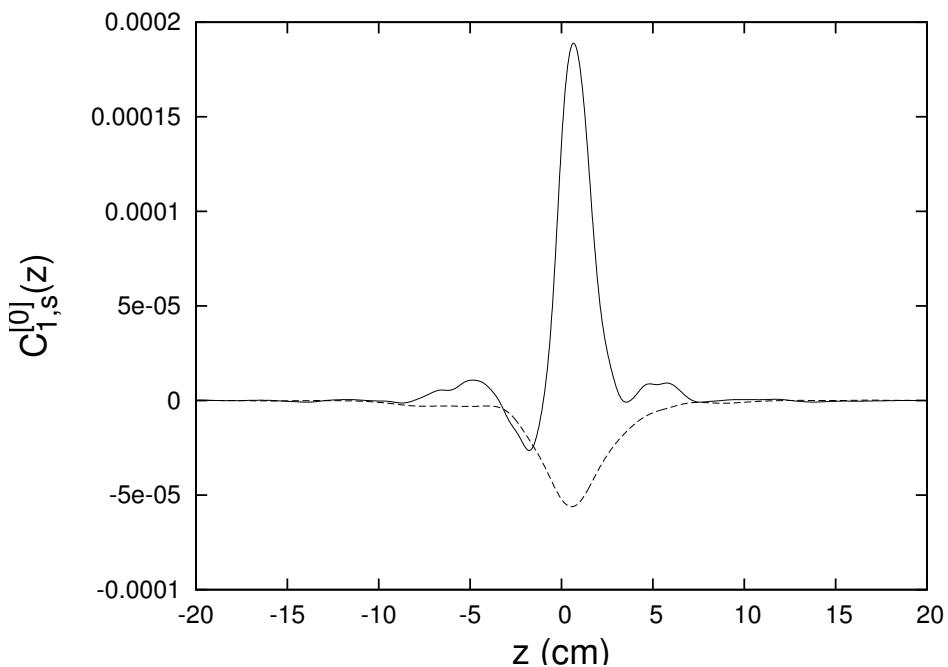


Figure 20.2.16: The functions  $C_{1,s}^{[0]}(z)$  produced by a pure noise field on circular cylinders having  $R = 1$  (solid line) and  $R = 2$  (dashed line) and arising from a second different random number seed.

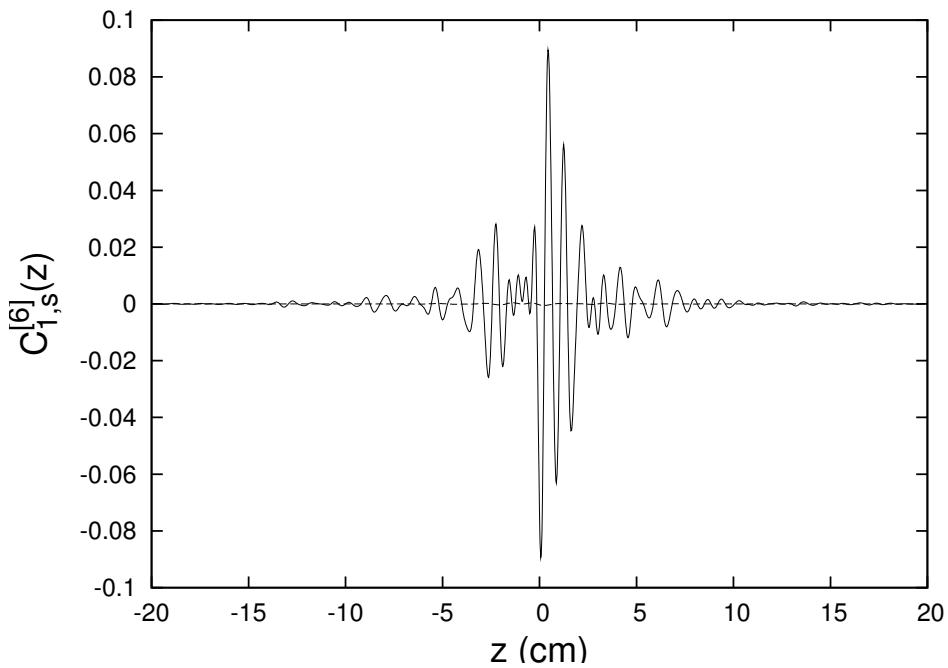


Figure 20.2.17: The functions  $C_{1,s}^{[6]}(z)$  produced by a pure noise field on circular cylinders having  $R = 1$  (solid line) and  $R = 2$  (dashed line).

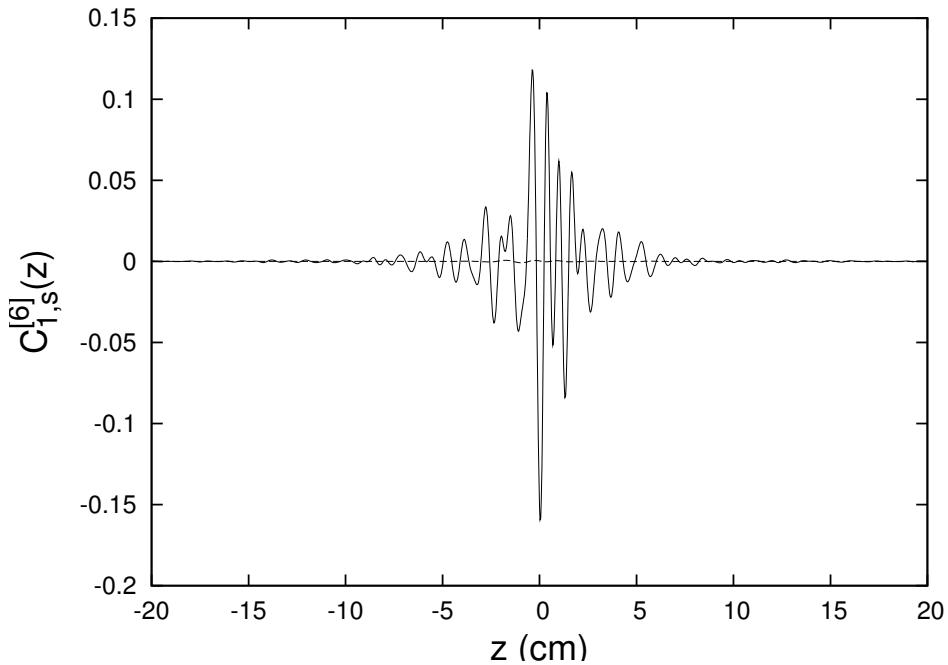


Figure 20.2.18: The functions  $C_{1,s}^{[6]}(z)$  produced by a pure noise field on circular cylinders having  $R = 1$  (solid line) and  $R = 2$  (dashed line) and arising from a second different random number seed.

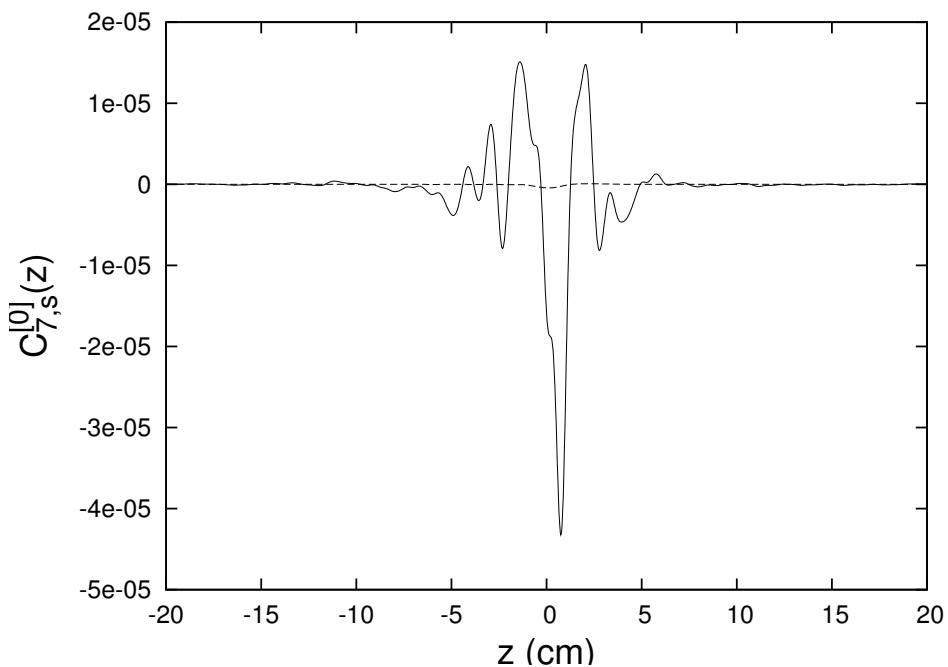


Figure 20.2.19: The functions  $C_{7,s}^{[0]}(z)$  produced by a pure noise field on circular cylinders having  $R = 1$  (solid line) and  $R = 2$  (dashed line).

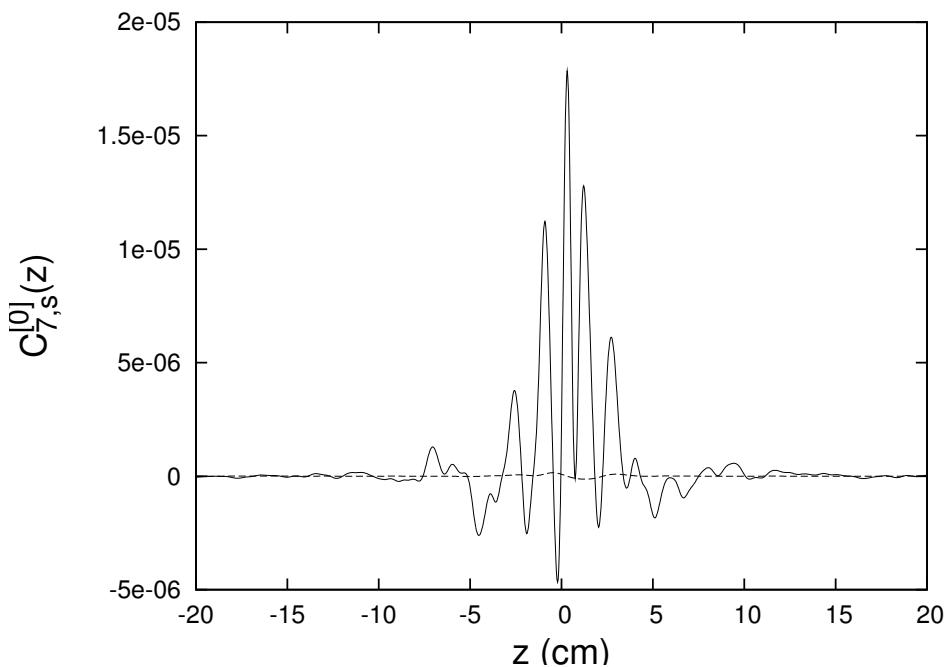


Figure 20.2.20: The functions  $C_{7,s}^{[0]}(z)$  produced by a pure noise field on circular cylinders having  $R = 1$  (solid line) and  $R = 2$  (dashed line) and arising from a second different random number seed.

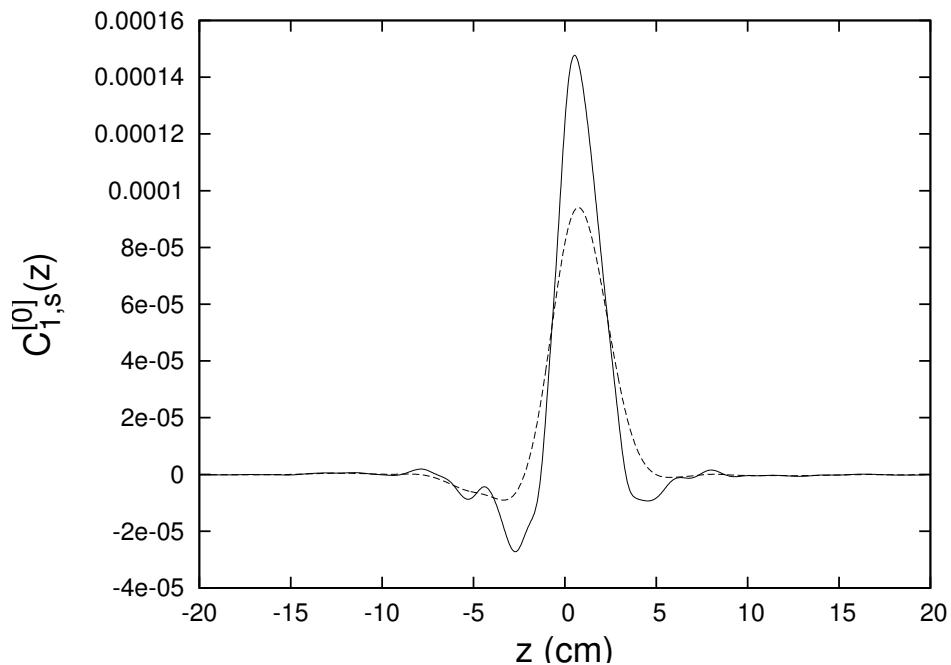


Figure 20.2.21: The functions  $C_{1,s}^{[0]}(z)$  produced by nearly the same pure noise fields on circular cylinders having  $R = 1$  (solid line) and  $R = 2$  (dashed line).

## 20.3 Elliptic Cylinders

The discussion of this section parallels that of the previous section, but now deals with the monopole-doublet test case and the use of the surface of an elliptic cylinder as described in Section 16.2.6. We use the same noise model as that of the previous section.

Now the first step in the purely numerical calculation is to interpolate the field onto the surface of the elliptic cylinder and to find the function  $F(U, v, z)$  given by (14.4.67). Figures 3.1 and 3.2 show the resulting  $F(U, v, z = 0)$  for two different random number seeds. Compare with Figure 16.2.28. Correspondingly, Figures 3.3 and 3.4 show  $F(U, v = \pi/2, z)$ . In view of (14.4.66), this quantity is proportional to the normal component of  $B$  when  $\phi = \pi/2$  so that these figures should be compared with Figure 13.7.7. Observe that the surface field is noisy as expected.

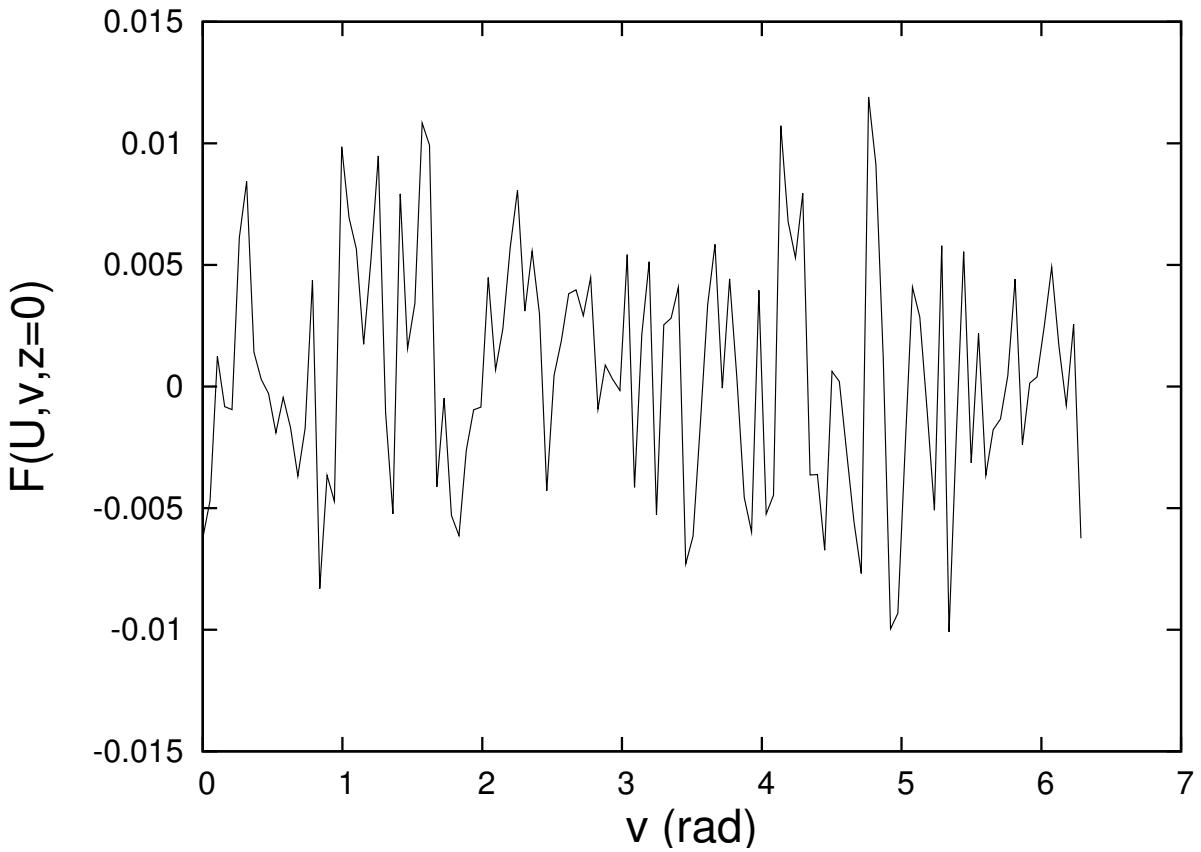


Figure 20.3.1: The function  $F(U, v, z = 0)$  produced by a pure noise field.

According to (14.4.72) the second step is to perform a Fourier transform to produce  $\tilde{F}(v, k)$ . Figures 3.5 and 3.6 display the real parts of  $\tilde{F}(v = \pi/2, k)$  for the two different seeds. The imaginary parts are comparable. Compare with Figure 16.2.3. In both cases  $\tilde{F}(v = \pi/2, k)$  has support for large  $|k|$  as is expected for noisy data.

The third step is to compute the Mathieu coefficients defined by (16.2.20) and (16.2.21). As described in Section 16.2.1, for the monopole doublet we are particularly interested in the coefficients  $\tilde{\tilde{F}}_r^s(k)$  for odd  $r$ . Figures 3.7 and 3.8 display the real parts of the first few

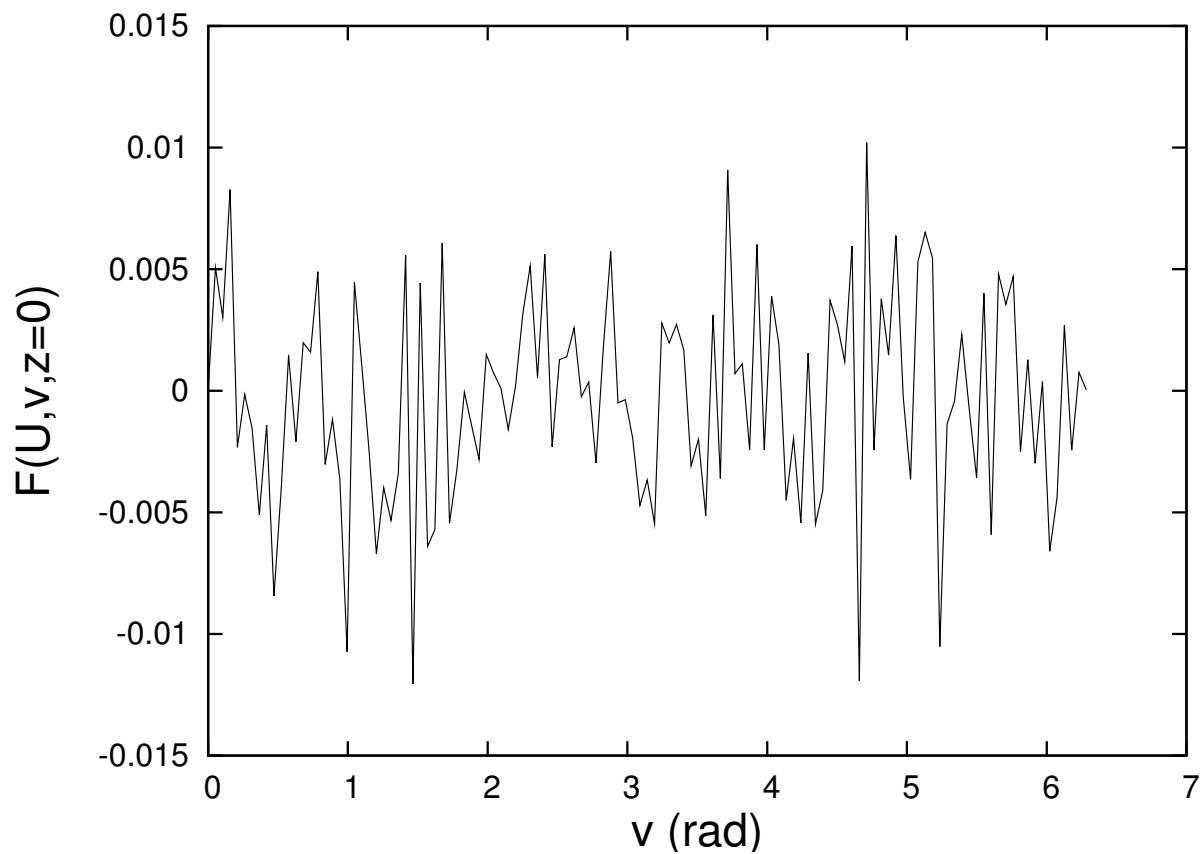


Figure 20.3.2: The function  $F(U, v, z = 0)$  produced by a pure noise field arising from a second different random number seed.

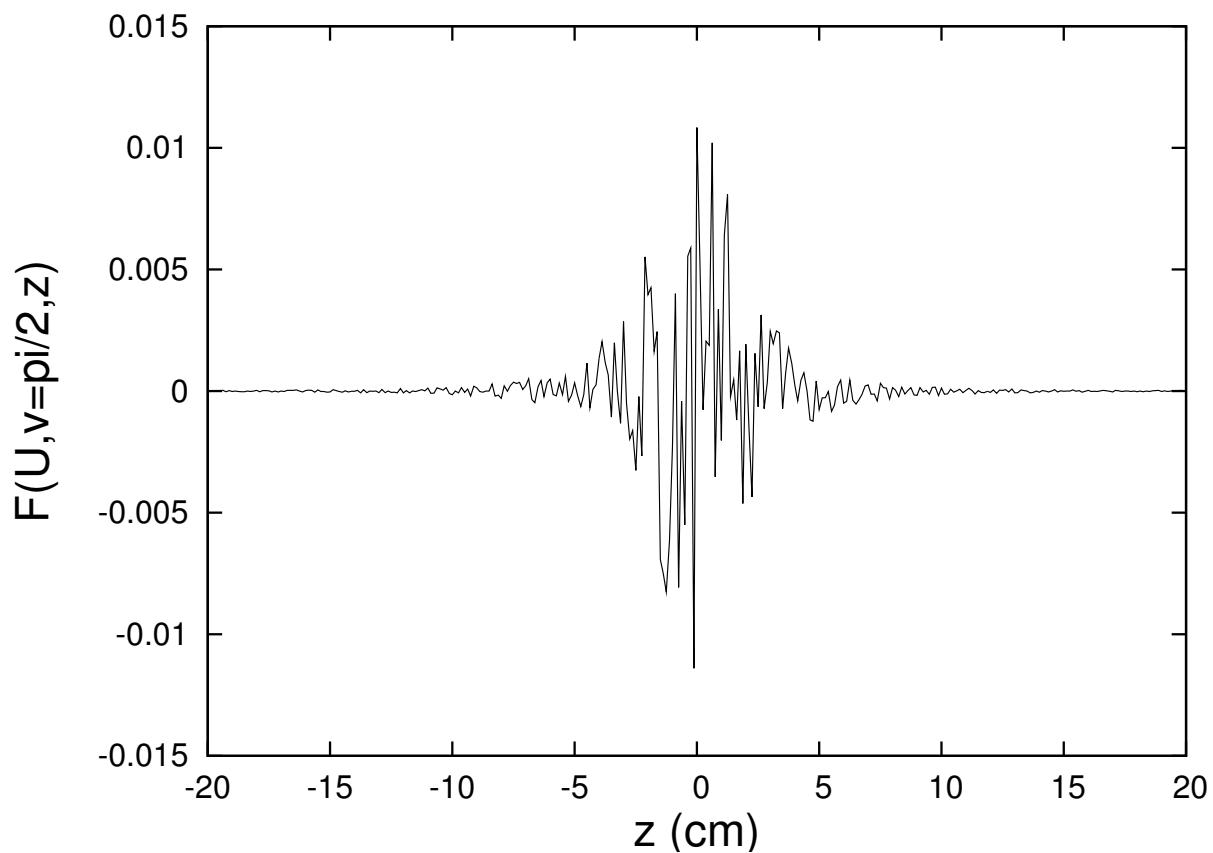


Figure 20.3.3: The function  $F(U, v = \pi/2, z)$  produced by a pure noise field.

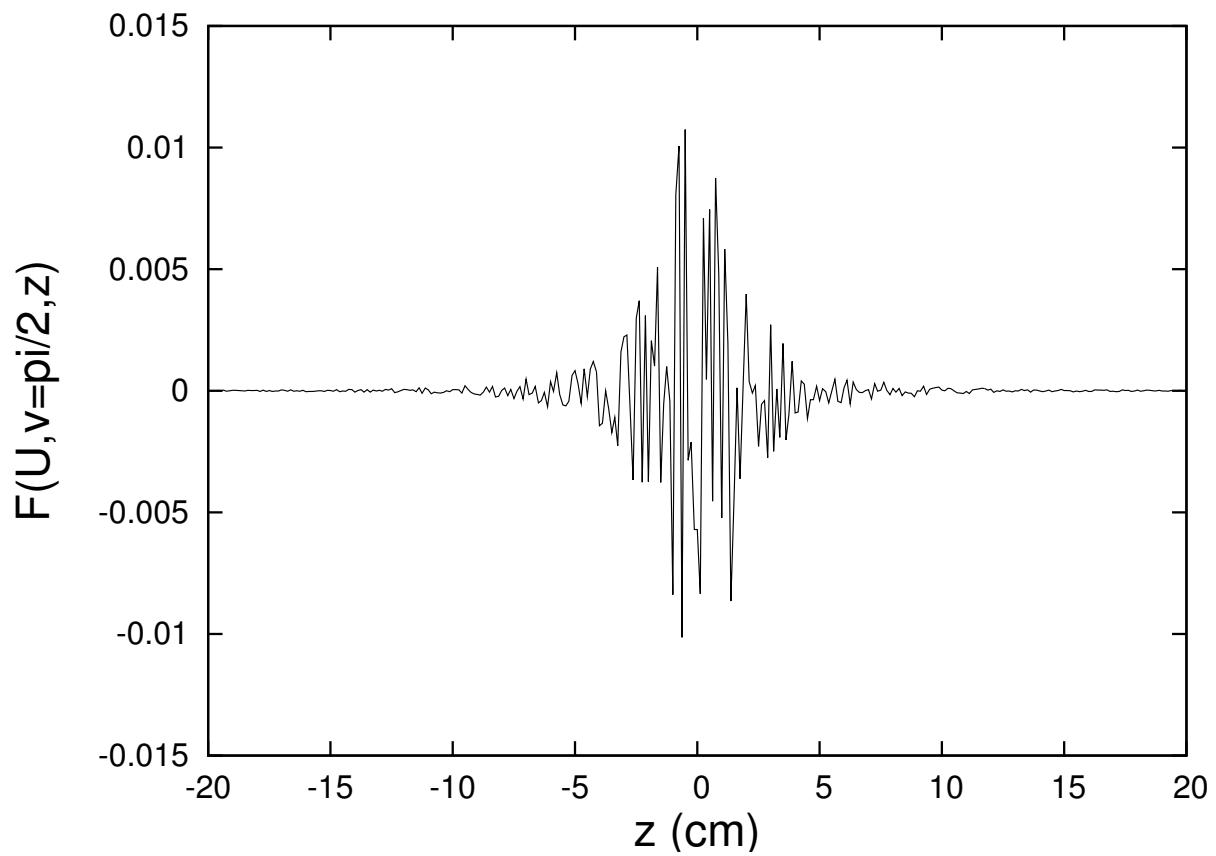


Figure 20.3.4: The function  $F(U, v = \pi/2, z)$  produced by a pure noise field arising from a second different random number seed.

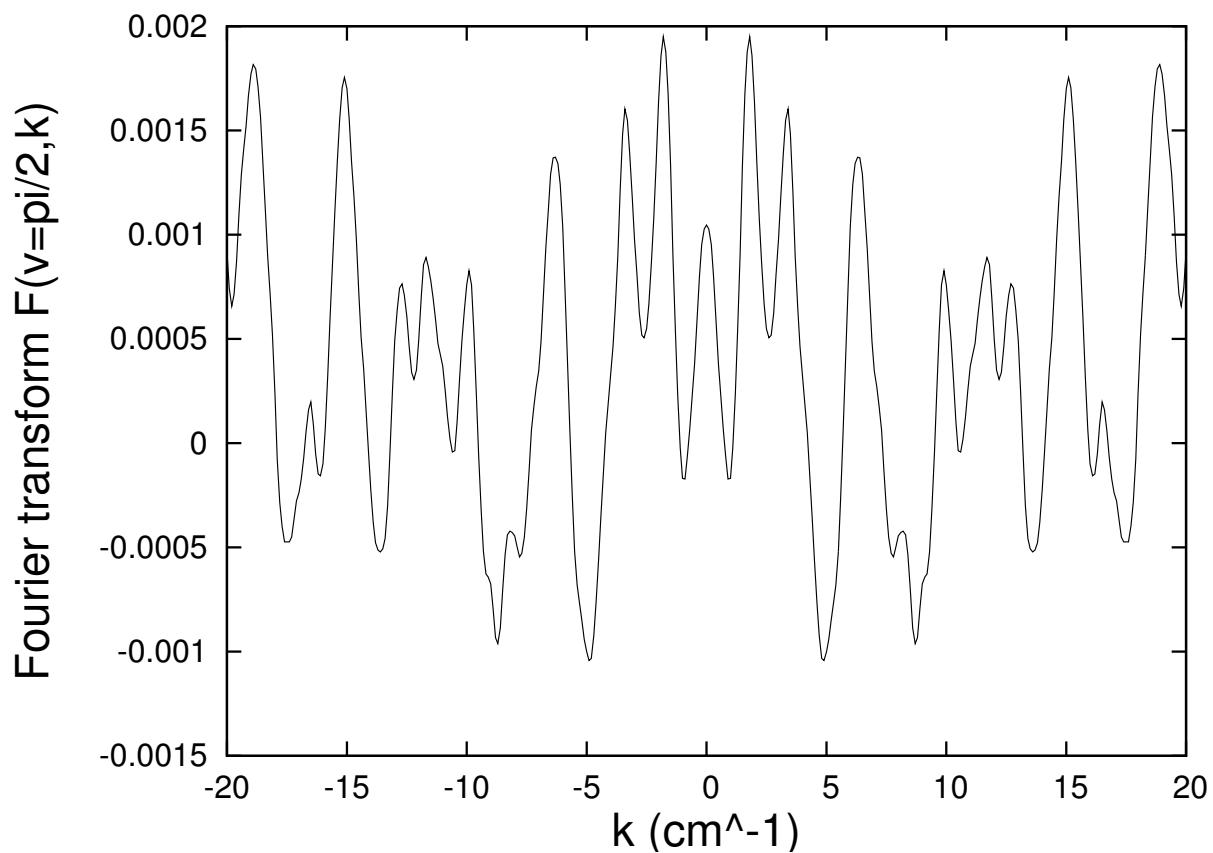


Figure 20.3.5: Real part of  $\tilde{F}(v = \pi/2, k)$  produced by a pure noise field. The imaginary part is comparable.

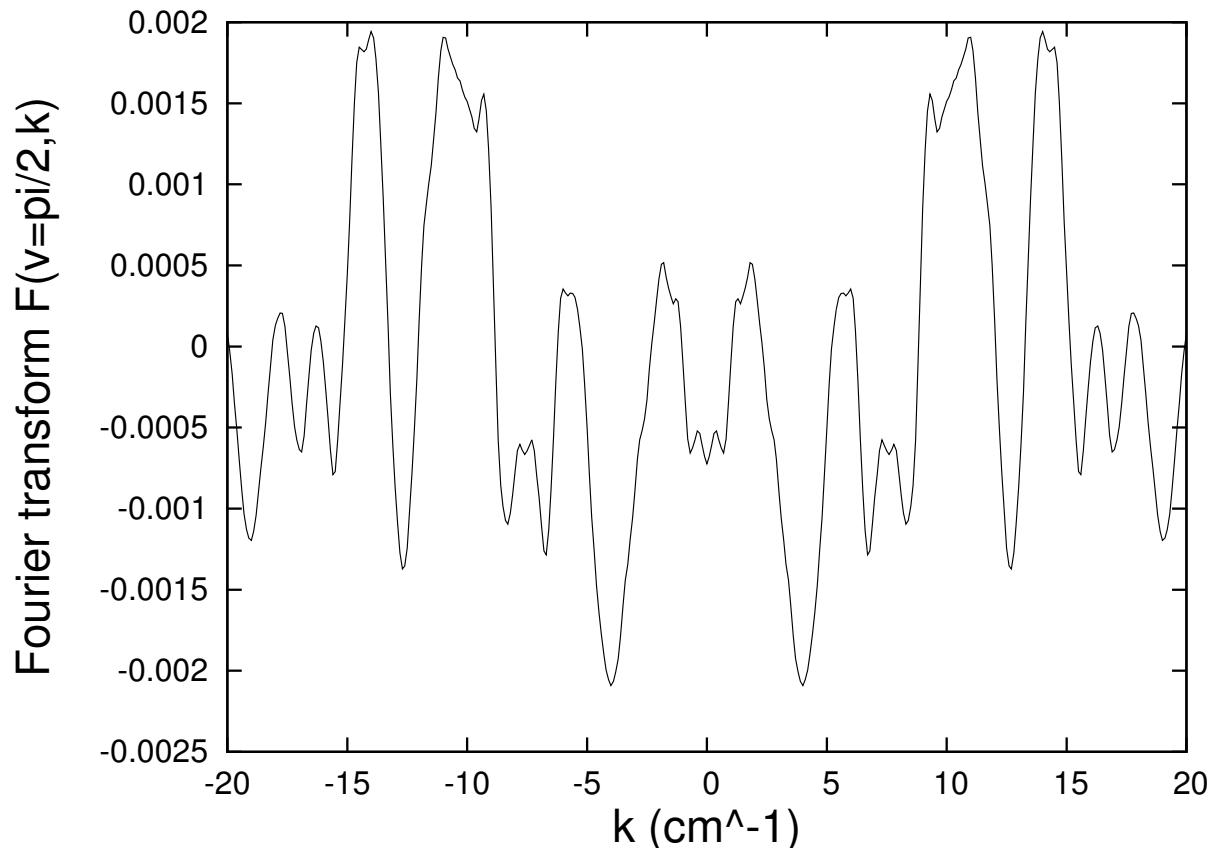


Figure 20.3.6: Real part of  $\tilde{F}(v = \pi/2, k)$  produced by a pure noise field arising from a second different random number seed. The imaginary part is comparable.

of these, namely those for  $r = 1, 3$ , and  $5$ , for the two different seeds. Compare with Figure 16.2.4. Note that they also have support for large  $|k|$ .

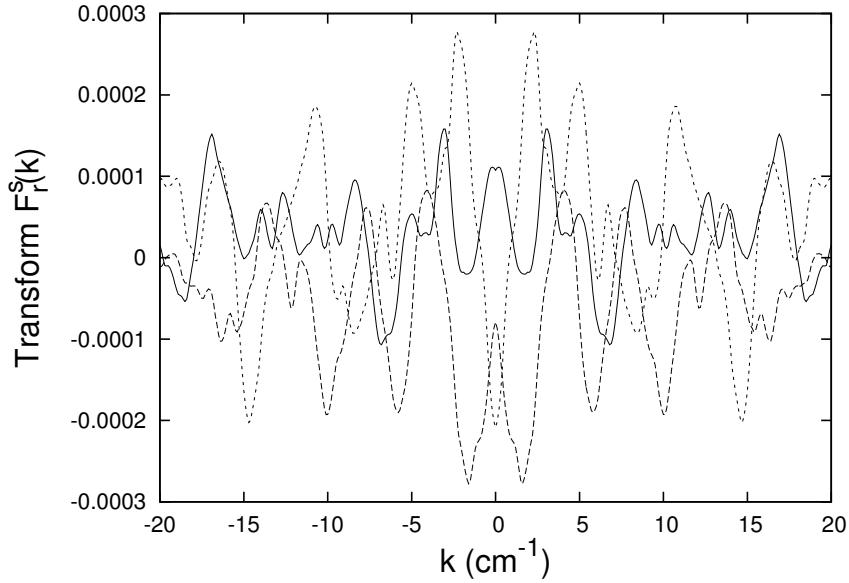


Figure 20.3.7: Real parts of the first few functions  $\tilde{F}_r^s(k)$ , those for  $r = 1, 3$ , and  $5$ , produced by a pure noise field. The imaginary parts are comparable.

Suppose, for example, that we now again wish to compute the  $C_{1,s}^{[0]}(z)$  produced by the pure noise field. Then, according to (14.8.6) and (14.8.4), we must compute the quantity  $kG_{1,s}(k)$  by multiplying the  $\tilde{F}_r^s(k)$  by the kernels  $k\beta_1^r(k)/Se_r'(U, q)$  shown in Figure 16.2.7, and then summing over  $r$ . As described in Section 16.2.3, this sum is terminated at  $r_{\max}(1) = 11$ . Figures 3.9 and 3.10 display the real parts of the  $kG_{1,s}(k)$  for the two different choices of random number seed. The imaginary parts are comparable. Note that, like the circular cylinder case, the kernels effectively filter out all the high frequency components.

This is a good place to compare the filtering provided by the use of an elliptic cylinder with that provided by the use of a circular cylinder. Figure 3.11 shows the circular cylinder  $m = 1$  kernel  $[1/I_1'(kR)]$  and the first few elliptical cylinder  $m = 1$  kernels  $[k\beta_1^r(k)/Se_r'(U, q)]$ , and Figure 3.12 shows the same kernels all normalized to 1 at  $k = 0$ . From Figure 3.12 we see that the elliptic kernels for  $r = 1$  and  $r = 3$  fall off more rapidly with  $|k|$  than the circular kernel, but that they fall off less rapidly for  $r \geq 5$ . The  $r$  value for which this transition occurs depends on the eccentricity of the ellipse: the larger the eccentricity the larger the transition  $r$ . Moreover, we see from Figure 3.11 that the elliptic kernels for small  $r$  dominate.

Finally, we need to carry out the integration in (14.4.86). Figures 3.13 and 3.14 show, as dashed lines, the  $C_{1,s}^{[0]}(z)$  so obtained for each noise realization. Also shown, as solid lines, are the  $C_{1,s}^{[0]}(z)$  obtained using a circular cylinder. (See Figures 2.9 and 2.10). Evidently use of the elliptic cylinder, in this case, has reduced the effect of noise by another factor of  $\sim 2.5$  compared to the use of a circular cylinder. Comparison of these figures with Figure 16.1.8 shows that in this study a 1% noise in field data produces about .01% error in  $C_{1,s}^{[0]}(z)$  when

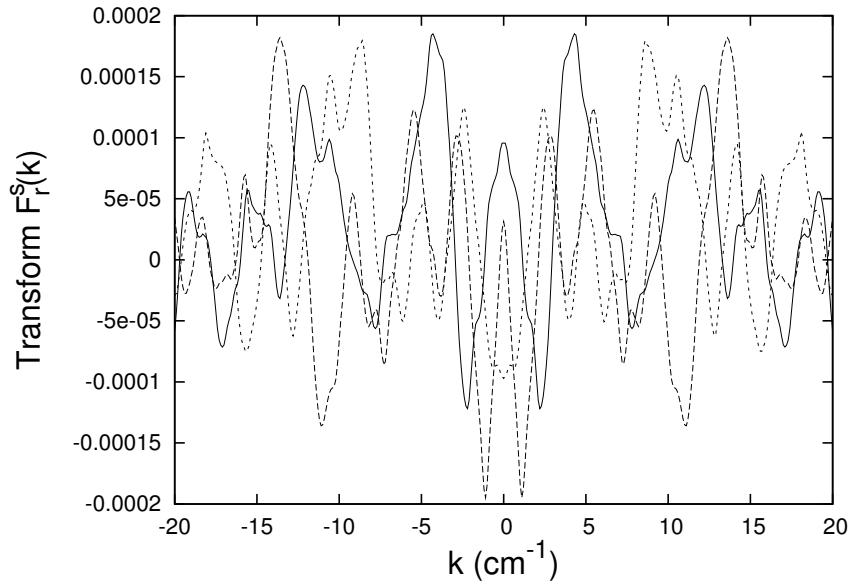


Figure 20.3.8: Real parts of the first few functions  $\tilde{F}_r^s(k)$ , those for  $r = 1, 3$ , and  $5$ , produced by a pure noise field arising from a second different random number seed. The imaginary parts are comparable.

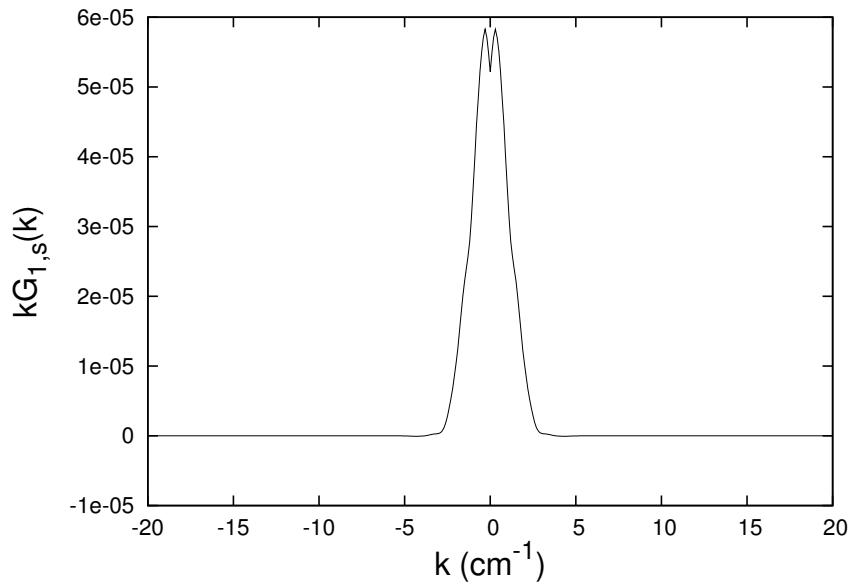


Figure 20.3.9: Real part of  $kG_{1,s}(k)$  computed from  $\tilde{F}_r^s(k)$  associated with the first seed. The imaginary part is comparable.

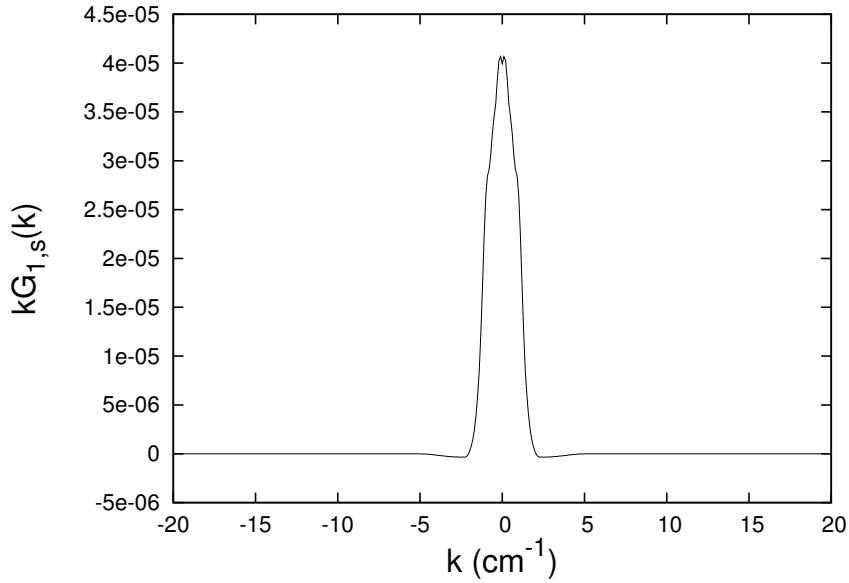


Figure 20.3.10: Real part of  $kG_{1,s}(k)$  computed from  $\tilde{\tilde{F}}_r^s(k)$  associated with the second seed. The imaginary part is comparable.

the elliptic cylinder is used.

What can be said about the other  $C_{m,s}^{[n]}(z)$ ? Generally the use of an elliptic cylinder gives better results. But in some cases the circular and elliptic cylinder results are comparable, and sometimes the circular cylinder error is somewhat smaller. Results vary from seed to seed. For example, Figures 3.15 through 3.18 show the functions  $C_{1,s}^{[6]}(z)$  and  $C_{7,s}^{[0]}(z)$  computed using both elliptic and circular cylinders.

There are at least two remarks to be made. First, just as in the cases in Section 17.2 where the results of using circular cylinders with different radii were compared, statistical fluctuations can mask the effects of improved smoothing. Second, it is primarily the vertical magnetic field that contributes to the  $C_{1,s}^{[n]}(z)$ ,  $C_{3,s}^{[n]}(z)$ ,  $C_{5,s}^{[n]}(z)$  . . . . However, what enters our calculation is the component of the magnetic field that is perpendicular to the surface. Although the elliptical cylinder surface has points that are farther from the axis than the circular cylinder, at these points the normal to the surface is primarily in the horizontal direction. Consequently, the points on the elliptical surface for which the field values actually contribute to the  $C_{m,\alpha}^{[n]}(z)$  thus far examined are not very much farther from the axis than points on the circular surface.

We should also examine, for example, the quantities  $C_{1,c}^{[n]}(z)$ ,  $C_{3,c}^{[n]}(z)$ ,  $C_{5,c}^{[n]}(z)$  . . . for which the horizontal magnetic field makes substantial contributions. For these quantities we expect that noise results for circular and elliptic cylinders should be noticeably different. As a first exploratory step, let us again examine the relevant kernels. Figure 3.19 shows the circular cylinder  $m = 1$  kernel [ $1/I'_1(kR)$ ] and the first few elliptical cylinder  $m = 1$  kernels [ $k\alpha_r^r(k)/Ce_r'(U,q)$ ], and Figure 3.20 shows the same kernels all normalized to 1 at  $k = 0$ . From Figure 3.20 we see that the elliptic kernels for  $r = 1$  and  $r = 3$  fall off more rapidly with  $|k|$  than the circular kernel, but that they fall off less rapidly for  $r \geq 5$ . The  $r$  value for which this transition occurs depends on the eccentricity of the ellipse: the larger the

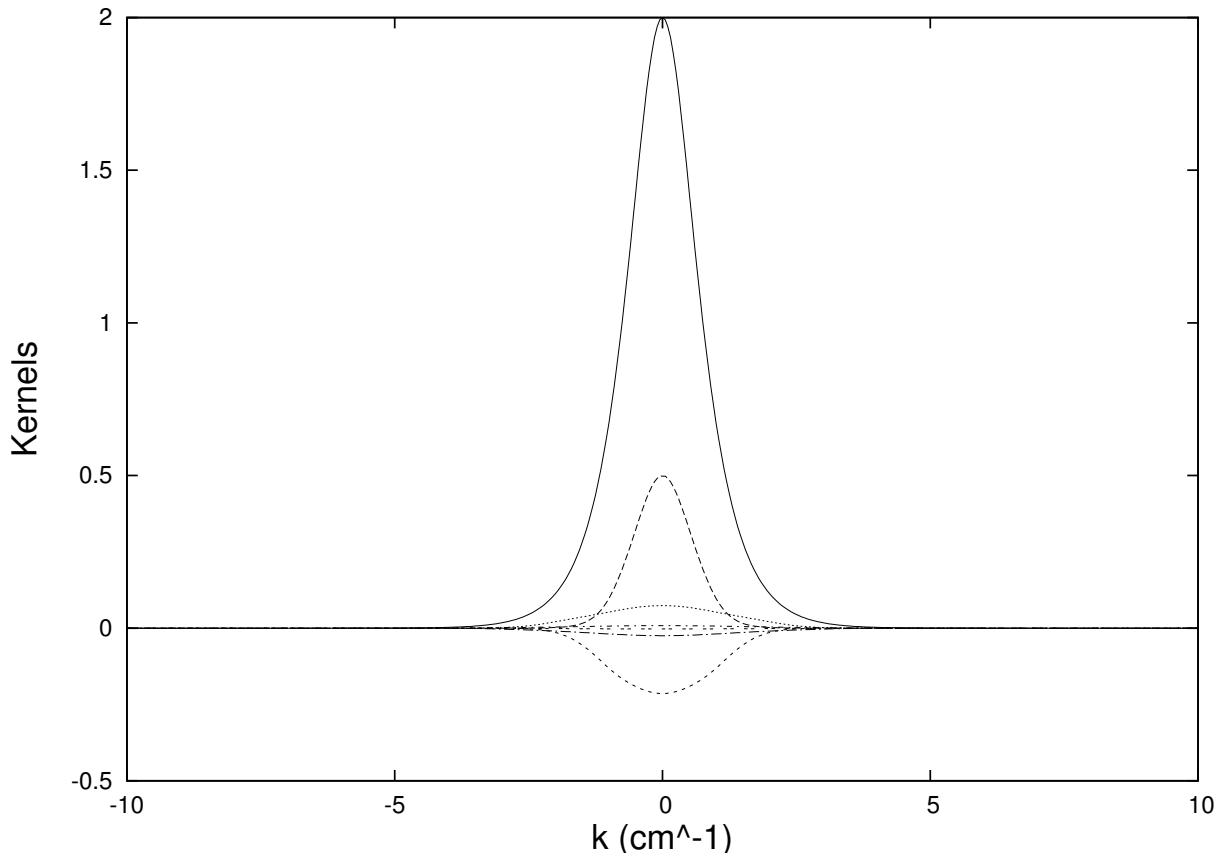


Figure 20.3.11: A comparison of the circular cylinder  $m = 1$  kernel  $[1/I'_1(kR)]$ , shown as a solid line, and the first few relevant elliptical cylinder  $m = 1$  kernels  $[k\beta_1^r(k)/Se'_r(U, q)]$ , namely those for  $r = 1, 3, 5, 7, 9$ , and  $11$ , shown as dashed lines. The elliptic kernels alternate in sign, and their magnitude at  $k = 0$  decreases with increasing  $r$ . See Figure 16.2.7.

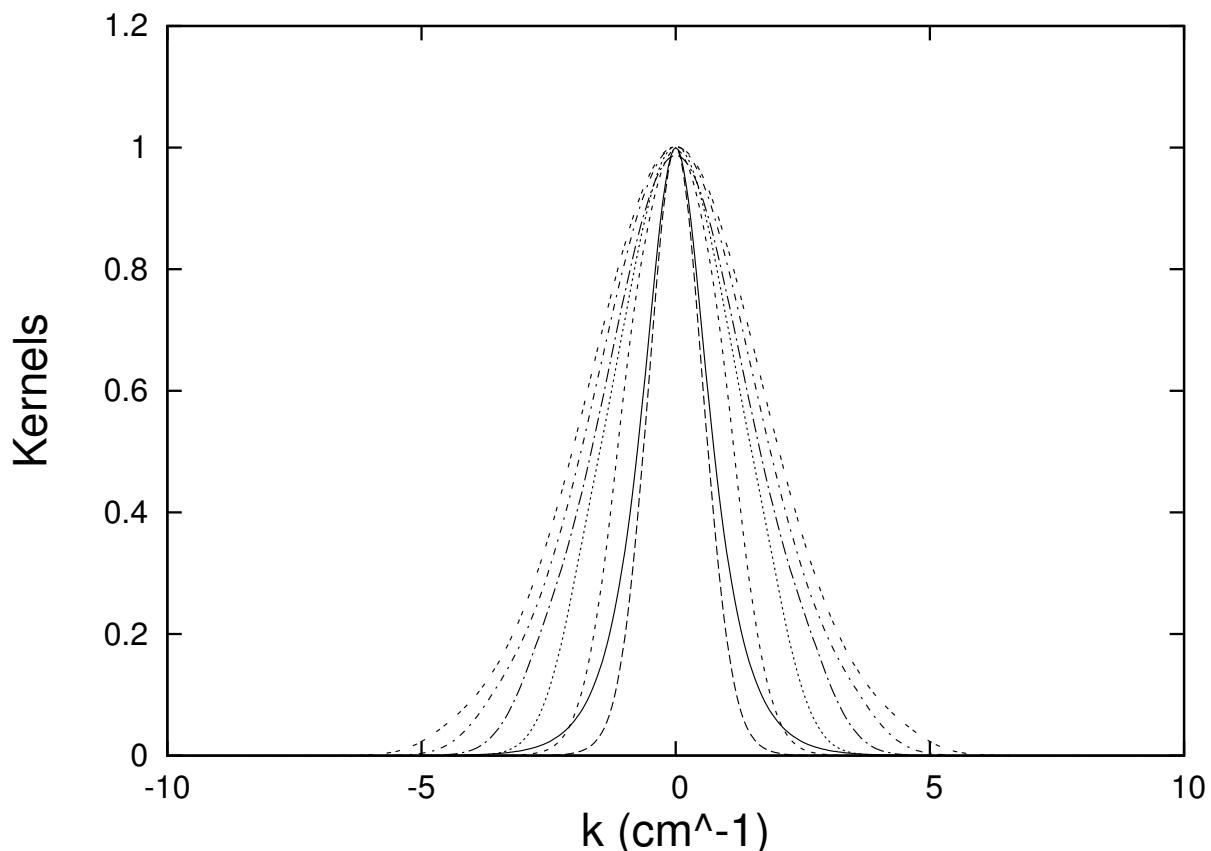


Figure 20.3.12: A comparison of the circular cylinder  $m = 1$  kernel  $[1/I'_1(kR)]$  and the first few elliptic cylinder  $m = 1$  kernels  $[k\beta_1^r(k)/Se'_r(U, q)]$ , all normalized to 1 at  $k = 0$ .

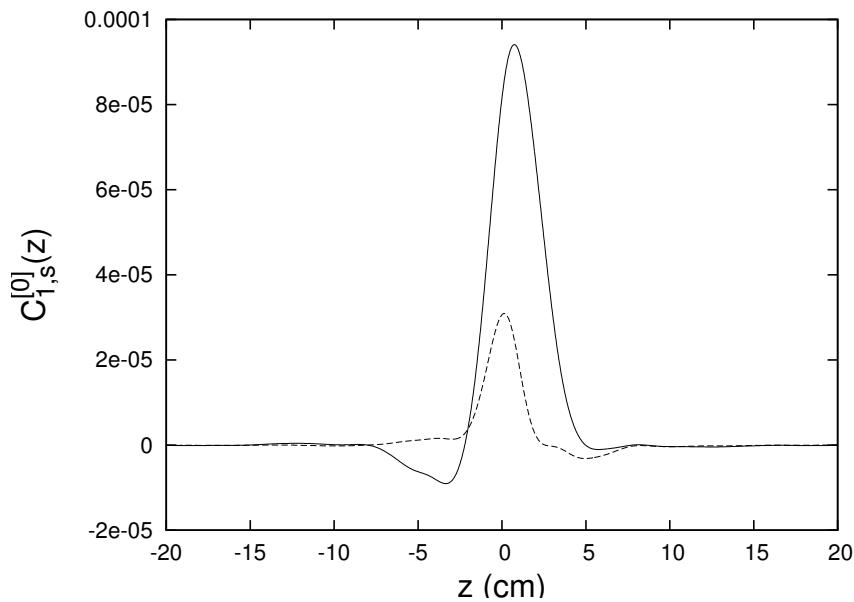


Figure 20.3.13: Dashed line: The function  $C_{1,s}^{[0]}(z)$  produced by a pure noise field and using an elliptic cylinder. Solid line: The function  $C_{1,s}^{[0]}(z)$  produced by a pure noise field and using a circular cylinder.

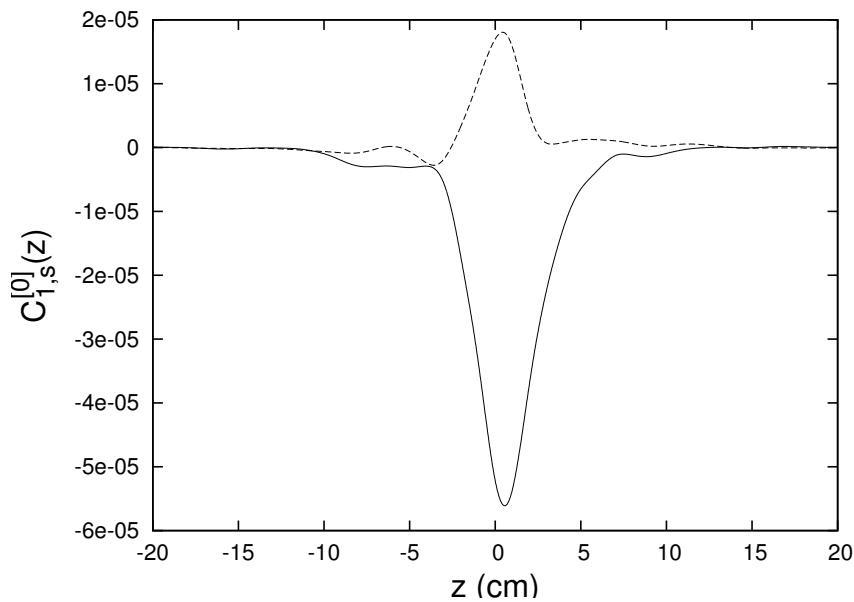


Figure 20.3.14: Results for the second random number seed. Dashed line: The function  $C_{1,s}^{[0]}(z)$  produced by a pure noise field and using an elliptic cylinder. Solid line: The function  $C_{1,s}^{[0]}(z)$  produced by a pure noise field and using a circular cylinder.

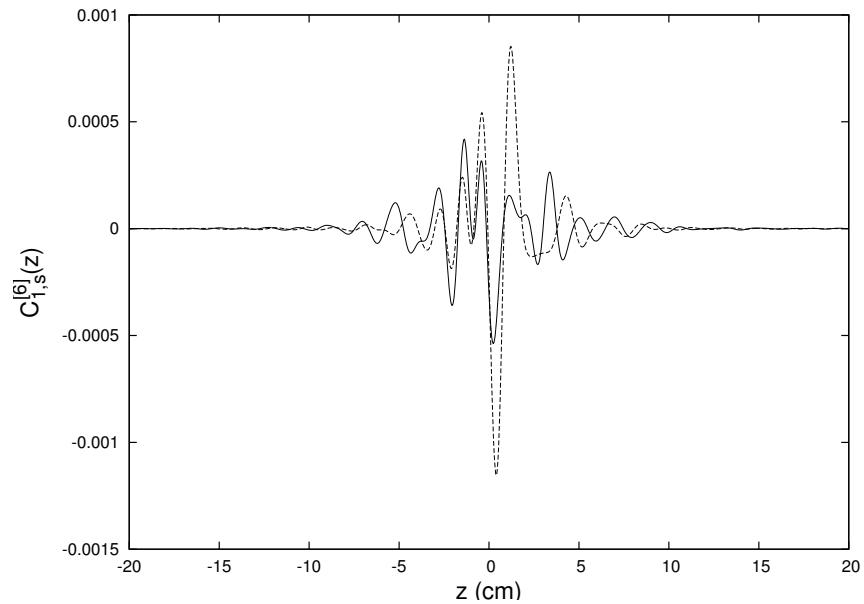


Figure 20.3.15: The function  $C_{1,s}^{[6]}(z)$  produced by a pure noise field. Dashed line: Elliptic cylinder result. Solid line: Circular cylinder result.

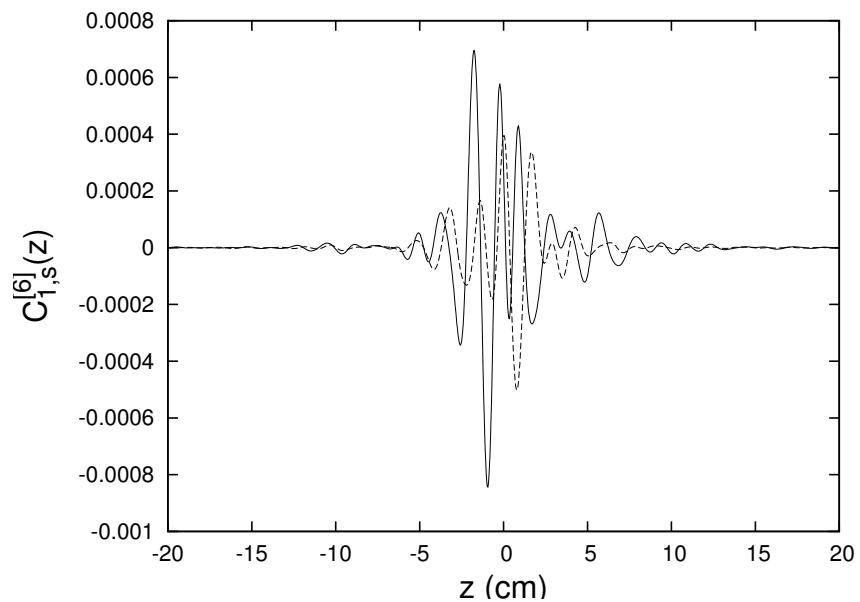


Figure 20.3.16: The function  $C_{1,s}^{[6]}(z)$  produced by a pure noise field arising from a second different random number seed. Dashed line: Elliptic cylinder result. Solid line: Circular cylinder result.

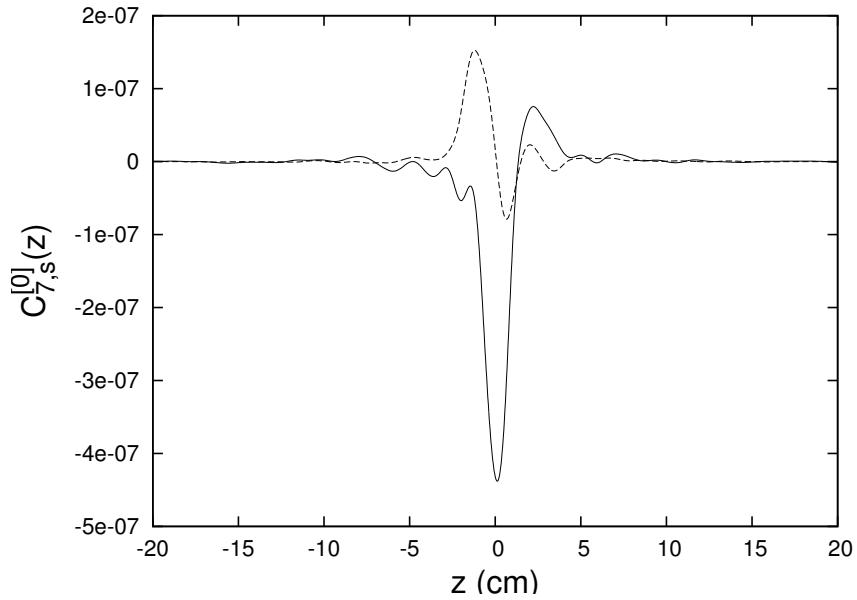


Figure 20.3.17: The function  $C_{7,s}^{[0]}(z)$  produced by a pure noise field. Dashed line: Elliptic cylinder result. Solid line: Circular cylinder result.

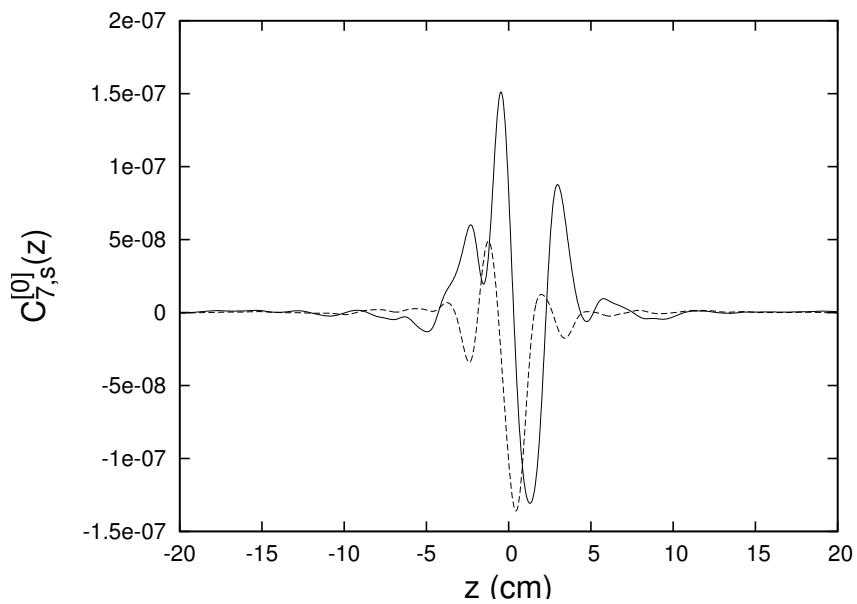


Figure 20.3.18: The function  $C_{7,s}^{[0]}(z)$  produced by a pure noise field arising from a second different random number seed. Dashed line: Elliptic cylinder result. Solid line: Circular cylinder result.

eccentricity the larger the transition  $r$ . Moreover, we see from Figure 3.19 that the elliptic kernels for small  $r$  dominate.

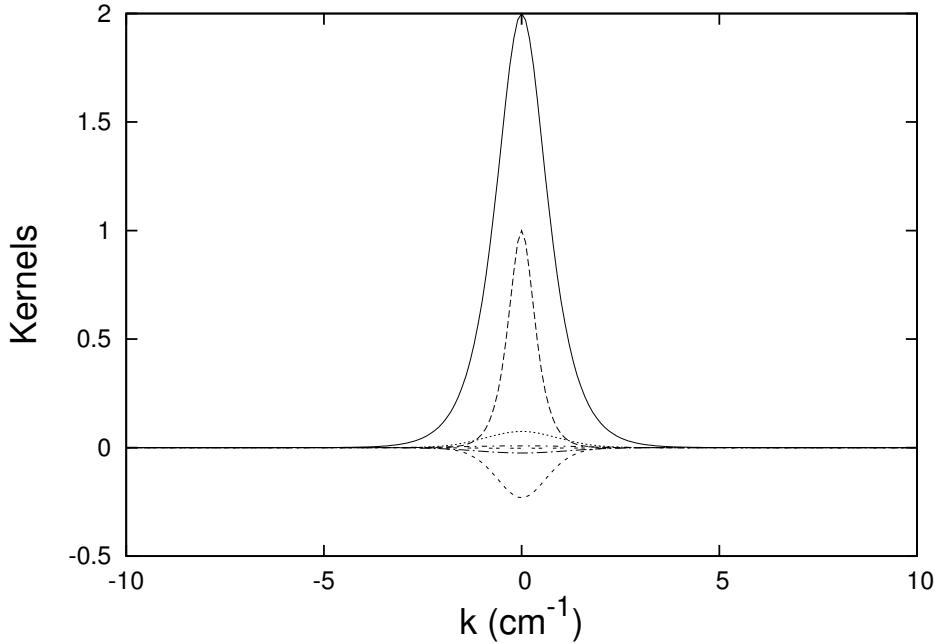


Figure 20.3.19: A comparison of the circular cylinder  $m = 1$  kernel  $[1/I_1'(kR)]$ , shown as a solid line, and the first few relevant elliptical cylinder  $m = 1$  kernels  $[k\alpha_1^r(k)/Ce_r'(U, q)]$ , namely those for  $r = 1, 3, 5, 7, 9$ , and  $11$ , shown as a dashed lines. The elliptic kernels alternate in sign, and their magnitude at  $k = 0$  decreases with increasing  $r$ .

Now we are prepared to examine the  $C_{m,c}^{[n]}(z)$ . Figures 3.21 and 3.22 show  $C_{1,c}^{[0]}(z)$  for each noise realization. And Figures 3.23 through 3.26 show the associated quantities  $C_{1,c}^{[6]}(z)$  and  $C_{7,c}^{[0]}(z)$ . Again there are statistical fluctuations, but the general trend is that the use of an elliptic cylinder yields on-axis gradients that have less sensitivity to errors in the grid data.

To study the problem more thoroughly, we should examine the results for a large number of seeds. We expect that when such results are examined, the effect of noise will average to zero (because the effect of noise can have either sign), but there will be a nonzero variance. What should be verified is that the variance is smaller for elliptic cylinders than for circular cylinders.

Figures 3.27 and 3.28 display circular and elliptical cylinder results for the  $C_{1,c}^{[0]}(z)$  obtained for 12 seeds. They also show  $\langle C_{1,c}^{[0]}(z) \rangle$ , the average of these results. Figures 3.29 through 3.32 do the same for  $C_{1,c}^{[6]}(z)$  and  $C_{7,c}^{[0]}(z)$ . Evidently the averaged results are smaller than the individual results, thereby indicating that the average does indeed approach zero as the number of seeds is increased.

Figures 3.33 through 3.35 show the quantities  $\{\langle [C_{1,c}^{[0]}(z)]^2 \rangle\}^{1/2}$ ,  $\{\langle [C_{1,c}^{[6]}(z)]^2 \rangle\}^{1/2}$ , and  $\{\langle [C_{7,c}^{[0]}(z)]^2 \rangle\}^{1/2}$ , the root-mean-square values based on these 12 seeds. Results are shown for both the circular and elliptic cylinder. We see that, for the two cases of

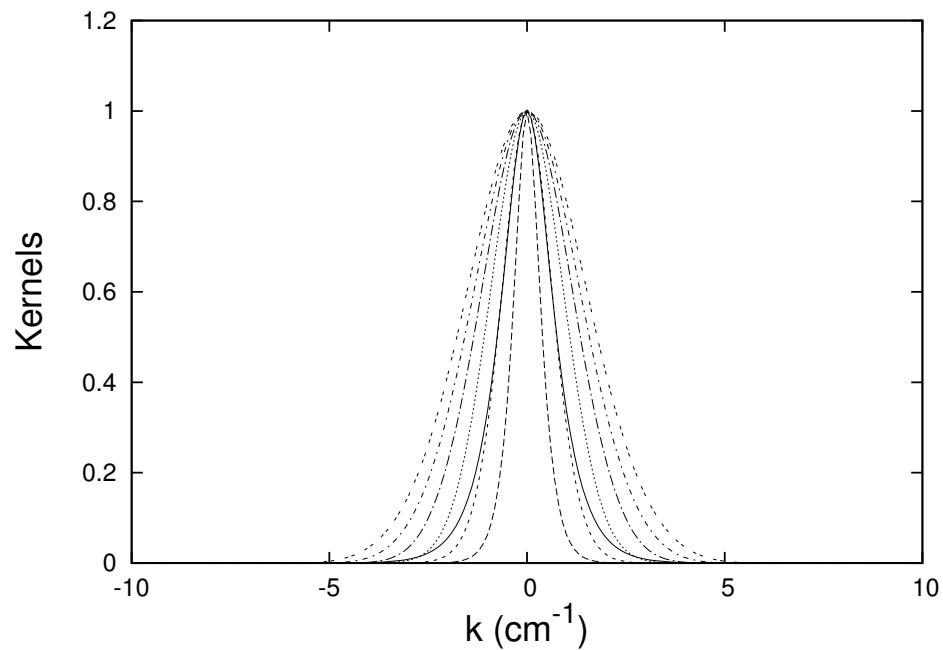


Figure 20.3.20: A comparison of the circular cylinder  $m = 1$  kernel  $[1/I'_1(kR)]$  and the first few elliptic cylinder  $m = 1$  kernels  $[k\alpha_1^r(k)/Ce_r'(U, q)]$ , all normalized to 1 at  $k = 0$ .

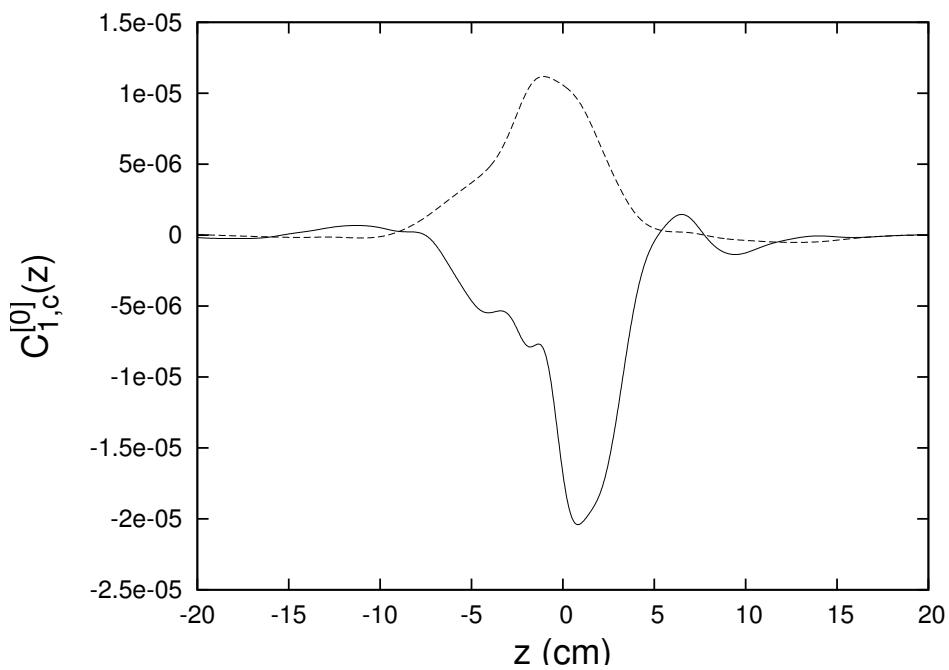


Figure 20.3.21: Dashed line: The function  $C_{1,c}^{[0]}(z)$  produced by a pure noise field and using an elliptic cylinder. Solid line: The function  $C_{1,c}^{[0]}(z)$  produced by a pure noise field and using a circular cylinder.

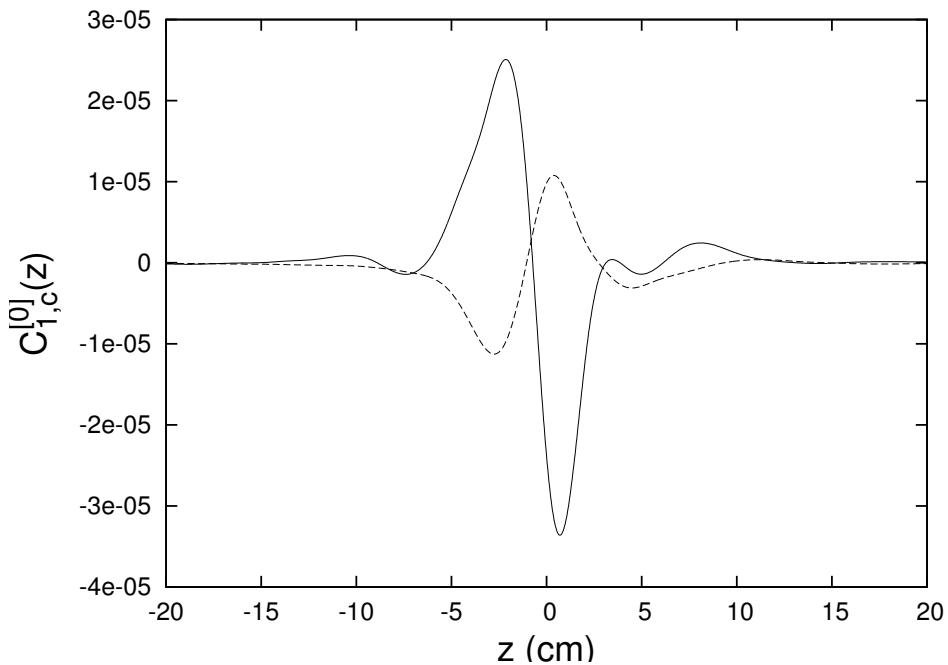


Figure 20.3.22: Results for the second random number seed. Dashed line: The function  $C_{1,c}^{[0]}(z)$  produced by a pure noise field and using an elliptic cylinder. Solid line: The function  $C_{1,c}^{[0]}(z)$  produced by a pure noise field and using a circular cylinder.

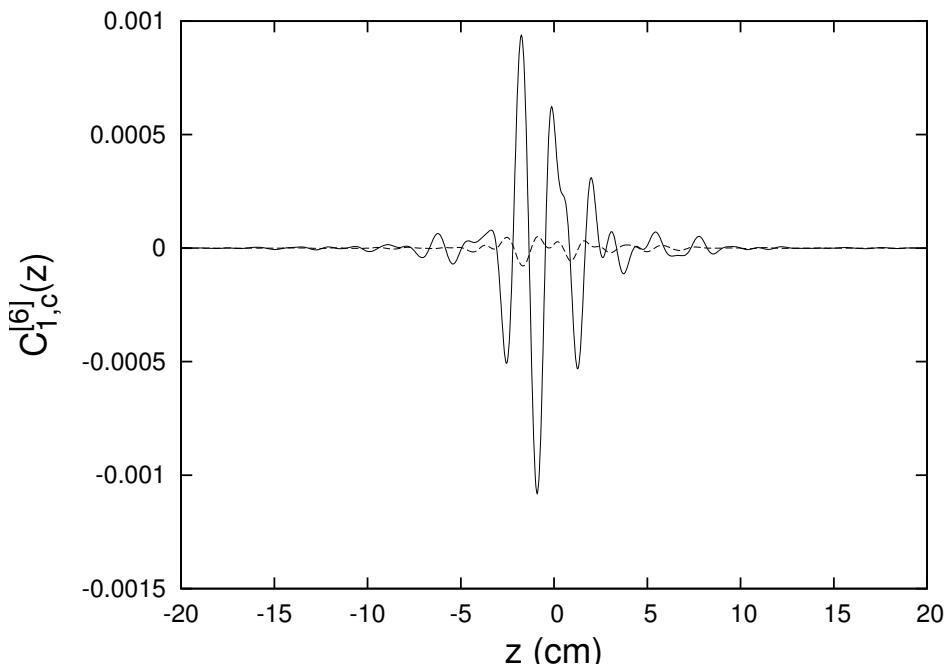


Figure 20.3.23: The function  $C_{1,c}^{[6]}(z)$  produced by a pure noise field. Dashed line: Elliptic cylinder result. Solid line: Circular cylinder result.

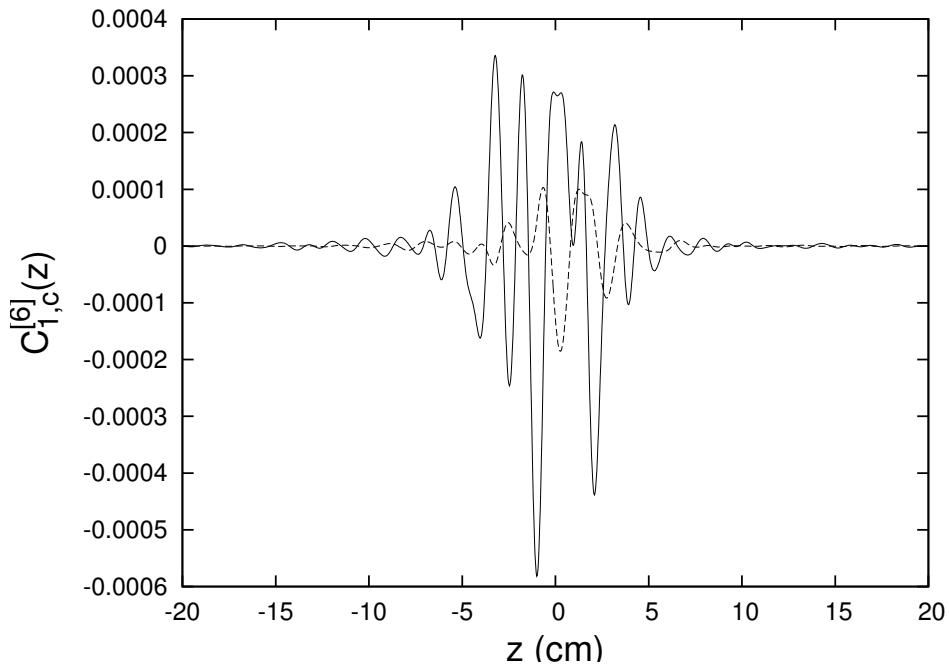


Figure 20.3.24: The function  $C_{1,c}^{[6]}(z)$  produced by a pure noise field arising from a second different random number seed. Dashed line: Elliptic cylinder result. Solid line: Circular cylinder result.

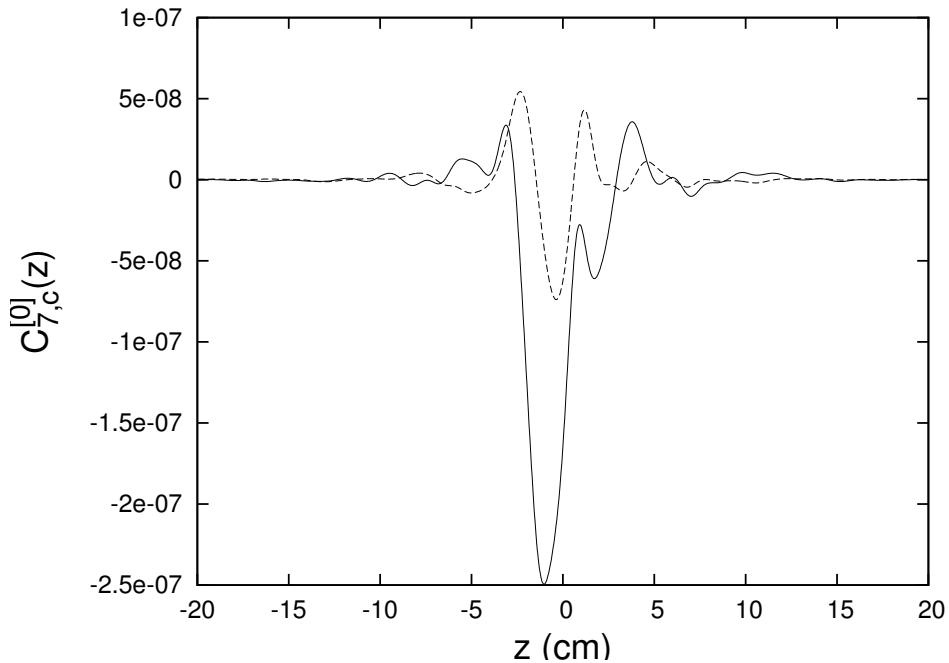


Figure 20.3.25: The function  $C_{7,c}^{[0]}(z)$  produced by a pure noise field. Dashed line: Elliptic cylinder result. Solid line: Circular cylinder result.

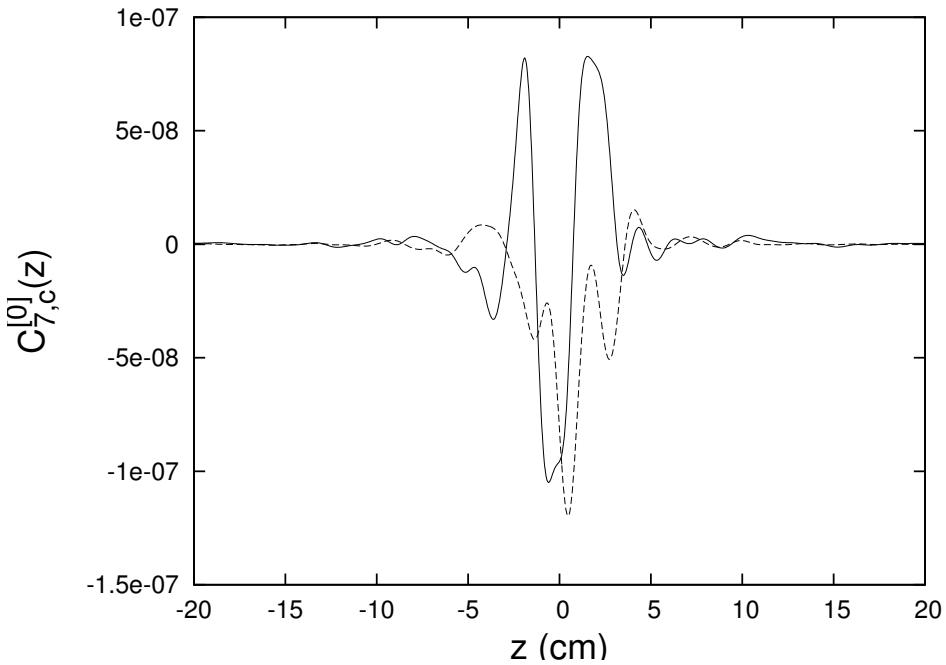


Figure 20.3.26: The function  $C_{7,c}^{[0]}(z)$  produced by a pure noise field arising from a second different random number seed. Dashed line: Elliptic cylinder result. Solid line: Circular cylinder result.

$\{\langle [C_{1,c}^{[6]}(z)]^2 \rangle\}^{1/2}$  and  $\{\langle [C_{7,c}^{[0]}(z)]^2 \rangle\}^{1/2}$ , the root-mean-square values are indeed smaller for the elliptic cylinder compared to the circular cylinder.

The case for  $\{\langle [C_{1,c}^{[0]}(z)]^2 \rangle\}^{1/2}$  is inconclusive. It appears that the number of samples is still too small so that statistical fluctuations still overwhelm the expected effect. This hypothesis is validated by Figure 3.36. It shows the functions  $C_{1,c}^{[0]}(z)$  produced by assigning a nonzero field value to only a *single* grid point. Consider the field value

$$(B_x, B_y, B_z) = (.01 \text{ T}, 0, 0). \quad (20.3.1)$$

For the elliptic cylinder case we assign this field value to the grid point

$$(x, y, z) = (4 \text{ cm}, 0, 0). \quad (20.3.2)$$

And for the circular cylinder case we assign this field value to the grid point

$$(x, y, z) = (2 \text{ cm}, 0, 0). \quad (20.3.3)$$

All other grid points are assigned vanishing field values. As the figure shows, the elliptic cylinder result for  $C_{1,c}^{[0]}(z)$  is indeed smaller than the circular cylinder result.

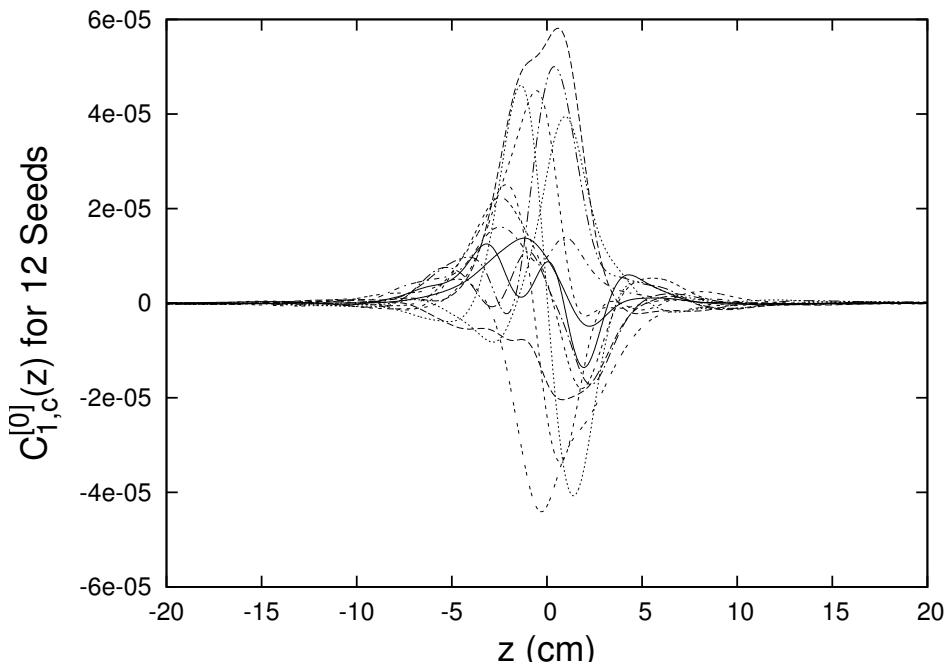


Figure 20.3.27: The functions  $C_{1,c}^{[0]}(z)$  produced by pure noise fields generated by 12 seeds using data on a circular cylinder. Broken lines: Results from individual seeds. Solid line: Averaged results.

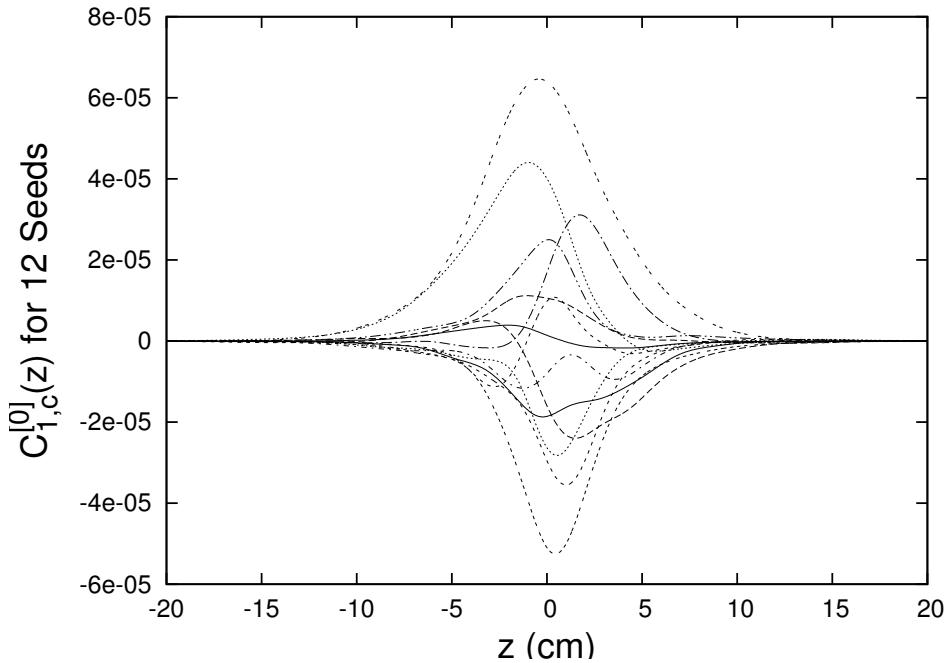


Figure 20.3.28: The functions  $C_{1,c}^{[0]}(z)$  produced by pure noise fields generated by 12 seeds using data on an elliptical cylinder. Broken lines: Results from individual seeds. Solid line: Averaged results.

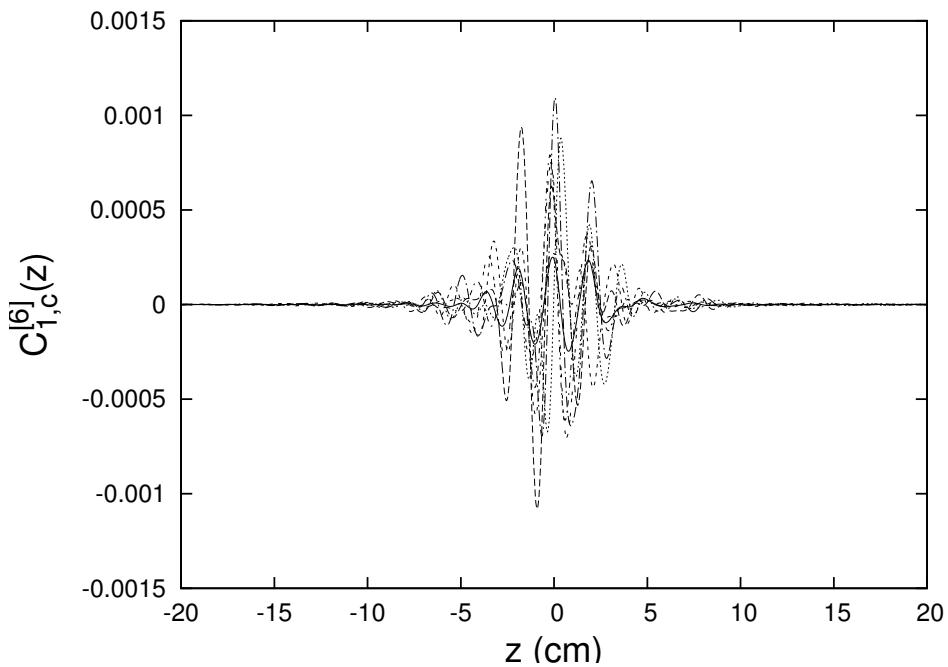


Figure 20.3.29: The functions  $C_{1,c}^{[6]}(z)$  produced by pure noise fields generated by 12 seeds using data on a circular cylinder. Broken lines: Results from individual seeds. Solid line: Averaged results.

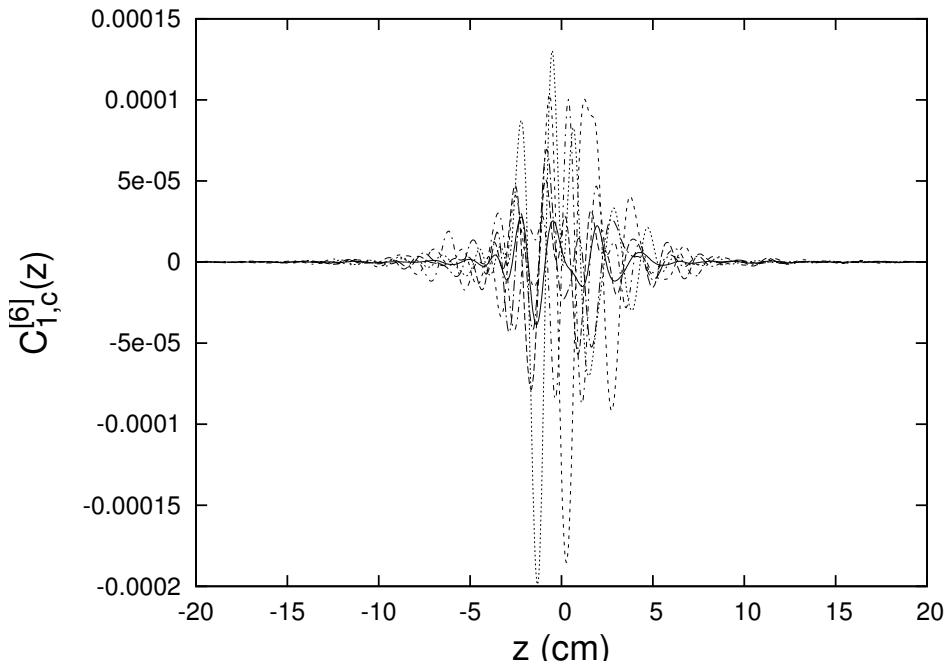


Figure 20.3.30: The functions  $C_{1,c}^{[6]}(z)$  produced by pure noise fields generated by 6 seeds using data on an elliptical cylinder. Broken lines: Results from individual seeds. For clarity, in this graphic only results for 6 seeds are shown. Solid line: Averaged results. As in other related figures, results for 12 seeds were used in computing the average.

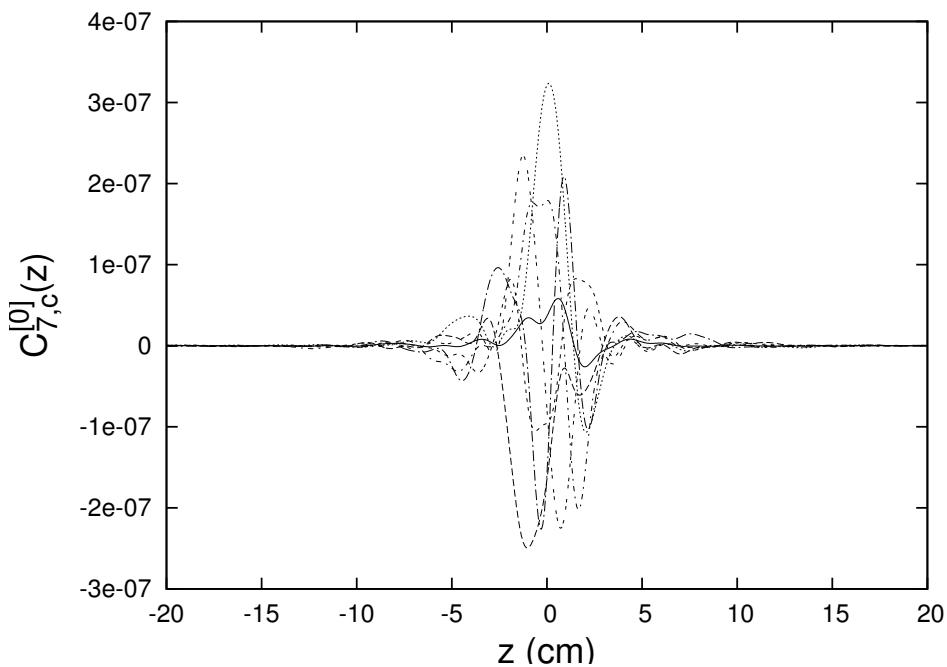


Figure 20.3.31: The functions  $C_{7,c}^{[0]}(z)$  produced by pure noise fields generated by 12 seeds using data on a circular cylinder. Broken lines: Results from individual seeds. Solid line: Averaged results.

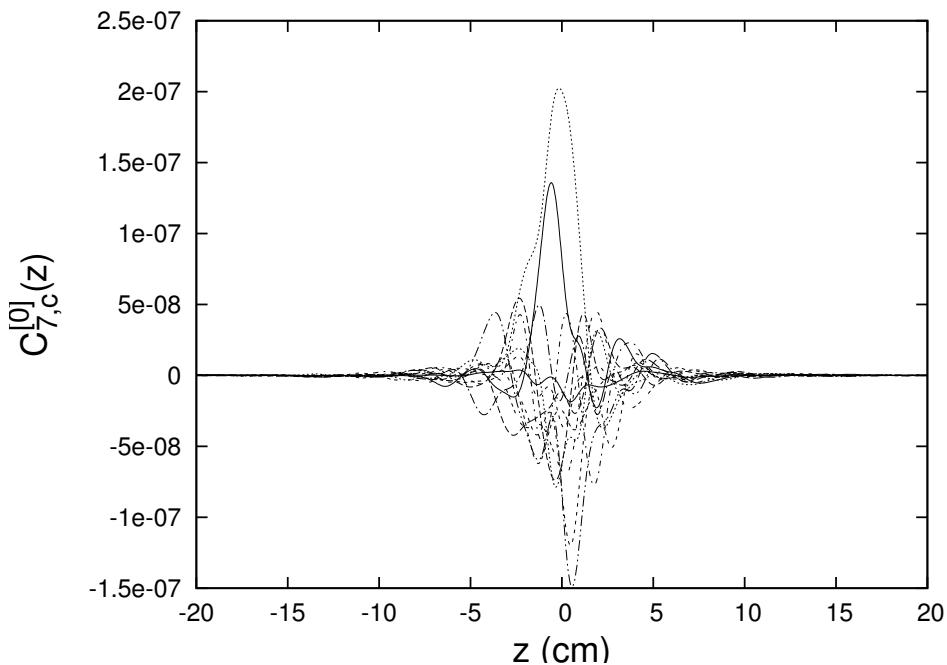


Figure 20.3.32: The functions  $C_{7,c}^{[0]}(z)$  produced by pure noise fields generated by 12 seeds using data on an elliptical cylinder. Broken lines: Results from individual seeds. Solid line: Averaged results.

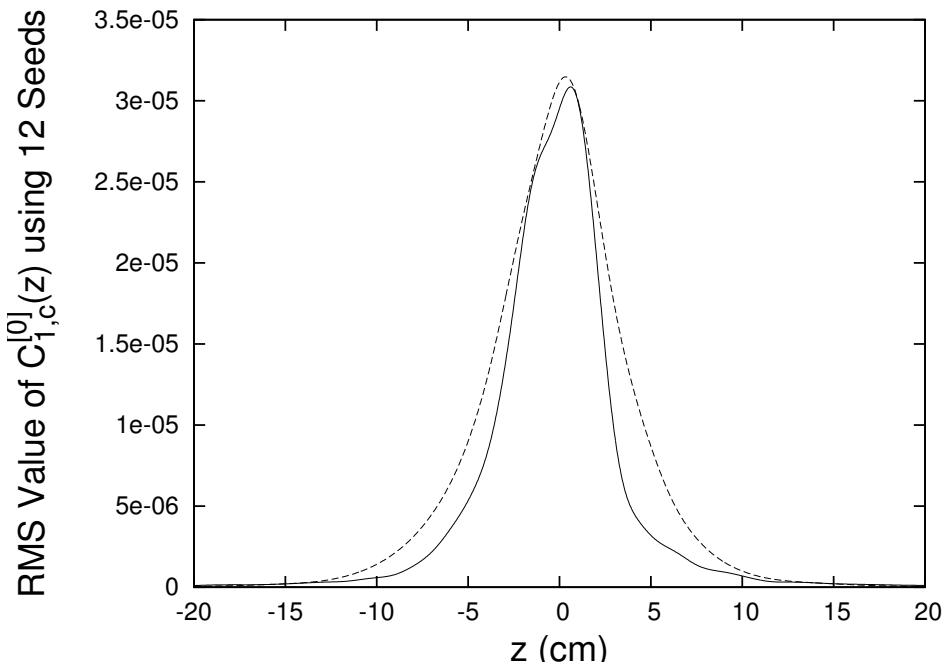


Figure 20.3.33: The function  $\{\langle [C_{1,c}^{[0]}(z)]^2 \rangle\}^{1/2}$  produced by 12 pure noise fields. Dashed line: Result from using an elliptic cylinder. Solid line: Result from using a circular cylinder.

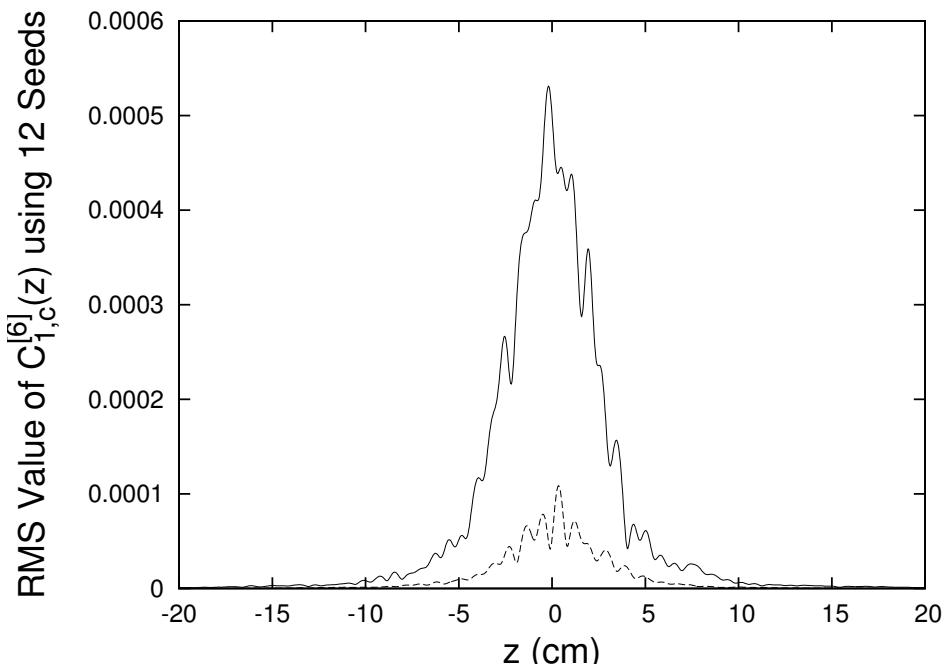


Figure 20.3.34: The function  $\{\langle [C_{1,c}^{[6]}(z)]^2 \rangle\}^{1/2}$  produced by 12 pure noise fields. Dashed line: Result from using an elliptic cylinder. Solid line: Result from using a circular cylinder.

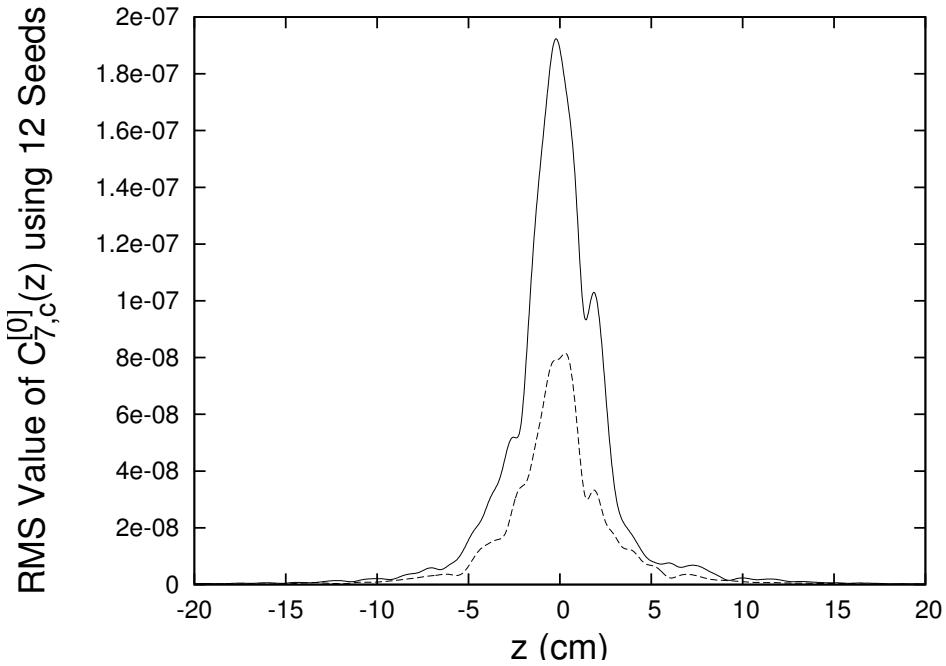


Figure 20.3.35: The function  $\{\langle [C_{7,c}^{[0]}(z)]^2 \rangle\}^{1/2}$  produced by 12 pure noise fields. Dashed line: Result from using an elliptic cylinder. Solid line: Result from using a circular cylinder.

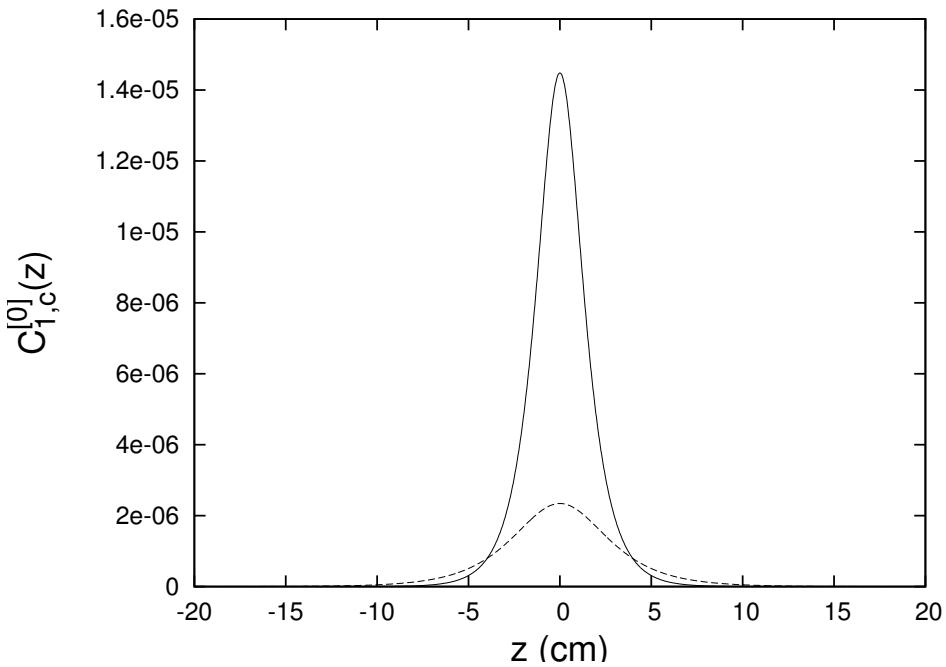


Figure 20.3.36: The functions  $C_{1,c}^{[0]}(z)$  produced by assigning a non-zero field value to only a single grid point. Dashed line: Result from using an elliptic cylinder. Solid line: Result from using a circular cylinder.

## 20.4 Rectangular Cylinders



# Bibliography

## General References

- [1] M. Venturini and A. Dragt, “Accurate Computation of Transfer Maps from Magnetic Field Data”, *Nuclear Instruments and Methods* **A427** (1991), p. 387.
- [2] M. Venturini, Transfer Map for Printed-Circuit Magnetic Quadrupoles, Technical Note, Dept. of Physics, Univ. of Maryland (1995).
- [3] C. Mitchell, “Calculation of Realistic Charged-Particle Transfer Maps”, University of Maryland Physics Department Ph.D. Thesis (2007).
- [4] C. Mitchell and A. Dragt, “Accurate transfer maps for realistic beam-line elements: Part I, straight elements”, *Physical Review Special Topics - Accelerators and Beams* **13** 064001 (2010).



# Chapter 21

## Realistic Transfer Maps for General Straight Beam-Line Elements

Chapter 15 described cylindrical harmonic expansions for straight elements. This chapter utilizes these expansions and applies surface methods to several common magnetic beam-line elements.<sup>1</sup> It also summarizes various cases in which fields can be computed analytically. Particular attention is devoted to the way in which fringe fields fall off with increasing distance.

### 21.1 Solenoids

#### 21.1.1 Preliminaries

A solenoid is a beam-line element whose field is described by a cylindrical harmonic expansion that contains (ideally) only an  $m = 0$  term. We recall from Section 15.2.3 that in this case the magnetic scalar potential  $\psi$  has the expansion

$$\psi(x, y, z) = \psi_0(x, y, z) = C_0^{[0]}(z) - (1/4)(x^2 + y^2)C_0^{[2]}(z) + \dots \quad (21.1.1)$$

See (15.2.57). Correspondingly, the associated magnetic field has the expansion

$$B_x = \partial_x \psi_0 = -(1/2)x C_0^{[2]}(z) + \dots, \quad (21.1.2)$$

$$B_y = \partial_y \psi_0 = -(1/2)y C_0^{[2]}(z) + \dots, \quad (21.1.3)$$

$$B_z = \partial_z \psi_0 = C_0^{[1]}(z) - (1/4)(x^2 + y^2)C_0^{[3]}(z) + \dots. \quad (21.1.4)$$

Finally, according to Section (15.4.1), there is a suitable associated vector potential  $\hat{\mathbf{A}}^0$  given by the relation

$$\hat{A}_x^0 = -yU, \quad (21.1.5)$$

$$\hat{A}_y^0 = xU, \quad (21.1.6)$$

$$\hat{A}_z^0 = 0, \quad (21.1.7)$$

---

<sup>1</sup>Electrostatic beam-line elements can be treated in an analogous way.

where  $U$  is defined to be

$$U(\rho, z) = (1/2) \sum_{\ell=0}^{\infty} (-1)^{\ell} \frac{1}{2^{2\ell} \ell! (\ell + 1)!} C_0^{[2\ell+1]}(z) \rho^{2\ell} \quad (21.1.8)$$

with

$$\rho^2 = x^2 + y^2. \quad (21.1.9)$$

From (1.2) through (1.9) we see that both  $\mathbf{B}$  and  $\hat{\mathbf{A}}^0$  are completely specified in terms of a single “master” function  $C_0^{[1]}(z)$  and its derivatives. Moreover, according to (1.4), the on-axis field is given by the relation

$$B_z(0, 0, z) = C_0^{[1]}(z). \quad (21.1.10)$$

We will next see what can be said about the master function  $C_0^{[1]}(z)$  in various cases.

### 21.1.2 Qualitatively Correct Iron-Dominated Solenoid Model

We next consider the case of an iron-dominated solenoid. Suppose an approximating signum function is defined by the rule

$$\text{sgn}(z, a) = \tanh(z/a) \quad (21.1.11)$$

instead of the rule (1.21), and this approximating signum function is employed to define a soft-edge bump function with the use of (1.20). Then the relation (1.14) continues to hold, but with the soft-edge bump function defined in terms of (1.29) and (1.20).

In this case it can be verified that (1.15) through (1.18) remain true. See Exercise 1.2. However (1.26) and (1.27) are replaced by the asymptotic relations

$$\text{sgn}(z, a) = 1 - 2 \exp(-2z/a) + O[\exp(-4z/a)] \text{ as } z \rightarrow \infty, \quad (21.1.12)$$

$$\text{bump}(z, a, L) = [\exp(2L/a) - 1] \exp(-2|z|/a) + O[\exp(-4|z|/a)] \text{ as } |z| \rightarrow \infty. \quad (21.1.13)$$

Consequently in this model (1.28) is replaced by the *exponential* fall off relation

$$C_0^{[1]}(z) = B(0, 0, z) = B[\exp(2L/a) - 1] \exp(-2|z|/a) + O[\exp(-4|z|/a)] \text{ as } |z| \rightarrow \infty. \quad (21.1.14)$$

The characteristic fall-off length is again governed by  $a$ .

This model does not correspond to any easily-described iron and current distribution. But it does have the property of exponential fall off, which is characteristic of iron-dominated solenoids when the coils are buried deep within the iron. See the next section. It therefore may be of some use in preliminary modeling studies of transfer maps for iron-dominated solenoids.

## Exercises

**21.1.1.** Verify that  $B_z(0, 0, z)$  as given by (1.11) describes the on-axis field of a simple air-core solenoid.

**21.1.2.** The purpose of this exercise is to verify the relations (1.15) through (1.18). Show that the approximating signum function (1.21) has the properties

$$\operatorname{sgn}(-z, a) = -\operatorname{sgn}(z, a), \quad (21.1.15)$$

$$\lim_{z \rightarrow \pm\infty} \operatorname{sgn}(z, a) = \pm 1. \quad (21.1.16)$$

Sketch  $\operatorname{sgn}(z, a)$ ,  $-\operatorname{sgn}(z - L, a)$ , and  $\operatorname{bump}(z, a, L)$  as given by (1.20) to verify the relations (1.15) through (1.17).

What remains is to prove the relation (1.18). Begin by writing

$$\int_{-\infty}^{\infty} \operatorname{bump}(z, a, L) dz = \lim_{w \rightarrow \infty} \int_{-w}^w \operatorname{bump}(z, a, L) dz. \quad (21.1.17)$$

Next verify from the representation (1.20) that

$$\int_{-w}^w \operatorname{bump}(z, a, L) dz = (1/2) \int_{-w}^w \operatorname{sgn}(z, a) dz - (1/2) \int_{-w}^w \operatorname{sgn}(z - L, a) dz. \quad (21.1.18)$$

Show that the first integral on the right side of (1.36) vanishes because of (1.33). Show that by making the change of variables  $x = z - L$  the second integral on the right side of (1.36) becomes

$$\begin{aligned} -(1/2) \int_{-w}^w \operatorname{sgn}(z - L, a) dz &= -(1/2) \int_{-w-L}^{w-L} \operatorname{sgn}(x, a) dx \\ &= -(1/2) \int_{-w-L}^{w+L} \operatorname{sgn}(x, a) dx + (1/2) \int_{w-L}^{w+L} \operatorname{sgn}(x, a) dx. \end{aligned} \quad (21.1.19)$$

Verify that the first integral in the second line of (1.37) vanishes, again because of (1.33). It follows that there is the result

$$\int_{-w}^w \operatorname{bump}(z, a, L) dz = (1/2) \int_{w-L}^{w+L} \operatorname{sgn}(x, a) dx. \quad (21.1.20)$$

Show from (1.34) that there is the result

$$\lim_{w \rightarrow \infty} (1/2) \int_{w-L}^{w+L} \operatorname{sgn}(x, a) dx = (1/2)2L = L. \quad (21.1.21)$$

Put all your intermediate results together to obtain the final result

$$\int_{-\infty}^{\infty} \operatorname{bump}(z, a, L) dz = L, \quad (21.1.22)$$

as desired. Note that the proof of this result has depended only on the representation (1.20) and properties (1.33) and (1.34), which are required properties of any approximating signum function.

**21.1.3.** Verify the limiting behavior (1.22). Verify the fall-off relations (1.26) through (1.28).

**21.1.4.** Verify for a long simple air-core solenoid that the on-axis field at either end ( $z = 0$  or  $z = L$ ) is  $B/2$ . Verify that the same is true for a long solenoid described by the tanh model (1.29), and for any bump function model that is constructed from approximating signum functions.

**21.1.5.** Verify the fall-off relations (1.30) through (1.32).

### 21.1.3 Improved Model for Iron-Dominated Solenoid

Chapter 16 described the computation of transfer maps for straight magnetic beam-line elements based on the *normal* component of the field on the surface of a cylinder. In this subsection we will use instead the *tangential*  $B_z$  component of the field on the surface of a cylinder. For simplicity, we will use the surface of a circular cylinder. This approach of employing the  $B_z$  component is particularly useful in the case of a solenoid.

From (15.2.13) we find the result

$$\begin{aligned} B_z(x, y, z) &= B_z(\rho, \phi, z) = \partial\psi(x, y, z)/\partial z \\ &= \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} dk G_m(k)(ik) \exp(ikz) \exp(im\phi) I_m(k\rho). \end{aligned} \quad (21.1.23)$$

Now integrate both sides of (2.1) over  $\phi \in [0, 2\pi]$  to find the relation

$$\tilde{B}_z(\rho, 0, z) = \int_{-\infty}^{\infty} dk G_0(k)(ik) \exp(ikz) I_0(k\rho) \quad (21.1.24)$$

where

$$\tilde{B}_z(\rho, 0, z) = [1/(2\pi)] \int_0^{2\pi} d\phi B_z(\rho, \phi, z). \quad (21.1.25)$$

Next, from the uniqueness of the Fourier transform, it follows that

$$G_0(k)(ik) I_0(kR) = \tilde{B}_z(R, 0, k) \quad (21.1.26)$$

where

$$\tilde{B}_z(R, 0, k) = [1/(2\pi)] \int_{-\infty}^{\infty} dz \exp(-ikz) \tilde{B}_z(R, 0, z). \quad (21.1.27)$$

Inserting (2.4) into (2.2) gives the result

$$\tilde{B}_z(\rho, 0, z) = \int_{-\infty}^{\infty} dk \exp(ikz) \tilde{B}_z(R, 0, k) I_0(k\rho) / I_0(kR). \quad (21.1.28)$$

Finally, use of (15.2.71) or (15.4.36) gives the relation

$$C_0^{[1]}(z) = \int_{-\infty}^{\infty} dk \exp(ikz) \tilde{B}_z(R, 0, k) / I_0(kR) \quad (21.1.29)$$

from which it follows that

$$C_0^{[n]}(z) = \int_{-\infty}^{\infty} dk \exp(ikz) \tilde{B}_z(R, 0, k) [(ik)^{n-1} / I_0(kR)]. \quad (21.1.30)$$

We have found expressions for the  $m = 0$  on-axis gradients in terms of the  $B_z$  component of the magnetic field on the surface of a circular cylinder. If desired, we could also find expressions for the  $m \neq 0$  on-axis gradients in terms of  $B_z$  on the surface.<sup>2</sup>

As a simple application, the representation (2.7) can be used to find an *approximation* to the on-axis gradients (and hence the magnetic field) in the case that the field is produced by an iron-dominated magnetic solenoid of bore radius  $R$  with a small pole gap of length  $L$  centered at  $z = 0$ . Solenoids for use in electron microscopes are often of this type. See Figure 1.1. In this case we may make the approximation

$$\begin{aligned} B_z(R, \phi, z) &= B_{\text{gap}} \text{ for } z \in (-L/2, L/2), \\ B_z(R, \phi, z) &= 0 \text{ elsewhere.} \end{aligned} \quad (21.1.31)$$

That is, the tangential surface field exists only in the gap, and is constant there both in  $\phi$  and in  $z$ .<sup>3</sup> With the assumption (2.9), the relations (2.3) and (2.5) become

$$\tilde{B}_z(R, 0, k) = [B_{\text{gap}}/(2\pi)] \int_{-L/2}^{L/2} dz \exp(-ikz) = [B_{\text{gap}}/(\pi k)] \sin(kL/2). \quad (21.1.32)$$

Correspondingly, we find for the on-axis gradient the result

$$C_0^{[1]}(z) = [B_{\text{gap}}/(\pi)] \int_{-\infty}^{\infty} dk \exp(ikz) \sin(kL/2) / [kI_0(kR)]. \quad (21.1.33)$$

Examination of the integral representation (2.11) for  $C_0^{[1]}(z)$  reveals that it depends on the dimensionless quantities  $z/R$  and  $L/R$ . Indeed, upon making the change of integration variable given by

$$\lambda = kR, \quad (21.1.34)$$

(2.11) takes the form

$$C_0^{[1]}(z) = B_{\text{gap}}(L/R) F(z/R, L/R) \quad (21.1.35)$$

where  $F$  is a *profile function* given by

$$\begin{aligned} F(z/R, L/R) &= (1/\pi)(R/L) \int_{-\infty}^{\infty} d\lambda \exp(i\lambda z/R) \sin[(\lambda/2)(L/R)] / [\lambda I_0(\lambda)] \\ &= [1/(2\pi)] \int_{-\infty}^{\infty} d\lambda \exp(i\lambda z/R) \frac{\sin[(\lambda/2)(L/R)]}{[(\lambda/2)(L/R)]} [1/I_0(\lambda)] \\ &= [1/(2\pi)] \int_{-\infty}^{\infty} d\lambda \exp(i\lambda z/R) \{\text{sinc}[(\lambda/2)(L/R)]\} [1/I_0(\lambda)]. \end{aligned} \quad (21.1.36)$$

---

<sup>2</sup>There might also be occasions in which one might want to use the azimuthal component  $B_\phi$  on the surface.

<sup>3</sup>That the tangential magnetic field should be zero outside the gap follows from the idealized boundary condition for the interface between vacuum and a medium with infinite magnetic permeability. That the field should be constant in the gap is a further idealization.

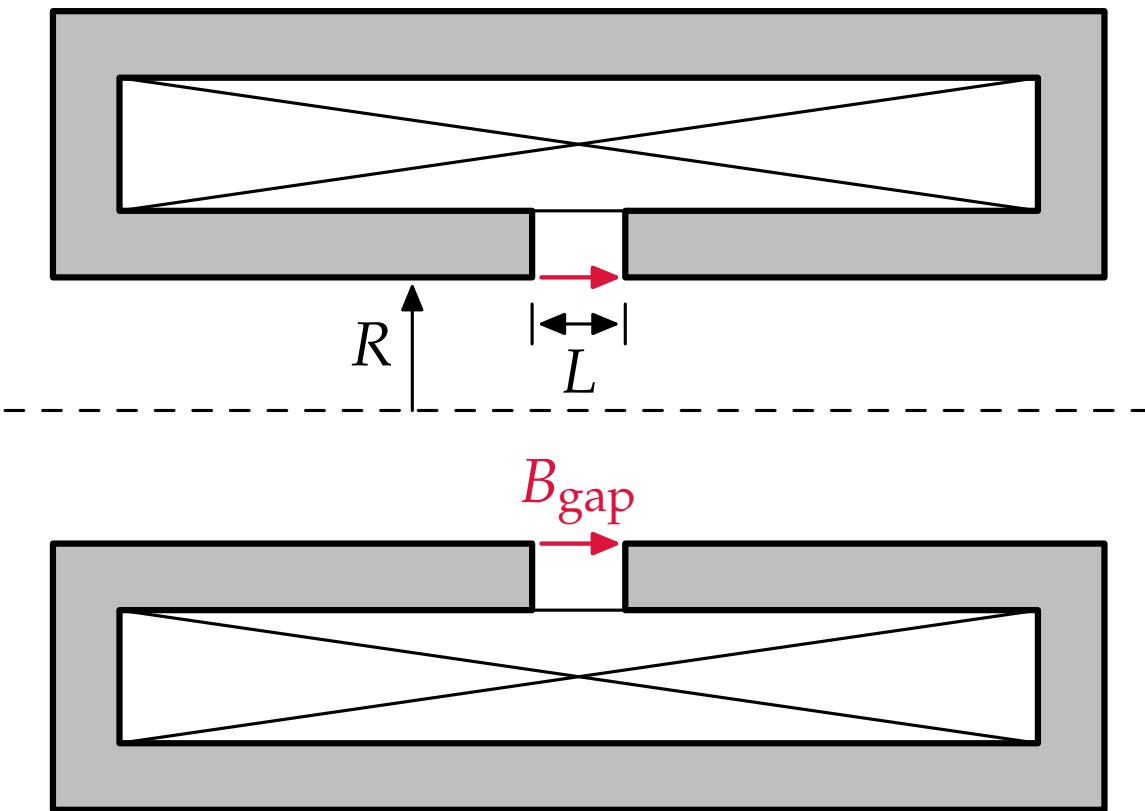


Figure 21.1.1: Schematic of an iron-dominated solenoid with an inter-pole gap  $L$  substantially smaller than the bore radius  $R$ .

Suppose that  $L$  is small compared to  $R$  so that  $L/R$  is small. Then, when the argument of the sine or sinc function in (2.14) differs significantly from zero,  $[I_0(\lambda)]^{-1}$  will be essentially zero. Therefore, we can expand the sine or sinc function in a Taylor series and integrate term by term. We also note that the surface field model (2.9) is only reasonable in the limit of small  $L/R$ . Upon making the Taylor expansion just described, one finds that  $F(z/R, L/R)$  differs from  $F(z/R, 0)$  only by terms of order  $(L/R)^2$  and

$$F(z/R, 0) = [1/(2\pi)] \int_{-\infty}^{\infty} d\lambda \exp(i\lambda z/R) / I_0(\lambda). \quad (21.1.37)$$

Therefore, for small  $L/R$ , it is useful to make the approximation

$$C_0^{[1]}(z) \simeq B_{\text{gap}}(L/R) F(z/R, 0). \quad (21.1.38)$$

For example, Figure 1.2 displays  $F(z/R, L/R)$  as a function of  $z/R$  for the two values  $L/R = 0$  and  $L/R = 1/2$ . Evidently the two profiles nearly agree when  $z = 0$ , and are essentially identical away from  $z = 0$ .

Let us make the further and more drastic approximation of replacing  $I_0(\lambda)$ , the denominator in (2.15), by  $\cosh(\lambda)$ . In this approximation the integral (2.11) can be evaluated analytically to give the result

$$F(z/R, 0) \approx G(z/R) = [1/(2\pi)] \int_{-\infty}^{\infty} d\lambda \exp(i\lambda z/R) / \cosh(\lambda) = 1/\{2 \cosh[\pi z/(2R)]\}. \quad (21.1.39)$$

Figure 1.2 also displays the approximate profile function  $G(z/R)$ .

There are three things that we can learn/observe from this approximation. First, although it is not a particularly good approximation,  $G$  becomes singular when  $z = \pm iR$ , which agrees with the discussion of analytic properties given in Subsection 19.1.2. [Note also that (1.12) is singular when  $z = \pm ia$ .] Second,  $G$  falls off *exponentially* as  $\exp[-\pi|z|/(2R)]$  for large  $|z|$ . Third, as Figure 1.2 illustrates,  $F(z/R, L/R)$  and correspondingly  $C_0^{[1]}(z)$  essentially fall off for large  $|z|$  in the same way as  $G$ . Thus, for example, at a distance of one bore diameter away from the center of the solenoid, when  $|z| = 2R$ , the on-axis field will have fallen from its maximum value by approximately a factor of  $\exp(-\pi) \simeq .04$ .

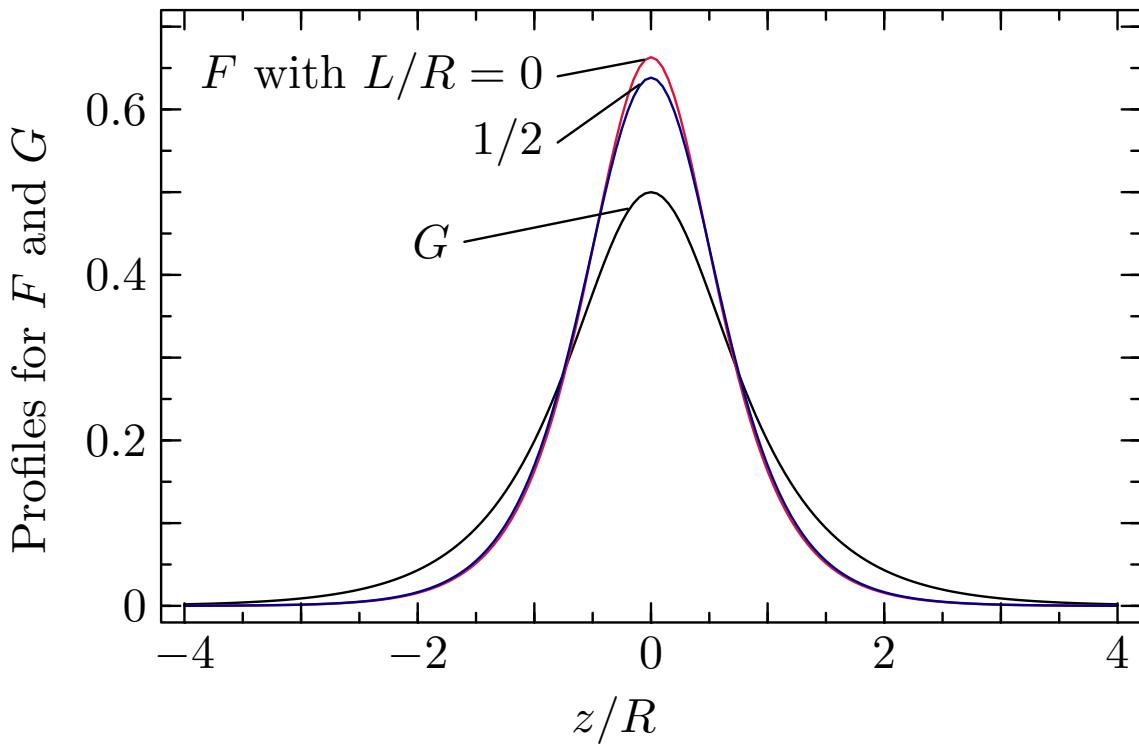


Figure 21.1.2: The profile function  $F(z/R, L/R)$  as given by (2.14) in the cases  $L/R = 0$  and  $L/R = 1/2$ , and the approximate profile function  $G(z/R)$ . The two curves that nearly agree are those for  $F$ , with the highest curve being that for  $F$  when  $L/R = 0$ . The third curve is that for  $G$ .

#### 21.1.4 Quantitatively Correct Iron-dominated Solenoid

### 21.2 Realistic Wigglers/Undulators

#### 21.2.1 An Iron-Dominated Superconducting Wiggler/Undulator

### 21.3 Quadrupoles

#### 21.3.1 Validation of Circular Cylinder Surface Method

This subsection describes numerical tests performed for the case of a Lambertson type quadrupole. In this case the on-axis gradients and their derivatives can be determined analytically. The surface data  $B_\rho(\rho = R, \phi, z)$  can also be found directly using the Biot-Savart law. We will first show that the gradients computed from the surface data following the method of Section 4 agree with the gradients determined analytically. Next we will add noise to the surface data, and again apply the method of Section 4 to this noisy data. We will find that this noise has no undue effect on the computed gradients. Finally, we will show that the noise also has no undue effect on the transfer map  $\mathcal{M}$ .

The method described in Section 4 has been implemented in the code MARYLIE 5.0 [?] as a user-defined routine. The routine reads from an external file the functions  $a_m(R, z)$  and  $b_m(R, z)$ , see (4.1), evaluated on a discrete set of points  $z_i$ . It then generates the corresponding transfer map by using the built-in routine GENMAP to integrate the map equations (1.2). Since MARYLIE 5.0 is a 5<sup>th</sup> order code, only the multipoles through  $m = 6$  need be considered.

The Fourier transforms (4.4) and (4.5) are calculated from the read-in values of  $a_m(R, z)$  and  $b_m(R, z)$  using Filon's method [3] for various values of  $k$  in the interval  $[-k_{max}, k_{max}]$  where  $k_{max}$  is a suitable  $k$  cut-off for the integrals (4.2) and (4.3). For the cases described below, we have used the value  $Rk_{max} = 20$ . Filon's method requires interpolation of the functions  $a_m(R, z)$  and  $b_m(R, z)$ ; and for this purpose we use local parabolic fits.

The integration algorithm of GENMAP is based on a 11<sup>th</sup> order multistep (Adams) method. Because the algorithm uses a fixed step size, one needs to provide values of the on-axis gradients and their derivatives only at the predetermined locations in  $z$  required by GENMAP. The integrals (4.2) and (4.3) that provide the on-axis gradients and their derivations are evaluated at the values of  $z$  needed by GENMAP, again using Filon's method. We emphasize that no interpolation of the on-axis gradients is required by GENMAP.

We have tested the routine by treating the case of a Lambertson quadrupole [8, 9]. For this case only  $b_2(R, z)$  and  $b_6(R, z)$  are nonzero. Correspondingly only the functions  $C_{2,s}^{[0]}$ ,  $C_{2,s}^{[1]}$ ,  $C_{2,s}^{[2]}$ ,  $C_{2,s}^{[3]}$ ,  $C_{2,s}^{[4]}$ , and  $C_{6,s}^{[0]}$  are required. The use of this case as an example has the virtue that the various  $C$  functions can be found exactly [10].

Also, the surface data  $B_\rho(\rho = R, \phi, z)$  can be found directly using the Biot-Savart law, and this data can be integrated over  $\phi$  to yield  $b_2(R, z)$  and  $b_6(R, z)$ . In our test we evaluated  $B_\rho(\rho = R, \phi, z)$  for 279 equally spaced  $z$  values within the interval

$$z \in [z_{min}, z_{max}] = [-7r, 7r] \quad (21.3.1)$$

according to the rule

$$z_i = z_{min} + \Delta(i - 1) \text{ for } i = 1, 2, \dots, 279. \quad (21.3.2)$$

The cylinder on which we evaluated  $B_\rho$  had the radius

$$R = .75r. \quad (21.3.3)$$

Here  $r = .128$  m is the radius of the quadrupole itself, and the length of the quadrupole is  $2r$ . Corresponding,  $\Delta = 14r/278 = 6.44$  mm. The relatively large values of  $z_{min}$  and  $z_{max}$  were necessary because the large radius-to-length ratio of the quadrupole makes the fringe fields very extended.

For each  $z$  value the quantity  $B_\rho(\rho = R, \phi, z)$  was evaluated for 256 equally spaced angles over the interval  $[0, 2\pi]$ , and these  $B_\rho$  values were used to compute the integrals

$$a_m(R, z) = \frac{1}{\pi} \int_0^{2\pi} d\phi \cos(m\phi) B_\rho(\rho = R, \phi, z), \quad (21.3.4)$$

$$b_m(R, z) = \frac{1}{\pi} \int_0^{2\pi} d\phi \sin(m\phi) B_\rho(\rho = R, \phi, z). \quad (21.3.5)$$

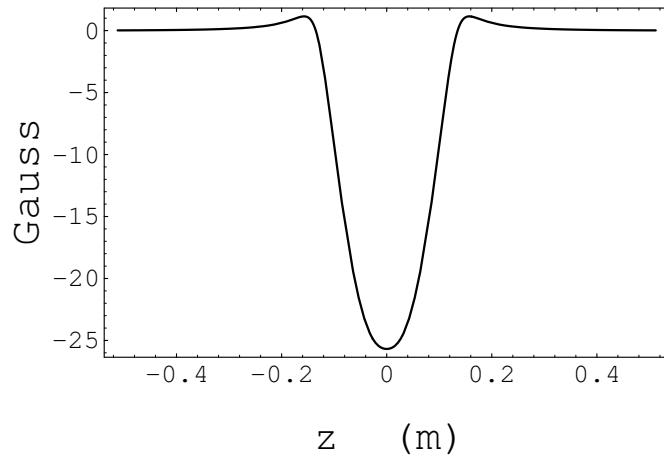


Figure 21.3.1: The angle integrated surface data  $b_2(R, z)$ . The magnet occupies the interval  $z \in [-0.128 \text{ m}, 0.128 \text{ m}]$ .

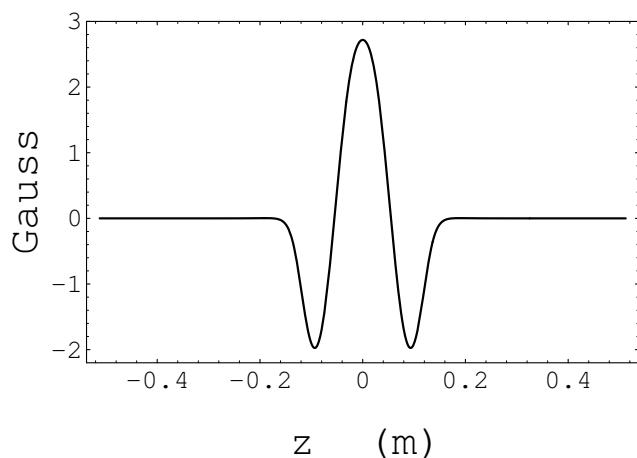


Figure 21.3.2: The angle integrated surface data  $b_6(R, z)$ .

Because of the symmetries of a (normal) quadrupole only the functions  $b_2$  and  $b_6$  are non vanishing for  $m \leq 6$ . The net results of the steps just described are the values of these functions at the points (8.2). These functions are shown in Figures 8.1 and 8.2.

Figure 8.3 shows  $C_{2,s}^{[0]}$  determined both analytically and computed numerically from  $b_2(R, z)$  using the method of Section 4 as described above. Evidently the agreement is excellent. Figure 8.4 shows analytic values of  $C_{6,s}^{[0]}$  and values computed numerically from  $b_6(R, z)$ . Again the agreement is excellent. The most stringent test is a comparison of analytic values of  $C_{2,s}^{[4]}$  with values computed numerically from  $b_2(R, z)$ . This comparison is given in Figure 8.5. Again the agreement is excellent.

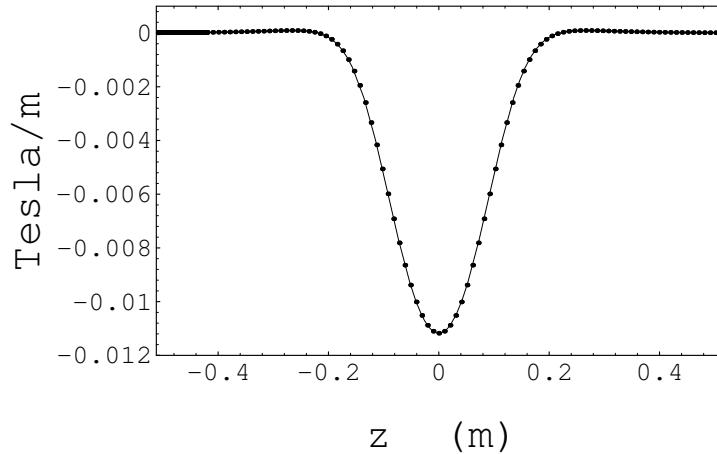


Figure 21.3.3: The function  $C_{2,s}^{[0]}(z)$  as calculated numerically from surface data (dots) and analytically (solid line).

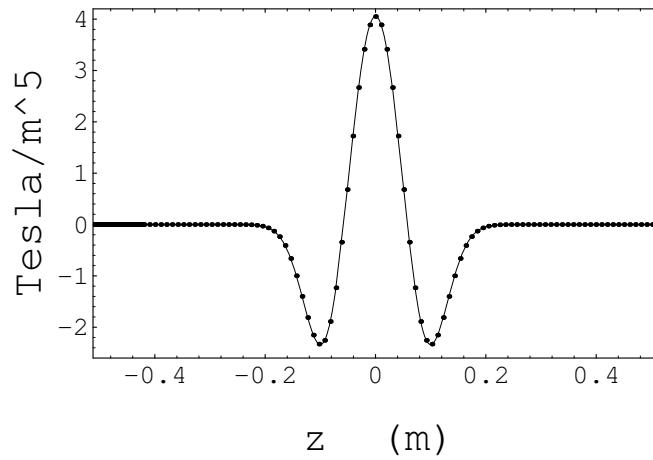


Figure 21.3.4: The function  $C_{6,s}^{[0]}(z)$  as calculated numerically from surface data (dots) and analytically (solid line).

As a final test of our routines, we compared the transfer maps for our Lambertson quadrupole obtained using either the analytically known on-axis gradients or on-axis gradients computed numerically from surface data. Table 8.1 shows that the (relative) difference

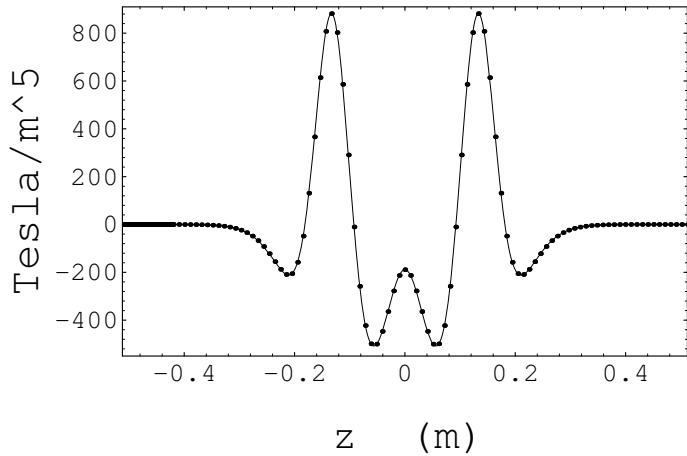


Figure 21.3.5: The function  $C_{2,s}^{[4]}(z)$  as calculated numerically from surface data (dots) and analytically (solid line).

in the surface-data-based map, as compared to the exact map, is very small. Of course, apart from roundoff problems, we expect this difference will vanish as the number of sampling points in  $z$  and  $\phi$  is made arbitrarily large.

Table 21.3.1: Relative difference between the surface-data-based map and the exact map.

map generators	relative difference
$\mathcal{R}_2$	$< 10^{-6}$
$f_3, f_4$	$< 10^{-5}$
$f_5, f_6$	$< 10^{-4}$

Now that the method of Section 4 has been verified to work, we will study the sensitivity of transfer map calculations to the presence of random errors (noise) in the surface data. As a simple model, consider the perturbed functions

$$b_2^{rnd}(R, z_i) = b_2(R, z_i)[1 + \epsilon_2(z_i)], \quad (21.3.6)$$

$$b_6^{rnd}(R, z_i) = b_6(R, z_i)[1 + \epsilon_6(z_i)], \quad (21.3.7)$$

where the  $\epsilon_2(z_i), \epsilon_6(z_i)$  are random variables uniformly distributed in the interval  $[-\epsilon/2, \epsilon/2]$ , and  $b_2(R, z_i), b_6(R, z_i)$  are the same as before. What effect do these errors have on the on-axis gradients computed from the (noisy) surface data?

Figure 8.6 shows  $C_{2,s}^{[4]}$  for a particular realization of the error distribution (seed #2) and  $\epsilon = 10^{-2}$ . The solid line shows analytic results (the same as those of Figure 8.5) and the dots show results computed numerically from the noisy surface data. Close inspection of the figure shows that the points no longer fall exactly on the curve, as is to be expected in the case of noise. However, the size of the deviations from the curve is comparable to the size of the noise, and not unduly larger. To facilitate closer comparison, Figure 8.7 shows the difference between the analytic results and results computed numerically from the noisy surface data. Evidently the deviation is generally on the order of 1% or less, which is comparable with  $\epsilon = 10^{-2}$ .

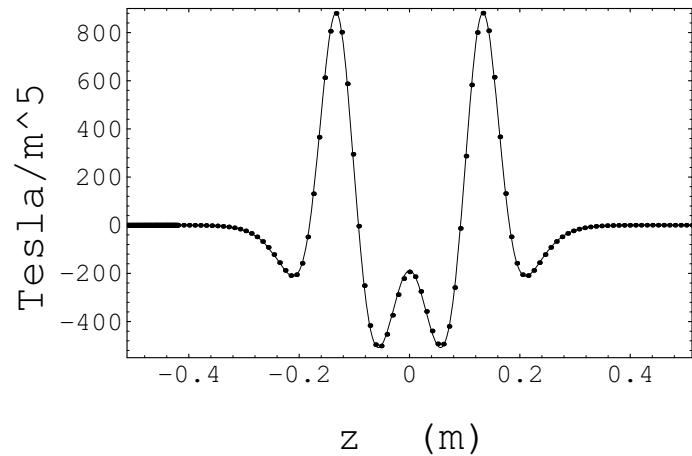


Figure 21.3.6: The function  $C_{2,s}^{[4]}(z)$  as calculated numerically from noisy (seed #2) surface data (dots), and analytically (solid line).

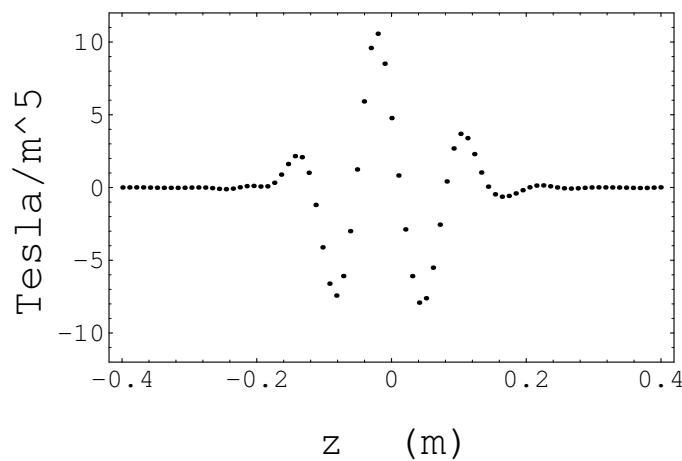


Figure 21.3.7: Difference between the solid line and dots of Figure 8.6.

Since, as we have seen, the computation of the on-axis gradients and their derivatives is not unduly affected by noise in the surface data, we expect the same will be true for the transfer map. This is indeed the case. Table 8.2 shows that the (relative) error in the noisy surface data based map (as compared to the exact map) is, at worst, on the order of the noise.

Table 21.3.2: Relative error of the noisy surface data based map compared to the exact map.

map generators	seed #1	seed#2	seed#3
$\mathcal{R}_2$	$< 3 \times 10^{-4}$	$< 8 \times 10^{-4}$	$< 5 \times 10^{-4}$
$f_3, f_4$	$< 10^{-3}$	$< 1.6 \times 10^{-3}$	$< 1.6 \times 10^{-3}$
$f_5, f_6$	$< 10^{-2}$	$< 1.5 \times 10^{-2}$	$< 1.3 \times 10^{-2}$

One might imagine that the introduction of noise would damage the differentiability (with respect to  $z$ ) of the surface data  $a_m(R, z)$  and  $b_m(R, z)$ . This is indeed true. Consider, for example, the function  $b_2(R, z)$  of the previous section. Figure 9.1 shows its Fourier transform  $\tilde{b}_2(R, k)$ . Evidently the spectrum of  $b_2$  cuts off beyond a wave number  $k_{max} \simeq 150 \text{ m}^{-1}$ . For comparison, Figure 9.2 shows the Fourier transform of  $\epsilon_2(z)$  for the case of seed #2. Note that the noise spectrum extends to  $k'_{max} \simeq 600 \text{ m}^{-1}$ . This is to be expected since  $\pi/\Delta \simeq 490 \text{ m}^{-1}$ . Finally, Figure 9.3 shows the Fourier transform of  $b_2\epsilon_2$ . Its spectrum extends to  $k''_{max} \simeq 750 \text{ m}^{-1}$ . We see that  $k''_{max} \simeq (k_{max} + k'_{max})$ , as is also to be expected.

Although noise does damage the differentiability of surface data, it has considerably less effect on the on-axis gradients derived from the (noisy) surface data. Mathematically, this pleasant result arises from the spectral ‘‘cutoff’’ provided by the kernel  $[k^{m+n-1}/I'_m(kR)]$  that occurs in (4.2) and (4.3). For example, Figure 9.4 shows the factor  $[k^5/I'_2(kR)]$  that is relevant to the computation of  $C_{2,s}^{[4]}$  for the quadrupole example of the last section. We see that it peaks at  $k \simeq 50 \text{ m}^{-1}$ , and falls off rapidly beyond  $k_{max} \simeq 200 \text{ m}^{-1}$ . Finally, Figure 9.5 shows the product of the two functions presented in Figures 9.3 and 9.4. From Figures 9.4 and 9.5 we see that the high wave-number part of the spectrum in Figure 9.3 is effectively filtered out. Correspondingly, as already seen in the previous section, noise in the surface data has no undue effect on the function  $C_{2,s}^{[4]}$ . We remark, as is obvious from our considerations, that noise has even less effect on the functions  $C_{2,s}^{[n]}$  with  $n < 4$ . It also has no undue effect on  $C_{6,s}^{[0]}$ . We also note that the condition  $Rk_{max} = 20$  used in Section 8 corresponds to  $k_{max} \simeq 208$ .

In summary, we have found that noise introduces high wave-number contributions to the spectrum of  $b_m^{rnd}$  where they are absent in the spectrum of  $b_m$ . These potentially damaging contributions are filtered by the kernel  $[k^{m+n-1}/I'_m(kR)]$ . Evidently this filtering becomes less effective with increasing  $n$ , and is improved by making  $R$  as large as possible. For the example studied, we found that smoothing was satisfactory for the selected value of  $R$  and the  $n$  values required for 5th-order calculations.

Finally we remark that, by making more detailed numerical studies, it should be possible to find the partial derivatives

$$\partial \mathcal{M} / \partial b_m(R, z_i) = \partial [\text{any selected coefficient of any generator for } \mathcal{M}] / \partial b_m(R, z_i). \quad (21.3.8)$$

That is, we can evaluate numerically how the map changes when the value of the surface

data at any point is varied. In this way, we can get precise and complete information about the effect of possible noise.

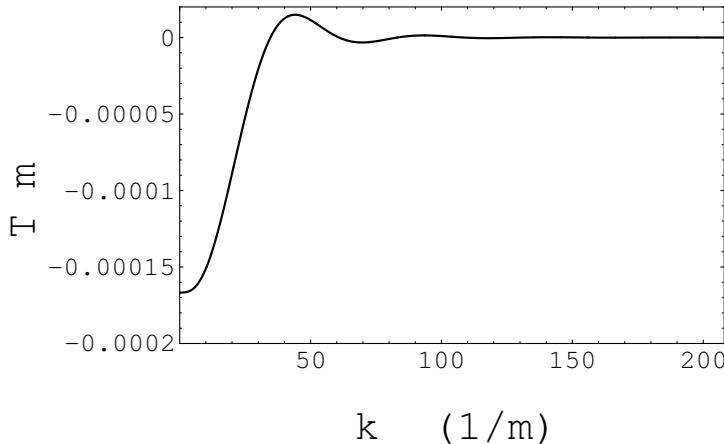


Figure 21.3.8: Real part of the function  $\tilde{b}_2(R, k)$ . The imaginary part vanishes.

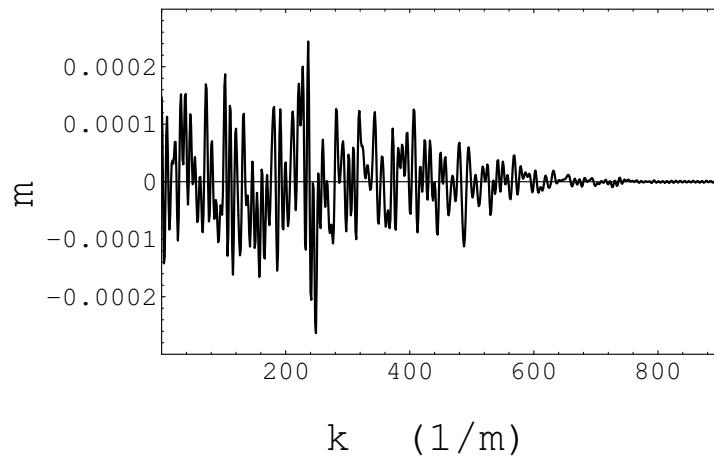


Figure 21.3.9: Real part of the function  $\tilde{\epsilon}_2(k)$ . The imaginary part has similar features.

### 21.3.2 Final Focus Quadrupoles

## 21.4 Closely Adjacent Quadrupoles and Sextupoles

## 21.5 Application to Radio-Frequency Cavities

### Acknowledgement

We have benefitted greatly from many conversations with Peter Walstrom and the reading of some of his internal Technical Notes. We are also grateful to Thomas Mottershead, Filippo

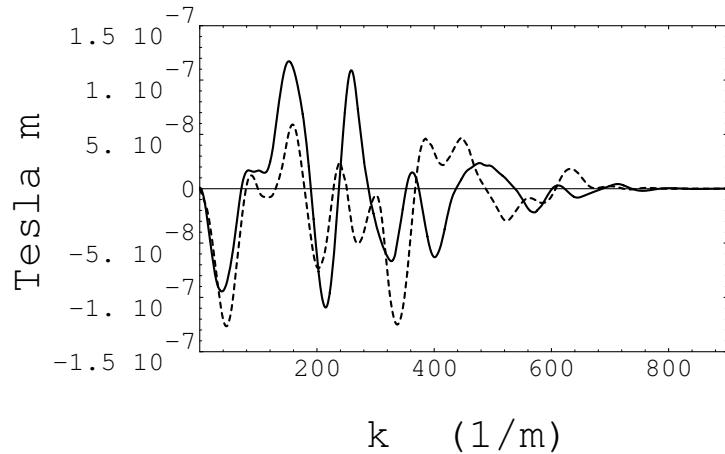


Figure 21.3.10: Real (solid line) and imaginary part (dashed line) of the Fourier transform of the function  $b_2(R, z)\epsilon_2(z)$ .

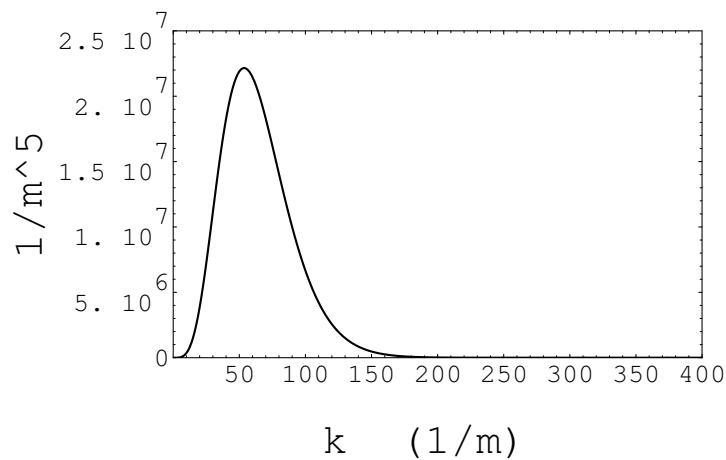


Figure 21.3.11: The factor  $[k^5/I'_2(kR)]$  that appears in the calculation of  $C_{2,s}^{[4]}$ .

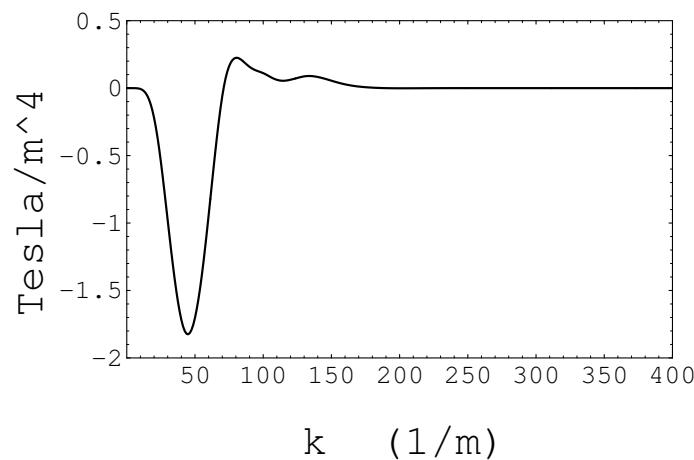


Figure 21.3.12: A plot of the real part of the product of the two functions of Figures 9.3 and 9.4. The imaginary part has similar features.

Neri, and Peter Walstrom for their many contributions to MARYLIE 5.0.



# Bibliography

## General References

- [1] M. Venturini, “Lie Methods, Exact Map computation, and the Problem of Dispersion in space Charge Dominated Beams”, University of Maryland College Park Physics Department Ph. D. thesis (1998).
- [2] M. Venturini and A. Dragt, “Accurate Computation of Transfer Maps from Magnetic Field Data”, *Nuclear Instruments and Methods* **A427**, p. 387 (1999).
- [3] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, Dover (1972).  
Also available on the Web by Googling “abramowitz and stegun 1972”.

## Solenoids

- [4] A. El-Kareh and J. El-Kareh, *Electron Beams, Lenses, and Optics*, Vols. 1 and 2, Academic Press (1970).
- [5] A. Dragt, “Numerical third-order transfer map for solenoid”, *Nuclear Instruments and Methods in Physics Research* **A298**, p. 441 (1990).
- [6] P. W. Hawkes and E. Kasper, *Principles of Electron Optics*, Vols. 1 through 3, Academic Press (1996).

## Dipoles

- [7] P. L. Walstrom, “Dipole-magnet field models based on a conformal map”, *Physical Review Special Topics-Accelerators and Beams* **15**, 102401 (2012).

## Wigglers

## Quadrupoles

- [8] R.P. Avery, B.R. Lambertson, C.D. Pike, PAC 1971 Proceedings, p. 885.
- [9] T.G. Godlove, S. Bernal, M. Reiser, Printed Circuit Quadrupole Design, PAC 1995 Proceedings, p. 2117
- [10] M. Venturini, Transfer Map for Printed-Circuit Magnetic Quadrupoles, Technical Note, Dept. of Physics, Univ. of Maryland (1995).

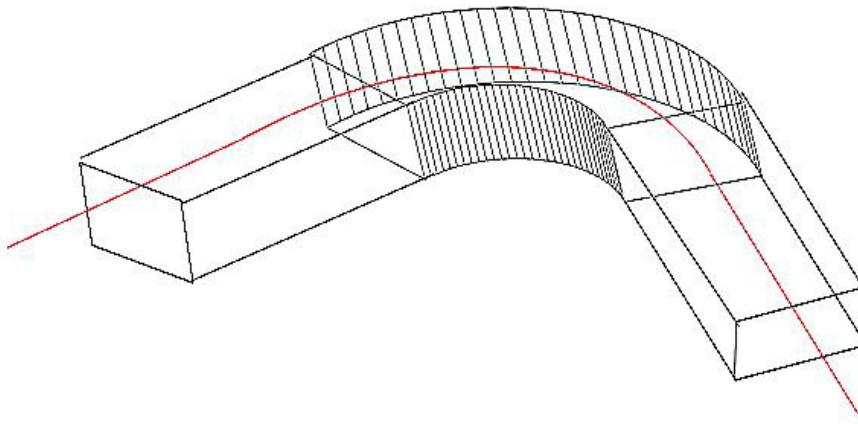
- [11] K. Halbach, "Physical and Optical Properties of Rare Earth Cobalt Magnets", *Nuclear Instruments and Methods* **187**, pp. 109-117 (1981).

# Chapter 22

## Realistic Transfer Maps for General Curved Beam-Line Elements: Theory

### 22.1 Introduction

Surface methods based on the use of cylinders are appropriate for straight beam-line elements or for bent elements with small sagitta. However, cylinders cannot be employed for elements with large sagitta, such as dipoles, where no straight cylinder would fit within the aperture. For such cases more complicated surfaces are required. For example, Figure 1.1 shows a bent box with straight end legs. Its surface could be used to treat a dipole with large sagitta. In this case, the bent part of the box would lie within the body of the dipole, and the straight legs would enclose the fringe-field regions.



---

Figure 22.1.1: A bent box with straight end legs.

But now there is a complication: The *straight* cylinder methods succeeded because Laplace's equation is separable in circular, elliptical, and rectangular cylinder coordinates. Consequently, we were able to find a kernel that related the interior vector potential to the

normal component of the surface magnetic field. However, there is no *bent* coordinate system with straight ends for which Laplace's equation is separable. The method of cylindrical multiples and on-axis gradients is only applicable to straight elements.

This problem can in principle be overcome if *both* the normal component of the magnetic field and the scalar potential for the magnetic field are known on the surface. (Note that a knowledge of the scalar potential on the surface is equivalent to a knowledge of the tangential component of the field on the surface.) Such data are in fact provided on a mesh by some 3-dimensional field solvers, and these data can be interpolated onto the surface.

Let  $V$  be some volume in three-dimensional space bounded by a surface  $S$ . Suppose that the magnetic field  $\mathbf{B}(\mathbf{r})$  is source free when  $\mathbf{r}$  is within  $V$ . That is, for  $\mathbf{r}$  within  $V$ ,  $\mathbf{B}(\mathbf{r})$  satisfies the requirements

$$\nabla \cdot \mathbf{B}(\mathbf{r}) = 0, \quad (22.1.1)$$

$$\nabla \times \mathbf{B}(\mathbf{r}) = 0. \quad (22.1.2)$$

This will be the case for the magnetic field in an evacuated beam pipe. For a Hamiltonian treatment of trajectories, we need a vector potential  $\mathbf{A}(\mathbf{r})$  such that

$$\mathbf{B}(\mathbf{r}) = \nabla \times \mathbf{A}(\mathbf{r}). \quad (22.1.3)$$

Let  $\mathbf{n}'(\mathbf{r}')$  be the outward normal to  $S$  at the point  $\mathbf{r}' \in S$ . Then the normal component of  $\mathbf{B}$  on  $S$  is given by the definition

$$B_n(\mathbf{r}') = \mathbf{n}'(\mathbf{r}') \cdot \mathbf{B}(\mathbf{r}'). \quad (22.1.4)$$

Also, let  $\psi(\mathbf{r}')$  be the value of the magnetic scalar potential at the point  $\mathbf{r}' \in S$ . It satisfies the relation

$$\mathbf{B}(\mathbf{r}') = \nabla' \psi(\mathbf{r}'). \quad (22.1.5)$$

Then, with the aid of the vector potential for Dirac magnetic monopoles and Helmholtz's theorem, it can be shown that there are *kernels*  $\mathbf{G}^n$  and  $\mathbf{G}^t$  such that a suitable interior vector potential  $\mathbf{A}(\mathbf{r})$  for  $\mathbf{r}$  within  $V$  is given by the relation

$$\mathbf{A}(\mathbf{r}) = \mathbf{A}^n(\mathbf{r}) + \mathbf{A}^t(\mathbf{r}) \quad (22.1.6)$$

with

$$\mathbf{A}^n(\mathbf{r}) = \int_S dS' B_n(\mathbf{r}') \mathbf{G}^n(\mathbf{r}, \mathbf{r}') \quad (22.1.7)$$

and

$$\mathbf{A}^t(\mathbf{r}) = \int_S dS' \psi(\mathbf{r}') \mathbf{G}^t(\mathbf{r}, \mathbf{r}'). \quad (22.1.8)$$

(Here the superscripts  $n$  and  $t$  denote *normal* and *tangential*, respectively, and indicate the contributions to the vector potential made by the normal and tangential components of  $\mathbf{B}$  on  $S$ .) Moreover, the constituents of  $\mathbf{A}(\mathbf{r})$ , and hence  $\mathbf{A}(\mathbf{r})$  itself, satisfy the Coulomb gauge condition,

$$\nabla \cdot \mathbf{A}^n(\mathbf{r}) = \nabla \cdot \mathbf{A}^t(\mathbf{r}) = \nabla \cdot \mathbf{A}(\mathbf{r}) = 0. \quad (22.1.9)$$

A detailed exposition of this method, including expected accuracy and insensitivity to noise in the surface data, is the subject of this and the next few chapters. Thus, taken

together, Chapters 15 through 21 and this chapter and Chapters 23 through 25 are intended to provide an extensive description of, and associated robust numerical algorithms for, the computation of transfer maps, including all fringe-field and higher-order multipole effects, for realistic beam-line elements having arbitrary geometry.<sup>1</sup>

Section 2 of this chapter describes the mathematical tools required to treat general geometries. These tools are Dirac's magnetic monopole vector potential and Helmholtz's theorem. Sections 3 and 4 derive the relations (1.3) through (1.9), find the kernels  $\mathbf{G}^n$  and  $\mathbf{G}^t$ , and describe their properties.

Before continuing on, we pause to advertise some of the virtues of what can be achieved with the use of general surface methods.

- The constituents  $\mathbf{A}^n(\mathbf{r})$  and  $\mathbf{A}^t(\mathbf{r})$  of  $\mathbf{A}(\mathbf{r})$ , and hence  $\mathbf{A}(\mathbf{r})$  itself, are analytic functions of  $\mathbf{r}$  for  $\mathbf{r}$  within  $V$ , even when there are errors in the surface fields  $B_n$  and  $\psi$ , and no matter how poorly the integrals (1.7) and (1.8) are evaluated.
- The Maxwell equations for  $\mathbf{B}(\mathbf{r})$ , and the Coulomb gauge condition for  $\mathbf{A}(\mathbf{r})$  and its constituents, are satisfied exactly even when there are errors in the surface fields  $B_n$  and  $\psi$ , and no matter how poorly the integrals (1.7) and (1.8) are evaluated.
- The kernels  $\mathbf{G}^n$  and  $\mathbf{G}^t$  are smoothing. Consequently, the  $\mathbf{A}(\mathbf{r})$  given by (1.6) through (1.8) is relatively insensitive to noise in the surface fields  $B_n$  and  $\psi$ .

We hasten to add that the first two items above should not be taken to mean that there is no need to take care to evaluate integrals well. They just indicate that the worst disasters have been avoided. Subsequently we will learn that the kernels  $\mathbf{G}^n$  and  $\mathbf{G}^t$ , and their  $\mathbf{r}$  derivatives, can be strongly peaked in  $\mathbf{r}'$  when  $\mathbf{r}$  is near  $S$ . To obtain accurate results, this behavior of the kernels must be taken into account when integrating, with respect to  $\mathbf{r}'$ , over the surface  $S$ .

## 22.2 Mathematical Tools

### 22.2.1 Electric Dirac Strings

In this subsection we will motivate the subject of magnetic Dirac strings by treating the simpler electric case. Suppose  $\mathbf{E}(\mathbf{r})$  is a vector field that obeys the equations

$$\nabla \times \mathbf{E} = 0, \tag{22.2.1}$$

$$\nabla \cdot \mathbf{E} = \rho. \tag{22.2.2}$$

From (2.1) we know there is a scalar potential  $\phi$  such that

$$\mathbf{E} = -\nabla\phi, \tag{22.2.3}$$

---

<sup>1</sup>In this sentence we have used the term *multipole* loosely to refer, simply, to nonlinear terms arising from nonlinear magnetic field variations. As already emphasized earlier, the concept of cylindrical multipoles only applies to *straight* elements.

and from (2.2) it follows that

$$\nabla^2 \phi = -\rho. \quad (22.2.4)$$

Introduce the notation

$$|\mathbf{r} - \mathbf{r}'| = ||\mathbf{r} - \mathbf{r}'|| = [(x - x')^2 + (y - y')^2 + (z - z')^2]^{1/2}. \quad (22.2.5)$$

Consider the function  $1/|\mathbf{r} - \mathbf{r}'|$ . It satisfies the relation

$$\nabla^2 [1/|\mathbf{r} - \mathbf{r}'|] = -4\pi\delta_3(\mathbf{r} - \mathbf{r}') \quad (22.2.6)$$

where the indicated derivatives are to be taken with respect to the components of  $\mathbf{r}$ . Assuming that  $\rho(\mathbf{r})$  falls off sufficiently rapidly at infinity, it follows that a solution to (2.4) is given by the relation

$$\phi(\mathbf{r}) = [1/(4\pi)] \int d^3 \mathbf{r}' \rho(\mathbf{r}') / |\mathbf{r} - \mathbf{r}'|. \quad (22.2.7)$$

Moreover, (2.7) is the unique solution that vanishes at infinity.

For our discussion we will need some knowledge of low-order (spherical) multipole expansions, which we review briefly here. Suppose that the charge distribution  $\rho$  is nonzero only in some volume  $V$  surrounding the point  $\mathbf{r}_d$ . (Here the subscript  $d$  stands for *distribution*, and will later stand for *dipole*.) Then (2.7) becomes

$$\phi(\mathbf{r}) = [1/(4\pi)] \int_V d^3 \mathbf{r}' \rho(\mathbf{r}') / |\mathbf{r} - \mathbf{r}'|. \quad (22.2.8)$$

Suppose also that  $\mathbf{r}$  lies outside  $V$  so that the denominator in (2.8) never vanishes. Make the change of variables

$$\mathbf{r}' = \mathbf{r}_d + \boldsymbol{\xi} \quad (22.2.9)$$

so that (2.8) becomes

$$\phi(\mathbf{r}) = [1/(4\pi)] \int_{V_0} d^3 \boldsymbol{\xi} \rho(\mathbf{r}_d + \boldsymbol{\xi}) / |(\mathbf{r} - \mathbf{r}_d) - \boldsymbol{\xi}| \quad (22.2.10)$$

where  $V_0$  is a volume surrounding the origin. Under the assumption that  $\mathbf{r} \notin V$ , the denominator factor in (2.10) can be expanded as a power series in the components of  $\boldsymbol{\xi}$ ,

$$1/|(\mathbf{r} - \mathbf{r}_d) - \boldsymbol{\xi}| = [1/|\mathbf{r} - \mathbf{r}_d|][1 + \boldsymbol{\xi} \cdot (\mathbf{r} - \mathbf{r}_d)/|\mathbf{r} - \mathbf{r}_d|^2 + O(\xi^2)]. \quad (22.2.11)$$

Put this expansion into the integral (2.10) to yield the result

$$\begin{aligned} \phi(\mathbf{r}) &= [1/(4\pi)][1/|\mathbf{r} - \mathbf{r}_d|] \int_{V_0} d^3 \boldsymbol{\xi} \rho(\mathbf{r}_d + \boldsymbol{\xi}) \\ &\quad + [1/(4\pi)][1/|\mathbf{r} - \mathbf{r}_d|^3](\mathbf{r} - \mathbf{r}_d) \cdot \int_{V_0} d^3 \boldsymbol{\xi} \boldsymbol{\xi} \rho(\mathbf{r}_d + \boldsymbol{\xi}) + O(\xi^2). \end{aligned} \quad (22.2.12)$$

The integrals in (2.12) can be manipulated to bring them to the forms

$$\int_{V_0} d^3 \boldsymbol{\xi} \rho(\mathbf{r}_d + \boldsymbol{\xi}) = \int_V d^3 \mathbf{r}' \rho(\mathbf{r}') = Q, \quad (22.2.13)$$

$$\int_{V_0} d^3 \boldsymbol{\xi} \, \boldsymbol{\xi} \cdot \rho(\mathbf{r}_d + \boldsymbol{\xi}) = \int_V d^3 \mathbf{r}' \, (\mathbf{r}' - \mathbf{r}_d) \cdot \rho(\mathbf{r}') = \mathbf{p}_d. \quad (22.2.14)$$

Here  $Q$ , the total charge in  $V$ , is the monopole moment. And  $\mathbf{p}_d$  is the dipole moment (with respect to the point  $\mathbf{r}_d$ ) of the charge distribution in  $V$ . Thus, we find that

$$\phi(\mathbf{r}) = [Q/(4\pi)][1/|\mathbf{r} - \mathbf{r}_d|] + [1/(4\pi)][\mathbf{p}_d \cdot (\mathbf{r} - \mathbf{r}_d)]/|\mathbf{r} - \mathbf{r}_d|^3 + O(\xi^2). \quad (22.2.15)$$

That is, the potential arising from a charge distribution, at a point  $\mathbf{r}$  outside the distribution, is a sum of monopole, dipole, and higher-order multipole contributions.

We recall that the prototypical example of a dipole consists of two opposite charges  $\pm q$  separated by a distance  $2\epsilon$  in the limit that  $\epsilon \rightarrow 0$  and  $q \rightarrow \infty$  in such a way that the product  $2q\epsilon$  remains constant. For example, suppose a charge  $+q$  is placed at the location  $\mathbf{r}_d + \boldsymbol{\epsilon}$  and a charge  $-q$  is placed at the location  $\mathbf{r}_d - \boldsymbol{\epsilon}$ . Then we find that the potential due to this two-charge combination is given by the relation

$$\phi(\mathbf{r}, \mathbf{r}_d) = [1/(4\pi)][q/|\mathbf{r} - (\mathbf{r}_d + \boldsymbol{\epsilon})| - q/|\mathbf{r} - (\mathbf{r}_d - \boldsymbol{\epsilon})|]. \quad (22.2.16)$$

Expansion of (2.16) in powers of  $\boldsymbol{\epsilon}$  gives the result

$$\phi(\mathbf{r}, \mathbf{r}_d) = [1/(4\pi)](2q\boldsymbol{\epsilon}) \cdot (\mathbf{r} - \mathbf{r}_d)/|\mathbf{r} - \mathbf{r}_d|^3 + O(q\epsilon^2). \quad (22.2.17)$$

Now let  $\boldsymbol{\epsilon} \rightarrow 0$  and  $q \rightarrow \infty$  in such a way that

$$2q\boldsymbol{\epsilon} \rightarrow \mathbf{p}_d. \quad (22.2.18)$$

In this limit (2.17) becomes

$$\phi_d(\mathbf{r}, \mathbf{r}_d) = [1/(4\pi)][\mathbf{p}_d \cdot (\mathbf{r} - \mathbf{r}_d)]/|\mathbf{r} - \mathbf{r}_d|^3, \quad (22.2.19)$$

in agreement with the second term in (2.15). We note, with the convention  $q > 0$ , that the dipole moment vector  $\mathbf{p}_d$  points from the location of  $-q$  to the location of  $+q$ .

We also note, for future use, that the field  $\mathbf{E}_d(\mathbf{r}, \mathbf{r}_d)$  at the point  $\mathbf{r}$  arising from a dipole at the point  $\mathbf{r}_d$  (with  $\mathbf{r} \neq \mathbf{r}_d$ ) is given by the relation

$$\begin{aligned} \mathbf{E}_d(\mathbf{r}, \mathbf{r}_d) &= -\nabla \phi_d(\mathbf{r}, \mathbf{r}_d) \\ &= -[1/(4\pi)][\mathbf{p}_d/|\mathbf{r} - \mathbf{r}_d|^3] + [3/(4\pi)][(\mathbf{r} - \mathbf{r}_d)[\mathbf{p}_d \cdot (\mathbf{r} - \mathbf{r}_d)]/|\mathbf{r} - \mathbf{r}_d|^5]. \end{aligned} \quad (22.2.20)$$

We will now use the expression for the potential of a dipole, namely (2.19), to carry out an instructive construction and calculation. Suppose  $\mathbf{r}_A$  and  $\mathbf{r}_B$  are the locations of two points  $A$  and  $B$ . Imagine these two points to be joined by a line (path, *string*)  $L$  starting at  $\mathbf{r}_A$  and ending at  $\mathbf{r}_B$ . See Figure 2.1. Divide the path into  $N$  segments, each of length  $\Delta s$ , and place a dipole of magnitude  $g\Delta s$  at the center of each segment with the dipole moment vector pointing along the path at each point. Here  $g$  is some constant. Thus, the dipole moment  $\Delta \mathbf{p}_d$  of each segment is given by the expression

$$\Delta \mathbf{p}_d = g\Delta s(\Delta \mathbf{r}/|\Delta \mathbf{r}|) = g\Delta \mathbf{r} \quad (22.2.21)$$

since  $|\Delta\mathbf{r}| = \Delta s$ . Let us compute the potential  $\phi_s(\mathbf{r})$  produced by this *string* of dipoles. It will be the sum of the potentials of the individual dipoles. In the limit  $\Delta s \rightarrow 0$  and  $N \rightarrow \infty$  it is given by the integral

$$\begin{aligned}\phi_s(\mathbf{r}) &= [1/(4\pi)] \int_L d\mathbf{p}_d \cdot (\mathbf{r} - \mathbf{r}_d) / |\mathbf{r} - \mathbf{r}_d|^3 \\ &= [g/(4\pi)] \int_{\mathbf{r}_A}^{\mathbf{r}_B} d\mathbf{r}_d \cdot (\mathbf{r} - \mathbf{r}_d) / |\mathbf{r} - \mathbf{r}_d|^3.\end{aligned}\quad (22.2.22)$$

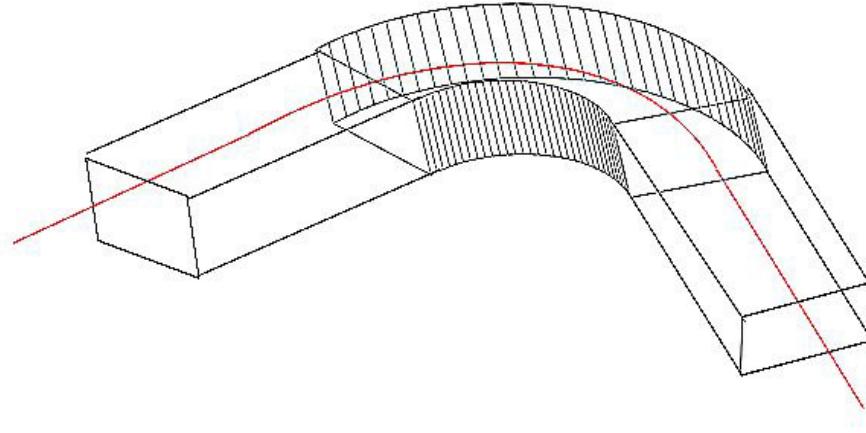


Figure 22.2.1: (Place Holder) A path  $L$  from the point  $A$  to the point  $B$ . Dipoles are laid out and aligned along the path to form a string.

Can the integral (2.22) be evaluated? Recall the identity

$$\nabla^d(1/|\mathbf{r} - \mathbf{r}_d|) = (\mathbf{r} - \mathbf{r}_d)/|\mathbf{r} - \mathbf{r}_d|^3 \quad (22.2.23)$$

where  $\nabla^d$  denotes differentiation with respect to the components of  $\mathbf{r}_d$ . This identity may be employed in (2.22) to yield the result

$$\begin{aligned}\phi_s(\mathbf{r}) &= [g/(4\pi)] \int_{\mathbf{r}_A}^{\mathbf{r}_B} d\mathbf{r}_d \cdot (\mathbf{r} - \mathbf{r}_d) / |\mathbf{r} - \mathbf{r}_d|^3 \\ &= [g/(4\pi)] \int_{\mathbf{r}_A}^{\mathbf{r}_B} d\mathbf{r}_d \cdot [\nabla^d(1/|\mathbf{r} - \mathbf{r}_d|)] \\ &= [g/(4\pi)] \{ [(1/|\mathbf{r} - \mathbf{r}_B|)] - [(1/|\mathbf{r} - \mathbf{r}_A|)] \}.\end{aligned}\quad (22.2.24)$$

We see that the potential  $\phi_s(\mathbf{r})$  resulting from a string of dipoles is the same as the potential produced by a charge  $-g$  located at  $\mathbf{r}_A$  and a charge  $+g$  located at  $\mathbf{r}_B$ . This mathematically derived result is also intuitive because we expect, for a string of dipoles arrayed head-to-tail, that adjacent head-tail pairs would cancel so all that would be left would be the initial negative tail and the final positive head.

Note that, as it stands, (2.22) is undefined for points  $\mathbf{r} \in L$ . However, since the integrand in (2.22) is a perfect differential, see (2.23), the path can be deformed at will to avoid any possible vanishings of the denominator in (2.22) without changing the value of the integral. Indeed, (2.24) shows that  $\phi_s(\mathbf{r})$  depends only on the endpoints of the path, and is otherwise path independent.

### 22.2.2 Magnetic Dirac Strings

#### The General Case

In analogy to the work of the previous subsection, this subsection will describe calculations for the complementary case of a vector field  $\mathbf{B}(\mathbf{r})$  that obeys the equations

$$\nabla \times \mathbf{B} = \mathbf{J}, \quad (22.2.25)$$

$$\nabla \cdot \mathbf{B} = 0. \quad (22.2.26)$$

Note, in order for (2.25) to make sense, we must require that

$$\nabla \cdot \mathbf{J} = \nabla \cdot (\nabla \times \mathbf{B}) = 0. \quad (22.2.27)$$

(Recall that the divergence of a curl vanishes.)

In the case of (2.25) and (2.26) it is often assumed that there is a vector potential  $\mathbf{A}(\mathbf{r})$  such that

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (22.2.28)$$

because (2.26) will then be satisfied automatically. Let us verify that this Ansatz is possible by construction. Substitution of (2.28) into (2.25) yields the hypothesis

$$\nabla \times (\nabla \times \mathbf{A}) = \mathbf{J}. \quad (22.2.29)$$

Recall the vector identity

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} \quad (22.2.30)$$

where here it is essential that Cartesian components be employed. Let us make the further Coulomb gauge assumption

$$\nabla \cdot \mathbf{A} = 0. \quad (22.2.31)$$

In this circumstance (2.29) and (2.30) become

$$\nabla^2 \mathbf{A} = -\mathbf{J}. \quad (22.2.32)$$

Thanks to (2.6), equation (2.32) has the immediate solution

$$\mathbf{A}(\mathbf{r}) = [1/(4\pi)] \int d^3 \mathbf{r}' \mathbf{J}(\mathbf{r}') / |\mathbf{r} - \mathbf{r}'|. \quad (22.2.33)$$

Moreover, (2.33) is the unique solution that vanishes at infinity. But wait, we must also verify that (2.33) also satisfies (2.31). It does, as you will have the pleasure of showing in Exercise 2.4.

Next suppose that the current distribution  $\mathbf{J}$  is nonzero only in some volume  $V$  surrounding the point  $\mathbf{r}_d$ . Then (2.33) becomes

$$\mathbf{A}(\mathbf{r}) = [1/(4\pi)] \int_V d^3\mathbf{r}' \mathbf{J}(\mathbf{r}') / |\mathbf{r} - \mathbf{r}'|. \quad (22.2.34)$$

Suppose also that  $\mathbf{r}$  lies outside  $V$  so that the denominator in (2.34) never vanishes. Make the change of variables (2.9) so that (2.34) can be rewritten in the form

$$\mathbf{A}(\mathbf{r}) = [1/(4\pi)] \int_{V_0} d^3\boldsymbol{\xi} \mathbf{J}(\mathbf{r}_d + \boldsymbol{\xi}) / |(\mathbf{r} - \mathbf{r}_d) - \boldsymbol{\xi}| \quad (22.2.35)$$

where  $V_0$  is a volume surrounding the origin. As before, make the expansion (2.11) so that (2.35) can be written in the form

$$\begin{aligned} \mathbf{A}(\mathbf{r}) &= [1/(4\pi)][1/|\mathbf{r} - \mathbf{r}_d|] \int_{V_0} d^3\boldsymbol{\xi} \mathbf{J}(\mathbf{r}_d + \boldsymbol{\xi}) \\ &\quad + [1/(4\pi)][1/|\mathbf{r} - \mathbf{r}_d|^3] \int_{V_0} d^3\boldsymbol{\xi} [(\mathbf{r} - \mathbf{r}_d) \cdot \boldsymbol{\xi}] \mathbf{J}(\mathbf{r}_d + \boldsymbol{\xi}) + O(\xi^2). \end{aligned} \quad (22.2.36)$$

The integrals in (2.36) can again be manipulated to bring them to more convenient forms. For the first integral we find that

$$\int_{V_0} d^3\boldsymbol{\xi} \mathbf{J}(\mathbf{r}_d + \boldsymbol{\xi}) = \int_V d^3\mathbf{r}' \mathbf{J}(\mathbf{r}') = 0. \quad (22.2.37)$$

Here use has been made of (2.27). See Exercise 2.5. The second integral can be brought to the form

$$\begin{aligned} \int_{V_0} d^3\boldsymbol{\xi} [(\mathbf{r} - \mathbf{r}_d) \cdot \boldsymbol{\xi}] \mathbf{J}(\mathbf{r}_d + \boldsymbol{\xi}) &= \int_V d^3\mathbf{r}' [(\mathbf{r} - \mathbf{r}_d) \cdot (\mathbf{r}' - \mathbf{r}_d)] \mathbf{J}(\mathbf{r}') \\ &= \mathbf{m}_d \times (\mathbf{r} - \mathbf{r}_d). \end{aligned} \quad (22.2.38)$$

Here use has again been made of (2.27), and  $\mathbf{m}_d$  is the magnetic dipole moment defined by the integral

$$\mathbf{m}_d = (1/2) \int_V d^3\mathbf{r}' [(\mathbf{r}' - \mathbf{r}_d) \times \mathbf{J}(\mathbf{r}')]. \quad (22.2.39)$$

See Exercise 2.6. Thus, we find that

$$\mathbf{A}(\mathbf{r}) = \mathbf{A}_d(\mathbf{r}, \mathbf{r}_d) + O(\xi^2) \quad (22.2.40)$$

where

$$\mathbf{A}_d(\mathbf{r}, \mathbf{r}_d) = [1/(4\pi)][\mathbf{m}_d \times (\mathbf{r} - \mathbf{r}_d)] / |\mathbf{r} - \mathbf{r}_d|^3. \quad (22.2.41)$$

We see that the vector potential arising from a current distribution, at a point  $\mathbf{r}$  outside the distribution, is a sum of dipole and higher order multipole contributions. Unlike the

electric case, there is no monopole contribution. We also remark that  $\mathbf{A}(\mathbf{r}, \mathbf{r}_d)$  satisfies the Coulomb gauge condition (2.31),

$$\nabla \cdot \mathbf{A}(\mathbf{r}, \mathbf{r}_d) = 0. \quad (22.2.42)$$

See Exercise 2.7.

We recall that the prototypical example of a magnetic dipole consists of a small circular and planar ring of radius  $R$ , surrounding an area  $A$  and carrying a current  $I$ , in the limit that  $A \rightarrow 0$  and  $I \rightarrow \infty$  in such a way that the product  $AI$  remains constant. For example, suppose the ring is placed in the  $x, y$  plane and centered around the origin. Suppose also that the current  $I$  circulates in the counterclockwise direction when viewed from above (looking down from positive  $z$  toward the origin). Then we find that (2.39) takes the form

$$\mathbf{m}_d = (1/2) \int_V d^3\mathbf{r}' [\mathbf{r}' \times \mathbf{J}(\mathbf{r}')] = AI\mathbf{e}_z. \quad (22.2.43)$$

For the field  $\mathbf{B}_d(\mathbf{r}, \mathbf{r}_d)$  at the point  $\mathbf{r}$  arising from a magnetic dipole at the point  $\mathbf{r}_d$  (with  $\mathbf{r} \neq \mathbf{r}_d$ ) we find the result

$$\begin{aligned} \mathbf{B}_d(\mathbf{r}, \mathbf{r}_d) &= \nabla \times \mathbf{A}_d(\mathbf{r}, \mathbf{r}_d) \\ &= -[1/(4\pi)][\mathbf{m}_d/|\mathbf{r} - \mathbf{r}_d|^3] + [3/(4\pi)](\mathbf{r} - \mathbf{r}_d)[\mathbf{m}_d \cdot (\mathbf{r} - \mathbf{r}_d)]/|\mathbf{r} - \mathbf{r}_d|^5. \end{aligned} \quad (22.2.44)$$

Note that the right sides of (2.20) and (2.44) agree if  $\mathbf{p}_d = \mathbf{m}_d$ . Thus, we have the key mathematical relation

$$-\nabla\phi_d(\mathbf{r}, \mathbf{r}_d) = \nabla \times \mathbf{A}_d(\mathbf{r}, \mathbf{r}_d) \text{ when } \mathbf{p}_d = \mathbf{m}_d \text{ and } \mathbf{r} \neq \mathbf{r}_d. \quad (22.2.45)$$

In analogy to what was done in the previous subsection for a string of electric dipoles, let us compute the vector potential  $\mathbf{A}_s(\mathbf{r})$  arising from a string of magnetic dipoles. Again we will initially divide the path into  $N$  equal segments, and the magnetic dipole moment of each segment will be given by the relation

$$\Delta\mathbf{m}_d = g\Delta s(\Delta\mathbf{r}/|\Delta\mathbf{r}|) = g\Delta\mathbf{r}. \quad (22.2.46)$$

In the limit  $\Delta s \rightarrow 0$  and  $N \rightarrow \infty$  the vector potential due to the string is given by the integral

$$\begin{aligned} \mathbf{A}_s(\mathbf{r}) &= [1/(4\pi)] \int_L d\mathbf{m}_d \times (\mathbf{r} - \mathbf{r}_d)/|\mathbf{r} - \mathbf{r}_d|^3 \\ &= [g/(4\pi)] \int_{\mathbf{r}_A}^{\mathbf{r}_B} d\mathbf{r}_d \times (\mathbf{r} - \mathbf{r}_d)/|\mathbf{r} - \mathbf{r}_d|^3. \end{aligned} \quad (22.2.47)$$

Recall (2.41). Note that, as it stands, (2.47) is undefined for points  $\mathbf{r} \in L$ . As before, the path can be deformed to avoid any possible vanishings of the denominator. However, unlike the electric case and as will soon be seen, so doing changes the value of  $\mathbf{A}_s(\mathbf{r})$ . We also note that the current distribution associated with a string of magnetic dipoles (all aligned along the string) is that of an infinitesimally thin solenoid bent into the shape of the string.

What is the nature of the magnetic field  $\mathbf{B}_s(\mathbf{r})$  given by

$$\mathbf{B}_s(\mathbf{r}) = \nabla \times \mathbf{A}_s(\mathbf{r})? \quad (22.2.48)$$

We claim, for  $\mathbf{r} \notin L$ , that

$$\nabla \times \mathbf{A}_s(\mathbf{r}) = -\nabla \phi_s(\mathbf{r}). \quad (22.2.49)$$

We will prove this assertion shortly. Assuming it is true, the right side of (2.48) can be evaluated easily using (2.49). In view of (2.24), there is the relation

$$-\nabla \phi_s(\mathbf{r}) = [g/(4\pi)][(\mathbf{r} - \mathbf{r}_B)/|\mathbf{r} - \mathbf{r}_B|^3] - [g/(4\pi)][(\mathbf{r} - \mathbf{r}_A)/|\mathbf{r} - \mathbf{r}_A|^3]. \quad (22.2.50)$$

It follows that

$$\mathbf{B}_s(\mathbf{r}) = [g/(4\pi)][(\mathbf{r} - \mathbf{r}_B)/|\mathbf{r} - \mathbf{r}_B|^3] - [g/(4\pi)][(\mathbf{r} - \mathbf{r}_A)/|\mathbf{r} - \mathbf{r}_A|^3]. \quad (22.2.51)$$

We see that the field  $\mathbf{B}_s(\mathbf{r})$  is that produced by two magnetic *monopoles*, one located at  $\mathbf{r}_B$  with strength  $g$ , and a second located at  $\mathbf{r}_A$  with strength  $-g$ .

At this juncture two comments are in order. First, the  $\mathbf{B}_s(\mathbf{r})$  given by (2.51) evidently *is not* divergence free at the points  $\mathbf{r}_A$  and  $\mathbf{r}_B$ . But the  $\mathbf{B}_s(\mathbf{r})$  given by (2.48) is a curl, and we again recall the theorem that a curl *is* divergence free. The resolution to this apparent paradox is that  $\mathbf{A}_s(\mathbf{r})$  is singular for  $\mathbf{r} \in L$ , and every neighborhood of the points  $\mathbf{r}_A$  and  $\mathbf{r}_B$  contains such singular points, and therefore the conditions for the theorem are not met. Correspondingly, (2.51) holds only for points  $\mathbf{r} \notin L$ .

The second comment is equally subtle. Suppose two different strings  $s$  and  $s'$  (but with the same endpoints) are used to compute  $\mathbf{B}_s(\mathbf{r})$  and  $\mathbf{B}_{s'}(\mathbf{r})$ . Then, according to (2.51), these fields should agree except possibly at the points for which  $\mathbf{r} \in L$  and/or  $\mathbf{r} \in L'$ . Thus, we have the relation

$$\nabla \times [\mathbf{A}_s(\mathbf{r}) - \mathbf{A}_{s'}(\mathbf{r})] = 0 \text{ for } \mathbf{r} \notin L \text{ and } \mathbf{r} \notin L'. \quad (22.2.52)$$

Let  $\Sigma$  be some surface spanning the two strings  $s$  and  $s'$ . See Figure 2.2. Three-dimensional Euclidean space with the surface  $\Sigma$  excluded is still simply connected. It follows that there is a function  $\psi_{ss'}(\mathbf{r})$  such that

$$\mathbf{A}_s(\mathbf{r}) - \mathbf{A}_{s'}(\mathbf{r}) = \nabla \psi_{ss'}(\mathbf{r}) \text{ for } \mathbf{r} \notin \Sigma. \quad (22.2.53)$$

That is, the vector potentials associated with two different strings (but with the same endpoints) are related by a gauge transformation. From (2.53) we see that  $\psi_{ss'}(\mathbf{r})$  will be singular for both  $\mathbf{r} \in L$  and  $\mathbf{r} \in L'$ . It can be shown that  $\psi_{ss'}(\mathbf{r})$  is also harmonic,

$$\nabla^2 \psi_{ss'}(\mathbf{r}) = 0 \text{ for } \mathbf{r} \notin L \text{ and } \mathbf{r} \notin L'. \quad (22.2.54)$$

See Exercise 2.13.

Finally, suppose we let  $\mathbf{r}_B \rightarrow \infty$ . In this limit, the first term on the right side of (2.51) vanishes, and we have the result

$$\mathbf{B}_s(\mathbf{r}) = -[g/(4\pi)][(\mathbf{r} - \mathbf{r}_A)/|\mathbf{r} - \mathbf{r}_A|^3], \quad (22.2.55)$$

which is the field of a monopole located at  $\mathbf{r}_A$  and having strength  $-g$ . Correspondingly, the upper limit in the integral (2.47) is infinite, and the string  $s$ , which we will call a *half-infinite Dirac string*, extends from  $\mathbf{r}_A$  to infinity. And the field (2.55) may be viewed as that of a *Dirac* magnetic monopole.

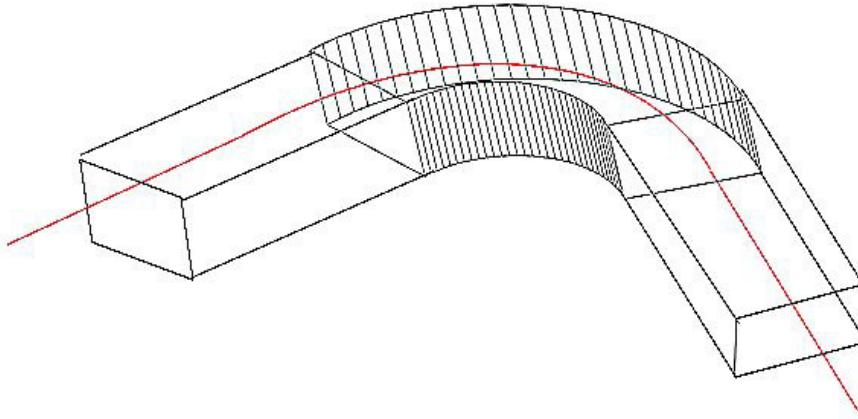


Figure 22.2.2: (Place holder.) A surface  $\Sigma$  spanning the two strings  $s$  and  $s'$ .

### Straight Half-Infinite Strings

For future use, there is a special class of half-infinite strings that is particularly convenient. Let  $\mathbf{m}$  be some unit vector. Consider the straight string (path) from  $\mathbf{r}_A$  to infinity parameterized as

$$\mathbf{r}_d(\lambda) = \mathbf{r}_A + \lambda\mathbf{m} \text{ with } \lambda \in [0, \infty]. \quad (22.2.56)$$

See Figure 2.3. Then, on this path,  $\mathbf{m}_d$  is in the direction of  $\mathbf{m}$ , and we also have the relation

$$d\mathbf{r}_d(\lambda) = \mathbf{m}d\lambda. \quad (22.2.57)$$

For this class of strings the integral (2.47) can be evaluated analytically. We begin by rewriting (2.47) in the form

$$\mathbf{A}_s(\mathbf{r}; \mathbf{r}_A, \mathbf{m}) = [g/(4\pi)] \int_0^\infty d\lambda \mathbf{m} \times [\mathbf{r} - \mathbf{r}_d(\lambda)] / |\mathbf{r} - \mathbf{r}_d(\lambda)|^3. \quad (22.2.58)$$

From (2.56) we see that

$$\mathbf{r} - \mathbf{r}_d(\lambda) = \mathbf{r} - \mathbf{r}_A - \lambda\mathbf{m} \quad (22.2.59)$$

and therefore

$$\mathbf{m} \times [\mathbf{r} - \mathbf{r}_d(\lambda)] = \mathbf{m} \times (\mathbf{r} - \mathbf{r}_A). \quad (22.2.60)$$

Consequently, the integral (2.58) simplifies to the form

$$\mathbf{A}_s(\mathbf{r}; \mathbf{r}_A, \mathbf{m}) = [g/(4\pi)][\mathbf{m} \times (\mathbf{r} - \mathbf{r}_A)] \int_0^\infty d\lambda / |\mathbf{r} - \mathbf{r}_d(\lambda)|^3. \quad (22.2.61)$$

As shown in Exercise 2.14, the integral appearing in (2.61) can be evaluated to yield the result

$$\begin{aligned} \int_0^\infty d\lambda / |\mathbf{r} - \mathbf{r}_d(\lambda)|^3 &= \int_0^\infty d\lambda / |\mathbf{r} - \mathbf{r}_A - \lambda\mathbf{m}|^3 \\ &= 1 / \{|\mathbf{r} - \mathbf{r}_A| [|\mathbf{r} - \mathbf{r}_A| - \mathbf{m} \cdot (\mathbf{r} - \mathbf{r}_A)]\}. \end{aligned} \quad (22.2.62)$$

Therefore,  $\mathbf{A}_s(\mathbf{r}; \mathbf{r}_A, \mathbf{m})$  takes the final explicit form

$$\mathbf{A}_s(\mathbf{r}; \mathbf{r}_A, \mathbf{m}) = [g/(4\pi)][\mathbf{m} \times (\mathbf{r} - \mathbf{r}_A)]/\{|\mathbf{r} - \mathbf{r}_A|[\|\mathbf{r} - \mathbf{r}_A\| - \mathbf{m} \cdot (\mathbf{r} - \mathbf{r}_A)]\}. \quad (22.2.63)$$

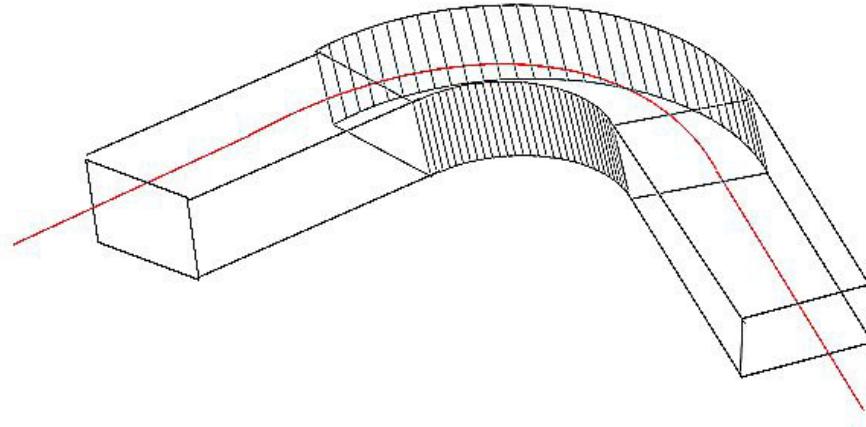


Figure 22.2.3: (Place holder.) A straight half-infinite string extending from  $\mathbf{r}_A$  to infinity in the direction  $\mathbf{m}$ .

### Remaining Verifications

It remains to be verified that (2.49) holds. Suppose that (2.19) is written in the form

$$\phi_d(\mathbf{r}, \mathbf{r}_d; |\mathbf{p}_d|, \mathbf{n}_d) = [1/(4\pi)][|\mathbf{p}_d| \mathbf{n}_d \cdot (\mathbf{r} - \mathbf{r}_d)]/|\mathbf{r} - \mathbf{r}_d|^3 \quad (22.2.64)$$

where  $\mathbf{n}_d$  is the unit vector in the direction of  $\mathbf{p}_d$ . Then (2.22) takes the form

$$\begin{aligned} \phi_s(\mathbf{r}) &= [1/(4\pi)] \int_L d\mathbf{p}_d \cdot (\mathbf{r} - \mathbf{r}_d)/|\mathbf{r} - \mathbf{r}_d|^3 \\ &= [g/(4\pi)] \int_{\mathbf{r}_A}^{\mathbf{r}_B} d\mathbf{r}_d \cdot (\mathbf{r} - \mathbf{r}_d)/|\mathbf{r} - \mathbf{r}_d|^3 \\ &= \int_L \phi_d(\mathbf{r}, \mathbf{r}_d; gds, d\mathbf{r}_d/|d\mathbf{r}_d|), \end{aligned} \quad (22.2.65)$$

and therefore

$$-\nabla \phi_s(\mathbf{r}) = \int_L -\nabla \phi_d(\mathbf{r}, \mathbf{r}_d; gds, d\mathbf{r}_d/|d\mathbf{r}_d|). \quad (22.2.66)$$

Suppose also that (2.41) is written in the form

$$\mathbf{A}_d(\mathbf{r}, \mathbf{r}_d; |\mathbf{m}_d|, \mathbf{n}_d) = [1/(4\pi)][\mathbf{m}_d \mathbf{n}_d \times (\mathbf{r} - \mathbf{r}_d)]/|\mathbf{r} - \mathbf{r}_d|^3. \quad (22.2.67)$$

Then (2.47) takes the form

$$\begin{aligned}\mathbf{A}_s(\mathbf{r}) &= [1/(4\pi)] \int_L d\mathbf{m}_d \times (\mathbf{r} - \mathbf{r}_d)/|\mathbf{r} - \mathbf{r}_d|^3 \\ &= [g/(4\pi)] \int_{\mathbf{r}_A}^{\mathbf{r}_B} d\mathbf{r}_d \times (\mathbf{r} - \mathbf{r}_d)/|\mathbf{r} - \mathbf{r}_d|^3 \\ &= \int_L \mathbf{A}_d(\mathbf{r}, \mathbf{r}_d; gds, d\mathbf{r}_d/|d\mathbf{r}_d|),\end{aligned}\quad (22.2.68)$$

and therefore

$$\nabla \times \mathbf{A}_s(\mathbf{r}) = \int_L \nabla \times \mathbf{A}_d(\mathbf{r}, \mathbf{r}_d; gds, d\mathbf{r}_d/|d\mathbf{r}_d|). \quad (22.2.69)$$

Now compare the integrands on the right sides of (2.66) and (2.69). We see that they have identical arguments. Consequently, by (2.45), they are equal. It follows that the left sides of (2.66) and (2.69) are equal, and therefore (2.49) is correct.

There are still two final matters. First, (2.68) shows that  $\mathbf{A}_s(\mathbf{r})$  is a superposition (integration over  $\mathbf{r}_d$ ) of the  $\mathbf{A}_d(\mathbf{r}, \mathbf{r}_d)$  and, for each  $\mathbf{A}_d(\mathbf{r}, \mathbf{r}_d)$ , we know that the relation (2.42) holds. It follows that  $\mathbf{A}_s(\mathbf{r})$  also satisfies the Coulomb gauge condition,

$$\nabla \cdot \mathbf{A}_s(\mathbf{r}) = 0. \quad (22.2.70)$$

In particular, there is the relation

$$\nabla \cdot \mathbf{A}_s(\mathbf{r}; \mathbf{r}_A, \mathbf{m}) = 0. \quad (22.2.71)$$

Second, since  $\mathbf{A}_s(\mathbf{r}; \mathbf{r}_A, \mathbf{m})$ , is a magnetic monopole vector potential, there is the relation

$$\nabla \times \mathbf{A}_s(\mathbf{r}; \mathbf{r}_A, \mathbf{m}) = -[g/(4\pi)][(\mathbf{r} - \mathbf{r}_A)/|\mathbf{r} - \mathbf{r}_A|^3] = [g/(4\pi)]\nabla(1/|\mathbf{r} - \mathbf{r}_A|). \quad (22.2.72)$$

It follows that

$$\nabla \times [\nabla \times \mathbf{A}_s(\mathbf{r}; \mathbf{r}_A, \mathbf{m})] = 0. \quad (22.2.73)$$

The relations (2.71) and (2.73) will be of subsequent use.

### Fully Infinite (Two) String Monopole Vector Potential

The previous discussion treated the half-infinite string vector potential for a magnetic monopole. In particular, (2.63) gives the vector potential  $\mathbf{A}_s(\mathbf{r}; \mathbf{r}_A, \mathbf{m})$  for a magnetic monopole of strength  $-g$ , see (2.55), located at  $\mathbf{r}_A$  with a straight string extending from  $\mathbf{r}_A$  to  $\infty$  in the direction  $\mathbf{m}$ . This vector potential is singular on the line

$$\mathbf{r} = \mathbf{r}_A + \lambda\mathbf{m} \text{ with } \lambda \in [0, \infty]. \quad (22.2.74)$$

For completeness we will now describe what we will call the *fully infinite* string monopole vector potential.<sup>2</sup> Suppose we form the average of  $\mathbf{A}_s(\mathbf{r}; \mathbf{r}_A, \mathbf{m})$  and  $\mathbf{A}_s(\mathbf{r}; \mathbf{r}_A, -\mathbf{m})$  by writing

$$\mathbf{A}_{2s}(\mathbf{r}; \mathbf{r}_A, \mathbf{m}) = (1/2)[\mathbf{A}_s(\mathbf{r}; \mathbf{r}_A, \mathbf{m}) + \mathbf{A}_s(\mathbf{r}; \mathbf{r}_A, -\mathbf{m})]. \quad (22.2.75)$$

---

<sup>2</sup>The fully infinite string monopole vector potential is sometimes called the *Schwinger potential*.

This vector potential, which we will also call a *two-string monopole* vector potential, will (by superposition) also produce the monopole field (2.55), and will be singular along the full line

$$\mathbf{r} = \mathbf{r}_A + \lambda \mathbf{m} \text{ with } \lambda \in [-\infty, \infty]. \quad (22.2.76)$$

See Figure 2.4.

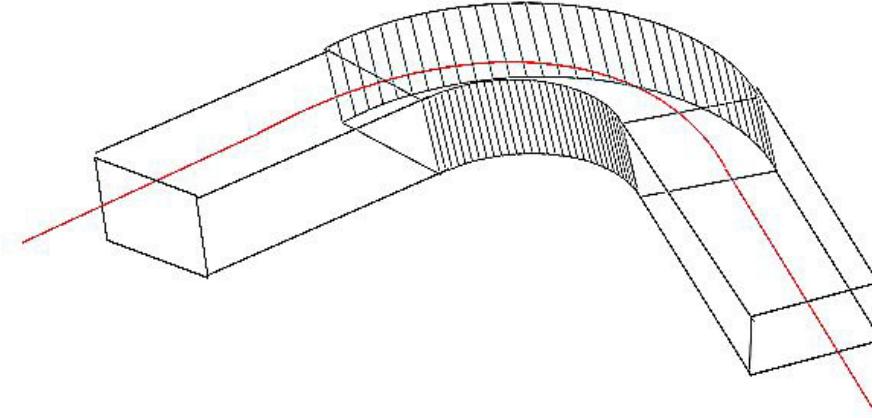


Figure 22.2.4: (Place holder.) A straight full infinite string extending from  $\mathbf{r}_A$  to infinity in the directions  $\pm \mathbf{m}$ .

By superposition, and the use of (2.71) and (2.73), the fully infinite string monopole vector potential satisfies the analogous relations

$$\nabla \cdot \mathbf{A}_{2s}(\mathbf{r}; \mathbf{r}_A, \mathbf{m}) = 0 \quad (22.2.77)$$

and

$$\nabla \times [\nabla \times \mathbf{A}_{2s}(\mathbf{r}; \mathbf{r}_A, \mathbf{m})] = 0, \quad (22.2.78)$$

provided  $\mathbf{r}$  is not on the line (2.76). Finally, from (2.63) and the definition (2.75), we find that  $\mathbf{A}_{2s}(\mathbf{r}; \mathbf{r}_A, \mathbf{m})$  has the explicit form

$$\begin{aligned} \mathbf{A}_{2s}(\mathbf{r}; \mathbf{r}_A, \mathbf{m}) &= [g/(8\pi)][\mathbf{m} \times (\mathbf{r} - \mathbf{r}_A)]/\{|\mathbf{r} - \mathbf{r}_A|[\|\mathbf{r} - \mathbf{r}_A\| - \mathbf{m} \cdot (\mathbf{r} - \mathbf{r}_A)]\} \\ &\quad - [g/(8\pi)][\mathbf{m} \times (\mathbf{r} - \mathbf{r}_A)]/\{|\mathbf{r} - \mathbf{r}_A|[\|\mathbf{r} - \mathbf{r}_A\| + \mathbf{m} \cdot (\mathbf{r} - \mathbf{r}_A)]\} \\ &= \frac{[g/(4\pi)][\mathbf{m} \times (\mathbf{r} - \mathbf{r}_A)][\mathbf{m} \cdot (\mathbf{r} - \mathbf{r}_A)]}{|\mathbf{r} - \mathbf{r}_A|\{|\mathbf{r} - \mathbf{r}_A|^2 - [\mathbf{m} \cdot (\mathbf{r} - \mathbf{r}_A)]^2\}} \\ &= \frac{[g/(4\pi)][\mathbf{m} \times (\mathbf{r} - \mathbf{r}_A)][\mathbf{m} \cdot (\mathbf{r} - \mathbf{r}_A)]}{|\mathbf{r} - \mathbf{r}_A||\mathbf{m} \times (\mathbf{r} - \mathbf{r}_A)|^2}. \end{aligned} \quad (22.2.79)$$

### 22.2.3 Helmholtz Decomposition

Suppose  $V$  is some simply connected volume in 3-dimensional space bounded by a surface  $S$ , and suppose  $\mathbf{F}(\mathbf{r})$  is some 3-dimensional vector field defined in  $V$ . Then, according to a

theorem of *Helmholtz*, there are scalar and vector potentials  $\phi(\mathbf{r})$  and  $\mathbf{A}(\mathbf{r})$  such that

$$\mathbf{F}(\mathbf{r}) = -\nabla\phi(\mathbf{r}) + \nabla \times \mathbf{A}(\mathbf{r}) \text{ for } \mathbf{r} \in V. \quad (22.2.80)$$

Moreover,  $\mathbf{A}(\mathbf{r})$  will have the property

$$\nabla \cdot \mathbf{A}(\mathbf{r}) = 0 \text{ for } \mathbf{r} \in V. \quad (22.2.81)$$

Finally, let Let  $G(\mathbf{r}, \mathbf{r}')$  be the function

$$G(\mathbf{r}, \mathbf{r}') = 1/|\mathbf{r} - \mathbf{r}'|. \quad (22.2.82)$$

Then, the scalar and vector potentials are given in terms of  $\mathbf{F}(\mathbf{r})$ , with  $\mathbf{r} \in V$ , by the relations

$$\phi(\mathbf{r}) = -[1/(4\pi)] \int_S dS' \mathbf{n}' \cdot \mathbf{F}(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') + [1/(4\pi)] \int_V d^3 \mathbf{r}' G(\mathbf{r}, \mathbf{r}') \nabla' \cdot \mathbf{F}(\mathbf{r}'), \quad (22.2.83)$$

$$\mathbf{A}(\mathbf{r}) = -[1/(4\pi)] \int_S dS' [\mathbf{n}' \times G(\mathbf{r}, \mathbf{r}') \mathbf{F}(\mathbf{r}')] + [1/(4\pi)] \int_V d^3 \mathbf{r}' G(\mathbf{r}, \mathbf{r}') \nabla' \times \mathbf{F}(\mathbf{r}'). \quad (22.2.84)$$

Here  $\mathbf{n}'$  is the outward normal to  $S$  at the point  $\mathbf{r}'$ . We emphasize, as is evident from (2.80), (2.83), and (2.84), that for  $\mathbf{r} \in V$  the vector field  $\mathbf{F}(\mathbf{r})$  is completely specified in terms of the divergence and curl of  $\mathbf{F}$  within  $V$  and the values of  $\mathbf{F}$  on the bounding surface  $S$ . No information is required outside of  $V$ .

We will derive this result in stages. Before doing so, some remarks are in order. There two cases of special interest. If  $\mathbf{F}(\mathbf{r})$  is globally defined and falls off at infinity at least as fast as  $1/|\mathbf{r}|^2$ , then we may take the surface  $S$  to infinity and find that the surface integrals vanish. This result shows that, with suitable boundary conditions (fall off) imposed at infinity,  $\mathbf{F}(\mathbf{r})$  is completely specified in terms of its divergence and curl. That the operations of divergence and curl are necessary and sufficient to determine  $\mathbf{F}(\mathbf{r})$  is a consequence of two things: the fact that we are working in *three* dimensions and certain properties of the Euclidean group in three dimensions. See Exercise 2.21.

The second case, of special interest for our purposes, is that for which  $\mathbf{F}(\mathbf{r})$  is divergence and curl free (source free) in  $V$ ,

$$\nabla \cdot \mathbf{F}(\mathbf{r}) = 0 \text{ for } \mathbf{r} \in V \quad (22.2.85)$$

and

$$\nabla \times \mathbf{F}(\mathbf{r}) = 0 \text{ for } \mathbf{r} \in V. \quad (22.2.86)$$

In this case, only the surface terms appear in (2.83) and (2.84), and we obtain the results

$$\phi(\mathbf{r}) = -[1/(4\pi)] \int_S dS' \mathbf{n}' \cdot \mathbf{F}(\mathbf{r}') G(\mathbf{r}, \mathbf{r}'), \quad (22.2.87)$$

$$\mathbf{A}(\mathbf{r}) = -[1/(4\pi)] \int_S dS' [\mathbf{n}' \times \mathbf{F}(\mathbf{r}')] G(\mathbf{r}, \mathbf{r}'). \quad (22.2.88)$$

We will eventually apply these results to the case of a magnetic field  $\mathbf{B}(\mathbf{r})$  that is assumed to be source free within  $V$ , as in (1.1) and (1.2). We take the opportunity at this point to note that  $G(\mathbf{r}, \mathbf{r}')$  as given by (2.82), and for fixed  $\mathbf{r}'$ , is an *analytic* function of the components of  $\mathbf{r}$  for  $\mathbf{r} \neq \mathbf{r}'$ . It follows from the representations (2.87) and (2.88), under very mild assumptions on the surface behavior of  $\mathbf{F}(\mathbf{r})$ , boundedness and continuity will do, that  $\phi(\mathbf{r})$  and  $\mathbf{A}(\mathbf{r})$  are analytic functions of the components of  $\mathbf{r}$  for  $\mathbf{r}$  within  $V$ . Correspondingly, from (2.80),  $\mathbf{F}(\mathbf{r})$  must then also be analytic for  $\mathbf{r}$  within  $V$ .

We begin the proof of Helmholtz's theorem by noting that  $G(\mathbf{r}, \mathbf{r}')$  has the properties

$$\nabla G(\mathbf{r}, \mathbf{r}') = -\nabla' G(\mathbf{r}, \mathbf{r}') = -(\mathbf{r} - \mathbf{r}')/|\mathbf{r} - \mathbf{r}'|^3, \quad (22.2.89)$$

$$\nabla^2 G(\mathbf{r}, \mathbf{r}') = (\nabla')^2 G(\mathbf{r}, \mathbf{r}') = -4\pi\delta_3(\mathbf{r} - \mathbf{r}'), \quad (22.2.90)$$

where  $\nabla'$  denotes differentiation with respect to the components of  $\mathbf{r}'$ . As a result of (2.90) there is, for  $\mathbf{r} \in V$ , the identity

$$\begin{aligned} \mathbf{F}(\mathbf{r}) &= \int_V d^3\mathbf{r}' \delta_3(\mathbf{r} - \mathbf{r}') \mathbf{F}(\mathbf{r}') \\ &= -[1/(4\pi)] \int_V d^3\mathbf{r}' \mathbf{F}(\mathbf{r}') \nabla^2 G(\mathbf{r}, \mathbf{r}') \\ &= -[1/(4\pi)] \nabla^2 \int_V d^3\mathbf{r}' \mathbf{F}(\mathbf{r}') G(\mathbf{r}, \mathbf{r}') \\ &= -\nabla^2 \mathbf{H}(\mathbf{r}) \end{aligned} \quad (22.2.91)$$

where

$$\mathbf{H}(\mathbf{r}) = [1/(4\pi)] \int_V d^3\mathbf{r}' \mathbf{F}(\mathbf{r}') G(\mathbf{r}, \mathbf{r}'). \quad (22.2.92)$$

Invoke again the vector identity

$$-\nabla^2 \mathbf{H}(\mathbf{r}) = \nabla \times [\nabla \times \mathbf{H}(\mathbf{r})] - \nabla [\nabla \cdot \mathbf{H}(\mathbf{r})]. \quad (22.2.93)$$

It follows that

$$\mathbf{F}(\mathbf{r}) = \nabla \times [\nabla \times \mathbf{H}(\mathbf{r})] - \nabla [\nabla \cdot \mathbf{H}(\mathbf{r})], \quad (22.2.94)$$

and therefore (2.80) holds with the definitions

$$\phi(\mathbf{r}) = \nabla \cdot \mathbf{H}(\mathbf{r}), \quad (22.2.95)$$

$$\mathbf{A}(\mathbf{r}) = \nabla \times \mathbf{H}(\mathbf{r}). \quad (22.2.96)$$

It remains to work out computationally useful expressions for  $\phi(\mathbf{r})$  and  $\mathbf{A}(\mathbf{r})$ . Doing so requires a flurry of vector manipulations. Begin with  $\phi(\mathbf{r})$ . According to (2.92) and (2.95) it can be written as

$$\phi(\mathbf{r}) = [1/(4\pi)] \int_V d^3\mathbf{r}' \nabla \cdot [\mathbf{F}(\mathbf{r}') G(\mathbf{r}, \mathbf{r}')]. \quad (22.2.97)$$

Manipulate the integrand in (2.97) to find the result

$$\begin{aligned}\nabla \cdot [\mathbf{F}(\mathbf{r}')G(\mathbf{r}, \mathbf{r}')] &= \mathbf{F}(\mathbf{r}') \cdot \nabla G(\mathbf{r}, \mathbf{r}') = -\mathbf{F}(\mathbf{r}') \cdot \nabla' G(\mathbf{r}, \mathbf{r}') \\ &= -\nabla' \cdot [\mathbf{F}(\mathbf{r}')G(\mathbf{r}, \mathbf{r}')] + G(\mathbf{r}, \mathbf{r}')\nabla' \cdot \mathbf{F}(\mathbf{r}').\end{aligned}\quad (22.2.98)$$

Employ this result in (2.97) to rewrite it in the form

$$\phi(\mathbf{r}) = [1/(4\pi)] \int_V d^3\mathbf{r}' \{-\nabla' \cdot [\mathbf{F}(\mathbf{r}')G(\mathbf{r}, \mathbf{r}')] + G(\mathbf{r}, \mathbf{r}')\nabla' \cdot \mathbf{F}(\mathbf{r}')\}. \quad (22.2.99)$$

Finally, use the divergence theorem to transform the first term on the right side of (2.99) to yield the result

$$\phi(\mathbf{r}) = -[1/(4\pi)] \int_S dS' \mathbf{n}' \cdot \mathbf{F}(\mathbf{r}')G(\mathbf{r}, \mathbf{r}') + [1/(4\pi)] \int_V d^3\mathbf{r}' G(\mathbf{r}, \mathbf{r}')\nabla' \cdot \mathbf{F}(\mathbf{r}'), \quad (22.2.100)$$

in agreement with (2.83).

The case of  $\mathbf{A}(\mathbf{r})$  requires somewhat more effort. Combining (2.92) and (2.96) gives the result

$$\mathbf{A}(\mathbf{r}) = [1/(4\pi)] \int_V d^3\mathbf{r}' \nabla \times [\mathbf{F}(\mathbf{r}')G(\mathbf{r}, \mathbf{r}')]. \quad (22.2.101)$$

Manipulate the integrand in (2.101) to find the result

$$\begin{aligned}\nabla \times [\mathbf{F}(\mathbf{r}')G(\mathbf{r}, \mathbf{r}')] &= [\nabla G(\mathbf{r}, \mathbf{r}')] \times \mathbf{F}(\mathbf{r}') = -\mathbf{F}(\mathbf{r}') \times \nabla G(\mathbf{r}, \mathbf{r}') \\ &= \mathbf{F}(\mathbf{r}') \times \nabla' G(\mathbf{r}, \mathbf{r}').\end{aligned}\quad (22.2.102)$$

There is also the vector identity

$$\nabla' \times [\mathbf{F}(\mathbf{r}')G(\mathbf{r}, \mathbf{r}')] = G(\mathbf{r}, \mathbf{r}')\nabla' \times \mathbf{F}(\mathbf{r}') - \mathbf{F}(\mathbf{r}') \times \nabla' G(\mathbf{r}, \mathbf{r}'). \quad (22.2.103)$$

Combining (2.102) and (2.103) gives the result

$$\nabla \times [\mathbf{F}(\mathbf{r}')G(\mathbf{r}, \mathbf{r}')] = G(\mathbf{r}, \mathbf{r}')\nabla' \times \mathbf{F}(\mathbf{r}') - \nabla' \times [\mathbf{F}(\mathbf{r}')G(\mathbf{r}, \mathbf{r}')]. \quad (22.2.104)$$

Employ this result in (2.101) to rewrite it in the form

$$\mathbf{A}(\mathbf{r}) = [1/(4\pi)] \int_V d^3\mathbf{r}' G(\mathbf{r}, \mathbf{r}')\nabla' \times \mathbf{F}(\mathbf{r}') - [1/(4\pi)] \int_V d^3\mathbf{r}' \nabla' \times [\mathbf{F}(\mathbf{r}')G(\mathbf{r}, \mathbf{r}')]. \quad (22.2.105)$$

Now work on the second integral appearing on the right side of (2.105). Let  $\mathbf{c}$  be any constant vector. By the divergence theorem there is the relation

$$\int_V d^3\mathbf{r}' \nabla' \cdot [\mathbf{c} \times G(\mathbf{r}, \mathbf{r}')\mathbf{F}(\mathbf{r}')] = \int_S dS' \mathbf{n}' \cdot [\mathbf{c} \times G(\mathbf{r}, \mathbf{r}')\mathbf{F}(\mathbf{r}')]. \quad (22.2.106)$$

There is also the vector identity

$$\begin{aligned}\mathbf{n}' \cdot [\mathbf{c} \times G(\mathbf{r}, \mathbf{r}')\mathbf{F}(\mathbf{r}')] &= -\mathbf{n}' \cdot [G(\mathbf{r}, \mathbf{r}')\mathbf{F}(\mathbf{r}') \times \mathbf{c}] \\ &= -[\mathbf{n}' \times G(\mathbf{r}, \mathbf{r}')\mathbf{F}(\mathbf{r}')] \cdot \mathbf{c} \\ &= -\mathbf{c} \cdot [\mathbf{n}' \times G(\mathbf{r}, \mathbf{r}')\mathbf{F}(\mathbf{r}')].\end{aligned}\quad (22.2.107)$$

Consequently, (2.106) can be rewritten in the form

$$\int_V d^3 \mathbf{r}' \nabla' \cdot [\mathbf{c} \times G(\mathbf{r}, \mathbf{r}') \mathbf{F}(\mathbf{r}')] = -\mathbf{c} \cdot \int_S dS' [\mathbf{n}' \times G(\mathbf{r}, \mathbf{r}') \mathbf{F}(\mathbf{r}')]. \quad (22.2.108)$$

Next manipulate the integrand on the left side of (2.108) to find the result

$$\begin{aligned} \nabla' \cdot [\mathbf{c} \times G(\mathbf{r}, \mathbf{r}') \mathbf{F}(\mathbf{r}')] &= -\nabla' \cdot [G(\mathbf{r}, \mathbf{r}') \mathbf{F}(\mathbf{r}') \times \mathbf{c}] \\ &= -\{\nabla' \times [G(\mathbf{r}, \mathbf{r}') \mathbf{F}(\mathbf{r}')]\} \cdot \mathbf{c} \\ &= -\mathbf{c} \cdot \{\nabla' \times [G(\mathbf{r}, \mathbf{r}') \mathbf{F}(\mathbf{r}')]\}. \end{aligned} \quad (22.2.109)$$

Therefore (2.108) can be rewritten as

$$-\mathbf{c} \cdot \int_V d^3 \mathbf{r}' \nabla' \times [G(\mathbf{r}, \mathbf{r}') \mathbf{F}(\mathbf{r}')] = -\mathbf{c} \cdot \int_S dS' [\mathbf{n}' \times G(\mathbf{r}, \mathbf{r}') \mathbf{F}(\mathbf{r}')], \quad (22.2.110)$$

from which it follows, because  $\mathbf{c}$  is arbitrary, that

$$\int_V d^3 \mathbf{r}' \nabla' \times [G(\mathbf{r}, \mathbf{r}') \mathbf{F}(\mathbf{r}')] = \int_S dS' [\mathbf{n}' \times G(\mathbf{r}, \mathbf{r}') \mathbf{F}(\mathbf{r}')]. \quad (22.2.111)$$

The last step is to employ (2.111) in (2.105) to obtain the final result

$$\mathbf{A}(\mathbf{r}) = [1/(4\pi)] \int_V d^3 \mathbf{r}' G(\mathbf{r}, \mathbf{r}') \nabla' \times \mathbf{F}(\mathbf{r}') - [1/(4\pi)] \int_S dS' [\mathbf{n}' \times G(\mathbf{r}, \mathbf{r}') \mathbf{F}(\mathbf{r}')], \quad (22.2.112)$$

in agreement with (2.84).

It still remains to be shown that, for the definitions made,  $\nabla \cdot \mathbf{A}(\mathbf{r}) = 0$ . Look at (2.96). Since the divergence of a curl vanishes, when suitable smoothness conditions are met by the functions involved, it follows that under these conditions  $\mathbf{A}(\mathbf{r})$  as given by (2.96), and therefore also by (2.112), is indeed divergence free. From (2.92) we see that the analytic properties of  $\mathbf{H}(\mathbf{r})$  are determined by those of  $\mathbf{F}(\mathbf{r})$ . In general  $\mathbf{H}(\mathbf{r})$  will be smoother than  $\mathbf{F}(\mathbf{r})$ . See Appendix F. Therefore, under mild conditions on  $\mathbf{F}(\mathbf{r})$ , the vector potential  $\mathbf{A}(\mathbf{r})$  will be divergence free.

## Exercises

**22.2.1.** Verify the expansions (2.11) and (2.17).

**22.2.2.** Verify (2.20).

**22.2.3.** Verify the identity (2.23) and its use to evaluate the integral (2.24).

**22.2.4.** Verify that  $\mathbf{A}(\mathbf{r})$  as given by (2.33) satisfies (2.31).

**22.2.5.** The purpose of this exercise is to verify (2.37) using (2.27).

**22.2.6.** The purpose of this exercise is to verify (2.38) using the definition (2.39).

**22.2.7.** The purpose of this exercise is to verify (2.42) using the definition (2.41).

**22.2.8.** Verify (2.43).

**22.2.9.** Verify (2.44).

**22.2.10.** Show that the integral (2.47) can be written in the form

$$[g/(4\pi)] \int_{\mathbf{r}_A}^{\mathbf{r}_B} d\mathbf{r}_d \times (\mathbf{r} - \mathbf{r}_d)/|\mathbf{r} - \mathbf{r}_d|^3 = -[g/(4\pi)] \int_{\mathbf{r}_A}^{\mathbf{r}_B} d\mathbf{r}_d \times \nabla [1/|\mathbf{r} - \mathbf{r}_d|]. \quad (22.2.113)$$

**22.2.11.** Verify (2.50).

**22.2.12.** Nature of thin solenoid and nature of field at the end of a thin solenoid.

**22.2.13.** The purpose of this exercise is to verify (2.54).

**22.2.14.** The purpose of this exercise is to verify (2.62).

**22.2.15.** Evaluate  $\mathbf{A}_s(\mathbf{r}; \mathbf{r}_A, \mathbf{m})$  as given by (2.63) for the case

$$\mathbf{r}_A = 0 \quad (22.2.114)$$

and

$$\mathbf{m} = \mathbf{e}_z. \quad (22.2.115)$$

Show, using spherical coordinates, that in this case  $\mathbf{A}_s(\mathbf{r}; \mathbf{r}_A, \mathbf{m})$  has only a  $\phi$  component  $A_\phi^s$  given by

$$A_\phi^s(\mathbf{r}; 0, \mathbf{e}_z) = [g/(4\pi)](1 + \cos \theta)/[r \sin \theta] = [g/(4\pi)](1/r) \cot(\theta/2) \quad (22.2.116)$$

Verify that  $A_\phi^s$  is singular on the positive  $z$  axis, but not on the negative  $z$  axis. Show, by explicit calculation, that

$$\nabla \times \mathbf{A}_s(\mathbf{r}; 0, \mathbf{e}_z) = -[g/(4\pi)][\mathbf{r}/|\mathbf{r}|^3], \quad (22.2.117)$$

as expected.

Repeat the above calculations for the case  $\mathbf{m} = -\mathbf{e}_z$ . Show that again  $\mathbf{A}_s(\mathbf{r}; \mathbf{r}_A, \mathbf{m})$  has only a  $\phi$  component  $A_\phi^s$  now given by

$$\begin{aligned} A_\phi^s(\mathbf{r}; 0, -\mathbf{e}_z) &= -[g/(4\pi)](1 - \cos \theta)/[r \sin \theta] = -[g/(4\pi)](1/r) \sin \theta / (1 + \cos \theta) \\ &= -[g/(4\pi)](1/r) \tan(\theta/2). \end{aligned} \quad (22.2.118)$$

Verify that this  $A_\phi^s$  is singular on the negative  $z$  axis, but not on the positive  $z$  axis. Show, by explicit calculation, that

$$\nabla \times \mathbf{A}_s(\mathbf{r}; 0, -\mathbf{e}_z) = -[g/(4\pi)][\mathbf{r}/|\mathbf{r}|^3], \quad (22.2.119)$$

again as expected.

Verify that  $\mathbf{A}_s(\mathbf{r}; 0, -\mathbf{e}_z)$  and  $\mathbf{A}_s(\mathbf{r}; 0, \mathbf{e}_z)$  are related by a gauge transformation,

$$\mathbf{A}_s(\mathbf{r}; 0, -\mathbf{e}_z) = \mathbf{A}_s(\mathbf{r}; 0, \mathbf{e}_z) + \nabla\chi \quad (22.2.120)$$

with

$$\chi = -[g/(2\pi)]\phi. \quad (22.2.121)$$

Form the fully infinite string vector potential  $\mathbf{A}_{2s}(\mathbf{r}; 0, \mathbf{e}_z)$  using (2.75). Show that  $\mathbf{A}_{2s}(\mathbf{r}; 0, \mathbf{e}_z)$  also has only a  $\phi$  component given by

$$\begin{aligned} A_\phi^{2s}(\mathbf{r}; 0, \mathbf{e}_z) &= (1/2)A_\phi^s(\mathbf{r}; 0, \mathbf{e}_z) + (1/2)A_\phi^s(\mathbf{r}; 0, -\mathbf{e}_z) \\ &= [g/(4\pi)](1/r)(\cot\theta). \end{aligned} \quad (22.2.122)$$

Verify that  $A_\phi^{2s}(\mathbf{r}; 0, \mathbf{e}_z)$  is singular everywhere on the  $z$  axis. Verify by explicit calculation that

$$\nabla \times \mathbf{A}_{2s}(\mathbf{r}; 0, \mathbf{e}_z) = -[g/(4\pi)][\mathbf{r}/|\mathbf{r}|^3], \quad (22.2.123)$$

also as expected.

According to Subsection 2.2, a magnetic Dirac string can be viewed as an infinitesimally thin solenoid. In the case that the string is straight, one can assign a definite vector to the string that points in the direction of current flow. Show that for a string directed along the  $z$  axis, as is the case for this exercise, the current is in the  $+$  (or perhaps  $-$ )  $\mathbf{e}_\phi$  direction, which is the same direction as the associated vector potential  $\mathbf{A}$ .

**22.2.16.** Exercise on the singularity structure of the vector potential for a straight half-infinite Dirac string.

**22.2.17.** Let  $\mathbf{A}_s(\mathbf{r}; \mathbf{r}_A, \mathbf{m})$  and  $\mathbf{A}_s(\mathbf{r}; \mathbf{r}_A, \mathbf{m}')$  be equal strength monopole vector potentials produced by straight-line strings both originating at  $\mathbf{r}_A$  but extending to infinity in the directions  $\mathbf{m}$  and  $\mathbf{m}'$ . See (2.63). Show that both produce the same magnetic field (2.55) at points off the strings. Show that these vector potentials are related by a gauge transformation.

**22.2.18.** Verify (2.72) and (2.73).

**22.2.19.** Show from (2.79) that

$$|\mathbf{A}_{2s}(\mathbf{r}; \mathbf{r}_A, \mathbf{m})| = \frac{|[g/(4\pi)][\mathbf{m} \cdot (\mathbf{r} - \mathbf{r}_A)]|}{|\mathbf{r} - \mathbf{r}_A||\mathbf{m} \times (\mathbf{r} - \mathbf{r}_A)|}. \quad (22.2.124)$$

Verify that  $\mathbf{A}_{2s}(\mathbf{r}; \mathbf{r}_A, \mathbf{m})$  is singular on, and only on, the line (2.76).

**22.2.20.** Suppose a vector field  $\mathbf{F}(\mathbf{r})$  is specified in some volume  $V$ . Surround this volume by a *thin shell*  $\Sigma$ . Extend  $\mathbf{F}(\mathbf{r})$  to all of space by requiring that it vanish outside  $\Sigma$  and go to zero smoothly within  $\Sigma$ . That is, on the boundary of  $V$ , which is the inner surface of  $\Sigma$ ,  $\mathbf{F}$  may have finite values; but within  $\Sigma$  it goes smoothly to zero so that it vanishes on the outer surface of  $\Sigma$  and beyond. It is a standard result in analysis that this can be done in such a way that  $\mathbf{F}(\mathbf{r})$  will have as many derivatives as desired in  $\Sigma$ . Find formulas for  $\phi(\mathbf{r})$  and  $\mathbf{A}(\mathbf{r})$  in this case. Now let the shell shrink to zero thickness while keeping  $V$  unchanged so that  $\Sigma$  becomes the surface  $S$ . Show that the relations (2.80), (2.83), and (2.84) continue to give  $\mathbf{F}(\mathbf{r})$  for  $\mathbf{r} \in V$ , and give  $\mathbf{F}(\mathbf{r}) = 0$  for  $\mathbf{r} \notin V$ .

**22.2.21.** Suppose a vector field  $\mathbf{F}(\mathbf{r})$  is globally defined and falls off at infinity at least as fast as  $1/|\mathbf{r}|^2$ . Show that, when the surface  $S$  in (2.83) and (2.84) is taken to infinity, the surface integrals then vanish. Consequently, (2.83) and (2.84) then take the form

$$\phi(\mathbf{r}) = [1/(4\pi)] \int d^3\mathbf{r}' G(\mathbf{r}, \mathbf{r}') \nabla' \cdot \mathbf{F}(\mathbf{r}'), \quad (22.2.125)$$

$$\mathbf{A}(\mathbf{r}) = [1/(4\pi)] \int d^3\mathbf{r}' G(\mathbf{r}, \mathbf{r}') \nabla' \times \mathbf{F}(\mathbf{r}'). \quad (22.2.126)$$

Thus, in view of (2.80), such a vector field is completely specified by a knowledge of its divergence and curl.

Why should this be the case? Suppose that  $\mathbf{F}(\mathbf{r})$  has the Fourier representation

$$\mathbf{F}(\mathbf{r}) = \int d^3\mathbf{k} \exp(i\mathbf{k} \cdot \mathbf{r}) \tilde{\mathbf{F}}(\mathbf{k}). \quad (22.2.127)$$

Such a representation is possible in any number of dimensions, and its existence is a consequence of the completeness of the unitary representations of the translation part of the Euclidean group. Show that there are the relations

$$\nabla \cdot \mathbf{F}(\mathbf{r}) = i \int d^3\mathbf{k} \exp(i\mathbf{k} \cdot \mathbf{r}) \mathbf{k} \cdot \tilde{\mathbf{F}}(\mathbf{k}), \quad (22.2.128)$$

$$\nabla \times \mathbf{F}(\mathbf{r}) = i \int d^3\mathbf{k} \exp(i\mathbf{k} \cdot \mathbf{r}) \mathbf{k} \times \tilde{\mathbf{F}}(\mathbf{k}). \quad (22.2.129)$$

Consequently, if the functions  $\nabla \cdot \mathbf{F}(\mathbf{r})$  and  $\nabla \times \mathbf{F}(\mathbf{r})$  are assumed known, then, by the Fourier inversion theorem, the functions  $\mathbf{k} \cdot \tilde{\mathbf{F}}(\mathbf{k})$  and  $\mathbf{k} \times \tilde{\mathbf{F}}(\mathbf{k})$  are also known. Recall the vector identity

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b}). \quad (22.2.130)$$

Use this identity to show that

$$\mathbf{k} \times (\mathbf{k} \times \tilde{\mathbf{F}}) = \mathbf{k}(\mathbf{k} \cdot \tilde{\mathbf{F}}) - \tilde{\mathbf{F}}(\mathbf{k} \cdot \mathbf{k}), \quad (22.2.131)$$

and therefore

$$\tilde{\mathbf{F}} = [1/(\mathbf{k} \cdot \mathbf{k})][\mathbf{k}(\mathbf{k} \cdot \tilde{\mathbf{F}})] - [1/(\mathbf{k} \cdot \mathbf{k})][\mathbf{k} \times (\mathbf{k} \times \tilde{\mathbf{F}})]. \quad (22.2.132)$$

Thus, the function  $\tilde{\mathbf{F}}(\mathbf{k})$  is known if the functions  $\mathbf{k} \cdot \tilde{\mathbf{F}}(\mathbf{k})$  and  $\mathbf{k} \times \tilde{\mathbf{F}}(\mathbf{k})$  are known. Correspondingly, the function  $\mathbf{F}(\mathbf{r})$  is determined if the functions  $\nabla \cdot \mathbf{F}(\mathbf{r})$  and  $\nabla \times \mathbf{F}(\mathbf{r})$  are assumed known. Finally, we note that the identity (2.130) may be viewed as a Lie algebraic relation for the cross-product Lie algebra. See Section 3.7.4. From Exercise 3.7.31 we know that the cross-product Lie algebra is equivalent to  $so(3)$ , and therefore (2.130) is also a property of  $so(3)$ . Finally,  $so(3)$  is a subalgebra of the Lie algebra of the three-dimensional Euclidean group. Thus, the fact that a vector field in three dimensions is specified, if its divergence and curl are known, is a consequence of the properties of the three-dimensional Euclidean group.

## 22.3 Construction of Kernels $G^n$ and $G^t$

### 22.3.1 Background

Let us apply the results of the previous section to the case of a magnetic field  $\mathbf{B}(\mathbf{r})$  in a volume  $V$  under the assumption that there are no sources in  $V$ . See (1.1) and (1.2). As stated earlier, this would be the case of interest for charged particles propagating through an evacuated beam pipe. In this circumstance we may use (2.80), (2.87), and (2.88) to write

$$\mathbf{B}(\mathbf{r}) = -\nabla\phi^n(\mathbf{r}) + \nabla \times \mathbf{A}^t(\mathbf{r}) \quad \text{for } \mathbf{r} \in V \quad (22.3.1)$$

with

$$\phi^n(\mathbf{r}) = -[1/(4\pi)] \int_S dS' \mathbf{n}' \cdot \mathbf{B}(\mathbf{r}') G(\mathbf{r}, \mathbf{r}'), \quad (22.3.2)$$

$$\mathbf{A}^t(\mathbf{r}) = -[1/(4\pi)] \int_S dS' [\mathbf{n}' \times \mathbf{B}(\mathbf{r}')] G(\mathbf{r}, \mathbf{r}'). \quad (22.3.3)$$

Here, as before, the superscripts  $n$  and  $t$  denote *normal* and *tangential* since the quantities so denoted involve normal and tangential components of  $\mathbf{B}$ .

The relations (3.1) through (3.3) could be employed if one wished to integrate Newton's equations of motion, and also find Taylor maps based on these equations, for all that would then be required is the magnetic field  $\mathbf{B}(\mathbf{r})$ . See, for example, the equations of motion (1.6.68) and (1.6.69), or (1.6.135) through (1.6.138) and (1.6.145) through (1.6.147). However, if one wishes instead to employ a Hamiltonian formulation in order to reap the benefits of symplectic symmetry, then it is necessary to have the magnetic field specified *entirely* in terms of a vector potential rather than in terms of both a scalar and vector potential as in (3.1). What we need is a vector potential  $\mathbf{A}^n(\mathbf{r})$  such that

$$\nabla \times \mathbf{A}^n(\mathbf{r}) = -\nabla\phi^n(\mathbf{r}). \quad (22.3.4)$$

Then, with the definition

$$\mathbf{A}(\mathbf{r}) = \mathbf{A}^n(\mathbf{r}) + \mathbf{A}^t(\mathbf{r}), \quad (22.3.5)$$

there would be the result

$$\mathbf{B}(\mathbf{r}) = \nabla \times \mathbf{A}(\mathbf{r}). \quad (22.3.6)$$

The construction of an  $\mathbf{A}^n(\mathbf{r})$  that satisfies (3.4) can be accomplished with the aid of the Dirac monopole vector potential. Inspection of  $\phi^n(\mathbf{r})$ , as given by (3.2), shows that it appears to arise from a distribution of magnetic monopoles described by a magnetic charge surface density spread over the surface  $S$ . Therefore, it should be possible to find an equivalent vector potential based on the vector potential for a magnetic monopole.

### 22.3.2 Construction of $G^n$ Using Half-Infinite String Monopoles

Let us make this idea precise. To do so, for simplicity, will use half-infinite string Dirac monopoles. (Fully infinite string Dirac monopoles can also be used. See Exercise 3.2.) Define  $\mathbf{B}^n$  by the rule

$$\mathbf{B}^n = -\nabla\phi^n \quad (22.3.7)$$

so that the  $\mathbf{A}^n$  that we seek satisfies

$$\nabla \times \mathbf{A}^n = \mathbf{B}^n. \quad (22.3.8)$$

Combining (3.2) and (3.7) gives the result

$$\mathbf{B}^n(\mathbf{r}) = [1/(4\pi)] \int_S dS' \mathbf{n}' \cdot \mathbf{B}(\mathbf{r}') \nabla G(\mathbf{r}, \mathbf{r}'). \quad (22.3.9)$$

From (2.89) we know that

$$\nabla G(\mathbf{r}, \mathbf{r}') = -(\mathbf{r} - \mathbf{r}')/|\mathbf{r} - \mathbf{r}'|^3. \quad (22.3.10)$$

But, from (2.72), we also have the relation

$$(4\pi/g) \nabla \times \mathbf{A}_s(\mathbf{r}; \mathbf{r}', \mathbf{m}') = -[(\mathbf{r} - \mathbf{r}')/|\mathbf{r} - \mathbf{r}'|^3]. \quad (22.3.11)$$

Define a quantity  $\mathbf{K}(\mathbf{r}; \mathbf{r}', \mathbf{m}')$  by the rule

$$\begin{aligned} \mathbf{K}(\mathbf{r}; \mathbf{r}', \mathbf{m}') &= (4\pi/g) \mathbf{A}_s(\mathbf{r}; \mathbf{r}', \mathbf{m}') \\ &= [\mathbf{m}' \times (\mathbf{r} - \mathbf{r}')]/\{|\mathbf{r} - \mathbf{r}'|[\|\mathbf{r} - \mathbf{r}'\| - \mathbf{m}' \cdot (\mathbf{r} - \mathbf{r}')]\}. \end{aligned} \quad (22.3.12)$$

See (2.63). In view of (3.10) through (3.12), we have established the key relation

$$\nabla G(\mathbf{r}, \mathbf{r}') = \nabla \times \mathbf{K}(\mathbf{r}; \mathbf{r}', \mathbf{m}'). \quad (22.3.13)$$

See Exercise 2.15 for a specific instance of this relation.

We are almost done. Insertion of (3.13) into (3.9) gives the result

$$\begin{aligned} \mathbf{B}^n &= [1/(4\pi)] \int_S dS' \mathbf{n}' \cdot \mathbf{B}(\mathbf{r}') \nabla \times \mathbf{K}(\mathbf{r}; \mathbf{r}', \mathbf{m}') \\ &= [1/(4\pi)] \nabla \times \int_S dS' \mathbf{n}' \cdot \mathbf{B}(\mathbf{r}') \mathbf{K}(\mathbf{r}; \mathbf{r}', \mathbf{m}'). \end{aligned} \quad (22.3.14)$$

Comparison of (3.8) and (3.14) shows that we may make the definition

$$\mathbf{A}^n(\mathbf{r}) = \mathbf{A}^{n1s}(\mathbf{r}) \quad (22.3.15)$$

with

$$\mathbf{A}^{n1s}(\mathbf{r}) = [1/(4\pi)] \int_S dS' \mathbf{n}' \cdot \mathbf{B}(\mathbf{r}') \mathbf{K}(\mathbf{r}; \mathbf{r}', \mathbf{m}'). \quad (22.3.16)$$

Here we have used the superscript  $n1s$  to indicate that the vector potential for *one* half-infinite Dirac *string* has been employed. Finally, we make the definitions

$$B_n(\mathbf{r}') = \mathbf{n}' \cdot \mathbf{B}(\mathbf{r}') \quad (22.3.17)$$

and

$$\begin{aligned} \mathbf{G}^{n1s}(\mathbf{r}; \mathbf{r}', \mathbf{m}') &= [1/(4\pi)] \mathbf{K}(\mathbf{r}; \mathbf{r}', \mathbf{m}') \\ &= \{\mathbf{m}'(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}')\}/\{4\pi|\mathbf{r} - \mathbf{r}'|[\|\mathbf{r} - \mathbf{r}'\| - \mathbf{m}'(\mathbf{r}') \cdot (\mathbf{r} - \mathbf{r}')]\}. \end{aligned} \quad (22.3.18)$$

Here  $\mathbf{n}'(\mathbf{r}')$  is the outward normal to  $S$  at the point  $\mathbf{r}'$ . With these definitions we have the result

$$\mathbf{A}^{n1s}(\mathbf{r}) = \int_S dS' B_n(\mathbf{r}') \mathbf{G}^{n1s}(\mathbf{r}; \mathbf{r}', \mathbf{m}'). \quad (22.3.19)$$

Together (3.15) and (3.17) through (3.19) provide a realization of the relation (1.7).

In evaluating the integral (3.19) it necessary to specify  $\mathbf{m}'(\mathbf{r}')$ , the direction of the straight half-infinite Dirac string, as  $\mathbf{r}'$  varies over  $S$ . There is considerable freedom in doing so, and different choices simply result in different gauges for  $\mathbf{A}^{n1s}(\mathbf{r})$ . There is only one major consideration. No string should intersect the volume  $V$  because it is desirable that  $\mathbf{A}^{n1s}(\mathbf{r})$  be analytic for  $\mathbf{r} \in V$ . For many geometries a convenient choice is to require that  $\mathbf{m}'(\mathbf{r}')$  be normal to and point outward from  $S$ ,

$$\mathbf{m}'(\mathbf{r}') = \mathbf{n}'(\mathbf{r}'). \quad (22.3.20)$$

Other choices may also be convenient and useful.

### 22.3.3 Discussion

Let  $\mathbf{A}^n(\mathbf{r})$  denote the  $\mathbf{A}^{n1s}(\mathbf{r})$  given by (3.19) and let  $\mathbf{G}^n(\mathbf{r}, \mathbf{r}')$  denote the  $\mathbf{G}^{n1s}(\mathbf{r}; \mathbf{r}', \mathbf{m}')$  given by (3.18). At this point we can take pleasure in observing that  $\mathbf{A}^n(\mathbf{r})$  and  $\mathbf{G}^n(\mathbf{r}, \mathbf{r}')$  have several desirable properties: First, as long as the Dirac strings for  $\mathbf{r}' \in S$  do not intersect  $V$ , the functions  $\mathbf{G}^n(\mathbf{r}, \mathbf{r}')$ , for every  $\mathbf{r}' \in S$ , are analytic in  $\mathbf{r}$  for all  $\mathbf{r} \in V$ . It follows from (3.19), under mild conditions on  $B_n(\mathbf{r}')$  for  $\mathbf{r}' \in S$ , that  $\mathbf{A}^n(\mathbf{r})$  is analytic in  $V$ . Second, since the kernel  $\mathbf{G}^n(\mathbf{r}, \mathbf{r}')$  is essentially the vector potential for a Dirac magnetic monopole, see (3.12) and (3.18), it has, for  $\mathbf{r} \in V$ , the properties

$$\nabla \cdot [\mathbf{G}^n(\mathbf{r}, \mathbf{r}')] = 0, \quad (22.3.21)$$

$$\nabla \times [\nabla \times \mathbf{G}^n(\mathbf{r}, \mathbf{r}')] = 0. \quad (22.3.22)$$

See (2.71) and (2.73). It follows from (3.19), again under mild conditions on  $B_n(\mathbf{r}')$ , that  $\mathbf{A}^n(\mathbf{r})$  has these same properties,

$$\nabla \cdot [\mathbf{A}^n(\mathbf{r})] = 0, \quad (22.3.23)$$

$$\nabla \times [\nabla \times \mathbf{A}^n(\mathbf{r})] = 0. \quad (22.3.24)$$

In practical applications, the surface values  $B_n(\mathbf{r}')$  will only be known approximately, and the integrals (3.19) may be evaluated numerically with limited precision. It is comforting to know that, nevertheless, the resulting  $\mathbf{A}^n(\mathbf{r})$  will be analytic in  $V$  and will satisfy the relations (3.23) and (3.24) exactly no matter what errors are present in the surface values  $B_n(\mathbf{r}')$  and no matter how poorly the integrals (3.19) are evaluated. All that matters is that the kernel  $\mathbf{G}^n$  be evaluated to high precision.

### 22.3.4 Construction of $\mathbf{G}^t$

What can be said about the properties of  $\mathbf{A}^t(\mathbf{r})$  as given by (3.3)? Just as is the case for  $\mathbf{A}^n(\mathbf{r})$ , we would like  $\mathbf{A}^t(\mathbf{r})$  to be analytic in  $V$  and to satisfy properties analogous to (3.23) and (3.24). That is, we desire the relations

$$\nabla \cdot [\mathbf{A}^t(\mathbf{r})] = 0, \quad (22.3.25)$$

$$\nabla \times [\nabla \times \mathbf{A}^t(\mathbf{r})] = 0, \quad (22.3.26)$$

and we would like to have them hold no matter how poorly the integral (3.3) is evaluated. As the expression (3.3) for  $\mathbf{A}^t(\mathbf{r})$  stands, this is not the case. However, we can transform (3.3) into a form that meets all our hopes.

Since, by assumption,  $\mathbf{B}(\mathbf{r}')$  is curl free for  $\mathbf{r}' \in V$ , there exists a scalar potential  $\psi(\mathbf{r}')$  such that

$$\mathbf{B}(\mathbf{r}') = +\nabla' \psi(\mathbf{r}'). \quad (22.3.27)$$

[Note, by convention, we have used a minus sign in (2.3) and a plus sign in (3.27). See also (15.2.1) and (15.2.6).] Consequently, (3.3) can be rewritten in the form

$$\mathbf{A}^t(\mathbf{r}) = -[1/(4\pi)] \int_S dS' [\mathbf{n}' \times \nabla' \psi(\mathbf{r}')] G(\mathbf{r}, \mathbf{r}'). \quad (22.3.28)$$

[Also note, as observed earlier, that a knowledge of the tangential component of  $\nabla' \psi(\mathbf{r}')$ , which is what is involved in (3.28) and is equivalent to a knowledge of  $\psi(\mathbf{r}')$  on  $S$ , is in turn equivalent to a knowledge of the tangential component of  $\mathbf{B}(\mathbf{r}')$  on  $S$  under the assumption that  $\mathbf{B}(\mathbf{r}')$  is curl free.] Next observe that there is the identity

$$[\nabla' \psi(\mathbf{r}')] G(\mathbf{r}, \mathbf{r}') = \nabla' [\psi(\mathbf{r}') G(\mathbf{r}, \mathbf{r}')] - \psi(\mathbf{r}') \nabla' G(\mathbf{r}, \mathbf{r}'). \quad (22.3.29)$$

Therefore (3.28) can also be written in the form

$$\begin{aligned} \mathbf{A}^t(\mathbf{r}) &= -[1/(4\pi)] \int_S dS' \{ \mathbf{n}' \times \nabla' [\psi(\mathbf{r}') G(\mathbf{r}, \mathbf{r}')] \} \\ &\quad + [1/(4\pi)] \int_S dS' \{ \mathbf{n}' \times [\psi(\mathbf{r}') \nabla' G(\mathbf{r}, \mathbf{r}')] \}. \end{aligned} \quad (22.3.30)$$

It can be shown that the first integral on the right side of (3.30) vanishes,

$$-[1/(4\pi)] \int_S dS' \{ \mathbf{n}' \times \nabla' [\psi(\mathbf{r}') G(\mathbf{r}, \mathbf{r}')] \} = 0. \quad (22.3.31)$$

See Exercise 3.1. Moreover, the second integral can be rewritten in the form

$$[1/(4\pi)] \int_S dS' \{ \mathbf{n}' \times [\psi(\mathbf{r}') \nabla' G(\mathbf{r}, \mathbf{r}')] \} = [1/(4\pi)] \int_S dS' \psi(\mathbf{r}') [\mathbf{n}' \times \nabla' G(\mathbf{r}, \mathbf{r}')]. \quad (22.3.32)$$

Consequently  $\mathbf{A}^t(\mathbf{r})$  can also be written in the form

$$\mathbf{A}^t(\mathbf{r}) = [1/(4\pi)] \int_S dS' \psi(\mathbf{r}') [\mathbf{n}' \times \nabla' G(\mathbf{r}, \mathbf{r}')]. \quad (22.3.33)$$

Finally, let  $\mathbf{G}^t(\mathbf{r}, \mathbf{r}')$  be the kernel

$$\mathbf{G}^t(\mathbf{r}, \mathbf{r}') = [1/(4\pi)][\mathbf{n}'(\mathbf{r}') \times \nabla' G(\mathbf{r}, \mathbf{r}')]. \quad (22.3.34)$$

With this definition,  $\mathbf{A}^t(\mathbf{r})$  takes the final form

$$\mathbf{A}^t(\mathbf{r}) = \int_S dS' \psi(\mathbf{r}') \mathbf{G}^t(\mathbf{r}, \mathbf{r}'). \quad (22.3.35)$$

And working out (3.34) explicitly gives the result

$$\mathbf{G}^t(\mathbf{r}, \mathbf{r}') = [\mathbf{n}'(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}')]/[4\pi|\mathbf{r} - \mathbf{r}'|^3]. \quad (22.3.36)$$

We have derived the relation (1.8) with  $\mathbf{G}^t$  given by (3.36).

At this point we should verify that we have achieved our desired goals. First, it is evident from (3.36) that  $\mathbf{G}^t(\mathbf{r}, \mathbf{r}')$  is analytic in the components of  $\mathbf{r}$  for  $\mathbf{r} \in V$  and  $\mathbf{r}' \in S$ . Therefore, from the representation (3.35), we see that, under mild conditions on  $\psi(\mathbf{r}')$ ,  $\mathbf{A}^t(\mathbf{r})$  will be analytic in  $V$ .

Next, let us compute  $\nabla \cdot \mathbf{G}^t(\mathbf{r}, \mathbf{r}')$  and  $\nabla \times [\nabla \times \mathbf{G}^t(\mathbf{r}, \mathbf{r}')]$ . We will see that they both vanish for  $\mathbf{r} \in V$ . Recall the vector identity

$$\nabla \cdot (\mathbf{C} \times \mathbf{D}) = \mathbf{D} \cdot (\nabla \times \mathbf{C}) - \mathbf{C} \cdot (\nabla \times \mathbf{D}). \quad (22.3.37)$$

From this identity, (2.89), (3.34), and the fact that the curl of a gradient vanishes, it follows that

$$\begin{aligned} \nabla \cdot \mathbf{G}^t(\mathbf{r}, \mathbf{r}') &= -[1/(4\pi)]\mathbf{n}'(\mathbf{r}') \cdot \{\nabla \times [\nabla' G(\mathbf{r}, \mathbf{r}')]\} \\ &= [1/(4\pi)]\mathbf{n}'(\mathbf{r}') \cdot \{\nabla \times [\nabla' G(\mathbf{r}, \mathbf{r}')]\} = 0. \end{aligned} \quad (22.3.38)$$

Also, it is evident from (2.90) and (3.34) that

$$\begin{aligned} \nabla^2 \mathbf{G}^t(\mathbf{r}, \mathbf{r}') &= [1/(4\pi)][\nabla^2][\mathbf{n}'(\mathbf{r}') \times \nabla' G(\mathbf{r}, \mathbf{r}')] \\ &= [1/(4\pi)]\{\mathbf{n}'(\mathbf{r}') \times \nabla'[\nabla^2 G(\mathbf{r}, \mathbf{r}')]\} \\ &= 0 \text{ for } \mathbf{r} \text{ within } V \text{ and } \mathbf{r}' \in S. \end{aligned} \quad (22.3.39)$$

Finally, again invoke the vector identity

$$\nabla \times (\nabla \times \mathbf{C}) = \nabla(\nabla \cdot \mathbf{C}) - \nabla^2 \mathbf{C}. \quad (22.3.40)$$

When applied to  $\mathbf{G}^t(\mathbf{r}, \mathbf{r}')$ , in view of (3.38) and (3.39), it yields the relation

$$\nabla \times [\nabla \times \mathbf{G}^t(\mathbf{r}, \mathbf{r}')] = 0 \text{ for } \mathbf{r} \text{ within } V \text{ and } \mathbf{r}' \in S. \quad (22.3.41)$$

We have seen that the kernel  $\mathbf{G}^t(\mathbf{r}, \mathbf{r}')$  satisfies the relations (3.38) and (3.41), and note that these relations are analogous to the relations (3.21) and (3.22) for  $\mathbf{G}^n(\mathbf{r}, \mathbf{r}')$ . It follows, by the same reasoning used in the case of  $\mathbf{G}^n(\mathbf{r}, \mathbf{r}')$  and  $\mathbf{A}^n(\mathbf{r})$ , that  $\mathbf{A}^t(\mathbf{r})$  satisfies the relations (3.25) and (3.26), and these relations hold exactly even in the presence of errors in the surface values  $\psi(\mathbf{r}')$  and no matter how poorly the integrals (3.35) are evaluated. Similar to the case of  $\mathbf{G}^n$ , all that matters is that the kernel  $\mathbf{G}^t$  be evaluated to high precision.

### 22.3.5 Final Discussion

Let us put together what we have learned about analyticity and “exactness”. Look at (3.5) and (3.6). Since  $\mathbf{A}^n(\mathbf{r})$  and  $\mathbf{A}^t(\mathbf{r})$  are both analytic in  $V$ ,  $\mathbf{A}(\mathbf{r})$  will be analytic in  $V$ . And since (3.23) through (3.26) hold, analogous results will hold for  $\mathbf{A}(\mathbf{r})$ ,

$$\nabla \cdot [\mathbf{A}(\mathbf{r})] = 0, \quad (22.3.42)$$

$$\nabla \times [\nabla \times \mathbf{A}(\mathbf{r})] = 0. \quad (22.3.43)$$

Moreover, analyticity and the relations (3.42) and (3.43) will still hold exactly even in the presence of errors in the surface values  $B_n$  and  $\psi$ , and no matter how poorly the relevant integrals are evaluated. Finally, in view of (3.6), the Maxwell equation

$$\nabla \cdot \mathbf{B} = 0 \quad (22.3.44)$$

will be satisfied exactly. And, in view of (3.6) and (3.43), the second Maxwell equation

$$\nabla \times \mathbf{B} = 0 \quad (22.3.45)$$

will also be satisfied exactly.

## Exercises

**22.3.1.** The purpose of this exercise is to verify the relation (3.31).

**22.3.2.** Subsection 3.2 described the construction of the kernel we called  $\mathbf{G}^{n1s}$  using the vector potential for a half-infinite string Dirac monopole. Another possibility is to use the vector potential for fully infinite string (two string) Dirac monopole to construct an analogous kernel we will call  $\mathbf{G}^{n2s}$ . The purpose of this exercise is of explore that possibility.

The vector potential  $\mathbf{A}_{2s}(\mathbf{r}; \mathbf{r}_A, \mathbf{m})$  given by (2.75) also produces the monopole field (2.55) so that there is the relation

$$(4\pi/g)\nabla \times \mathbf{A}_{2s}(\mathbf{r}; \mathbf{r}', \mathbf{m}') = -[(\mathbf{r} - \mathbf{r}')/|\mathbf{r} - \mathbf{r}'|^3]. \quad (22.3.46)$$

Now define a quantity  $\mathbf{K}(\mathbf{r}; \mathbf{r}', \mathbf{m}')$  by the rule

$$\begin{aligned} \mathbf{K}(\mathbf{r}; \mathbf{r}', \mathbf{m}') &= (4\pi/g)\mathbf{A}_{2s}(\mathbf{r}; \mathbf{r}', \mathbf{m}') \\ &= \frac{[\mathbf{m} \times (\mathbf{r} - \mathbf{r}_A)][\mathbf{m} \cdot (\mathbf{r} - \mathbf{r}_A)]}{|\mathbf{r} - \mathbf{r}_A||\mathbf{m} \times (\mathbf{r} - \mathbf{r}_A)|^2}. \end{aligned} \quad (22.3.47)$$

See (2.79). In view of (3.10), (3.21), and (3.22), we have also established the key relation

$$\nabla G(\mathbf{r}, \mathbf{r}') = \nabla \times \mathbf{K}(\mathbf{r}; \mathbf{r}', \mathbf{m}') \quad (22.3.48)$$

with  $\mathbf{K}$  now given by (3.22). Also see Exercise 2.15 for a specific instance of this relation.

Next, insertion of (3.23) into (3.9) gives the result

$$\begin{aligned}\mathbf{B}^n &= [1/(4\pi)] \int_S dS' \mathbf{n}' \cdot \mathbf{B}(\mathbf{r}') \nabla \times \mathbf{K}(\mathbf{r}; \mathbf{r}', \mathbf{m}') \\ &= [1/(4\pi)] \nabla \times \int_S dS' \mathbf{n}' \cdot \mathbf{B}(\mathbf{r}') \mathbf{K}(\mathbf{r}; \mathbf{r}', \mathbf{m}').\end{aligned}\quad (22.3.49)$$

Comparison of (3.8) and (3.24) shows that we may also make the definition

$$\mathbf{A}^n(\mathbf{r}) = \mathbf{A}^{n2s}(\mathbf{r}) \quad (22.3.50)$$

with

$$\mathbf{A}^{n2s}(\mathbf{r}) = [1/(4\pi)] \int_S dS' \mathbf{n}' \cdot \mathbf{B}(\mathbf{r}') \mathbf{K}(\mathbf{r}; \mathbf{r}', \mathbf{m}'). \quad (22.3.51)$$

Here we have used the superscript  $n2s$  to indicate that the vector potential for a fully infinite (2-sided) Dirac *string* has been employed. Finally we may write (3.26) in the form

$$\mathbf{A}^{n2s}(\mathbf{r}) = \int_S dS' B_n(\mathbf{r}') \mathbf{G}^{n2s}(\mathbf{r}; \mathbf{r}', \mathbf{m}') \quad (22.3.52)$$

where (3.17) is again employed and  $\mathbf{G}^{n2s}(\mathbf{r}; \mathbf{r}', \mathbf{m}')$  is the kernel

$$\begin{aligned}\mathbf{G}^{n2s}(\mathbf{r}; \mathbf{r}', \mathbf{m}') &= [1/(4\pi)] \mathbf{K}(\mathbf{r}; \mathbf{r}', \mathbf{m}') \\ &= \frac{[\mathbf{m}' \times (\mathbf{r} - \mathbf{r}_A)][\mathbf{m}' \cdot (\mathbf{r} - \mathbf{r}_A)]}{4\pi |\mathbf{r} - \mathbf{r}_A| |\mathbf{m}' \times (\mathbf{r} - \mathbf{r}_A)|^2}.\end{aligned}\quad (22.3.53)$$

Together (3.25), (3.27), and (3.28) provide another realization of the relation (1.7). Show that this fully infinite Dirac string kernel obeys relations analogous to (3.21) and (3.22) and therefore the relations analogous to (3.23) and (3.24) are also satisfied.

In evaluating the integral (3.28) it again necessary to specify  $\mathbf{m}'(\mathbf{r}')$ , now the direction of the straight fully infinite Dirac string, as  $\mathbf{r}'$  varies over  $S$ . As before, there is considerable freedom in doing so, and different choices simply result in different gauges for  $\mathbf{A}^{n1s}(\mathbf{r})$ . The major considerations are again that no string intersect the volume  $V$  and that the vector potential fall off rapidly in fringe-field regions. We also note that one may use  $\mathbf{A}^{n1s}(\mathbf{r})$  for some parts of  $S$  and  $\mathbf{A}^{n2s}(\mathbf{r})$  for other parts.

**22.3.3.** At the beginning of this section it was mentioned that (3.1) through (3.3) could be used to integrate Newton's equations of motion in terms of  $\mathbf{B}(\mathbf{r})$ . However the  $\mathbf{B}(\mathbf{r})$  obtained using (3.1) is not guaranteed to satisfy the Maxwell equations if there are errors in surface values and/or the integrals are not evaluated accurately. Verify that, in this regard, there is no difficulty in the use of (3.2) by showing that it is guaranteed to satisfy

$$\nabla^2 \phi^n(\mathbf{r}) = 0, \quad (22.3.54)$$

and therefore (3.43) is satisfied. Show that if (3.3) is replaced by (3.32), then (3.44) is also guaranteed.

**22.3.4.** Suppose  $\mathbf{B}(\mathbf{r})$  is source free in a volume  $V$  bounded by a surface  $S$ , as in (1.1) and (1.2), and suppose  $B_n(\mathbf{r}')$  and  $\psi(\mathbf{r}')$  are known on  $S$ . The aim of this exercise is to compute  $\mathbf{B}(\mathbf{r})$  in terms of  $B_n(\mathbf{r}')$  and  $\psi(\mathbf{r}')$  using the representation given by (1.3), (1.6) through (1.8), (1.10), and (1.11). Verify that

$$\nabla \times \mathbf{G}^n(\mathbf{r}, \mathbf{r}') = [1/(4\pi)] \nabla G(\mathbf{r}, \mathbf{r}') \quad (22.3.55)$$

from which it follows that

$$\begin{aligned} \mathbf{B}^n(\mathbf{r}) &= \nabla \times \mathbf{A}^n(\mathbf{r}) = \int_S dS' B_n(\mathbf{r}') \nabla \times \mathbf{G}^n(\mathbf{r}, \mathbf{r}') \\ &= [1/(4\pi)] \int_S dS' B_n(\mathbf{r}') \nabla G(\mathbf{r}, \mathbf{r}') \\ &= -[1/(4\pi)] \int_S dS' B_n(\mathbf{r}') (\mathbf{r} - \mathbf{r}') / |\mathbf{r} - \mathbf{r}'|^3, \end{aligned} \quad (22.3.56)$$

in accord with (3.9). Recall the vector identity

$$\nabla \times (\mathbf{C} \times \mathbf{D}) = (\mathbf{D} \cdot \nabla) \mathbf{C} + \mathbf{C} (\nabla \cdot \mathbf{D}) - (\mathbf{C} \cdot \nabla) \mathbf{D} - \mathbf{D} (\nabla \cdot \mathbf{C}). \quad (22.3.57)$$

Using (3.31) and (3.48), show that

$$\nabla \times \mathbf{G}^t(\mathbf{r}, \mathbf{r}') = -[1/(4\pi)] \mathbf{n}'(\mathbf{r}') / |\mathbf{r} - \mathbf{r}'|^3 + [3/(4\pi)] [\mathbf{n}'(\mathbf{r}') \cdot (\mathbf{r} - \mathbf{r}')] (\mathbf{r} - \mathbf{r}') / |\mathbf{r} - \mathbf{r}'|^5, \quad (22.3.58)$$

from which it follows that

$$\begin{aligned} \mathbf{B}^t(\mathbf{r}) &= \nabla \times \mathbf{A}^t(\mathbf{r}) = \int_S dS' \psi(\mathbf{r}') \nabla \times \mathbf{G}^t(\mathbf{r}, \mathbf{r}') \\ &= -[1/(4\pi)] \int_S dS' \psi(\mathbf{r}') \mathbf{n}'(\mathbf{r}') / |\mathbf{r} - \mathbf{r}'|^3 \\ &\quad + [3/(4\pi)] \int_S dS' \psi(\mathbf{r}') [\mathbf{n}'(\mathbf{r}') \cdot (\mathbf{r} - \mathbf{r}')] (\mathbf{r} - \mathbf{r}') / |\mathbf{r} - \mathbf{r}'|^5. \end{aligned} \quad (22.3.59)$$

Observe that, if we wish, we may define kernels  $\mathbf{K}^n(\mathbf{r}, \mathbf{r}')$  and  $\mathbf{K}^t(\mathbf{r}, \mathbf{r}')$  by the rules

$$\begin{aligned} \mathbf{K}^n(\mathbf{r}, \mathbf{r}') &= \nabla \times \mathbf{G}^n(\mathbf{r}, \mathbf{r}') = [1/(4\pi)] \nabla G(\mathbf{r}, \mathbf{r}') \\ &= -[1/(4\pi)] (\mathbf{r} - \mathbf{r}') / |\mathbf{r} - \mathbf{r}'|^3 \end{aligned} \quad (22.3.60)$$

and

$$\begin{aligned} \mathbf{K}^t(\mathbf{r}, \mathbf{r}') &= \nabla \times \mathbf{G}^t(\mathbf{r}, \mathbf{r}') \\ &= -[1/(4\pi)] \mathbf{n}'(\mathbf{r}') / |\mathbf{r} - \mathbf{r}'|^3 + [3/(4\pi)] [\mathbf{n}'(\mathbf{r}') \cdot (\mathbf{r} - \mathbf{r}')] (\mathbf{r} - \mathbf{r}') / |\mathbf{r} - \mathbf{r}'|^5. \end{aligned} \quad (22.3.61)$$

With the aid of these definitions, (3.47) and (3.50) take the form

$$\mathbf{B}^n(\mathbf{r}) = \int_S dS' B_n(\mathbf{r}') \mathbf{K}^n(\mathbf{r}, \mathbf{r}') \quad (22.3.62)$$

and

$$\mathbf{B}^t(\mathbf{r}) = \int_S dS' \psi(\mathbf{r}') \mathbf{K}^t(\mathbf{r}, \mathbf{r}'). \quad (22.3.63)$$

Finally, write

$$\mathbf{B}(\mathbf{r}) = \mathbf{B}^n(\mathbf{r}) + \mathbf{B}^t(\mathbf{r}). \quad (22.3.64)$$

Show that, for fixed  $\mathbf{r}'$ ,  $\mathbf{K}^n(\mathbf{r}, \mathbf{r}')$  falls off as  $1/r^2$  for large  $r = |\mathbf{r}|$  and  $\mathbf{K}^t(\mathbf{r}, \mathbf{r}')$  falls off as  $1/r^3$ .

**22.3.5.** Show that the Cartesian components of  $\mathbf{A}(\mathbf{r})$ , as given by (3.5), (3.17), and (3.32), are harmonic functions,

$$\nabla^2 \mathbf{A}(\mathbf{r}) = 0. \quad (22.3.65)$$

**22.3.6.** According to (3.12) and (3.18),  $\mathbf{G}^{n1s}$  and  $\mathbf{A}_s$  are proportional. Consequently, Exercise 2.15 provides a description of the direction of the vector  $\mathbf{G}^n$  and, and hence the associated  $\mathbf{A}^n$  produced using (1.7). The purpose of this exercise is to determine the direction of  $\mathbf{G}^t$ , and hence the associated  $\mathbf{A}^t$  produced using (1.8). Consider, for purposes of calculation, a small patch of surface  $\Delta S'$  located at the point

$$\mathbf{r}' = d\mathbf{e}_y \text{ with } d > 0 \quad (22.3.66)$$

and whose normal is given by the relation

$$\mathbf{n}'(\mathbf{r}') = \mathbf{e}_y. \quad (22.3.67)$$

Then, from (3.44), verify that there is the result

$$\mathbf{G}^t(\mathbf{r}, d\mathbf{e}_y) = [\mathbf{e}_y \times (\mathbf{r} - d\mathbf{e}_y)]/[4\pi|\mathbf{r} - d\mathbf{e}_y|^3]. \quad (22.3.68)$$

Also, verify the relation

$$\mathbf{e}_y \times (\mathbf{r} - d\mathbf{e}_y) = -x\mathbf{e}_z + z\mathbf{e}_x. \quad (22.3.69)$$

Show, therefore, that in this case,  $\mathbf{G}^t$  is given by the relation

$$\mathbf{G}^t = [1/(4\pi)](-x\mathbf{e}_z + z\mathbf{e}_x)/[x^2 + (y - d)^2 + z^2]^{3/2}. \quad (22.3.70)$$

Recall that in cylindrical coordinates there is the relation

$$\mathbf{r} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z = \rho \cos \phi \mathbf{e}_x + \rho \sin \phi \mathbf{e}_y + z\mathbf{e}_z. \quad (22.3.71)$$

See (13.2.3) and (13.2.4). Consequently there is the relation

$$\partial \mathbf{r} / \partial \phi = -\rho \sin \phi \mathbf{e}_x + \rho \cos \phi \mathbf{e}_y = -y\mathbf{e}_x + x\mathbf{e}_y. \quad (22.3.72)$$

We also know that

$$\mathbf{e}_\phi = (\partial \mathbf{r} / \partial \phi) / |\partial \mathbf{r} / \partial \phi| = (-y\mathbf{e}_x + x\mathbf{e}_y) / \rho. \quad (22.3.73)$$

In the case of cylindrical coordinates the vector  $\mathbf{e}_\phi$  circles around the  $z$  axis. Verify, by geometric analogy, that the vector  $\mathbf{G}^t$  given by (3.70) circles about the  $y$  axis. And, according to (3.67), this axis is the  $\mathbf{n}'(\mathbf{r}')$  axis.

## 22.4 Expansion of Kernels

### 22.4.1 Our Goal

### 22.4.2 Binomial Theorem

Since Newton's discovery we have known the binomial expansion

$$(1 + x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k. \quad (22.4.1)$$

Moreover, the binomial coefficients obey the recursion relations

$$\binom{\alpha}{0} = 1, \quad (22.4.2)$$

$$\binom{\alpha}{k+1} = [(\alpha - k)/(k + 1)] \binom{\alpha}{k}, \quad (22.4.3)$$

and therefore can easily be computed sequentially.

### 22.4.3 Expansion of $\mathbf{G}^t(\mathbf{r}, \mathbf{r}')$

### 22.4.4 Expansion of $\mathbf{G}^n(\mathbf{r}, \mathbf{r}')$



# Bibliography

## Electromagnetism and Magnetic Monopoles

- [1] J.D. Jackson, *Classical Electrodynamics*, John Wiley (1999).
- [2] J. Reitz and F. Milford, *Foundations of Electromagnetic Theory*, Second Edition, Addison-Wesley (1967).
- [3] R. Plonsey and R. Collin, *Principles and Applications of Electromagnetic Fields*, McGraw Hill (1961).
- [4] S. Russenschuck, *Field Computation for Accelerator Magnets: Analytical and Numerical Methods for Electromagnetic Design and Optimization*, Wiley (2010).
- [5] Ya. Shnir, *Magnetic Monopoles*, Springer (2005).
- [6] P. Dirac, “Quantized singularities in the electromagnetic field”, *Proceedings of the Royal Society of London A133*, pp. 60-72 (1931).
- [7] The idea of constructing kernels using Helmholtz’s theorem and Dirac’s magnetic monopole vector potential is due to Peter Walstrom.

## General References

- [8] C. Mitchell, “Calculation of Realistic Charged-Particle Transfer Maps”, University of Maryland Physics Department Ph.D. Thesis (2007).



# Chapter 23

## Realistic Transfer Maps for General Curved Beam-Line Elements: Exact Monopole Doublet Results

How do the surface methods for curved elements, described by relations (22.1.3) through (22.1.8) of Section 22.1 with the  $\mathbf{G}^n$  and  $\mathbf{G}^t$  found in Section 22.3, work in practice? The purpose of this chapter and the next two is to explore this question.

This chapter finds exact results for the case of a monopole doublet. Chapter 24 finds bent box monopole doublet results. Comparison of the results of these two chapters provides a benchmark for the accuracy of surface methods for curved beam-line elements. Chapter 25 applies surface methods to the case of a realistic storage-ring dipole.

### 23.1 Magnetic Monopole Doublet Vector Potential

Consider the monopole doublet magnetic field described by Equations (15.8.1) through (15.8.6) and Figures 15.8.1 through 15.8.5 of Section 15.8. In order to set up the Hamiltonian that will describe particle motion in this field, we need a vector potential  $\mathbf{A}(\mathbf{r})$  such that

$$\nabla \times \mathbf{A}(\mathbf{r}) = \nabla\psi(\mathbf{r}) \quad (23.1.1)$$

with  $\psi$  given by (15.8.3). For this purpose we will employ the string vector potential given by (22.2.63). The desired vector potential will describe two Dirac magnetic monopoles of opposite sign. The upper, with strength  $4\pi g$ , will be situated at  $\mathbf{r}^+ = a\mathbf{e}_y$ , and will be taken to have a half-infinite string extending from  $\mathbf{r}^+$  to infinity along the positive  $y$  axis. The lower, with strength  $-4\pi g$ , will be situated at  $\mathbf{r}^- = -a\mathbf{e}_y$ , and will be taken to have a half-infinite string extending from  $\mathbf{r}^-$  to infinity along the negative  $y$  axis. See (22.2.63) and Figure 1.1. Thus,  $\mathbf{A}(\mathbf{r})$  will be given by the relation

$$\mathbf{A}(\mathbf{r}) = \mathbf{A}^+(\mathbf{r}) + \mathbf{A}^-(\mathbf{r}) \quad (23.1.2)$$

with

$$\begin{aligned}\mathbf{A}^+(\mathbf{r}) &= -\mathbf{A}_s(\mathbf{r}; \mathbf{r}^+ \rightarrow +\infty \mathbf{e}_y) \\ &= -g[\mathbf{e}_y \times (\mathbf{r} - a\mathbf{e}_y)] / \{|\mathbf{r} - a\mathbf{e}_y| [|\mathbf{r} - a\mathbf{e}_y| - \mathbf{e}_y \cdot (\mathbf{r} - a\mathbf{e}_y)]\} \\ &= -g(\mathbf{e}_y \times \mathbf{r}) / \{|\mathbf{r} - a\mathbf{e}_y| [|\mathbf{r} - a\mathbf{e}_y| - y + a]\},\end{aligned}\quad (23.1.3)$$

and

$$\begin{aligned}\mathbf{A}^-(\mathbf{r}) &= -(-1)\mathbf{A}_s(\mathbf{r}; \mathbf{r}^- \rightarrow -\infty \mathbf{e}_y) \\ &= -(-g)[-\mathbf{e}_y \times (\mathbf{r} + a\mathbf{e}_y)] / \{|\mathbf{r} + a\mathbf{e}_y| [|\mathbf{r} + a\mathbf{e}_y| + \mathbf{e}_y \cdot (\mathbf{r} + a\mathbf{e}_y)]\} \\ &= -g(\mathbf{e}_y \times \mathbf{r}) / \{|\mathbf{r} + a\mathbf{e}_y| [|\mathbf{r} + a\mathbf{e}_y| + y + a]\}.\end{aligned}\quad (23.1.4)$$

Here we have used the notation  $\mathbf{r}^+ \rightarrow +\infty \mathbf{e}_y$  to denote a string extending from  $\mathbf{r}^+$  to infinity along the positive  $y$  axis, and have used the notation  $\mathbf{r}^- \rightarrow -\infty \mathbf{e}_y$  to denote a string extending from  $\mathbf{r}^-$  to infinity along the negative  $y$  axis. Also, as in Section 15.9.1, we have taken the monopoles to have strengths  $\pm 4\pi g$  so as to avoid the appearance of  $4\pi$  factors in subsequent formulas such as (1.3) and (1.4).

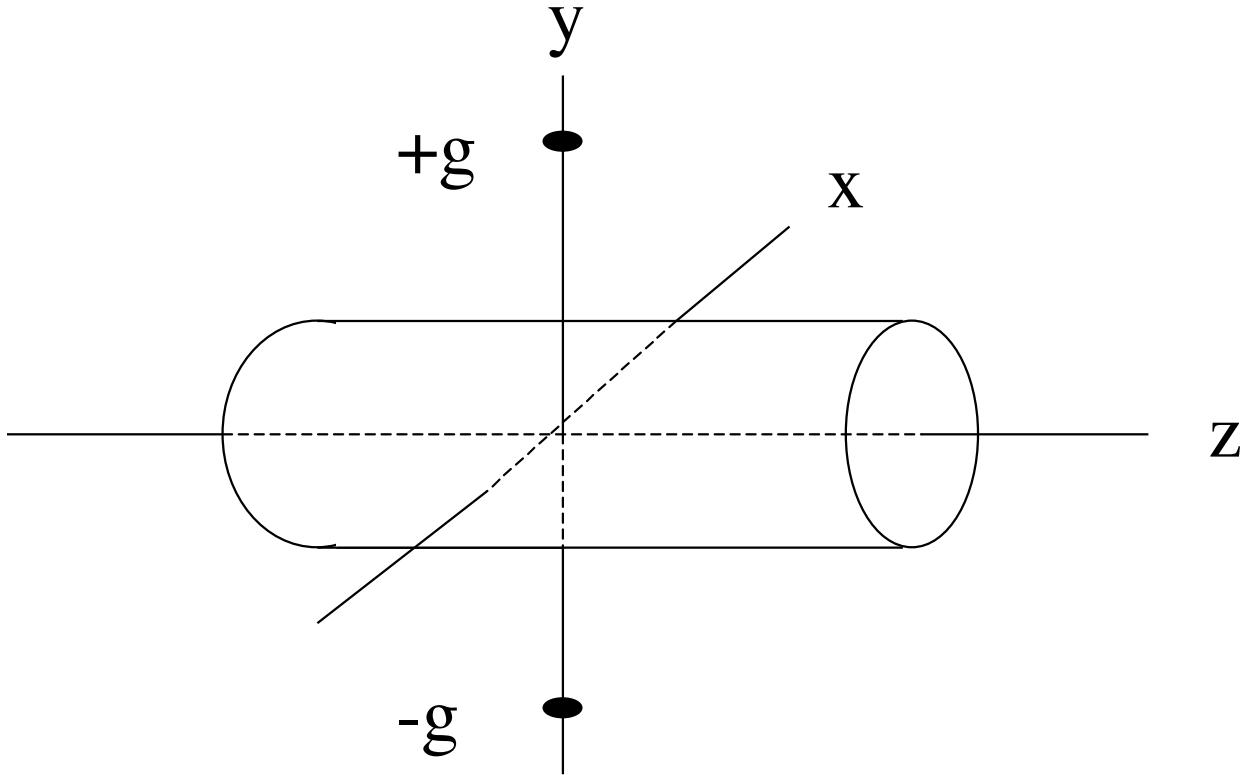


Figure 23.1.1: (Place holder) A monopole doublet consisting of two magnetic monopoles of equal and opposite sign placed on the  $y$  axis and centered on the origin. Also shown are half-infinite Dirac strings extending from the  $+g$  monopole along the positive  $y$  axis and from the  $-g$  monopole along the negative  $y$  axis.

Note that

$$\mathbf{e}_y \times \mathbf{r} = -x\mathbf{e}_z + z\mathbf{e}_x. \quad (23.1.5)$$

Therefore, in terms of components, the relation (1.2) takes the explicit form

$$\begin{aligned} A_x(x, y, z) &= -\frac{gz}{[x^2 + (y - a)^2 + z^2]^{1/2} \{ [x^2 + (y - a)^2 + z^2]^{1/2} - y + a \}} \\ &\quad - \frac{gz}{[x^2 + (y + a)^2 + z^2]^{1/2} \{ [x^2 + (y + a)^2 + z^2]^{1/2} + y + a \}}, \end{aligned} \quad (23.1.6)$$

$$A_y(x, y, z) = 0, \quad (23.1.7)$$

$$\begin{aligned} A_z(x, y, z) &= +\frac{gx}{[x^2 + (y - a)^2 + z^2]^{1/2} \{ [x^2 + (y - a)^2 + z^2]^{1/2} - y + a \}} \\ &\quad + \frac{gx}{[x^2 + (y + a)^2 + z^2]^{1/2} \{ [x^2 + (y + a)^2 + z^2]^{1/2} + y + a \}}. \end{aligned} \quad (23.1.8)$$

Examination of (1.6) reveals that  $A_x(x, y, z)$  is even in  $x$  and  $y$ , and odd in  $z$ ; and examination of (1.8) shows that  $A_z(x, y, z)$  is odd in  $x$  and even in  $y$  and  $z$ .

From (22.1.3) and (1.6) through (1.8), and with some algebraic effort, it can be checked that

$$\begin{aligned} B_x &= \partial_y A_z - \partial_z A_y = \partial_y A_z = \\ &= gx[x^2 + (y - a)^2 + z^2]^{-3/2} - gx[x^2 + (y + a)^2 + z^2]^{-3/2}, \end{aligned} \quad (23.1.9)$$

$$\begin{aligned} B_y &= \partial_z A_x - \partial_x A_z \\ &= g(y - a)[x^2 + (y - a)^2 + z^2]^{-3/2} - g(y + a)[x^2 + (y + a)^2 + z^2]^{-3/2}, \end{aligned} \quad (23.1.10)$$

$$\begin{aligned} B_z &= \partial_x A_y - \partial_y A_x = -\partial_y A_x = \\ &= gz[x^2 + (y - a)^2 + z^2]^{-3/2} - gz[x^2 + (y + a)^2 + z^2]^{-3/2}, \end{aligned} \quad (23.1.11)$$

in agreement with (15.9.4) through (15.9.6).

Figure 1.2 displays the quantity  $A_x(x, y, z)$  as a function of  $z$  along the line  $x = y = 0$ . Here, for convenience in plotting and as done before, we have used the values

$$a = 2.5 \text{ cm} = .025 \text{ m} \quad (23.1.12)$$

and

$$g = 1 \text{ Tesla (cm)}^2 = 1 \times 10^{-4} \text{ Tesla m}^2. \quad (23.1.13)$$

Evidently  $A_x$  along this line falls off very slowly with increasing  $|z|$ . Indeed, for large  $|z|$ , we see from (1.6) that  $A_x(x, y, z)$  has the asymptotic behavior

$$A_x(x, y, z) \simeq -2g/z. \quad (23.1.14)$$

Figure 1.3 displays  $A_z$  as a function of  $z$  along the line given by the conditions  $x = -1/2$  cm and  $y = 0$ . It falls off somewhat more rapidly. From (1.8) we see that, for large  $|z|$ , it has the asymptotic behavior

$$A_z(x, y, z) \simeq 2gx/z^2. \quad (23.1.15)$$

Neither  $A_x$  nor  $A_z$  falls off as rapidly as  $B_y(0, 0, z)$ , which falls off as  $1/|z|^3$  for large  $|z|$ . See Section 15.8.1 and Figure 15.8.3. We also note that if a cylindrical harmonic expansion is employed as in Section 16.3, which involves the use of on-axis gradients, then all components of the associated vector potential fall off as  $1/|z|^3$  for large  $|z|$ . What we are observing is that the asymptotic behavior of the vector potential depends on the choice of gauge. Why not, then, employ a cylindrical harmonic expansion for which the asymptotic behavior of the associated vector potential is optimal? The reason is that we wish to treat cases for which the design orbit is significantly bent so that on-axis expansions are not applicable.

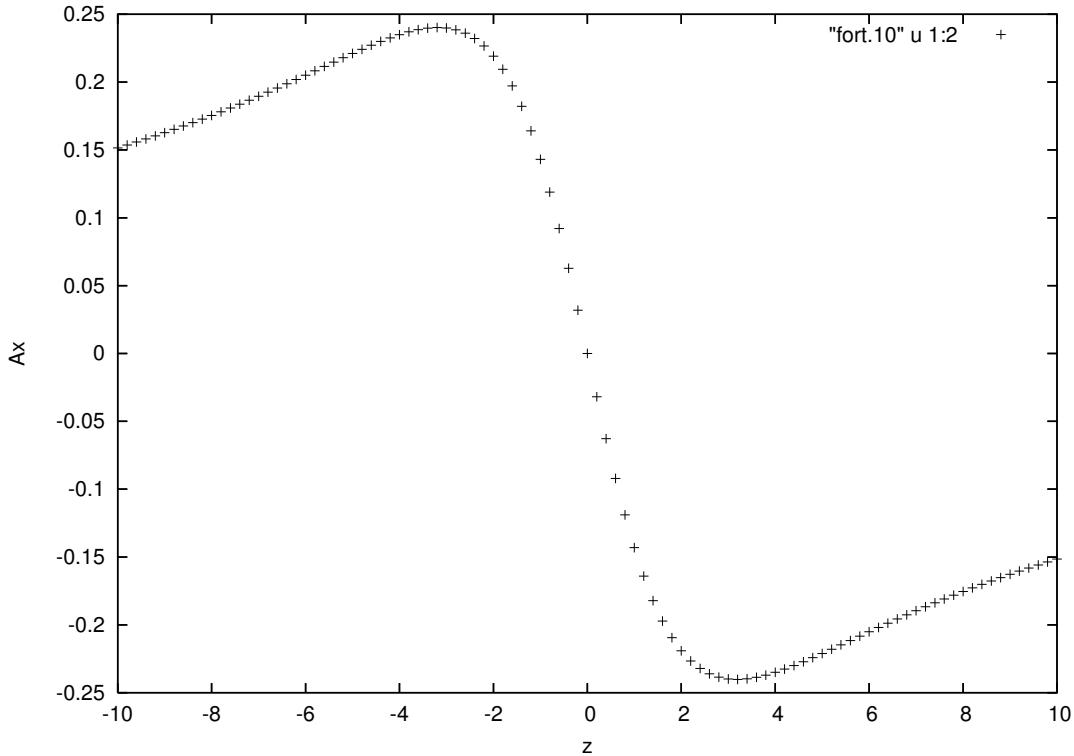


Figure 23.1.2: Behavior of  $A_x$  on the line  $(0, 0, z)$ . The quantity  $z$  is in cm.

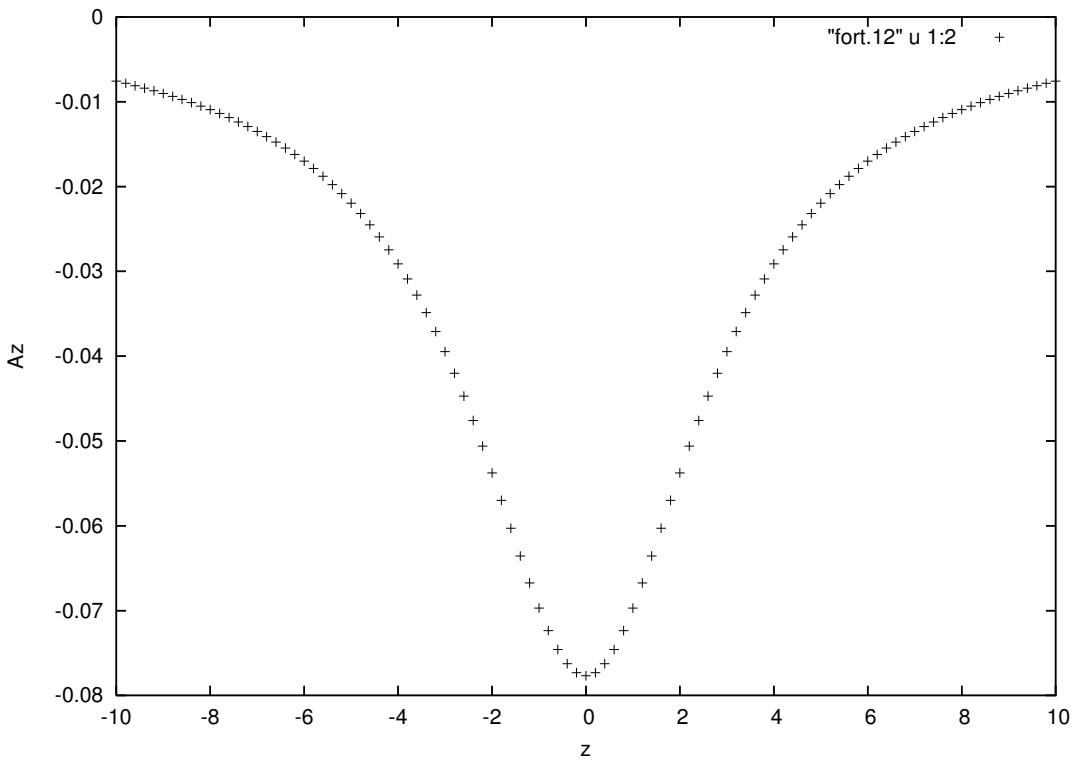


Figure 23.1.3: Behavior of  $A_z$  on the line  $(-1/2, 0, z)$ . The quantity  $z$  is in cm.

## Exercises

**23.1.1.** Using (1.6) through (1.8), verify (1.9) through (1.11).

## 23.2 Selection of Hamiltonian and Scaled Variables

To compute orbits (and maps) it is convenient to use  $z$  as the independent variable. In this case, and for the vector potential given by (1.6) through (1.8), the Hamiltonian becomes

$$K = -[p_t^2/c^2 - m^2c^2 - (p_x - qA_x)^2 - p_y^2]^{1/2} - qA_z. \quad (23.2.1)$$

See (1.6.16). Let  $\beta$  and  $\gamma$  be the usual relativistic factors defined by

$$\beta = v/c, \quad (23.2.2)$$

$$\gamma = (1 - \beta^2)^{-1/2} \quad (23.2.3)$$

where  $v$  is the particle velocity. Then the magnitude of the mechanical momentum is given by the relation

$$p = \gamma mv = \gamma\beta mc \quad (23.2.4)$$

and the quantity  $p_t$  has the value

$$p_t = -(m^2c^4 + p^2c^2)^{1/2} = -\gamma mc^2. \quad (23.2.5)$$

Since  $K$  is independent of  $t$ , the quantities  $p_t$  and  $p$  will be constants of motion. Finally, let  $p^0$  be the momentum for the design orbit.

At this point it is useful to introduce dimensionless/scaled variables by the rules

$$\hat{x} = x/\ell, \quad (23.2.6)$$

$$\hat{y} = y/\ell, \quad (23.2.7)$$

$$\tau = ct/\ell, \quad (23.2.8)$$

$$\hat{p}_x = p_x/p^0, \quad (23.2.9)$$

$$\hat{p}_y = p_y/p^0, \quad (23.2.10)$$

$$p_\tau = p_t/(p^0 c). \quad (23.2.11)$$

Here  $\ell$  is a convenient scale length, and is not to be confused with the path length introduced in Exercise 1.7.8.

The dimensionless variables satisfy the Poisson bracket rules

$$[\hat{x}, \hat{p}_x] = [\hat{y}, \hat{p}_y] = [\tau, p_\tau] = 1/(\ell p^0). \quad (23.2.12)$$

From now on we will redefine their Poisson brackets so that conjugate variables again have unity Poisson brackets. This is permissible providing the Hamiltonian  $K$  is replaced by a properly scaled new Hamiltonian  $H$  given by the relation

$$\begin{aligned} H &= -[1/(\ell p^0)]\{[(p^0 c)^2 p_\tau^2/c^2 - m^2 c^2 - (p^0 \hat{p}_x - q A_x)^2 - (p^0)^2 \hat{p}_y^2]^{1/2} + q A_z\} \\ &= -(1/\ell)\{p_\tau^2 - (mc/p^0)^2 - (\hat{p}_x - \mathcal{A}_x)^2 - \hat{p}_y^2\}^{1/2} + \mathcal{A}_z \end{aligned} \quad (23.2.13)$$

where

$$\mathcal{A}_x(\hat{x}, \hat{y}, z) = (q/p^0)A_x(\ell \hat{x}, \ell \hat{y}, z), \quad (23.2.14)$$

$$\mathcal{A}_z(\hat{x}, \hat{y}, z) = (q/p^0)A_z(\ell \hat{x}, \ell \hat{y}, z). \quad (23.2.15)$$

(See Appendix D.)

### 23.3 Design Orbit and Fields

How should we choose a design orbit? We would like it to lie in the  $y = 0$  plane, to pass through the origin, and to be symmetric about  $z = 0$ . How do we know that it is possible for there to be an orbit that lies in the  $y = 0$  plane? Let us evaluate (1.9) through (1.11) to find  $\mathbf{B}$  when  $y = 0$ . So doing gives the results

$$B_x(x, 0, z) = 0, \quad (23.3.1)$$

$$B_y(x, 0, z) = -2ga\{[x^2 + a^2 + z^2]^{-3/2}, \quad (23.3.2)$$

$$B_z(x, 0, z) = 0. \quad (23.3.3)$$

We see that if a particle is initially in the  $y = 0$  plane and moving with a velocity in this plane, then the Lorentz force is also in this plane: thus there is no force acting to accelerate the particle out of this plane and it must remain in this plane. Next observe from (1.6) through (1.8) that  $\mathbf{A}(\mathbf{r})$  vanishes at the origin,

$$\mathbf{A}(0, 0, 0) = 0. \quad (23.3.4)$$

Therefore the canonical and mechanical momenta agree at the origin. See (1.5.30). Consequently, and by symmetry, one way to achieve the desired design orbit is to select, for  $z = 0$ , the initial conditions

$$\hat{x} = \hat{y} = \tau = 0, \quad (23.3.5)$$

$$\hat{p}_x = \hat{p}_y = 0, \quad (23.3.6)$$

and then integrate both backward and forward in  $z$  to obtain the complete orbit. Note that for a orbit lying in the  $y = 0$  plane the relations

$$\hat{y} = \hat{p}_y = 0 \quad (23.3.7)$$

hold for all  $z$ .

What remains is to select the values of  $p_\tau$  and  $p^0$ . From (2.4) we see that for the design orbit there is the relation

$$p^0 = \gamma^0 \beta^0 mc. \quad (23.3.8)$$

From (2.5) we see that the energy on this orbit will be given by

$$p_t^0 = -\gamma^0 mc^2. \quad (23.3.9)$$

Therefore, on this orbit  $p_\tau$  has the value

$$p_\tau = p_\tau^0 = p_t^0 / (p^0 c) = -\gamma^0 mc^2 / (\gamma^0 \beta^0 mc) = -1/\beta^0. \quad (23.3.10)$$

And, with regard to the ingredients in (2.13), we see that

$$(p_\tau^0)^2 - (mc/p^0)^2 = (1/\beta^0)^2 - [1/(\gamma^0 \beta^0)^2] = 1. \quad (23.3.11)$$

Therefore, on the design orbit,  $H$  becomes

$$H = -(1/\ell) \{ [1 - (\hat{p}_x - \mathcal{A}_x)^2 - \hat{p}_y^2]^{1/2} + \mathcal{A}_z \}. \quad (23.3.12)$$

Finally, we should select (by trial and error) the quantity  $p^0$ , which now appears only in (2.14) and (2.15), in such a way that, for the specified values of  $a$  and  $g$ , the design orbit has some desired bend angle  $\phi_{\text{bend}}$ . For purposes of illustration, we will require that  $\phi_{\text{bend}}$  for an electron be approximately  $30^\circ$ .

Let us work out the spatial equations of motion associated with  $H$  as given by (3.12). For convenience we will take the scale length to have the value

$$\ell = 1 \text{ cm.} \quad (23.3.13)$$

We find the results

$$\hat{x}' = \partial H / \partial \hat{p}_x = (\hat{p}_x - \mathcal{A}_x) / [1 - (\hat{p}_x - \mathcal{A}_x)^2 - \hat{p}_y^2]^{1/2}, \quad (23.3.14)$$

$$\hat{y}' = \partial H / \partial \hat{p}_y = \hat{p}_y / [1 - (\hat{p}_x - \mathcal{A}_x)^2 - \hat{p}_y^2]^{1/2}, \quad (23.3.15)$$

$$\hat{p}'_x = -\partial H / \partial \hat{x} = (\partial \mathcal{A}_x / \partial \hat{x})(\hat{p}_x - \mathcal{A}_x) / [1 - (\hat{p}_x - \mathcal{A}_x)^2 - \hat{p}_y^2]^{1/2} + (\partial \mathcal{A}_z / \partial \hat{x}), \quad (23.3.16)$$

$$\hat{p}'_y = -\partial H / \partial \hat{y} = (\partial \mathcal{A}_x / \partial \hat{y})(\hat{p}_x - \mathcal{A}_x) / [1 - (\hat{p}_x - \mathcal{A}_x)^2 - \hat{p}_y^2]^{1/2} + (\partial \mathcal{A}_z / \partial \hat{y}). \quad (23.3.17)$$

Here a prime denotes  $d/dz$ .

We have already remarked that for this vector potential  $\mathcal{A}_x$  and  $\mathcal{A}_z$  are even in  $y$ , and therefore we may write

$$\mathcal{A}_x(\hat{x}, -\hat{y}, z) = \mathcal{A}_z(\hat{x}, \hat{y}, z), \quad (23.3.18)$$

$$\mathcal{A}_z(\hat{x}, -\hat{y}, z) = \mathcal{A}_z(\hat{x}, \hat{y}, z). \quad (23.3.19)$$

See (1.6) through (1.8). It follows that

$$[\partial \mathcal{A}_x(\hat{x}, \hat{y}, z) / \partial \hat{y}]|_{\hat{y}=0} = [\partial \mathcal{A}_z(\hat{x}, \hat{y}, z) / \partial \hat{y}]|_{\hat{y}=0} = 0. \quad (23.3.20)$$

Upon combining the information provided by (3.20) with the  $(\hat{y}, \hat{p}_y)$  equations of motion (3.15) and (3.17) we see that there are orbits, one of which will be the design orbit, that satisfy the conditions (3.7) for all  $z$ . Moreover, on these orbits, the  $(\hat{x}, \hat{p}_x)$  equations of motion take the form

$$\hat{x}' = (\hat{p}_x - \mathcal{A}_x) / [1 - (\hat{p}_x - \mathcal{A}_x)^2]^{1/2}, \quad (23.3.21)$$

$$\hat{p}'_x = (\partial \mathcal{A}_x / \partial \hat{x})(\hat{p}_x - \mathcal{A}_x) / [1 - (\hat{p}_x - \mathcal{A}_x)^2]^{1/2} + (\partial \mathcal{A}_z / \partial \hat{x}), \quad (23.3.22)$$

and it is only this pair we need integrate. For the record we note that, on the design orbit so that (3.7) holds, there are the relations

$$\mathcal{A}_x|_{\hat{y}=0} = \mathcal{A}_x(\hat{x}, 0, z) = -\frac{(2gq/p^0)z}{(\hat{x}^2 + a^2 + z^2)^{1/2}[(\hat{x}^2 + a^2 + z^2)^{1/2} + a]}, \quad (23.3.23)$$

$$\mathcal{A}_z|_{\hat{y}=0} = \mathcal{A}_z(\hat{x}, 0, z) = +\frac{(2gq/p^0)\hat{x}}{(\hat{x}^2 + a^2 + z^2)^{1/2}[(\hat{x}^2 + a^2 + z^2)^{1/2} + a]}, \quad (23.3.24)$$

$$(\partial \mathcal{A}_x / \partial \hat{x})|_{\hat{y}=0} = +\frac{(2gq/p^0)(\hat{x}z)[2(\hat{x}^2 + a^2 + z^2)^{1/2} + a]}{(\hat{x}^2 + a^2 + z^2)^{3/2}[(\hat{x}^2 + a^2 + z^2)^{1/2} + a]^2}, \quad (23.3.25)$$

$$\begin{aligned} (\partial \mathcal{A}_z / \partial \hat{x})|_{\hat{y}=0} &= +\frac{(2gq/p^0)}{(\hat{x}^2 + a^2 + z^2)^{1/2}[(\hat{x}^2 + a^2 + z^2)^{1/2} + a]} \\ &\quad - \frac{(2gq/p^0)(\hat{x}^2)[2(\hat{x}^2 + a^2 + z^2)^{1/2} + a]}{(\hat{x}^2 + a^2 + z^2)^{3/2}[(\hat{x}^2 + a^2 + z^2)^{1/2} + a]^2}. \end{aligned} \quad (23.3.26)$$

See (1.6) through (1.8), (2.14), and (2.15). Finally, imposing the initial conditions (3.5) and (3.6) and a suitable value for  $p^0$  yield the design orbit.

Figures 3.1 and 3.2 display, in canonical coordinates, the design orbit that results from integrating the equations of motion (3.21) and (3.22), with the initial conditions (3.5) and (3.6), when the design momentum  $p^0$  is selected to satisfy the relation

$$qg/p^0 = -.3291331 \text{ cm.} \quad (23.3.27)$$

(Recall that  $q < 0$  for an electron.) Figure 3.1 shows the spatial part of the design orbit. Figure 3.2 displays the canonical momentum  $\hat{p}_x$  on this orbit. Note that the canonical momentum depends on the choice of gauge.

To provide further insight, Figure 3.3 displays the *mechanical* scaled momentum  $\hat{p}_x^{\text{mech}}$  related to the canonical momentum by the rule

$$\hat{p}_x^{\text{mech}} = \hat{p}_x - \mathcal{A}_x. \quad (23.3.28)$$

Note that the mechanical momentum does not depend on the choice of gauge. Finally observe that, with the aid of (3.28), the relation (3.21) can be rewritten in the form

$$\hat{x}' = \hat{p}_x^{\text{mech}} / [1 - (\hat{p}_x^{\text{mech}})^2]^{1/2}. \quad (23.3.29)$$

Figure 3.4 displays  $\hat{x}'(z)$ . We also reiterate that  $\hat{y} = 0$  and  $\hat{p}_y = 0$  on a design orbit.

On this design orbit there are, for  $z = \mp 20$  cm, the end values

$$\hat{x}(\mp 20) = -4.75976218485406, \quad (23.3.30)$$

$$\hat{p}_x(\mp 20) = \pm .?, \quad (23.3.31)$$

$$\hat{p}_x^{\text{mech}}(\mp 20) = \pm .2588190579162489, \quad (23.3.32)$$

$$\hat{x}'(\mp 20) = \pm .2679492066493081. \quad (23.3.33)$$

Correspondingly, we find that over the interval  $z \in [-20, 20]$  the bend angle has the value

$$\phi_{\text{bend}} = 30.000001520142693^\circ. \quad (23.3.34)$$

See Exercise 3.2.

For the design orbit the magnetic rigidity has the value

$$\begin{aligned} p^0/|q| &= g/(.3291331 \text{ cm}) = 1 \text{ Tesla (cm)}^2/(.3291331 \text{ cm}) \\ &= 3.0382845116458963 \text{ Tesla cm} \\ &= 3.0382845116458963 \times 10^{-2} \text{ Tesla m.} \end{aligned} \quad (23.3.35)$$

See (1.6.116). Correspondingly, we find the values

$$p^0 = 9.108547817 \text{ MeV/c,} \quad (23.3.36)$$

$$p_t^0 = -9.122870347 \text{ MeV,} \quad (23.3.37)$$

$$\begin{aligned} \text{kinetic energy} &= -p_t^0 - m_e c^2 \\ &= (\gamma^0 - 1)m_e c^2 \\ &= 8.611871287313742 \text{ MeV,} \end{aligned} \quad (23.3.38)$$

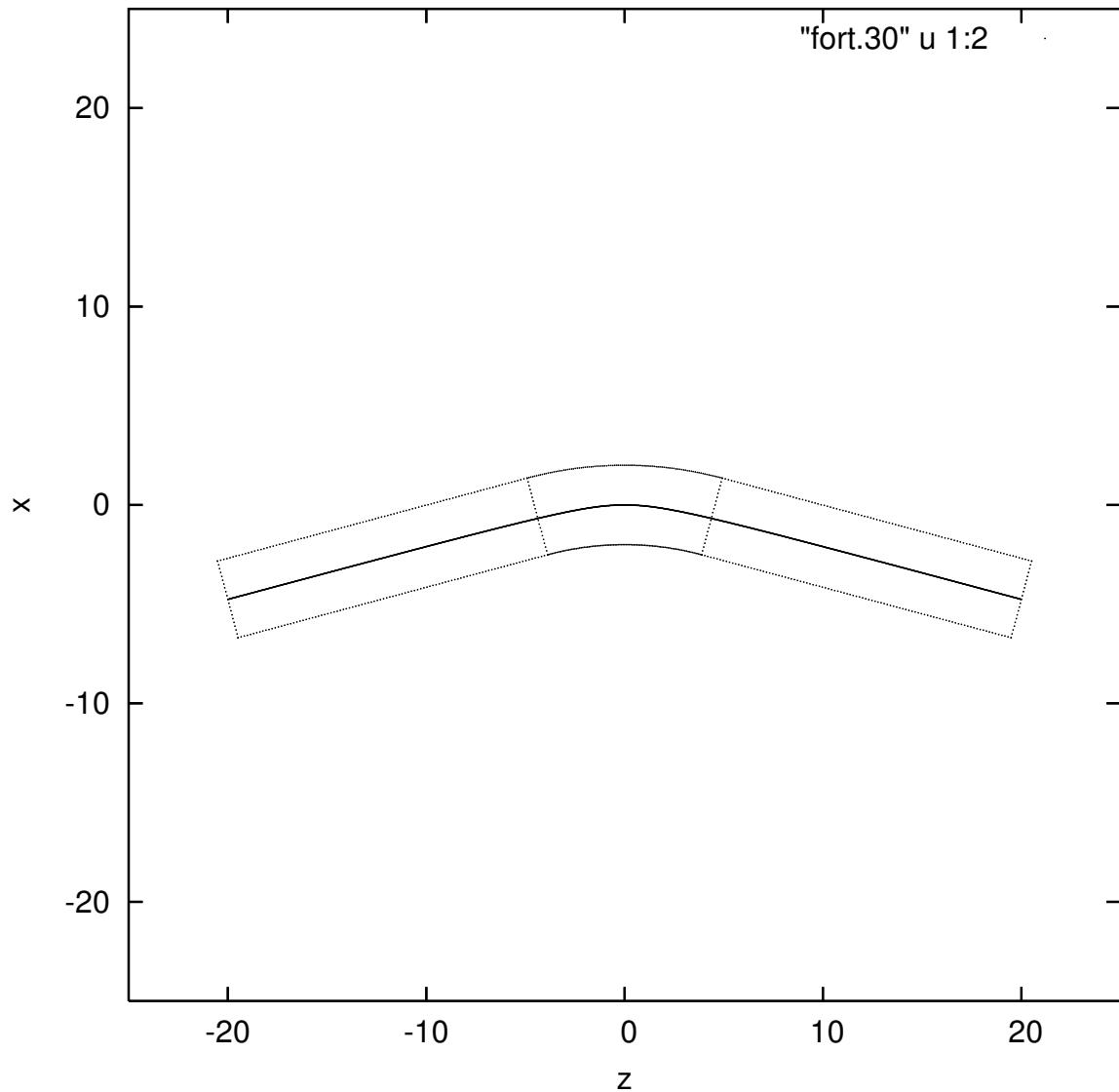


Figure 23.3.1: Design orbit  $x(z) = \hat{x}(z)$ . Also shown is a surrounding bent box with straight end legs. It will be employed in Chapter 24. The center curve is the design orbit. The outer curves are the boundary of the surrounding bent box with straight end legs. For ease of visualization, the seams between the bent box and the straight end legs are also shown. The quantities  $x$  and  $z$  are in cm.

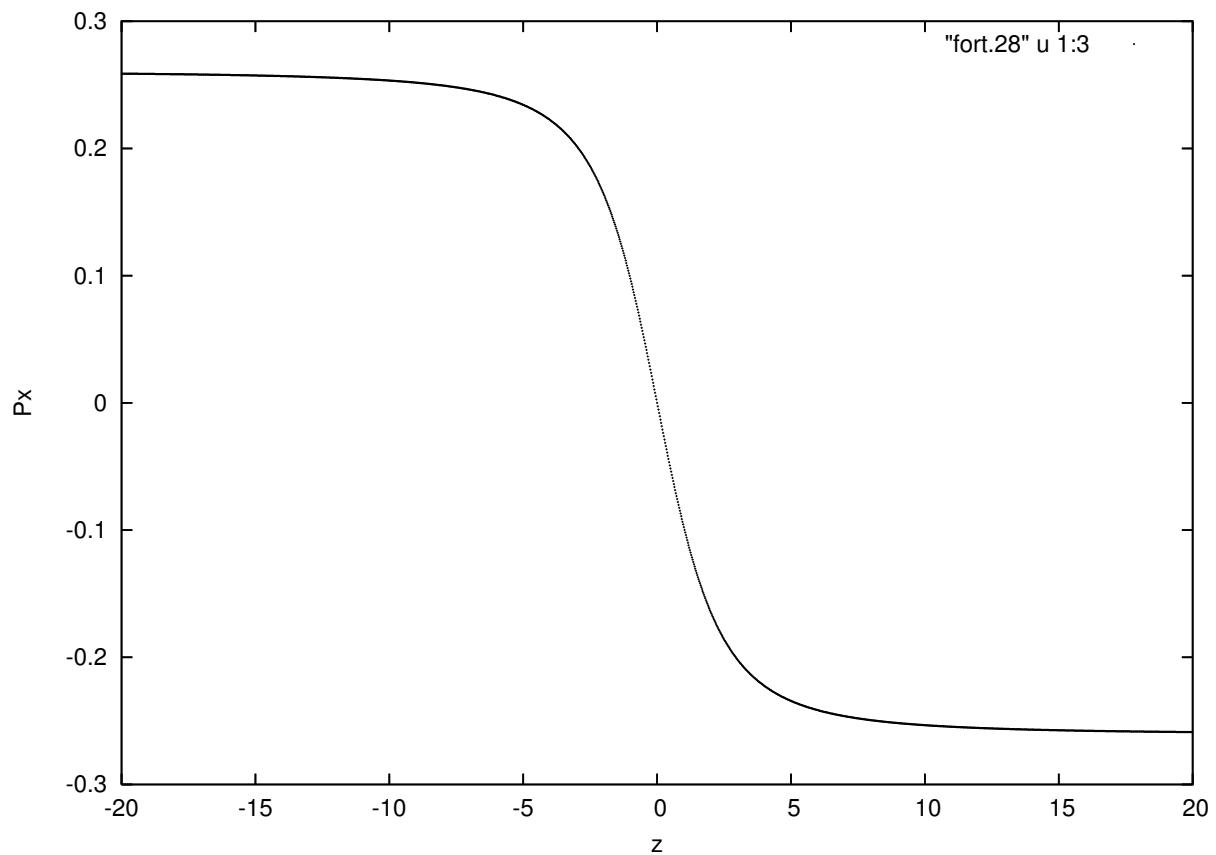


Figure 23.3.2: (Place holder) The canonical momentum  $\hat{p}_x(z)$  on the design orbit. The quantity  $z$  is in cm.

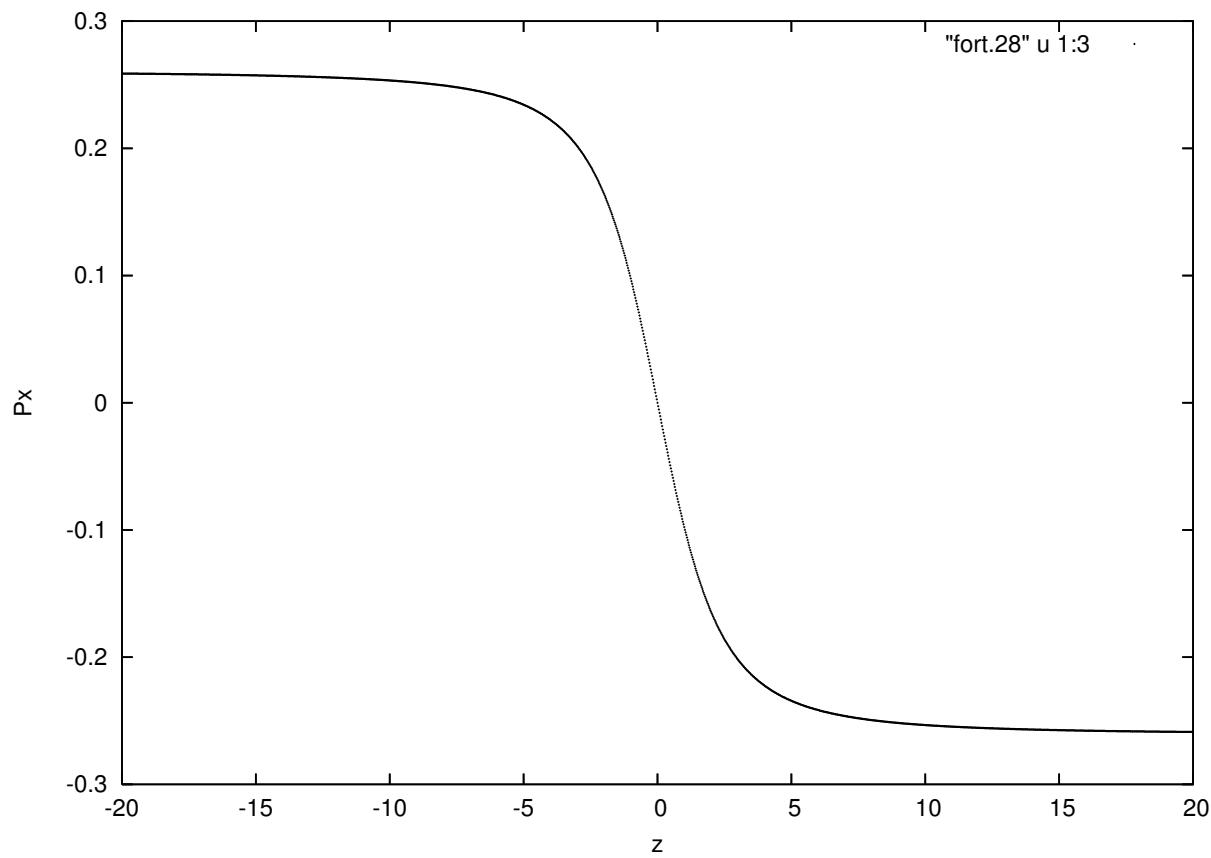


Figure 23.3.3: The scaled mechanical momentum  $\hat{p}_x^{\text{mech}}(z)$  on the design orbit. The quantity  $z$  is in cm.

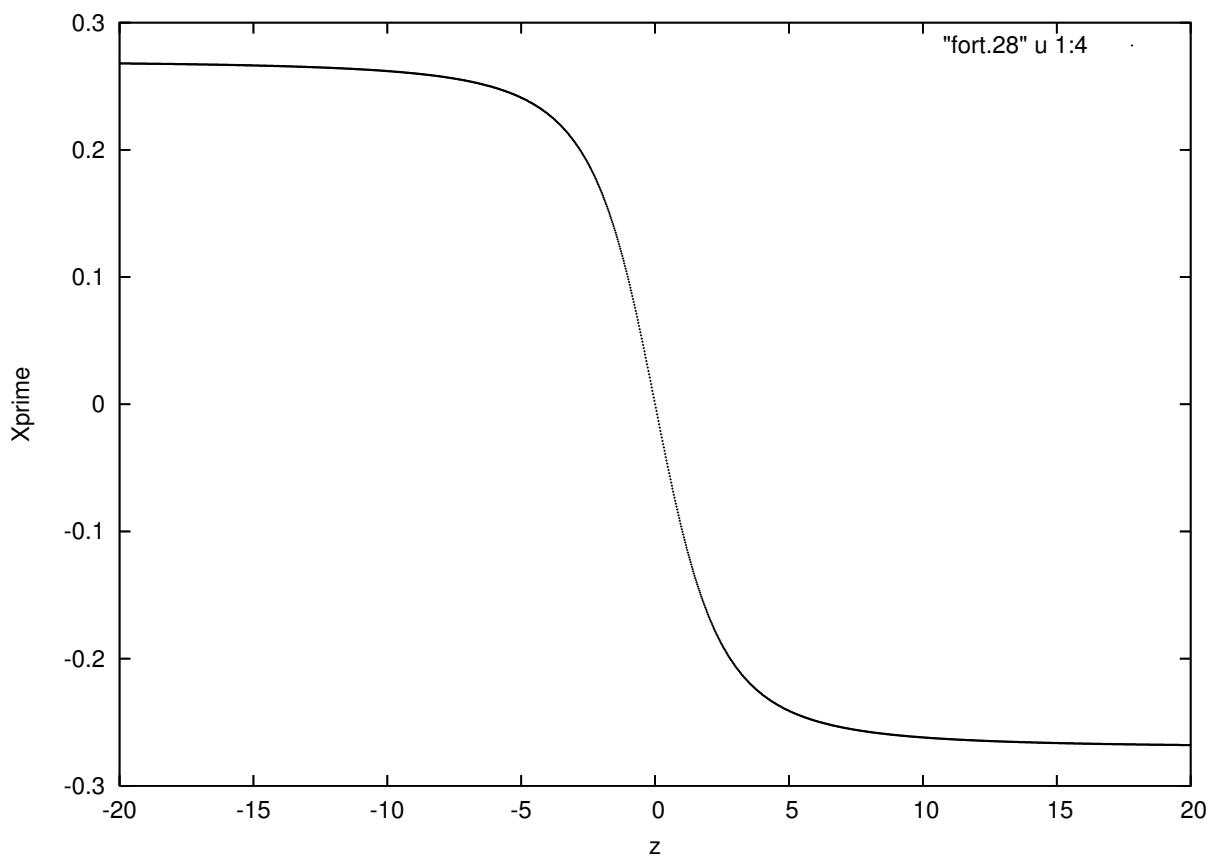


Figure 23.3.4: The quantity  $\hat{x}'(z)$  on the design orbit. The quantity  $z$  is in cm.

$$\beta^0 = .9984300412295174, \quad (23.3.39)$$

$$\gamma^0 = 17.853008080511426. \quad (23.3.40)$$

Here we have used the values

$$m_e c^2 = .51099906 \text{ MeV}, \quad (23.3.41)$$

$$-q = e = 1.60217733 \times 10^{-19} \text{ coulomb}, \quad (23.3.42)$$

$$c = 2.99792458 \times 10^8 \text{ m/s}. \quad (23.3.43)$$

It is also useful to have graphics of the quantities  $B_y$ ,  $\mathcal{A}_x$ , and  $\mathcal{A}_z$  along the design orbit. They are displayed in Figures 3.5 through 3.7. Note that  $B_y$  falls off quite rapidly with increasing  $|z|$ , i.e.  $\sim 1/|z|^3$ , as expected for a monopole doublet. However,  $\mathcal{A}_x$  and  $\mathcal{A}_z$  fall off less rapidly on the design orbit. From (1.14) we expect for  $\mathcal{A}_x$  a fall off  $\sim 1/|z|$ . And, from Figure 3.1 we see that on the design orbit  $|x|$  grows linearly with  $|z|$  for large  $|z|$ . Therefore, if (1.15) provides any indication, we expect that  $\mathcal{A}_z$  will also fall off only as  $1/|z|$  for large  $|z|$ .

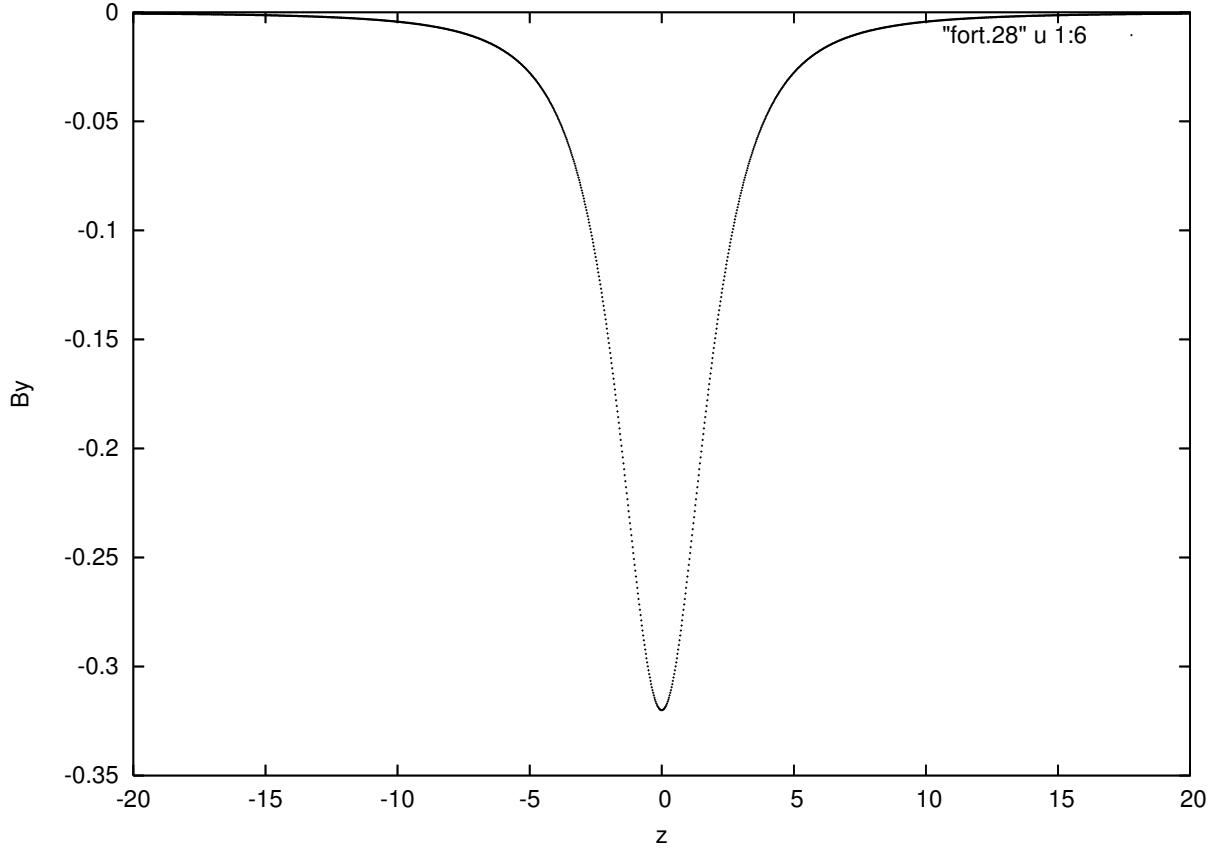


Figure 23.3.5: The quantity  $B_y$  on the design orbit. The quantity  $z$  is in cm.

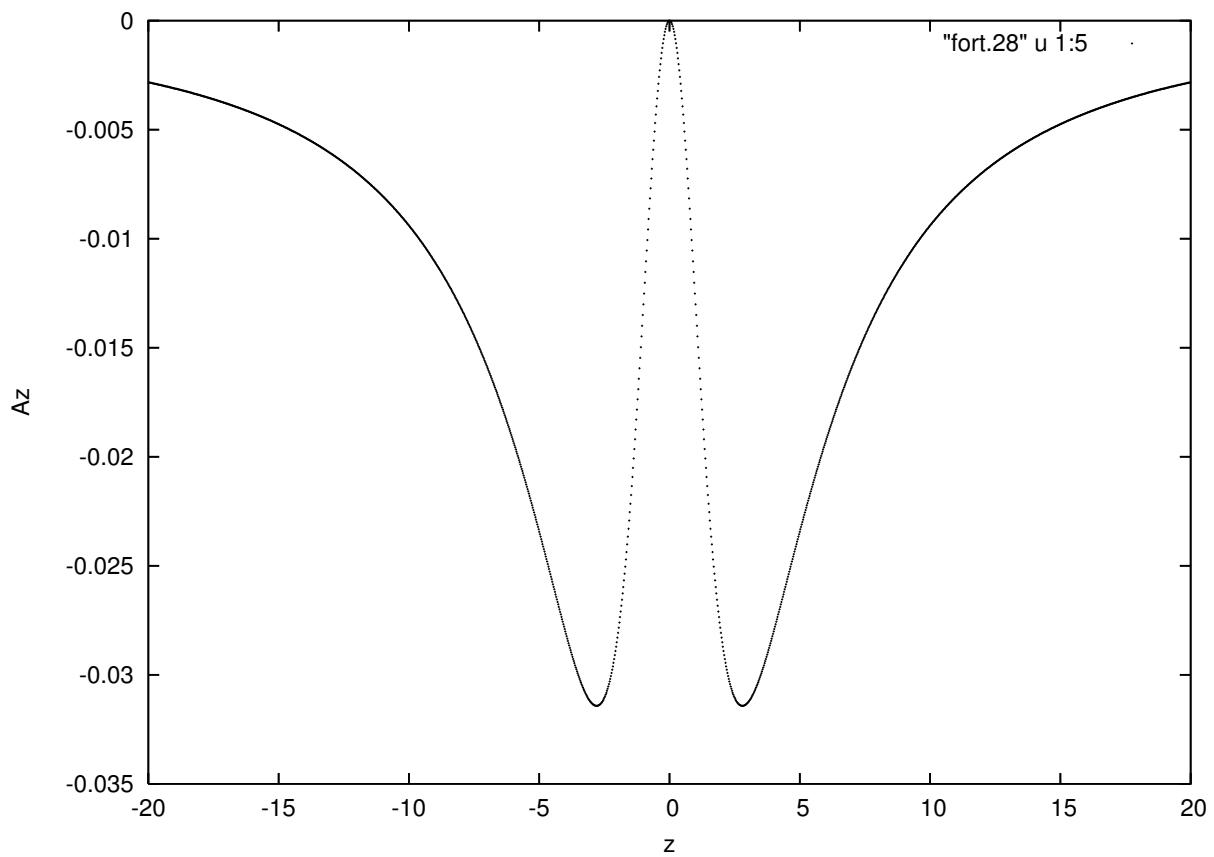


Figure 23.3.6: (Place holder) The quantity  $\mathcal{A}_x$  on the design orbit. The quantity  $z$  is in cm.

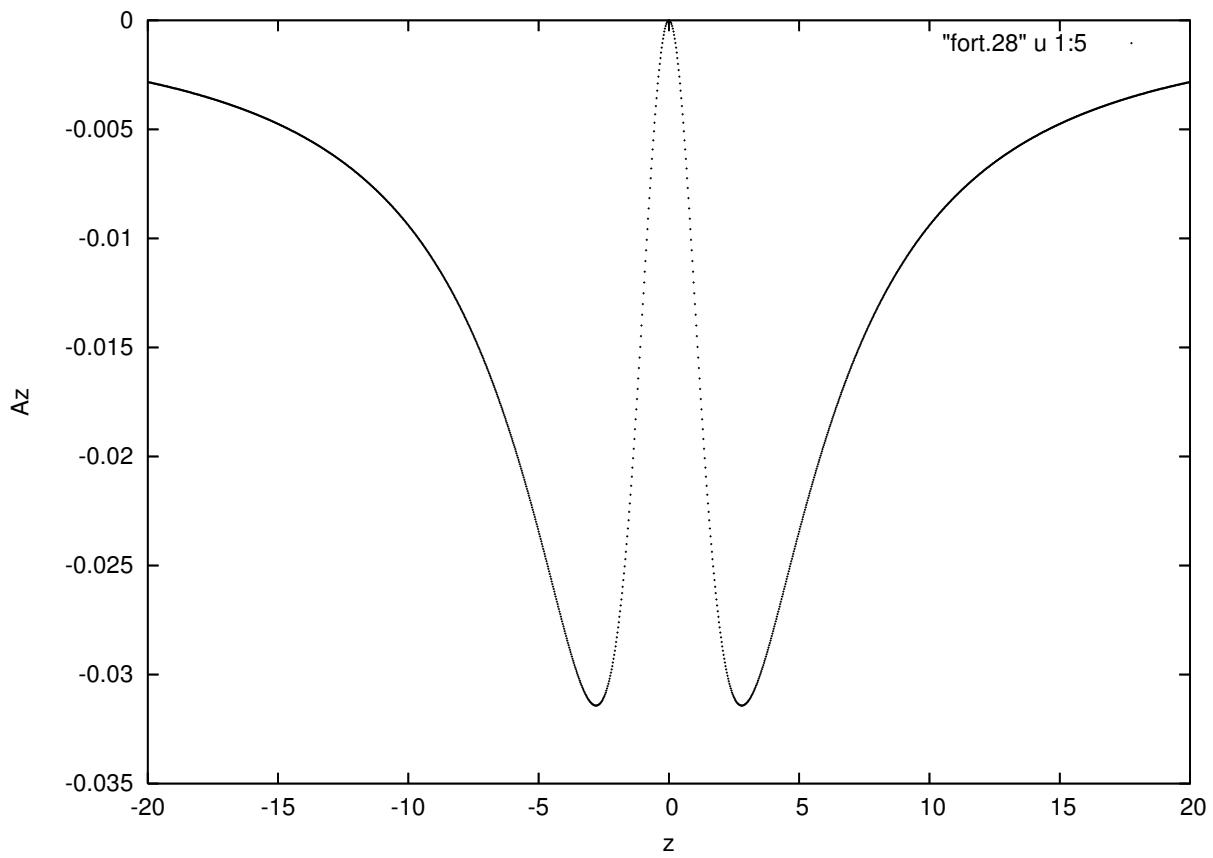


Figure 23.3.7: (Place Holder?) The quantity  $A_z$  on the design orbit. The quantity  $z$  is in cm.

Precise field values at some key points on the design orbit are given by the relations

$$B_y(-4.7597 \dots, 0, \mp 20) = -5.6290 \dots \times 10^{-4}, \quad (23.3.44)$$

$$B_y(0, 0, 0) = -32, \quad (23.3.45)$$

$$B_y(-4.7597 \dots, 0, \mp 20)/B_y(0, 0, 0) \simeq 1.8 \times 10^{-3}, \quad (23.3.46)$$

$$\mathcal{A}_x(-4.7597 \dots, 0, \mp 20) = \pm 2.78 \dots \times 10^{-2}, \quad (23.3.47)$$

$$\mathcal{A}_z(-4.7597 \dots, 0, \mp 20) = -6.6 \dots \times 10^{-3}. \quad (23.3.48)$$

From (3.45) we see that  $B_y$  on the design orbit has fallen by a factor of  $\simeq 1.8 \times 10^{-3}$  at the end points. By contrast, comparison of (3.31) and (3.36) shows that the vector potential for the half-infinite string choice of gauge still makes a significant contribution to the canonical momentum  $\hat{p}_x$  at the end points. Compare also (3.30) and (3.31). Therefore it is important to use some other gauge for end-field termination.

## Exercises

**23.3.1.** In Section 3 the design orbit was found by integrating the canonical pair (3.21) and (3.22). This exercise describes an alternate approach. It has the feature of illustrating that in mechanical variables the design orbit is manifestly gauge independent, as we know it should be.

Recall the relations (3.28) and (3.29). Suppose (3.28) is differentiated with respect to  $z$  and along the design orbit. Verify that doing so gives the result

$$(\hat{p}_x^{\text{mech}})' = \hat{p}'_x - (\partial \mathcal{A}_x / \partial \hat{x}) \hat{x}' - (\partial \mathcal{A}_x / \partial z). \quad (23.3.49)$$

Next employ (3.21) to rewrite the second term on the right of (3.49) in the form

$$-(\partial \mathcal{A}_x / \partial \hat{x}) \hat{x}' = -(\partial \mathcal{A}_x / \partial \hat{x})(\hat{p}_x - \mathcal{A}_x) / [1 - (\hat{p}_x - \mathcal{A}_x)^2]^{1/2}. \quad (23.3.50)$$

Observe that the right side of (3.50) agrees with the first term on the right side of (3.22) save for a sign. Show, therefore, that use of (3.22) in (3.49) yields, following a glorious cancellation, the simple result

$$(\hat{p}_x^{\text{mech}})' = (\partial \mathcal{A}_z / \partial \hat{x}) - (\partial \mathcal{A}_x / \partial z). \quad (23.3.51)$$

Also, from (1.10), (2.14), and (2.15), we have the result

$$[(\partial \mathcal{A}_z / \partial \hat{x}) - (\partial \mathcal{A}_x / \partial z)]|_{\dot{y}=0} = -(q/p^0) B_y(x, y, z)|_{y=0} \quad (23.3.52)$$

and, again from (1.10), we see that

$$B_y(x, y, z)|_{y=0} = -2ga/[x^2 + a^2 + z^2]^{3/2}. \quad (23.3.53)$$

It follows that (3.51) can be written in the final form

$$(\hat{p}_x^{\text{mech}})' = -(q/p^0) B_y(x, y, z)|_{y=0} = (q/p^0) 2ga / [\hat{x}^2 + a^2 + z^2]^{3/2}. \quad (23.3.54)$$

Together (3.29) and (3.54) form a convenient coupled set for numerical integration. Once the pair  $\{\hat{x}(z), \hat{p}_x^{\text{mech}}(z)\}$  has been found, the canonical momentum  $\hat{p}_x(z)$  is given by the relation (3.28) rewritten in the form

$$\hat{p}_x = \hat{p}_x^{\text{mech}} + \mathcal{A}_x|_{\dot{y}=0} \quad (23.3.55)$$

with  $\mathcal{A}_x|_{\dot{y}=0}$  given by (3.23). Verify that  $\mathcal{A}_x$  and  $\mathcal{A}_z$  on the design orbit are given by the relations (3.23) and (3.24). Verify (3.25) and (3.26).

**23.3.2.** Verify (3.34) based on (3.33).

## 23.4 Terminating End Fields

### 23.4.1 Minimum Vector Potential for End Fields

The first few terms in the expansion [about the point  $(X_0, 0, Y_0)$ ] of the minimum vector potential for a magnetic monopole doublet were found in Section 15.10. We recall the results

$$\mathbf{A}^{\min 1}(\mathbf{r}; X_0, Z_0) = [ga/(X_0^2 + Z_0^2 + a^2)^{3/2}](-z\mathbf{e}_x + x\mathbf{e}_z), \quad (23.4.1)$$

$$\begin{aligned} \mathbf{A}^{\min 2}(\mathbf{r}; X_0, Z_0) = & [-2ga/(X_0^2 + Z_0^2 + a^2)^{5/2}] \times \\ & [(Z_0y^2 - Z_0z^2 - X_0xz)\mathbf{e}_x + (X_0yz - Z_0xy)\mathbf{e}_y + (X_0x^2 + Z_0xz - X_0y^2)\mathbf{e}_z]. \end{aligned} \quad (23.4.2)$$

See (15.10.7) and (15.10.8). The still higher-order terms (the terms for  $n > 2$ ) can be found in an analogous way.

### 23.4.2 Associated Termination Error

Suppose we wish to initiate or terminate the magnetic field of a magnetic monopole doublet at the point  $(X_0, 0, Y_0)$ . Then we need to find the minimum vector potential expansion about this point in terms of variables appropriate to the relevant reference planes. As an example, supposed the field is initiated at the point  $(X_0 = -4.7597 \dots, 0, Z_0 = -20)$  corresponding to the beginning of the left leg of the bent box in Figure 3.1. Then the relevant reference plane would be the incoming face of the left leg.

To be more precise, let  $\mathbf{e}_\xi$  and  $\mathbf{e}_\eta$  be unit vectors in this reference plane, and let  $\mathbf{e}_\zeta$  be a unit vector perpendicular to this plane. These requirements can be met by making the definitions

$$\mathbf{e}_\xi = \cos \theta \mathbf{e}_x - \sin \theta \mathbf{e}_z, \quad (23.4.3)$$

$$\mathbf{e}_\eta = \mathbf{e}_y, \quad (23.4.4)$$

$$\mathbf{e}_\zeta = \sin \theta \mathbf{e}_x + \cos \theta \mathbf{e}_z. \quad (23.4.5)$$

Here  $\theta$  is the angle between the reference plane and the plane  $Y_0 = 0$ . See Figure 4.1. For the problem at hand,  $\theta$  is given by the relation

$$\theta = (1/2)\phi_{\text{bend}} \simeq 15^\circ. \quad (23.4.6)$$

As the notation is intended to convey, the vectors  $\mathbf{e}_\xi$ ,  $\mathbf{e}_\eta$ ,  $\mathbf{e}_\zeta$  comprise a right-handed orthonormal triad. Consequently, there are relations of the form

$$\mathbf{e}_\zeta \times \mathbf{e}_\xi = -\sin^2 \theta (\mathbf{e}_x \times \mathbf{e}_z) + \cos^2 \theta (\mathbf{e}_z \times \mathbf{e}_x) = \mathbf{e}_y = \mathbf{e}_\eta, \text{ etc.} \quad (23.4.7)$$

Next we observe that, associated with the unit vectors  $\mathbf{e}_\xi$ ,  $\mathbf{e}_\eta$ ,  $\mathbf{e}_\zeta$ , we may define local expansion coordinates  $\xi, \eta, \zeta$  by writing

$$\xi = x \cos \theta - z \sin \theta, \quad (23.4.8)$$

$$\eta = y, \quad (23.4.9)$$

$$\zeta = x \sin \theta + z \cos \theta. \quad (23.4.10)$$

Finally, the definitions (4.3) through (4.5) and (4.8) through (4.10) may be inverted to yield the relations

$$\mathbf{e}_x = \cos \theta \mathbf{e}_\xi + \sin \theta \mathbf{e}_\zeta, \quad (23.4.11)$$

$$\mathbf{e}_y = \mathbf{e}_\eta, \quad (23.4.12)$$

$$\mathbf{e}_z = -\sin \theta \mathbf{e}_\xi + \cos \theta \mathbf{e}_\zeta; \quad (23.4.13)$$

$$x = \xi \cos \theta + \zeta \sin \theta, \quad (23.4.14)$$

$$y = \eta, \quad (23.4.15)$$

$$z = -\xi \sin \theta + \zeta \cos \theta. \quad (23.4.16)$$

We also record that, as expected, there are the relations

$$\mathbf{r} = x \mathbf{e}_x + y \mathbf{e}_y + z \mathbf{e}_z = \xi \mathbf{e}_\xi + \eta \mathbf{e}_\eta + \zeta \mathbf{e}_\zeta. \quad (23.4.17)$$

With all these relations at hand, let us express the minimum vector potential for a magnetic monopole doublet in terms of the variables  $\xi, \eta, \zeta$  and their associated unit vectors. From (4.1) and using (4.11) through (4.17) we find the result

$$\mathbf{A}^{\min 1}(\xi, \eta, \zeta; X_0, Z_0) = \mathbf{A}^{\min 1}(\mathbf{r}; X_0, Z_0) = [ga/(X_0^2 + Z_0^2 + a^2)^{3/2}](-\zeta \mathbf{e}_\xi + \xi \mathbf{e}_\zeta). \quad (23.4.18)$$

And, from (4.2) and again using (4.11) through (4.17), we find the result

$$\begin{aligned} \mathbf{A}^{\min 2}(\xi, \eta, \zeta; X_0, Z_0) &= \mathbf{A}^{\min 2}(\mathbf{r}; X_0, Z_0) = -2ga/(X_0^2 + Z_0^2 + a^2)^{5/2}] \times \\ &[ (Z_0 y^2 - Z_0 z^2 - X_0 x z) \mathbf{e}_\xi + (X_0 y z - Z_0 x y) \mathbf{e}_\eta + (X_0 x^2 + Z_0 x z - X_0 y^2) \mathbf{e}_\zeta ]. \end{aligned} \quad (23.4.19)$$

Following the discussion of Section 16.1, what interests us with regard to the discontinuities in the transverse mechanical momenta associated with the field termination approximation are the  $\xi$  and  $\eta$  components of  $\mathbf{A}^{\min}$  evaluated at  $\zeta = 0$ . From (4.18) and (4.19) we see that the lowest order contributions to these discontinuities are given by the relations

$$A_\xi^{\min 1}(\xi, \eta, 0; X_0, Z_0) = A_\eta^{\min 1}(\xi, \eta, 0; X_0, Z_0) = 0, \quad (23.4.20)$$

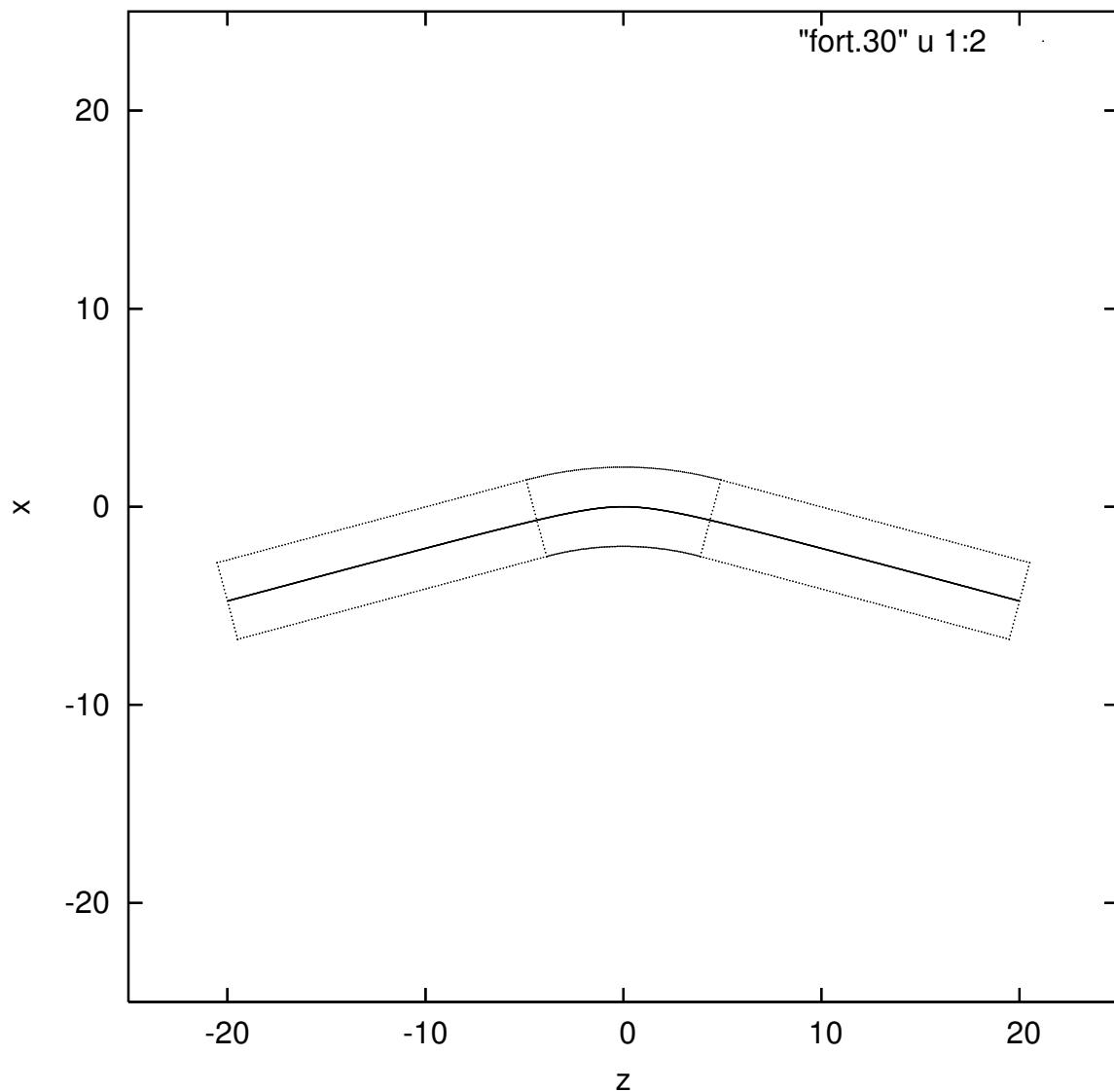


Figure 23.4.1: (Place Holder) The orthonormal triad  $e_\xi, e_\eta, e_\zeta$  and associated local deviation variables  $\xi, \eta, \zeta$  for the entry of the left leg of the bent box with legs.

$$A_{\xi}^{\min 2}(\xi, \eta, 0; X_0, Z_0) = [-2ga/(X_0^2 + Z_0^2 + a^2)^{5/2}](X_0 \sin \theta + Z_0 \cos \theta)\eta^2, \quad (23.4.21)$$

$$A_{\eta}^{\min 2}(\xi, \eta, 0; X_0, Z_0) = [2ga/(X_0^2 + Z_0^2 + a^2)^{5/2}](X_0 \sin \theta + Z_0 \cos \theta)\xi\eta. \quad (23.4.22)$$

We conclude that  $\mathbf{A}^{\min 1}$  makes no contributions to the discontinuities and that  $\mathbf{A}^{\min 2}$  makes contributions that are quadratic in the deviation variables  $\xi$  and  $\eta$ .

Let us examine the (upon entry) discontinuities associated with  $\mathbf{A}^{\min}$ . As a measure of these, define the *dimensionless* quantities  $\delta_{\xi}$  and  $\delta_{\eta}$  by the relations

$$\delta_{\xi, \eta} = (1/p^0)\Delta_{\xi, \eta}^{\text{mech}}. \quad (23.4.23)$$

See (16.1.30) and (16.1.31). Upon employing (16.1.30) and (16.1.31) in (4.23), and with the use of (4.21) and (4.22), we find the results

$$\delta_{\xi}^2(\xi, \eta, 0; X_0, Z_0) = (q/p^0)[-2ga/(X_0^2 + Z_0^2 + a^2)^{5/2}](X_0 \sin \theta + Z_0 \cos \theta)\eta^2, \quad (23.4.24)$$

$$\delta_{\eta}^2(\xi, \eta, 0; X_0, Z_0) = (q/p^0)[2ga/(X_0^2 + Z_0^2 + a^2)^{5/2}](X_0 \sin \theta + Z_0 \cos \theta)\xi\eta. \quad (23.4.25)$$

Let us evaluate these discontinuities when the transverse deviations from the design orbit have the substantial values  $\xi = \eta = 1$  cm. So doing, and recalling (3.23) and (4.26), we find the results

$$\delta_{\xi}^2(\xi = 1, \eta = 1, 0; X_0 = -4.7597 \dots, Z_0 = -20) = -8.8 \dots \times 10^{-6}, \quad (23.4.26)$$

$$\delta_{\eta}^2(\xi = 1, \eta = 1, 0; X_0 = -4.7597 \dots, Z_0 = -20) = 8.8 \dots \times 10^{-6}. \quad (23.4.27)$$

These numbers are pleasantly small, and we conclude that there is relatively little discontinuity error associated with terminating the field of the monopole doublet to the left of  $Z_0 = -20$  providing the minimal vector potential is employed. The same is true for termination to the right of  $Z_0 = +20$ .<sup>1</sup>

### 23.4.3 Taylor Expansion of String Vector Potential

$$\mathbf{A}^{\text{ex}}(\mathbf{r}; \mathbf{R}_0) = \mathbf{A}^{\text{ex}}(x, y, z; X_0, Z_0) = \mathbf{A}(\mathbf{R}_0 + \mathbf{r}) = \sum_{n=0}^{\infty} \mathbf{A}^{\text{ex } n}(x, y, z; X_0, Z_0) \quad (23.4.28)$$

where  $\mathbf{A}^{\text{ex } n}(x, y, z; X_0, Z_0)$  is a homogeneous polynomial vector field of degree  $n$  in the components of  $\mathbf{r}$ .

$$\begin{aligned} A_x(\mathbf{R}_0 + \mathbf{r}) &= \\ &- \frac{g(Z_0 + z)}{[(X_0 + x)^2 + (y - a)^2 + (Z_0 + z)^2]^{1/2}\{[(X_0 + x)^2 + (y - a)^2 + (Z_0 + z)^2]^{1/2} - y + a\}} \\ &- \frac{g(Z_0 + z)}{[(X_0 + x)^2 + (y + a)^2 + (Z_0 + z)^2]^{1/2}\{[(X_0 + x)^2 + (y + a)^2 + (Z_0 + z)^2]^{1/2} + y + a\}}, \end{aligned} \quad (23.4.29)$$

<sup>1</sup>However, one should not be overly sanguine. It turns out that the design orbit continues to bend by as much as a degree as one continues to the left of  $Z_0 = -20$  and the right of  $Z_0 = +20$ . That is, true *asymptopia* has not been reached even when  $Z_0 = \pm 20$  and (3.41) holds. The magnetic monopole doublet field, and the field of any iron-free dipole, are problematic to treat because of their slow fringe-field fall off.

$$A_y(\mathbf{R}_0 + \mathbf{r}) = 0, \quad (23.4.30)$$

$$\begin{aligned} A_z(\mathbf{R}_0 + \mathbf{r}) &= \\ &+ \frac{g(X_0 + x)}{[(X_0 + x)^2 + (y - a)^2 + (Z_0 + z)^2]^{1/2} \{[(X_0 + x)^2 + (y - a)^2 + (Z_0 + z)^2]^{1/2} - y + a\}} \\ &+ \frac{g(X_0 + x)}{[(X_0 + x)^2 + (y + a)^2 + (Z_0 + z)^2]^{1/2} \{[(X_0 + x)^2 + (y + a)^2 + (Z_0 + z)^2]^{1/2} + y + a\}}. \end{aligned} \quad (23.4.31)$$

$$A_x^{\text{ex } 0}(x, y, z; X_0, Z_0) = -\frac{2gZ_0}{[X_0^2 + a^2 + Z_0^2]^{1/2} \{[X_0^2 + a^2 + Z_0^2]^{1/2} + a\}}, \quad (23.4.32)$$

$$A_y^{\text{ex } 0}(x, y, z; X_0, Z_0) = 0, \quad (23.4.33)$$

$$A_z^{\text{ex } 0}(x, y, z; X_0, Z_0) = \frac{2gX_0}{[X_0^2 + a^2 + Z_0^2]^{1/2} \{[X_0^2 + a^2 + Z_0^2]^{1/2} + a\}}. \quad (23.4.34)$$

#### 23.4.4 Finding the Associated Gauge Function

### 23.5 Gauge Transformation Map

### 23.6 Pole Face Rotation

### 23.7 Computation of Transfer Map

#### Exercises

**23.7.1.** Using (1.7) through (1.9), verify (1.10) through (1.12).

**23.7.2.** Review the last paragraph of Exercise 22.2.15. Use the geometric insight provided in that paragraph to conclude that the direction of the vector potential found in Subsection 1.2 follows from the orientations of the Dirac strings assigned to the monopoles making up the monopole doublet.

### 23.8 Scraps

\*\*\*\*\*

$$\mathbf{B}^0(\mathbf{r}; X_0, Z_0) = -[2ga/(X_0^2 + Z_0^2 + a^2)^{3/2}] \mathbf{e}_\eta. \quad (23.8.1)$$

$$\begin{aligned}\mathbf{A}^{\min 1}(\xi, \eta, \zeta; X_0, Z_0) &= -(1/2)\mathbf{r} \times \mathbf{B}^0(\mathbf{r}; X_0, Z_0) \\ &= [ga/(X_0^2 + Z_0^2 + a^2)^{3/2}](-\zeta \mathbf{e}_\xi + \xi \mathbf{e}_\zeta),\end{aligned}\quad (23.8.2)$$

$$\mathbf{A}^{\min 1}(\xi, \eta, 0; X_0, Z_0) = [ga/(X_0^2 + Z_0^2 + a^2)^{3/2}](\xi \mathbf{e}_\zeta), \quad (23.8.3)$$

$$A_\xi^{\min 1}(\xi, \eta, 0; X_0, Z_0) = A_\eta^{\min 1}(\xi, \eta, 0; X_0, Z_0) = 0. \quad (23.8.4)$$

$$A_\xi^{\min 2}(\xi, \eta, 0; X_0, Z_0) = [-2ga/(X_0^2 + Z_0^2 + a^2)^{5/2}](X_0 \sin \theta + Z_0 \cos \theta)\eta^2, \quad (23.8.5)$$

$$A_\eta^{\min 2}(\xi, \eta, 0; X_0, Z_0) = [2ga/(X_0^2 + Z_0^2 + a^2)^{5/2}](X_0 \sin \theta + Z_0 \cos \theta)\xi\eta. \quad (23.8.6)$$

$$\begin{aligned}\mathbf{B}^1(\mathbf{r}; X_0, Z_0) &= [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}] \times \\ &\quad [(X_0x + Z_0z)\mathbf{e}_y + y(X_0\mathbf{e}_x + Z_0\mathbf{e}_z)].\end{aligned}\quad (23.8.7)$$

$$(X_0x + Z_0z)\mathbf{e}_y = [x_0(\xi \cos \theta + \zeta \sin \theta) + Z_0(-\xi \sin \theta + \zeta \cos \theta)]\mathbf{e}_\eta, \quad (23.8.8)$$

$$\begin{aligned}y(X_0\mathbf{e}_x + Z_0\mathbf{e}_z) &= \eta[X_0(\cos \theta \mathbf{e}_\xi + \sin \theta \mathbf{e}_\zeta) + Z_0(-\sin \theta \mathbf{e}_\xi + \cos \theta \mathbf{e}_\zeta)] \\ &= \eta[(X_0 \cos \theta - Z_0 \sin \theta)\mathbf{e}_\xi + (X_0 \sin \theta + Z_0 \cos \theta)\mathbf{e}_\zeta].\end{aligned}\quad (23.8.9)$$

$$\begin{aligned}\mathbf{B}^1(\xi, \eta, \zeta; X_0, Z_0) &= [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}] \times \\ &\quad \{[X_0(\xi \cos \theta + \zeta \sin \theta) + Z_0(-\xi \sin \theta + \zeta \cos \theta)]\mathbf{e}_\eta \\ &\quad + \eta[(X_0 \cos \theta - Z_0 \sin \theta)\mathbf{e}_\xi + (X_0 \sin \theta + Z_0 \cos \theta)\mathbf{e}_\zeta]\}.\end{aligned}\quad (23.8.10)$$

$$\begin{aligned}\mathbf{B}^1(\xi, \eta, 0; X_0, Z_0) &= [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}] \times \\ &\quad \{[X_0(\xi \cos \theta) + Z_0(-\xi \sin \theta)]\mathbf{e}_\eta \\ &\quad + \eta[(X_0 \cos \theta - Z_0 \sin \theta)\mathbf{e}_\xi + (X_0 \sin \theta + Z_0 \cos \theta)\mathbf{e}_\zeta]\}.\end{aligned}\quad (23.8.11)$$

$$\mathbf{r}(\xi, \eta, \zeta) = \xi \mathbf{e}_\xi + \eta \mathbf{e}_\eta + \zeta \mathbf{e}_\zeta. \quad (23.8.12)$$

$$\mathbf{r}(\xi, \eta, 0) = \xi \mathbf{e}_\xi + \eta \mathbf{e}_\eta. \quad (23.8.13)$$

$$\begin{aligned} \mathbf{e}_\eta \times \mathbf{B}^1(\xi, \eta, 0; X_0, Z_0) &= [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}] \times \\ &\quad \{\eta[(-X_0 \cos \theta + Z_0 \sin \theta)\mathbf{e}_\zeta + (X_0 \sin \theta + Z_0 \cos \theta)\mathbf{e}_\xi]\}. \end{aligned} \quad (23.8.14)$$

$$\begin{aligned} [\eta \mathbf{e}_\eta \times \mathbf{B}^1(\xi, \eta, 0; X_0, Z_0)]_\xi &= [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}] \times \\ &\quad \{\eta^2[(X_0 \sin \theta + Z_0 \cos \theta)]\}. \end{aligned} \quad (23.8.15)$$

$$\begin{aligned} \mathbf{e}_\xi \times \mathbf{B}^1(\xi, \eta, 0; X_0, Z_0) &= [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}] \times \\ &\quad \{[X_0(\xi \cos \theta) + Z_0(-\xi \sin \theta)]\mathbf{e}_\zeta \\ &\quad - \eta[(X_0 \sin \theta + Z_0 \cos \theta)\mathbf{e}_\eta]\}. \end{aligned} \quad (23.8.16)$$

$$\begin{aligned} [\xi \mathbf{e}_\xi \times \mathbf{B}^1(\xi, \eta, 0; X_0, Z_0)]_\eta &= [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}] \times \\ &\quad \{-\xi \eta[(X_0 \sin \theta + Z_0 \cos \theta)]\}. \end{aligned} \quad (23.8.17)$$

$$\mathbf{A}^{\min 2}(\xi, \eta, 0; X_0, Z_0) = -(1/3)\mathbf{r}(\xi, \eta, 0) \times \mathbf{B}^1(\xi, \eta, 0; X_0, Z_0). \quad (23.8.18)$$

$$A_\xi^{\min 2}(\xi, \eta, 0; X_0, Z_0) = [-2ga/(X_0^2 + Z_0^2 + a^2)^{5/2}](X_0 \sin \theta + Z_0 \cos \theta)\eta^2, \quad (23.8.19)$$

$$A_\eta^{\min 2}(\xi, \eta, 0; X_0, Z_0) = [2ga/(X_0^2 + Z_0^2 + a^2)^{5/2}](X_0 \sin \theta + Z_0 \cos \theta)\xi\eta. \quad (23.8.20)$$

# Chapter 24

## Realistic Transfer Maps for General Curved Beam-Line Elements: Bent Box Monopole Doublet Results

### 24.1 Choice of Surrounding Bent Box

Also shown in Figure 5.1 is the top view of a suitable bent box that surrounds this orbit. The top and bottom of the box are superposed in the figure, and lie in the planes  $y = \pm 2$  cm. The circular arcs that comprise the bent portion of the box have the common center

$$(x_c, z_c) = (-17 \text{ cm}, 0) \quad (24.1.1)$$

and have radii

$$r_{\text{out}} = 19 \text{ cm}, \quad (24.1.2)$$

$$r_{\text{in}} = 15 \text{ cm}. \quad (24.1.3)$$

Both subtend an angle of  $30^\circ$ , and are extended by straight lines thereby forming the straight ends of the box.

How was this bent box determined? Again by trial and error. Note that the construction of the bent box is not critical. All that is required is that the bent box well surround the design orbit. Consider all circular arcs that pass through the origin and are symmetric about the  $x$  axis. Such arcs will have their centers on the  $x$  axis. Also require that each arc subtend an angle of  $30^\circ$ . With these restrictions the only remaining quantity to be selected is the radius of an arc. Finally, require that the optimal arc, when extended by straight lines at both ends, well fit the design orbit. Figure 6.1 below shows that, for the problem at hand, a good fit occurs when the arc radius has the value

$$r_{\text{fit}} = 17 \text{ cm}. \quad (24.1.4)$$

By construction, the center of this arc is given by (5.55).

Now determine the outer and inner boundaries of the box by requiring that they also be circular arcs with straight-line extensions. Further require that both arcs have a common

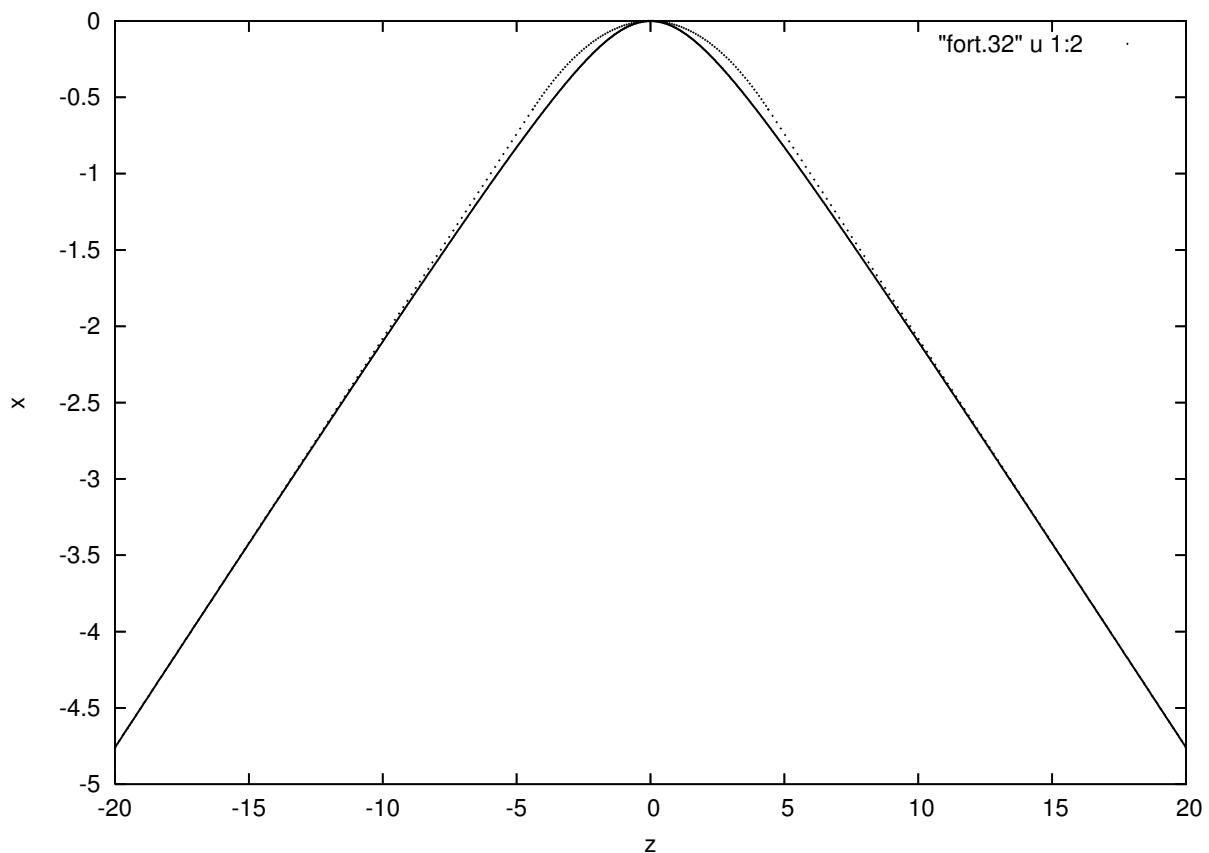


Figure 24.1.1: Design orbit  $x(z)$  and best approximating circular arc with straight-line extensions. The solid line is the design orbit, and the dotted line is the best approximating circular arc with straight-line extensions. The quantities  $x$  and  $z$  are in cm.

center given by (5.55), and that both arcs subtend an angle of  $30^\circ$ . The last step is to specify the radius of each arc. This is conveniently done by the prescription

$$r_{\text{out}} = r_{\text{fit}} + w, \quad (24.1.5)$$

$$r_{\text{in}} = r_{\text{fit}} - w, \quad (24.1.6)$$

where  $2w$  is the width of the box. For our illustration we have chosen the value

$$w = 2 \text{ cm}. \quad (24.1.7)$$

## 24.2 Comparison of Fields

How well do surface methods work for general geometries? In this subsection we will take magnetic field and scalar potential values for a monopole doublet, interpolate them onto the bent box surface found in the previous subsection, and then use surface methods to compute the interior field at various sample points. This computed interior field will then be compared with the actual monopole doublet field at these sample points, thereby providing a test of the accuracy of the method. Note that the bent box we have chosen has a square cross section with side 4 cm, and therefore is comparable in cross section to the 4 cm diameter cylinder used in Section 19.1.

### 24.2.1 Preliminaries

Since it is our intent to compute the interior field from surface values of  $B_n$  and  $\psi$ , it would be good to have some feel for how these quantities behave on the surface of the bent box with legs. Figure 5.7 displays  $B_n$  on the upper face,  $y = 2 \text{ cm}$ , of the bent box with legs directly above the design orbit. Figure 5.8 does the same for  $\psi$ . Up to signs, similar results hold for the bottom face,  $y = -2 \text{ cm}$ . The observation to be made is that  $B_n$  falls off fairly rapidly with increasing  $|z|$ , like  $\sim 1/|z|^3$ , and  $\psi$  falls off somewhat less rapidly, like  $\sim 1/|z|^2$ .

Something should also be said about the behavior of  $B_n$  and  $\psi$  on the sides of the box with legs. It is easily checked that, for a monopole doublet,  $B_x$ ,  $B_z$ , and  $\psi$  are odd functions of  $y$ , and therefore must vanish in the midplane  $y = 0$ . From this fact, and from considerations of field-line geometry for the case of a monopole doublet field, we conclude that the values of  $B_n$  and  $\psi$  on the sides of the box with legs will be comparable to, and usually smaller than, their values on the top and bottom faces.

Taking into account the behavior of  $B_n$  and  $\psi$  on the entire surface of the box with legs, we conclude that it is only necessary to integrate over some bounded portion of the surface in order to compute the interior vector potential  $\mathbf{A}$  accurately.

With this background in mind, we are prepared to make some numerical tests. We begin by imbedding the bent box with legs of Figure 5.1 (also see Figure 1.1) within a three-dimensional rectangular mesh,

$$x \in [x_{\min}, x_{\max}], \quad (24.2.1)$$

$$y \in [y_{\min}, y_{\max}], \quad (24.2.2)$$

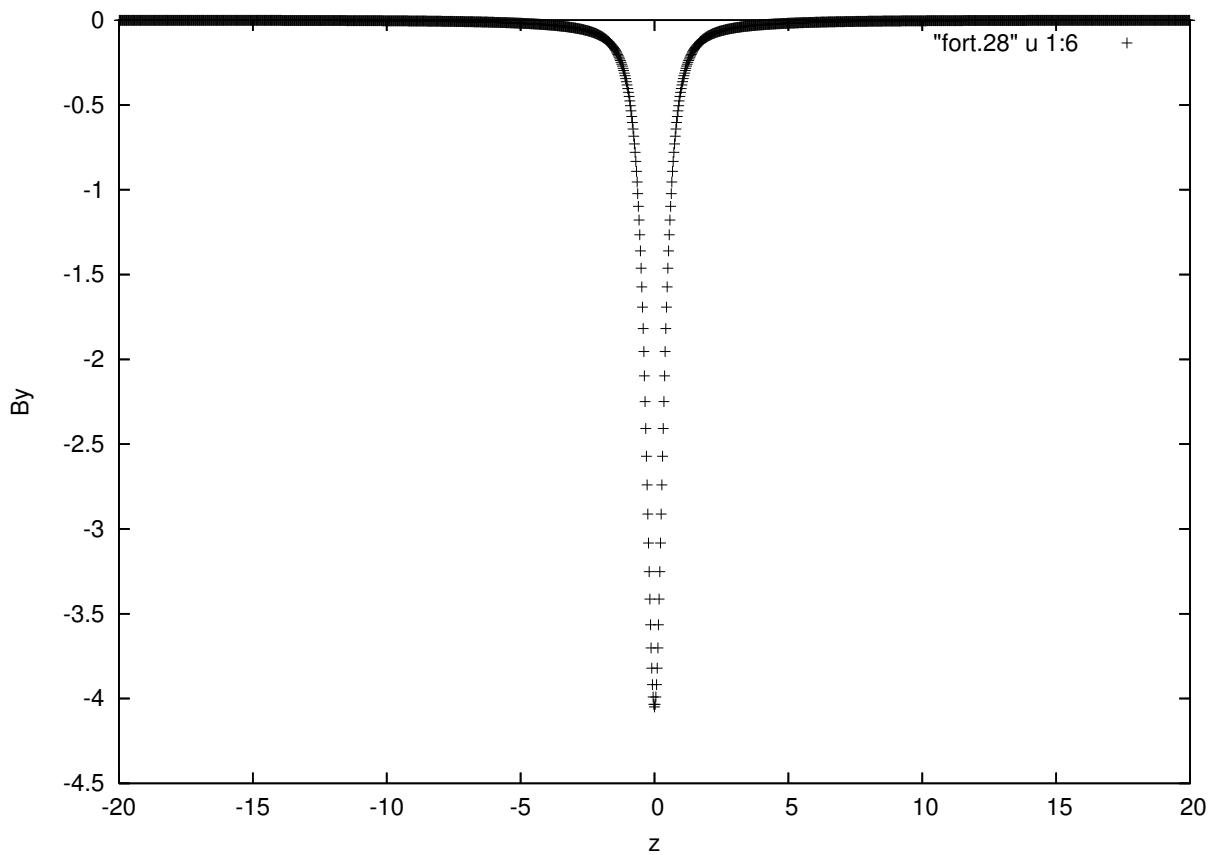


Figure 24.2.1: The quantity  $B_n = B_y$  on the upper face,  $y = 2$  cm, and directly above the design orbit. The quantity  $z$  is in cm.

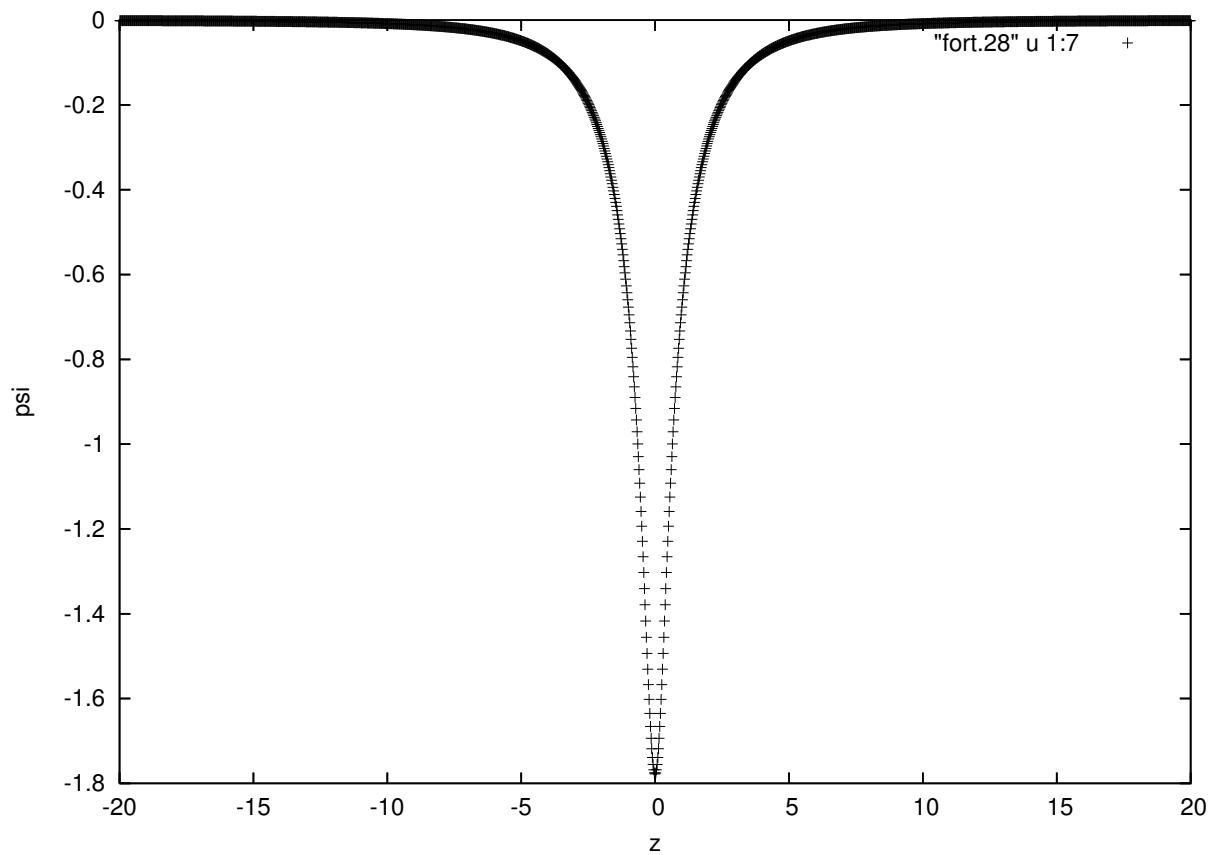


Figure 24.2.2: The quantity  $\psi$  on the upper face,  $y = 2$  cm, and directly above the design orbit. The quantity  $z$  is in cm.

$$z \in [z_{min}, z_{max}], \quad (24.2.3)$$

with mesh-point spacings  $h_x$ ,  $h_y$ , and  $h_z$ , respectively. By looking at Figure 5.1 we see that convenient end-point values are given by the relations

$$x_{min} = -7, \quad x_{max} = 3, \quad (24.2.4)$$

$$y_{min} = -3, \quad y_{max} = 3, \quad (24.2.5)$$

$$z_{min} = -25, \quad z_{max} = 25. \quad (24.2.6)$$

See also the corner coordinates given in (5.76) through (5.79). For the mesh-point spacings we take the values

$$h_x = h_y = h_z = .05. \quad (24.2.7)$$

At each of these grid points we compute and store the quantities  $\mathbf{B}$  and  $\psi$  for the monopole doublet field. It is data if this kind that could be expected for the output of some electromagnetic solver.

### 24.2.2 Evaluation of Surface Integrals

Our task now is to use this data to evaluate surface integrals of the kind (1.7) and (1.8). In this case the surface  $S$ , a bent box with legs, consists of a bent sector with straight end legs. When viewed from above, and as described earlier, the sector has inner and outer radii given by (5.59) and (5.60), and subtends an angle of  $30^\circ$ . It has corners at the locations

$$z_{\ell\ell}^s = -3.882285676537811, \quad x_{\ell\ell}^s = -2.511112605663975, \quad (24.2.8)$$

$$z_{u\ell}^s = -4.917561856947894, \quad x_{u\ell}^s = 1.352590699492298, \quad (24.2.9)$$

$$z_{ur}^s = 4.917561856947894, \quad x_{ur}^s = 1.352590699492298, \quad (24.2.10)$$

$$z_{\ell r}^s = 3.882285676537811, \quad x_{\ell r}^s = -2.511112605663975. \quad (24.2.11)$$

The left straight end leg, again when viewed from above, has leftmost corners at the locations

$$z_{\ell\ell}^\ell = -19.482361909794935, \quad x_{\ell\ell}^\ell = -6.69114043422916, \quad (24.2.12)$$

$$z_{u\ell}^\ell = -20.51763809020501, \quad x_{u\ell}^\ell = -2.8274371290729103. \quad (24.2.13)$$

The right straight end leg has rightmost corners at the locations

$$z_{ur}^r = 20.51763809020501, \quad x_{ur}^r = -2.8274371290729103, \quad (24.2.14)$$

$$z_{\ell r}^r = 19.482361909794935, \quad x_{\ell r}^r = -6.69114043422916. \quad (24.2.15)$$

Each straight end leg has a length of 16.150387336872548 cm.

We will decompose  $S$  into 12 pieces. The first 8 will be the top, bottom, inner, and outer faces of the two straight legs. The remaining 4 will be the top, bottom, inner, and outer faces of the bent sector. See Figures 1.1 and 5.1. The first 8 surfaces, those for the straight legs, are all rectangular, and can be conveniently integrated over using rectangular coordinates. Integrals over the remaining 4 surfaces, those for the bent sector, are most

easily evaluated using polar/cylindrical coordinates. Our task is to parameterize these 12 pieces and find expressions for  $dS'$  for each. In particular, we will convert the integration over each of these pieces into a related integration over a unit square.

Consider the straight legs. We will present results for the left leg. Results for the right leg are analogous.

Note that the right end of the left leg abuts the left end of the bent sector. Again see Figure 5.1. Consequently, the top face of the left leg has corners at  $(z_{\ell\ell}^\ell, x_{\ell\ell}^\ell)$ ,  $(z_{u\ell}^\ell, x_{u\ell}^\ell)$ ,  $(z_{u\ell}^s, x_{u\ell}^s)$ , and  $(z_{\ell\ell}^s, x_{\ell\ell}^s)$ . It can be described in terms of parameters  $u$  and  $v$  by writing

$$z^{\ell t}(u, v) = z_{\ell\ell}^\ell + (z_{\ell\ell}^s - z_{\ell\ell}^\ell)u + (z_{u\ell}^\ell - z_{\ell\ell}^\ell)v, \quad (24.2.16)$$

$$x^{\ell t}(u, v) = x_{\ell\ell}^\ell + (x_{\ell\ell}^s - x_{\ell\ell}^\ell)u + (x_{u\ell}^\ell - x_{\ell\ell}^\ell)v, \quad (24.2.17)$$

$$y^{\ell t} = 2, \quad (24.2.18)$$

with

$$u, v \in [0, 1]. \quad (24.2.19)$$

In this case one finds for the surface element the relation

$$dS' = dz dx = [(z_{\ell\ell}^s - z_{\ell\ell}^\ell)(x_{u\ell}^\ell - x_{\ell\ell}^\ell) - (z_{u\ell}^\ell - z_{\ell\ell}^\ell)(x_{\ell\ell}^s - x_{\ell\ell}^\ell)]du dv. \quad (24.2.20)$$

Similar results hold for the bottom face of the left leg.

The inner face of the left leg can be described in terms of parameters  $u$  and  $v$  by writing

$$z^{\ell i}(u, v) = z_{\ell\ell}^\ell + (z_{\ell\ell}^s - z_{\ell\ell}^\ell)u, \quad (24.2.21)$$

$$x^{\ell i}(u, v) = x_{\ell\ell}^\ell + (x_{\ell\ell}^s - x_{\ell\ell}^\ell)u, \quad (24.2.22)$$

$$y^{\ell i}(u, v) = -2 + 4v, \quad (24.2.23)$$

again with

$$u, v \in [0, 1]. \quad (24.2.24)$$

In this case one finds for the surface element the relation

$$dS' = [4/\cos(\pi/12)][(z_{\ell\ell}^s - z_{\ell\ell}^\ell)]du dv. \quad (24.2.25)$$

Similar results hold for the outer face of the left leg.

Consider the bent sector. The top face of the bent sector can be described in terms of cylindrical coordinates  $\rho$ ,  $\phi$ , and  $y$  by writing

$$z^{st} = \rho \sin \phi, \quad (24.2.26)$$

$$x^{st} = \rho \cos \phi - 17, \quad (24.2.27)$$

$$y^{st} = 2. \quad (24.2.28)$$

Introduce parameters  $u$  and  $v$  by writing

$$\rho(u, v) = 15 + 4v, \quad (24.2.29)$$

$$\phi(u, v) = -\pi/12 + (\pi/6)u, \quad (24.2.30)$$

again with the understanding (5.83). Combining (5.90) through (5.94) gives the results

$$z^{\text{st}}(u, v) = (15 + 4v) \sin(-\pi/12 + \pi u/6), \quad (24.2.31)$$

$$x^{\text{st}}(u, v) = (15 + 4v) \cos(-\pi/12 + \pi u/6) - 17, \quad (24.2.32)$$

$$y^{\text{st}}(u, v) = 2. \quad (24.2.33)$$

In this case one finds for the surface element of the top face of the bent sector the relation

$$dS' = \rho d\rho d\phi = (2\pi/3)(15 + 4v) du dv. \quad (24.2.34)$$

Similar results hold for the bottom face of the bent sector.

The inner face of the bent sector can be described in terms of relations analogous to (5.90) and (5.91) with  $\rho = 15$ ,

$$z^{\text{si}}(u, v) = 15 \sin(-\pi/12 + \pi u/6), \quad (24.2.35)$$

$$x^{\text{si}}(u, v) = 15 \cos(-\pi/12 + \pi u/6) - 17. \quad (24.2.36)$$

Here we have again used (5.94). We also write

$$y^{\text{si}}(u, v) = -2 + 4v. \quad (24.2.37)$$

In this case we find for the surface element the relation

$$dS' = \rho d\phi dy = 10\pi du dv. \quad (24.2.38)$$

Similar results hold for the outer face of the bent sector.

The result of the work so far is that the integrations over the 12 pieces of  $S$  have been converted into integrations over 12 unit squares of the form (5.83). For each piece, changes in  $u$  produce longitudinal displacements, and changes in  $v$  produce transverse displacements. We next select points within each unit square to be used in evaluating the various surface integrals numerically.

For the straight legs this is achieved as follows: Each unit square corresponding to a leg surface is decomposed into  $100 \times 160 = 16,000$  small rectangles by the prescription

$$h_u = 1/100, h_v = 1/160. \quad (24.2.39)$$

Thus, there are 100 subdivisions in the longitudinal direction and 160 subdivisions in the transverse directions.

For the surfaces of the bent sector each corresponding unit square is decomposed into  $160 \times 160 = 25,600$  small squares by the prescription

$$h_u = 1/160, h_v = 1/160. \quad (24.2.40)$$

Thus, for these surfaces there are 160 subdivisions for both the longitudinal and transverse directions.<sup>1</sup>

---

<sup>1</sup>More subdivisions are used for the bent sector surfaces because the fields are expected to vary more rapidly over these surfaces. See Figures 5.7 and 5.8.

The integral over each small rectangle or small square is approximated using a 7-point cubature formula. (For a discussion of cubature formulas, see Appendix T.) The values of the integrands at the cubature points are obtained from the values of  $\mathbf{B}$  and  $\psi$  at the grid points using 3-dimensional cubic spline interpolation.<sup>2</sup> Finally, all the small rectangle and small square results are summed to obtain the required integrals over  $S$ .

### 24.2.3 Resulting Vector Potential

Figures 5.9 and 5.10 show the components  $A_x^{sd}$  and  $A_z^{sd}$  of the vector potential computed along the design trajectory based on bent-box *surface data*.<sup>3</sup> The  $A_y^{sd}$  component vanishes in the midplane, and therefore is not shown. For the contribution from the  $B_n$  the kernel  $\mathbf{G}^{n2s}$  was used for the top and bottom faces, and the kernel  $\mathbf{G}^{n1s}$  was used for the side faces. In all cases the strings were taken to lie on lines parallel to the  $x$  axis. For the contribution from  $\psi$  the kernel  $\mathbf{G}^t$  was used.

Recall that the vector potential used in the previous subsection and given by (2.102) through (2.104) has no  $x$  component and only a  $z$  component on the design orbit. By contrast,  $A_x^{sd}$  as shown in Figure 5.9, although small, is not zero on the design orbit. Note also that  $A_z$  as displayed in Figure 5.4 and the  $A_z^{sd}$  displayed in Figure 5.10, while similar, are not the same. This apparent discrepancy arises from the fact that the vector potential given by (2.102) through (2.104) and the vector potential computed from surface data differ by a gauge transformation. We also remark that examination of the numerical results reveals that the nonzero contribution to  $A_x^{sd}$  arises from surface  $\psi$  values. See (3.43).

### 24.2.4 Comparison of Fields

If the interior vector potential  $\mathbf{A}^{sd}$  has been computed successfully using surface methods, so that it differs from the vector potential given by (2.102) through (2.104) at most only by a gauge transformation, in the interior of the box it should also give rise to the monopole-doublet  $\mathbf{B}$  field (2.105) through (2.107). Let  $\mathbf{B}^{sd}$  be the magnetic field given by

$$\mathbf{B}^{sd} = \nabla \times \mathbf{A}^{sd} \quad (24.2.41)$$

and let  $\mathbf{B}^e$  be the exact  $\mathbf{B}$  field. Then, for example, use of (5.105) to compute  $B_y^{sd}$  on the design orbit should produce a graphic similar to Figure 5.5.<sup>4</sup> This is indeed the case. A plot of  $B_y^{sd}$  on the design orbit is indistinguishable to the eye from Figure 5.5.

To give a better indication of the error involved, define a relative error  $\Delta$  by the relation

$$\Delta = (\mathbf{B}^{sd} - \mathbf{B}^e) / B_y^{maxmag}. \quad (24.2.42)$$

---

<sup>2</sup>Note that, like the case of cylindrical surfaces, the data at most of the data points on the grid are unused. For each cubature point on  $S$  there is an associated point in  $x,y,z$  space, and only data at the grid points near these points are actually used in interpolation.

<sup>3</sup>That is, as just described, grid data were manufactured and interpolated onto the 12 pieces of the surface  $S$  at the points required for the repeated use of a 7-point cubature formula. The results from these surface values were processed, by repeated application of this cubature formula, to find  $A_x^{sd}$  and  $A_z^{sd}$  along the design trajectory.

<sup>4</sup>Note that the use of (5.105) requires a knowledge of spatial derivatives of the components of  $\mathbf{A}$ . These derivatives are obtained by differentiating the kernels  $\mathbf{G}$  under the integral sign prior to carrying out the required surface integrations. See Subsections 4.3 and 4.4.

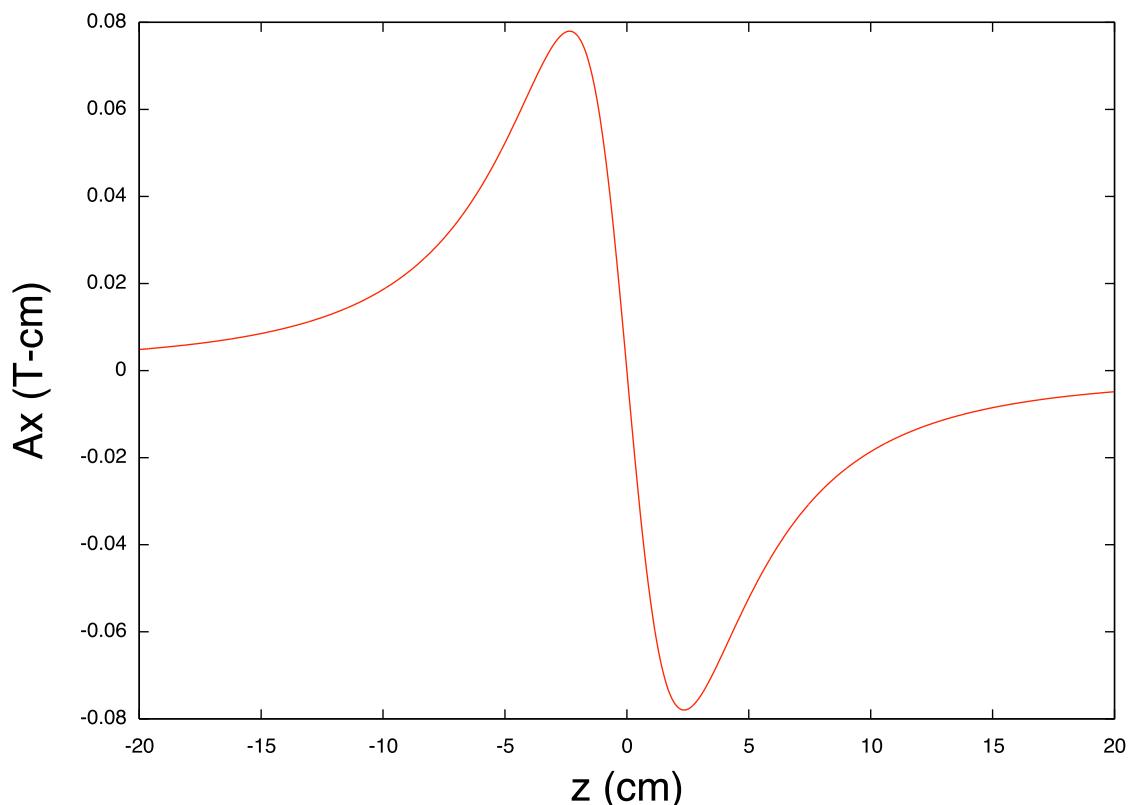


Figure 24.2.3: The quantity  $A_x^{sd}$  on the design orbit. The quantity  $z$  is in cm.

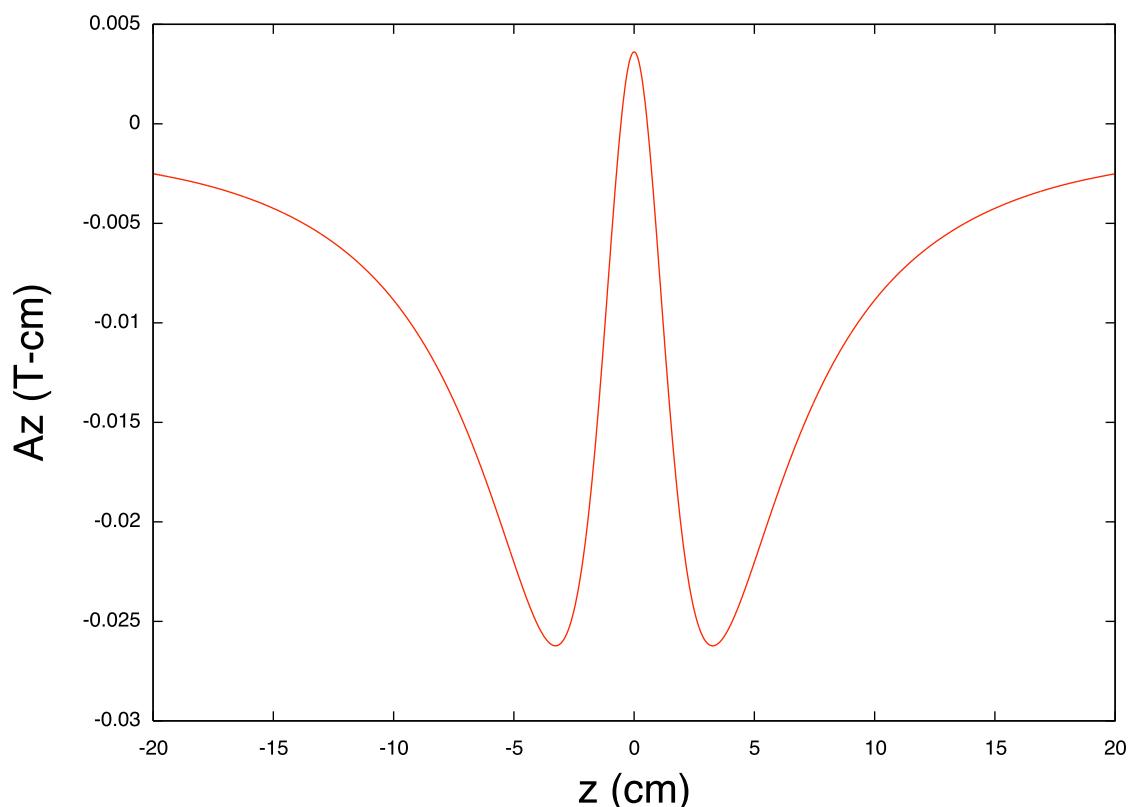


Figure 24.2.4: The quantity  $A_z^{sd}$  on the design orbit. The quantity  $z$  is in cm.

Here  $B_y^{maxmag}$  is the *maximum* value of the *magnitude* of  $B_y^e$  on the design orbit,

$$B_y^{maxmag} = |B_y(x = 0, y = 0, z = 0)| = 2g/a^2 = .32 \text{ Tesla.} \quad (24.2.43)$$

See (15.8.5) and Figure 15.8.3. Figure 5.11 displays the value of  $\Delta_y$  on the design orbit as a function of  $z$ . We see that  $\Delta_y$  is very small over most of the interval  $z \in [-20, 20]$ , but rises very rapidly to a value of  $\simeq 5 \times 10^{-4}$  at the endpoints.

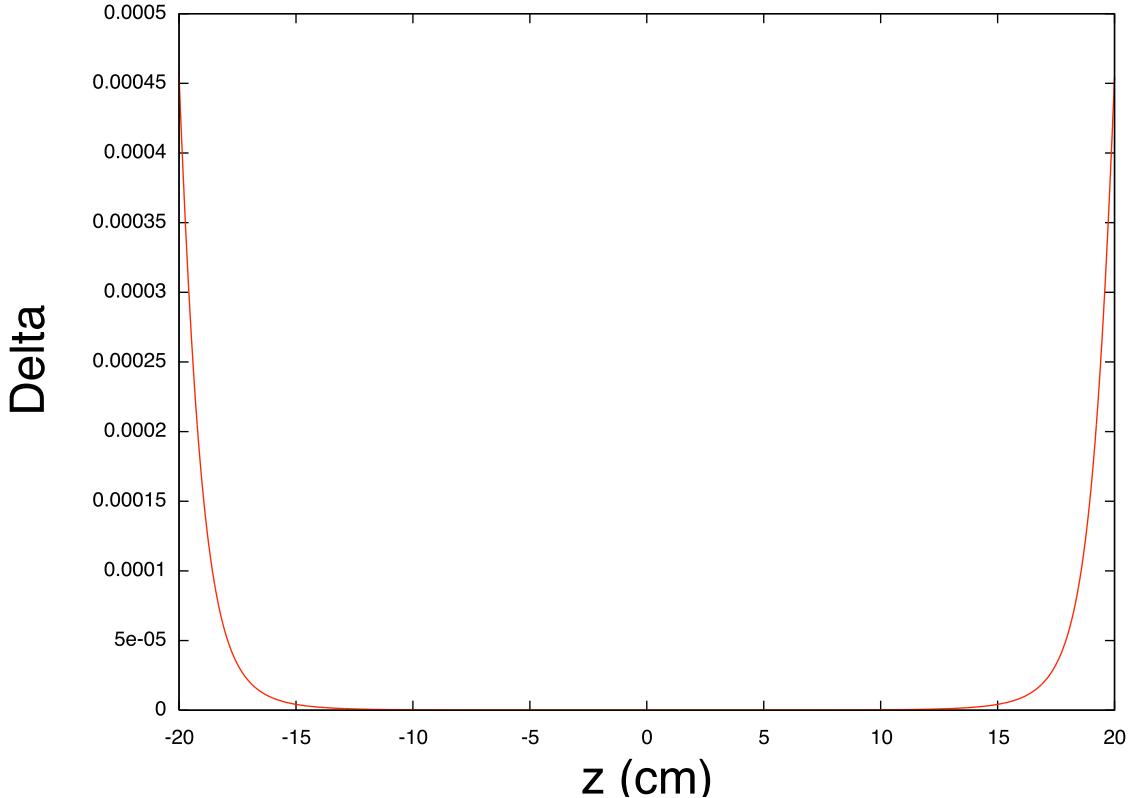


Figure 24.2.5: The relative error  $\Delta_y$  on the design orbit. The quantity  $z$  is in cm.

While the very small error for most of the interval error is very satisfying, the rapid increase of the error at the endpoints might seem alarming. It is not. The relative error  $\Delta_y$  remains bounded and eventually goes to zero as  $|z|$  goes to infinity. Moreover, both  $B_y^{sd}$  and  $B_y^e$  are small for  $|z| \geq 20$ , and go to zero as  $|z|$  goes to infinity.

To elaborate on these assertions, we begin by noting that both  $B_y^{sd}$  and  $B_y^e$  are negative. See Figure 5.5. But  $B_y^{sd}$  is slightly less negative than  $B_y^e$  because, by terminating the straight legs of the box at  $z = \pm 20$ , the surface fields that serve as a “source” for the interior field are effectively set to zero beyond  $z \in [-20, 20]$ . Correspondingly,  $\Delta_y$  is positive. Observe that, because  $B_y^{sd}$  is negative,  $\Delta_y$  always obeys the crude bound

$$\Delta_y < -B_y^e/B_y^{maxmag}. \quad (24.2.44)$$

At the end points the coordinates  $x, y, z$  have the values

$$x \simeq -4.76, y = 0, z = \pm 20. \quad (24.2.45)$$

See (5.45). Using these values in (15.8.5) we find that  $B_y^e$  has the value

$$B_y^e(x \simeq -4.76, y = 0, z = \pm 20) \simeq -5.63 \times 10^{-4} \text{ Tesla.} \quad (24.2.46)$$

Correspondingly, at worst and for  $|z| \geq 20$ ,  $\Delta_y$  can never exceed

$$5.63 \times 10^{-4} / .32 \simeq 1.8 \times 10^{-3}. \quad (24.2.47)$$

And since  $B_y^e$  for large  $|z|$  falls off as  $|z|^{-3}$ ,  $\Delta_y$  must eventually go to zero for large  $|z|$  as  $|z|^{-3}$ .

Figure 5.11 displays the relative error in the  $y$  component of  $\mathbf{B}^{sd}$  on the design orbit. We are also interested in examining the relative error in all the components of  $\mathbf{B}^{sd}$  in the vicinity of the design orbit. For this purpose it is convenient to introduce a deviation variable  $\xi$  by writing

$$x = x^d + \xi \quad (24.2.48)$$

where  $x^d$  is the design orbit shown in Figure 5.6. Figure 5.12 shows  $\Delta$ , the magnitude of  $\Delta$ , over the domain  $\xi \in [-1, 1], z \in [0, 20]$  in the plane  $y = 0$ . Figure 5.13 shows  $\Delta$  over the same domain in the plane  $y = 1$ . For ease of visualization, values are shown only for  $y \geq 0$  and  $z \geq 0$  since  $\Delta$  is even in these variables.

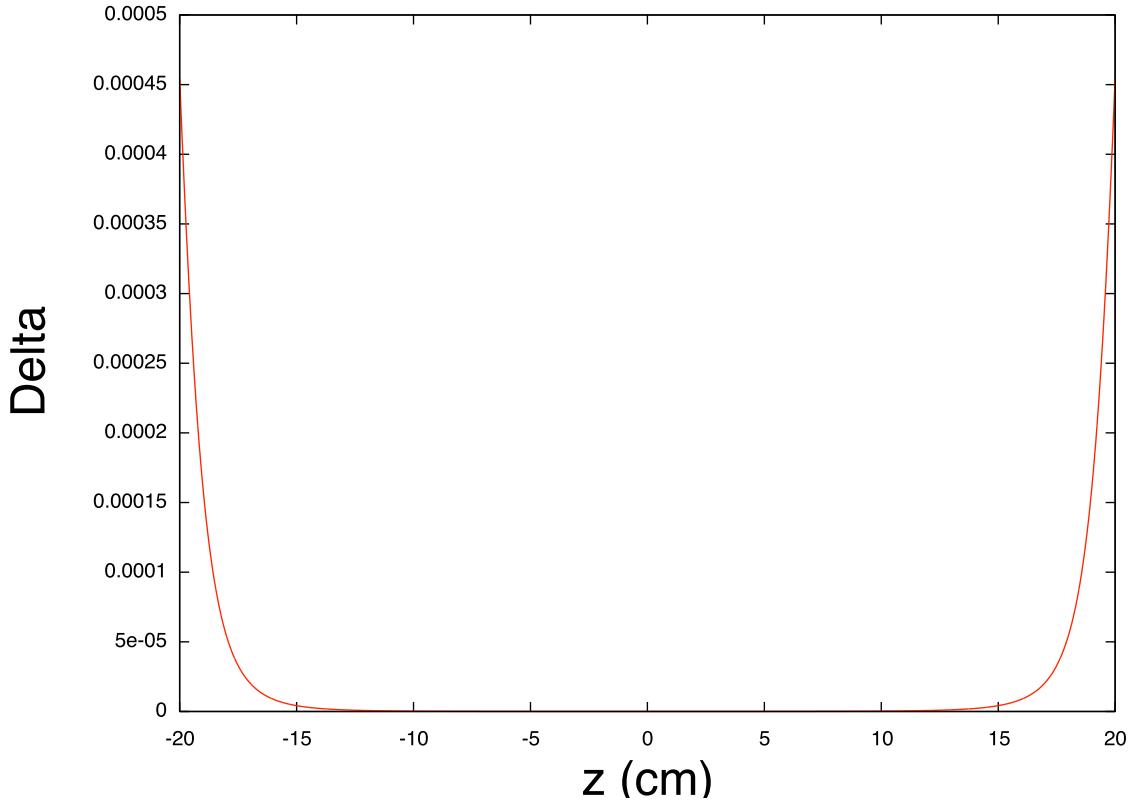


Figure 24.2.6: Place holder. The quantity  $\Delta = |\Delta|$  as a function of  $\xi$  and  $z$  in the vicinity of the design orbit and in the plane  $y = 0$ . The quantities  $\xi$ ,  $y$ , and  $z$  are in cm.

Upon examining these figures we see that . . . .

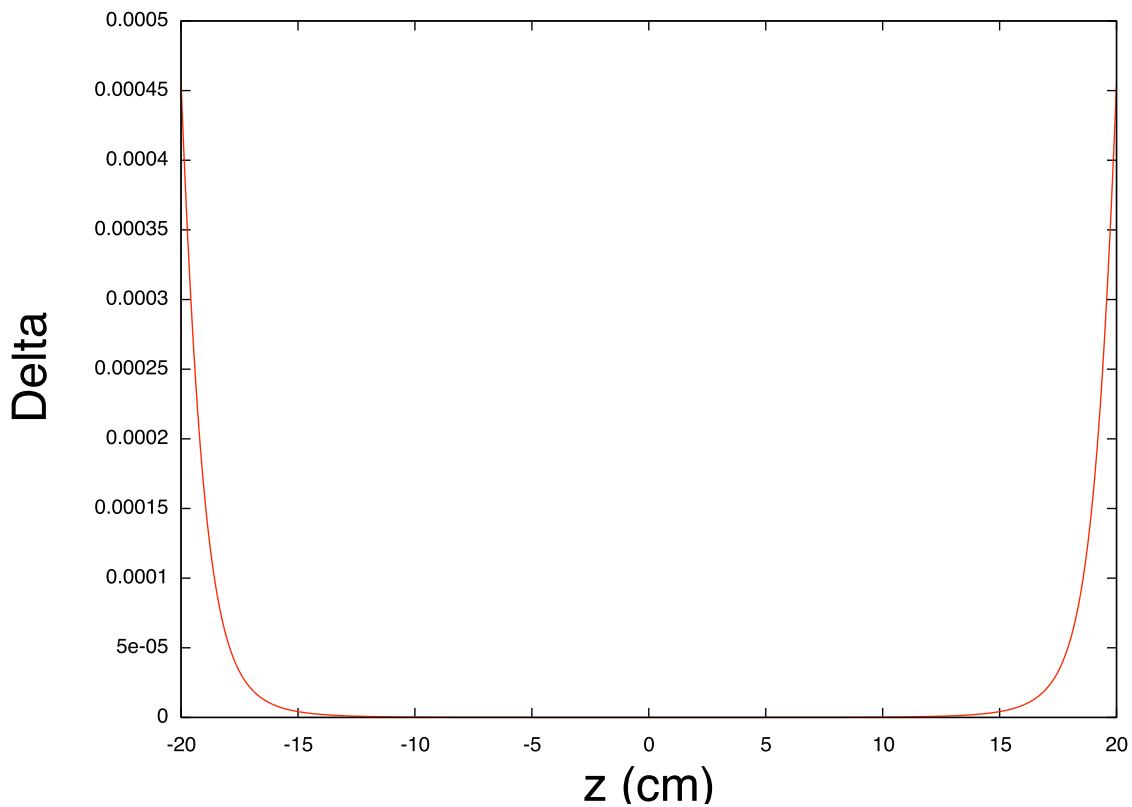


Figure 24.2.7: Place holder. The quantity  $\Delta = |\Delta|$  as a function of  $\xi$  and  $z$  in the vicinity of the design orbit and in the plane  $y = 1$ . The quantities  $\xi$ ,  $y$ , and  $z$  are in cm

## 24.3 Comparison of Design Orbits

How accurate are design orbits computed using surface methods? In this subsection we will use surface methods to compute a design orbit for the case of a magnetic monopole doublet. Comparison of this design orbit with the design orbit selected in Subsection 5.1 will provide a further indication of the accuracy of surface methods.

## 24.4 Terminating End Fields

Let us compute the magnetic field  $\mathbf{B}$  associated with the first two terms in (9.3). We find the result

$$\begin{aligned}\mathbf{B}(\mathbf{r}; X_0, Z_0) = & -[2ga/(X_0^2 + Z_0^2 + a^2)^{3/2}]\mathbf{e}_y \\ & + [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}](X_0x + Z_0z)\mathbf{e}_y \\ & + [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}][y(X_0\mathbf{e}_x + Z_0\mathbf{e}_z)].\end{aligned}\quad (24.4.1)$$

Next let us find the minimum vector potential  $\mathbf{A}^{\min}$  associated with the first two terms in (9.3). Begin by decomposing  $\mathbf{B}$  into homogeneous polynomials by rewriting (9.4) in the form (2.109) with

$$\mathbf{B}^0(\mathbf{r}; X_0, Z_0) = -[2ga/(X_0^2 + Z_0^2 + a^2)^{3/2}]\mathbf{e}_y \quad (24.4.2)$$

and

$$\mathbf{B}^1(\mathbf{r}; X_0, Z_0) = [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}][(X_0x + Z_0z)\mathbf{e}_y + y(X_0\mathbf{e}_x + Z_0\mathbf{e}_z)]. \quad (24.4.3)$$

The minimum vector potential associated with this magnetic field is given by the relations (2.109) through (2.111). Working out the indicated cross products yields the results

$$\mathbf{A}^{\min 1}(\mathbf{r}; X_0, Z_0) = [ga/(X_0^2 + Z_0^2 + a^2)^{3/2}](-ze_x + xe_z), \quad (24.4.4)$$

$$\begin{aligned}\mathbf{A}^{\min 2}(\mathbf{r}; X_0, Z_0) = & [-2ga/(X_0^2 + Z_0^2 + a^2)^{5/2}] \times \\ & [(Z_0y^2 - Z_0z^2 - X_0xz)\mathbf{e}_x + (X_0yz - Z_0xy)\mathbf{e}_y + (X_0x^2 + Z_0xz - X_0y^2)\mathbf{e}_z].\end{aligned}\quad (24.4.5)$$

Simple calculation verifies that there are the relations

$$\nabla \times \mathbf{A}^{\min 1}(\mathbf{r}; X_0, Z_0) = \mathbf{B}^0(\mathbf{r}; X_0, Z_0), \quad (24.4.6)$$

$$\nabla \times \mathbf{A}^{\min 2}(\mathbf{r}; X_0, Z_0) = \mathbf{B}^1(\mathbf{r}; X_0, Z_0), \quad (24.4.7)$$

as desired. We note that  $\mathbf{A}^{\min 1}$  falls off as  $(1/|X_0|)^3$  or  $(1/|Z_0|)^3$  for large  $|X_0|$  or  $|Z_0|$ , and  $\mathbf{A}^{\min 2}$  falls off as  $(1/|X_0|)^4$  or  $(1/|Z_0|)^4$ . In general, successive  $\mathbf{A}^{\min n}$  fall off with ever increasing powers of  $(1/|X_0|)$  or  $(1/|Z_0|)$ .

$$\mathbf{e}_\xi = \cos \theta \mathbf{e}_x - \sin \theta \mathbf{e}_z, \quad (24.4.8)$$

$$\mathbf{e}_\eta = \mathbf{e}_y, \quad (24.4.9)$$

$$\mathbf{e}_\zeta = \sin \theta \mathbf{e}_x + \cos \theta \mathbf{e}_z. \quad (24.4.10)$$

$$\mathbf{e}_\zeta \times \mathbf{e}_\xi = -\sin^2 \theta (\mathbf{e}_x \times \mathbf{e}_z) + \cos^2 \theta (\mathbf{e}_z \times \mathbf{e}_x) = \mathbf{e}_y = \mathbf{e}_\eta; \quad (24.4.11)$$

$$\mathbf{e}_x = \cos \theta \mathbf{e}_\xi + \sin \theta \mathbf{e}_\zeta, \quad (24.4.12)$$

$$\mathbf{e}_y = \mathbf{e}_\eta \quad (24.4.13)$$

$$\mathbf{e}_z = -\sin \theta \mathbf{e}_\xi + \cos \theta \mathbf{e}_\zeta. \quad (24.4.14)$$

$$\mathbf{r} = x \mathbf{e}_x + y \mathbf{e}_y + z \mathbf{e}_z = \xi \mathbf{e}_\xi + \eta \mathbf{e}_\eta + \zeta \mathbf{e}_\zeta. \quad (24.4.15)$$

$$\begin{aligned} x &= \mathbf{r} \cdot \mathbf{e}_x = (\xi \mathbf{e}_\xi + \eta \mathbf{e}_\eta + \zeta \mathbf{e}_\zeta) \cdot \mathbf{e}_x \\ &= \xi \mathbf{e}_\xi \cdot \mathbf{e}_x + \eta \mathbf{e}_\eta \cdot \mathbf{e}_x + \zeta \mathbf{e}_\zeta \cdot \mathbf{e}_x \\ &= \xi \cos \theta + \zeta \sin \theta. \end{aligned} \quad (24.4.16)$$

$$\begin{aligned} y &= \mathbf{r} \cdot \mathbf{e}_y = (\xi \mathbf{e}_\xi + \eta \mathbf{e}_\eta + \zeta \mathbf{e}_\zeta) \cdot \mathbf{e}_y \\ &= \xi \mathbf{e}_\xi \cdot \mathbf{e}_y + \eta \mathbf{e}_\eta \cdot \mathbf{e}_y + \zeta \mathbf{e}_\zeta \cdot \mathbf{e}_y \\ &= \eta. \end{aligned} \quad (24.4.17)$$

$$\begin{aligned} z &= \mathbf{r} \cdot \mathbf{e}_z = (\xi \mathbf{e}_\xi + \eta \mathbf{e}_\eta + \zeta \mathbf{e}_\zeta) \cdot \mathbf{e}_z \\ &= \xi \mathbf{e}_\xi \cdot \mathbf{e}_z + \eta \mathbf{e}_\eta \cdot \mathbf{e}_z + \zeta \mathbf{e}_\zeta \cdot \mathbf{e}_z \\ &= -\xi \sin \theta + \zeta \cos \theta. \end{aligned} \quad (24.4.18)$$

$$\mathbf{B}^0(\mathbf{r}; X_0, Z_0) = -[2ga/(X_0^2 + Z_0^2 + a^2)^{3/2}] \mathbf{e}_\eta. \quad (24.4.19)$$

$$\begin{aligned} \mathbf{A}^{\min 1}(\xi, \eta, \zeta; X_0, Z_0) &= -(1/2) \mathbf{r} \times \mathbf{B}^0(\mathbf{r}; X_0, Z_0) \\ &= [ga/(X_0^2 + Z_0^2 + a^2)^{3/2}](-\zeta \mathbf{e}_\xi + \xi \mathbf{e}_\zeta), \end{aligned} \quad (24.4.20)$$

$$\mathbf{A}^{\min 1}(\xi, \eta, 0; X_0, Z_0) = [ga/(X_0^2 + Z_0^2 + a^2)^{3/2}] (\xi \mathbf{e}_\zeta), \quad (24.4.21)$$

$$A_\xi^{\min 1}(\xi, \eta, 0; X_0, Z_0) = A_\eta^{\min 1}(\xi, \eta, 0; X_0, Z_0) = 0. \quad (24.4.22)$$

$$\begin{aligned} \mathbf{B}^1(\mathbf{r}; X_0, Z_0) &= [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}] \times \\ &\quad [(X_0 x + Z_0 z) \mathbf{e}_y + y(X_0 \mathbf{e}_x + Z_0 \mathbf{e}_z)]. \end{aligned} \quad (24.4.23)$$

$$(X_0 x + Z_0 z) \mathbf{e}_y = [x_0(\xi \cos \theta + \zeta \sin \theta) + Z_0(-\xi \sin \theta + \zeta \cos \theta)] \mathbf{e}_\eta, \quad (24.4.24)$$

$$\begin{aligned} y(X_0\mathbf{e}_x + Z_0\mathbf{e}_z) &= \eta[X_0(\cos\theta\mathbf{e}_\xi + \sin\theta\mathbf{e}_\zeta) + Z_0(-\sin\theta\mathbf{e}_\xi + \cos\theta\mathbf{e}_\zeta)] \\ &= \eta[(X_0\cos\theta - Z_0\sin\theta)\mathbf{e}_\xi + (X_0\sin\theta + Z_0\cos\theta)\mathbf{e}_\zeta]. \end{aligned} \quad (24.4.25)$$

$$\begin{aligned} \mathbf{B}^1(\xi, \eta, \zeta; X_0, Z_0) &= [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}] \times \\ &\quad \{[X_0(\xi\cos\theta + \zeta\sin\theta) + Z_0(-\xi\sin\theta + \zeta\cos\theta)]\mathbf{e}_\eta \\ &\quad + \eta[(X_0\cos\theta - Z_0\sin\theta)\mathbf{e}_\xi + (X_0\sin\theta + Z_0\cos\theta)\mathbf{e}_\zeta]\}. \end{aligned} \quad (24.4.26)$$

$$\begin{aligned} \mathbf{B}^1(\xi, \eta, 0; X_0, Z_0) &= [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}] \times \\ &\quad \{[X_0(\xi\cos\theta) + Z_0(-\xi\sin\theta)]\mathbf{e}_\eta \\ &\quad + \eta[(X_0\cos\theta - Z_0\sin\theta)\mathbf{e}_\xi + (X_0\sin\theta + Z_0\cos\theta)\mathbf{e}_\zeta]\}. \end{aligned} \quad (24.4.27)$$

$$\mathbf{r}(\xi, \eta, \zeta) = \xi\mathbf{e}_\xi + \eta\mathbf{e}_\eta + \zeta\mathbf{e}_\zeta. \quad (24.4.28)$$

$$\mathbf{r}(\xi, \eta, 0) = \xi\mathbf{e}_\xi + \eta\mathbf{e}_\eta. \quad (24.4.29)$$

$$\begin{aligned} \mathbf{e}_\eta \times \mathbf{B}^1(\xi, \eta, 0; X_0, Z_0) &= [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}] \times \\ &\quad \{\eta[(-X_0\cos\theta + Z_0\sin\theta)\mathbf{e}_\zeta + (X_0\sin\theta + Z_0\cos\theta)\mathbf{e}_\xi]\}. \end{aligned} \quad (24.4.30)$$

$$\begin{aligned} [\eta\mathbf{e}_\eta \times \mathbf{B}^1(\xi, \eta, 0; X_0, Z_0)]_\xi &= [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}] \times \\ &\quad \{\eta^2[(X_0\sin\theta + Z_0\cos\theta)]\}. \end{aligned} \quad (24.4.31)$$

$$\begin{aligned} \mathbf{e}_\xi \times \mathbf{B}^1(\xi, \eta, 0; X_0, Z_0) &= [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}] \times \\ &\quad \{[X_0(\xi\cos\theta) + Z_0(-\xi\sin\theta)]\mathbf{e}_\zeta \\ &\quad - \eta[(X_0\sin\theta + Z_0\cos\theta)\mathbf{e}_\eta]\}. \end{aligned} \quad (24.4.32)$$

$$\begin{aligned} [\xi\mathbf{e}_\xi \times \mathbf{B}^1(\xi, \eta, 0; X_0, Z_0)]_\eta &= [6ga/(X_0^2 + Z_0^2 + a^2)^{5/2}] \times \\ &\quad \{-\xi\eta[(X_0\sin\theta + Z_0\cos\theta)]\}. \end{aligned} \quad (24.4.33)$$

$$\mathbf{A}^{\min 2}(\xi, \eta, 0; X_0, Z_0) = -(1/3)\mathbf{r}(\xi, \eta, 0) \times \mathbf{B}^1(\xi, \eta, 0; X_0, Z_0). \quad (24.4.34)$$

$$A_{\xi}^{\min 2}(\xi, \eta, 0; X_0, Z_0) = [-2ga/(X_0^2 + Z_0^2 + a^2)^{5/2}](X_0 \sin \theta + Z_0 \cos \theta)\eta^2, \quad (24.4.35)$$

$$A_{\eta}^{\min 2}(\xi, \eta, 0; X_0, Z_0) = [2ga/(X_0^2 + Z_0^2 + a^2)^{5/2}](X_0 \sin \theta + Z_0 \cos \theta)\xi\eta. \quad (24.4.36)$$

## 24.5 Gauge Transformation Map

## 24.6 Pole Face Rotation

## 24.7 Comparison of Maps

How accurate are maps computed using surface methods? In this subsection we will use surface methods to compute the transfer map about the design orbit found in Subsection 5.3. We will also compute the exact transfer map for the case of a magnetic monopole doublet. Comparison of these maps will provide a final indication of the accuracy of surface methods.

## 24.8 Smoothing and Insensitivity to Errors

### Exercises

**24.8.1.** Show that any orbit having the initial conditions  $Y = 0$  and  $P_y = 0$  when  $z = 0$  must lie in the  $y = 0$  plane.

**24.8.2.** Show that, in the case of a  $30^\circ$  bend produced by a magnetic monopole doublet, one expects the asymptotic behavior

$$\lim_{z \rightarrow \mp\infty} X'(z) = \pm \tan(15^\circ) = \pm .267949 \dots \quad (24.8.1)$$

Actually, in the numerical computations for Section 20.5.1,  $p^0$  was chosen so that

$$X'(z = \mp 20) = \pm \tan(15^\circ) = \pm .267949 \dots \quad (24.8.2)$$

See (5.47) and (5.48). From Figure 5.3 we observe that “asymptopia” has essentially been achieved when  $|z| \geq 20$  so that the requirements (5.113) and (5.114) are nearly equivalent.

**24.8.3.** Consider  $A_z(x, y, z)$  as given by (2.104). Under the assumption that  $|x|$  increases linearly with  $|z|$ , as it does for large  $|z|$  on the design orbit shown in Figure 5.1, find the midplane,  $y = 0$ , asymptotic behavior of  $A_z(x, y, z)$  for large  $|z|$ . Do the same for  $B_y(x, y, z)$ . Verify that the results you obtain are consistent with Figures 5.4 and 5.5.

**24.8.4.** Verify the parameterizations (5.80) through (5.83), (5.85) through (5.88), (5.95) through (5.97), and (5.99) through (5.101). Verify the surface elements (5.84), (5.89), (5.98), and (5.102).



# Chapter 25

## Realistic Transfer Maps for General Curved Beam-Line Elements: Application to a Storage-Ring Dipole



# Bibliography

## General References

- [1] C. Mitchell, “Calculation of Realistic Charged-Particle Transfer Maps”, University of Maryland Physics Department Ph.D. Thesis (2007).



# Chapter 26

## The Euclidean Group and Error Effects

### 26.1 The Euclidean Group



# Chapter 27

## Representations of $sp(2n)$ and Related Matters

Historically there are several mathematical groups that have been studied in detail because of their relevance to our understanding of the physical world. A detailed knowledge of the 3-dimensional rotation group is of great use in many areas including rigid body dynamics, condensed matter physics, chemistry, atomic physics, nuclear physics, and elementary particle physics. Knowledge of the rotation-translation group leads to a classification of crystals and quasicrystals. Knowledge of the Lorentz group leads to the construction of spinors, 4-vectors, general tensors, and classical fields. Knowledge of the Poincaré group (the Lorentz group plus translations in space and time) leads to a classification of elementary particles and the construction of quantum fields. An understanding of the invariants of the full group of space-time diffeomorphisms plays a role in general relativity. Knowledge of various other groups, including  $E_8$ , facilitates many-body theory calculations. Finally, there are the various “internal” and/or gauge symmetry groups that play an important role in our current understanding of elementary particles and the fundamental forces.

We have seen that the symplectic group is the underlying group for Hamiltonian systems. Yet, in contrast to most of the groups just mentioned, almost nothing is commonly known or readily available about the symplectic group. For example, many readers will be familiar with some aspects of the rotation group including spin (irreducible representations and how they are labeled) and how spins couple and combine (the Clebsch-Gordan series and coefficients for the rotation group). Yet few have heard or read about representations of the symplectic group, knowledge of its Clebsch-Gordan series is not widespread, and little is known in detail about its Clebsch-Gordan coefficients.

The purpose of this chapter is to describe some aspects of the finite-dimensional representations of the first few symplectic groups with the hope that this knowledge, like that for the well-studied groups, will also ultimately prove useful. [What we will actually be finding are the finite dimensional irreducible unitary representations of  $usp(2n)$ , which is equivalent to  $sp(2n, \mathbb{R})$  over the complex field. See Sections 5.10 and 7.3.] Indeed, as a first consequence of this effort, we will find a symplectic classification of all analytic vector fields. Additional applications will be made in Chapters 29, 32, and 33.

## 27.1 Structure of $sp(2, \mathbb{R})$

The Lie algebra  $sp(2, \mathbb{R})$  is generated by the Lie operators associated with the quadratic polynomials  $q^2$ ,  $qp$ , and  $p^2$ . See Section 5.6. For present purposes it is convenient to introduce the basis polynomials

$$J_3 = -(i/4)(p^2 + q^2) \quad (27.1.1)$$

$$J_{\pm} = (1/4)(q \pm ip)^2. \quad (27.1.2)$$

They obey the Poisson bracket rules

$$[J_3, J_{\pm}] = \pm J_{\pm}, \quad (27.1.3)$$

$$[J_+, J_-] = 2J_3. \quad (27.1.4)$$

These rules are the familiar ones for angular momentum, and indicate that the Lie algebras  $so(3, \mathbb{R})$ ,  $su(2)$ ,  $sp(2, \mathbb{R})$ , and  $usp(2)$  are equivalent when one works over the complex field as in (1.1) and (1.2).

As a result of (7.3.14) through (7.3.16) we have the relations

$$: J_3 :^{\dagger} =: J_3 :, \quad (27.1.5)$$

$$: J_{\pm} :^{\dagger} =: J_{\mp} :. \quad (27.1.6)$$

Thus,  $: J_3 :$  is Hermitian. Finally, consider Lie transformations of the form

$$\mathcal{M}(\theta) = \exp(-i\theta : J_3 :) = \exp[-(\theta/4) : p^2 + q^2 :]. \quad (27.1.7)$$

These transformations are both real symplectic and unitary. Indeed, they have the periodicity property

$$\mathcal{M}(\theta + 4\pi) = \mathcal{M}(\theta), \quad (27.1.8)$$

and therefore form a *maximal torus* in  $Sp(2, \mathbb{R})$ . See Sections 3.9 and 7.2.

We next bring the rules (1.3) and (1.4) to Cartan form. For a review of the Cartan form for a Lie algebra, see Section 5.8. Let  $\mathbf{e}^1$  be a unit vector. For the case of  $sp(2)$  it is convenient to introduce root vectors  $\pm\boldsymbol{\alpha}$  by the relations

$$\pm \boldsymbol{\alpha} = \pm 2\mathbf{e}^1. \quad (27.1.9)$$

(Observe that they have length 2.) Then we have the normalization relation

$$\sum_{\boldsymbol{\mu}} (\mathbf{e}^1 \cdot \boldsymbol{\mu})(\boldsymbol{\mu} \cdot \mathbf{e}^1) = 8. \quad (27.1.10)$$

Compare with (5.8.21). Next introduce quantities  $c^1$  and  $r(\pm\boldsymbol{\alpha})$  by the relations

$$c^1 = 2J_3 = -(i/2)(p^2 + q^2), \quad (27.1.11)$$

$$r(\pm\boldsymbol{\alpha}) = \sqrt{2}J_{\pm} = (\sqrt{2}/4)(q \pm ip)^2. \quad (27.1.12)$$

They evidently obey the Poisson bracket rules

$$[c^1, r(\boldsymbol{\mu})] = (\mathbf{e}^1 \cdot \boldsymbol{\mu})r(\boldsymbol{\mu}), \quad (27.1.13)$$

$$[r(\boldsymbol{\mu}), r(-\boldsymbol{\mu})] = (\mathbf{e}^1 \cdot \boldsymbol{\mu})c^1. \quad (27.1.14)$$

These are the Cartan rules for  $sp(2)$ . Note that there are the conjugacy relations

$$: c^1 :^\dagger =: c^1 :, \quad (27.1.15)$$

$$: r(\boldsymbol{\mu}) :^\dagger =: r(-\boldsymbol{\mu}) :, \quad (27.1.16)$$

and  $: c^1 :$  is Hermitian as desired. The root vectors  $\pm\alpha$  are shown in Figure 1.1. Finally, we note the pleasing fact that for the scalar product (7.3.12) the basis elements  $c^1$  and  $r(\boldsymbol{\mu})$  all have unit norm and, indeed, are orthonormal:

$$\langle c^1, c^1 \rangle = \langle r(\boldsymbol{\mu}), r(\boldsymbol{\mu}) \rangle = 1, \quad (27.1.17)$$

$$\langle c^1, r(\boldsymbol{\mu}) \rangle = 0, \quad (27.1.18)$$

$$\langle r(\boldsymbol{\mu}), r(-\boldsymbol{\mu}) \rangle = 0. \quad (27.1.19)$$

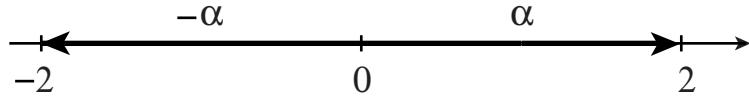


Figure 27.1.1: Root vectors for  $sp(2)$ .

## Exercises

**27.1.1.** Verify (1.3) through (1.8).

**27.1.2.** Define  $J_1$  and  $J_2$  by the rules

$$J_\pm = J_1 \pm iJ_2. \quad (27.1.20)$$

Verify the  $su(2)$  Poisson bracket rules

$$[J_1, J_2] = iJ_3, \text{ etc.} \quad (27.1.21)$$

**27.1.3.** Verify (1.10) and (1.13) through (1.19).

## 27.2 Representations of $sp(2, \mathbb{R})$

It is well known that irreducible representations of  $su(2)$  are labeled by a non-negative integer or half-integer  $j$ , and vectors within a representation are labeled by an integer or half-integer  $m$  that ranges between  $-j$  and  $j$  by integer steps. Let  $\hat{J}_3, \hat{J}_{\pm}$  be a set of irreducible matrices (with  $\hat{J}_3$  Hermitian) whose commutation rules are the same as the Poisson bracket rules (1.3) and (1.4). Then, in a representation labeled by  $j$ , there is a “highest” vector  $|j\rangle$  with  $m = j$  having the property

$$\begin{aligned}\hat{J}_3|j\rangle &= j|j\rangle, \text{ or} \\ (2\hat{J}_3)|j\rangle &= (2j)|j\rangle.\end{aligned}\tag{27.2.1}$$

In the terminology of Cartan, the vector  $|j\rangle$  is an eigenvector of highest weight with weight  $(2j)$ . [See (1.11).] It follows that the fundamental weight  $\phi^1$  for  $sp(2)$  is given by the relation

$$\phi^1 = e^1 = \alpha/2.\tag{27.2.2}$$

Correspondingly, the highest weight for a representation characterized by the non-negative integer  $n$ , with  $n = 2j$ , is given by the relation

$$w^h = n\phi^1, \quad n = 2j.\tag{27.2.3}$$

Call this representation  $\Gamma(n)$ .

Figure 2.1 shows the fundamental weight  $\phi^1$  along with the root vectors  $\pm\alpha$ . Figure 2.2 shows the weight diagrams for the first few representations. For  $su(2)$ , and hence  $sp(2, \mathbb{R})$ , each weight (vector within a representation) has unit multiplicity. It follows that the dimension of the representation  $\Gamma(n)$  is given by the relation

$$\dim \Gamma(n) = n + 1.\tag{27.2.4}$$

Sometimes we will label a representation by its dimension.

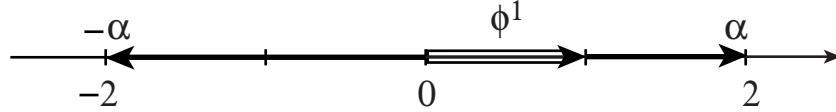


Figure 27.2.1: The fundamental weight  $\phi^1$  and the root vectors  $\pm\alpha$  for  $sp(2)$ .

Let  $\mathcal{P}_n$  denote the set of polynomials homogeneous of degree  $n$  in the variables  $q, p$ ; and let  $f_2$  be a quadratic polynomial in  $q, p$ . Then, in view of (7.6.14), we have the relation

$$:f_2:\mathcal{P}_n \subseteq \mathcal{P}_n.\tag{27.2.5}$$

It follows that the set of homogeneous polynomials of degree  $n$  forms a representation of  $sp(2, \mathbb{R})$ .

What representations occur? The study of this question is facilitated by using the map  $\mathcal{A}(\theta)$  defined by the equation

$$\mathcal{A}(\theta) = \exp(-i\theta :p^2 - q^2:).\tag{27.2.6}$$

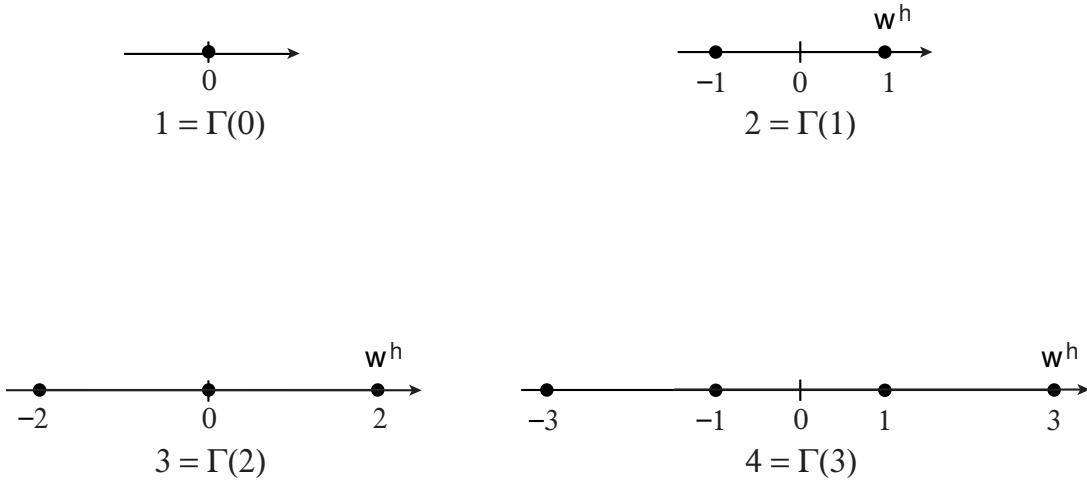


Figure 27.2.2: Weight diagrams for the  $sp(2)$  representations  $\Gamma(0)$ ,  $\Gamma(1)$ ,  $\Gamma(2)$ , and  $\Gamma(3)$ .

Evidently  $\mathcal{A}$  is *complex* symplectic and, in view of (7.3.14) through (7.3.16), it is also unitary. Calculation gives the results

$$\mathcal{A}(\theta)q = q \cos(2\theta) + ip \sin(2\theta), \quad (27.2.7)$$

$$\mathcal{A}(\theta)p = iq \sin(2\theta) + p \cos(2\theta). \quad (27.2.8)$$

In particular, there is the relation

$$\mathcal{A}(\pi/8)q = (1/\sqrt{2})(q + ip), \quad (27.2.9)$$

$$\mathcal{A}(\pi/8)p = i(1/\sqrt{2})(q - ip). \quad (27.2.10)$$

The map  $\mathcal{A}(\pi/8)$  is the operator analog of the matrix  $W$  given by (3.9.9). See Exercise 2.2. Also, since  $\mathcal{A}$  is a Lie transformation, there is the relation

$$\mathcal{A}(\pi/8)(q^r p^s) = [\mathcal{A}(\pi/8)q]^r [\mathcal{A}(\pi/8)p]^s = (1/\sqrt{2})^{r+s}(i)^s(q + ip)^r(q - ip)^s. \quad (27.2.11)$$

With the aid of  $\mathcal{A}$ , we define transformed Lie basis polynomials  $\tilde{c}^1$ ,  $\tilde{r}(\pm\alpha)$ , by the rule

$$\tilde{c}^1 = \mathcal{A}(-\pi/8)c^1 = -qp, \quad (27.2.12)$$

$$\tilde{r}(\alpha) = \mathcal{A}(-\pi/8)r(\alpha) = (1/\sqrt{2})q^2, \quad (27.2.13)$$

$$\tilde{r}(-\alpha) = \mathcal{A}(-\pi/8)r(-\alpha) = -(1/\sqrt{2})p^2. \quad (27.2.14)$$

Since  $\mathcal{A}$  is symplectic, the transformed basis polynomials obey the same Poisson bracket rules (1.13) and (1.14). See (5.4.14) and Section 6.3. Since  $\mathcal{A}$  is also unitary, as is easily verified, the transformed basis polynomials also satisfy the orthonormality relations (1.17) through (1.19) and the conjugacy relations (1.15) and (1.16).

Now consider the action of the Lie operators : $\tilde{c}^1$ : and : $\tilde{r}(\pm\alpha)$ : on the general monomial  $q^r p^s$ . Calculation gives the results

$$:\tilde{c}^1: q^r p^s = (r - s)q^r p^s, \quad (27.2.15)$$

$$:\tilde{r}(\alpha): q^r p^s = (\sqrt{2}) s q^{r+1} p^{s-1}, \quad (27.2.16)$$

$$:\tilde{r}(-\alpha): q^r p^s = (\sqrt{2}) r q^{r-1} p^{s+1}. \quad (27.2.17)$$

Evidently any monomial of a given degree can be transformed into any other monomial of the same degree with the aid of  $\tilde{r}(\pm\alpha)$ . Therefore  $sp(2)$  acts *irreducibly* on  $\mathcal{P}_n$ . Also  $q^n$  is the vector of highest weight in  $\mathcal{P}_n$ , and has the weight  $n\phi^1$ ,

$$:\tilde{c}^1: q^n = n q^n = (\mathbf{e}^1 \cdot n\phi^1) q^n. \quad (27.2.18)$$

We conclude that  $\mathcal{P}_n$  carries the representation  $\Gamma(n)$ . From (7.3.36) and (2.4) we find the result

$$\dim \mathcal{P}_n = N(n, 2) = n + 1 = \dim \Gamma(n). \quad (27.2.19)$$

The equality of  $\dim \mathcal{P}_n$  and  $\dim \Gamma(n)$  is to be expected from the fact that  $sp(2)$  acts irreducibly on  $\mathcal{P}_n$ . It can be shown that  $\Gamma(n)$  is self conjugate,

$$\bar{\Gamma}(n) = \Gamma(n). \quad (27.2.20)$$

See Exercise 3.7.36.

Let  $\mathcal{A}(\pi/8)$  act on both sides of (2.15) through (2.17). Then, for the left side of (2.15), we find the result

$$\begin{aligned} \mathcal{A}(\pi/8) : \tilde{c}^1 : q^r p^s &= \mathcal{A}(\pi/8)[\tilde{c}^1, q^r p^s] = [\mathcal{A}(\pi/8)\tilde{c}^1, \mathcal{A}(\pi/8)q^r p^s] \\ &= [c^1, \mathcal{A}(\pi/8)q^r p^s] =: c^1 : (1/\sqrt{2})^{r+s} (i)^s (q + ip)^r (q - ip)^s. \end{aligned} \quad (27.2.21)$$

For the right side we find

$$\mathcal{A}(\pi/8)(r-s)q^r p^s = (r-s)(1/\sqrt{2})^{r+s} (i)^s (q + ip)^r (q - ip)^s. \quad (27.2.22)$$

Therefore, after cancellation of common terms, (2.15) is transformed under the action of  $\mathcal{A}(\pi/8)$  to the relation

$$:(p^2 + q^2)/2 : (q + ip)^r (q - ip)^s = i(r-s)(q + ip)^r (q - ip)^s. \quad (27.2.23)$$

Similary, (2.16) and (2.17) are transformed to the relations

$$:(q + ip)^2 : (q + ip)^r (q - ip)^s = -4is(q + ip)^{r+1} (q - ip)^{s-1}, \quad (27.2.24)$$

$$:(q - ip)^2 : (q + ip)^r (q - ip)^s = 4ir(q + ip)^{r-1} (q - ip)^{s+1}. \quad (27.2.25)$$

The monomials  $q^r p^s$  (with  $r + s = n$ ) obviously form a basis for  $\mathcal{P}_n$ . The relations (2.9) through (2.11) show that the complex polynomials  $(q + ip)^r (q - ip)^s$  also form a basis for  $\mathcal{P}_n$  and, with the factors  $(1/\sqrt{2})^n (i^s)$ , the two bases are related by the symplectic and unitary transformations  $\mathcal{A}(\pm\pi/8)$ . According to (2.23) the polynomials  $(q + ip)^r (q - ip)^s$  are eigenfunctions of the harmonic oscillator Lie operator  $:(p^2 + q^2)/2:$ . For this reason, they will be referred to as the *resonance* basis. The utility of the resonance basis will become clear in Chapter 23. Since it is made from Cartesian components, the monomial basis  $q^r p^s$  will be referred to as the *Cartesian* basis.

## Exercises

**27.2.1.** The Lie algebras for  $su(2)$  and  $sp(2)$  are equivalent over the complex field. Yet, for purposes of observing how  $u(n)$  is embedded within  $sp(2n)$ , it is convenient to give  $su(2)$  and  $sp(2)$  different root vector structures. Specifically, for  $su(2)$  we define two root vectors  $\pm\alpha$  by the rules

$$\pm\alpha = \pm\sqrt{2}\mathbf{e}^1. \quad (27.2.26)$$

(Observe they have length  $\sqrt{2}$ ). Then the  $su(2)$  Lie algebra is spanned by the elements  $c^1$ ,  $r(\alpha)$ , and  $r(-\alpha)$ ; and the Cartan rules (1.13) and (1.14) give the results

$$[c^1, r(\pm\alpha)] = \{\mathbf{e}^1 \cdot (\pm\alpha)\}r(\pm\alpha) = \pm\sqrt{2}r(\pm\alpha), \quad (27.2.27)$$

$$[r(\alpha), r(-\alpha)] = (\mathbf{e}^1 \cdot \alpha)c^1 = \sqrt{2}c^1. \quad (27.2.28)$$

Consider the  $su(2)$  within  $sp(4)$  as described in Section 5.7. Upon making the identifications  $c \leftrightarrow c^1$  and  $r(\pm) \leftrightarrow r(\pm\alpha)$ , verify that the rules (5.7.10) and (5.7.11) are identical to (2.27) and (2.28). Show that  $C^1$  and  $R(\pm\alpha)$  as given by (5.8.8) and (5.8.11) satisfy analogous commutation rules, and therefore describe one of the  $su(2)$  subgroups within  $su(3)$ . Show that there are two other  $su(2)$  subgroups within  $su(3)$  corresponding to the use of  $R(\pm\beta)$  and  $R(\pm\gamma)$  and suitable linear combinations of the  $C^j$ . Note that all the  $su(3)$  root vectors in Figure 5.8.1 have length  $\sqrt{2}$ . With regard to representations of  $su(2)$ , call them  $\Gamma(n)$ , show that there is one such for each  $n$  value with  $n = 0, 1, 2, \dots$ . Show that the highest weight for  $\Gamma(n)$  is given by

$$\mathbf{w}^h = n\phi^1 \quad (27.2.29)$$

with

$$\phi^1 = (1/\sqrt{2})\mathbf{e}^1. \quad (27.2.30)$$

Draw  $su(2)$  weight diagrams for the first few representations. Verify that they are similar to those for  $sp(2)$ , see Figure 2.2, except that the spacing between dots is  $\sqrt{2}$  rather than 2. Examine the weight diagrams for  $su(3)$  as shown in Figures 5.8.3 through 5.8.8. Show that the spacing between the dots in the directions of the  $su(3)$  root vectors is  $\sqrt{2}$ . These dots describe  $su(2)$  representations within  $su(3)$ .

**27.2.2.** For a  $2n$ -dimensional phase space, let  $\mathcal{A}(\pi/8)$  be the map defined by the equation

$$\mathcal{A}(\pi/8) = \exp[-i(\pi/8) : (p_1^2 - q_1^2) + (p_2^2 - q_2^2) + \dots + (p_n^2 - q_n^2) :]. \quad (27.2.31)$$

Show that  $\mathcal{A}(\pi/8)$  has the property

$$\mathcal{A}(\pi/8)z_a = \sum_b W_{ab}z_b \quad (27.2.32)$$

where  $W$  is the matrix given by (3.9.9).

**27.2.3.** For first-order polynomials, and in analogy with (2.9) and (2.10), introduce the basis elements  $a^\pm$  defined by the relations

$$a^+ = (1/\sqrt{2})(p + iq) = \mathcal{A}(\pi/8)p, \quad (27.2.33)$$

$$a^- = (1/\sqrt{2})(p - iq) = -i\mathcal{A}(\pi/8)q. \quad (27.2.34)$$

Show that these elements satisfy the Poisson bracket relations

$$[a^+, a^-] = \mathcal{A}(\pi/8)[p, -iq] = i. \quad (27.2.35)$$

Let  $H$  be the harmonic oscillator Hamiltonian

$$H = (\omega/2)(p^2 + q^2). \quad (27.2.36)$$

Show that  $H$  can be written in the form

$$H = \omega a^+ a^-. \quad (27.2.37)$$

Use (2.35) to verify the equations of motion

$$\dot{a}^\pm = [a^\pm, H] = \omega[a^\pm, a^+ a^-] = \pm i\omega a^\pm. \quad (27.2.38)$$

Show that they have the solution

$$a^\pm(t) = a^\pm(0) \exp(\pm i\omega t). \quad (27.2.39)$$

From this result, find  $q(t)$  and  $p(t)$ .

### 27.3 Symplectic Classification of Analytic Vector Fields in Two Variables

Let  $\mathcal{L}_f$  be a general vector field in two variables  $z_1$  and  $z_2$  where  $f$  denotes the collection of 2 functions  $(f_1, f_2)$  as in Section 5.3. (The functions  $f_1$  and  $f_2$  may also depend on the time  $t$ , but for simplicity we will suppress this possible dependence in our notation because  $t$  only plays the role of a parameter.) Assume that  $f_1$  and  $f_2$  are analytic at some common point  $z_1^0, z_2^0$ . Without loss of generality we may take this point to be the origin. (If not, make a linear change of variables that sends  $z_1^0, z_2^0$  to the origin.) Then we may decompose the Taylor expansions of the components of  $f$  into sums of homogeneous polynomials and, in so doing, decompose  $\mathcal{L}_f$  into a sum of vector fields of the form  $\mathcal{L}_{f^n}$  where the components of  $f^n$  are homogeneous polynomials of degree  $n$ :

$$\mathcal{L}_f = \sum_{n=0}^{\infty} \mathcal{L}_{f^n}. \quad (27.3.1)$$

The homogeneous vector fields  $\mathcal{L}_{f^n}$  can now be considered individually.

Let  $\Sigma$  denote the vector field

$$\Sigma = \sum_a z_a (\partial/\partial z_a) = z_1 (\partial/\partial z_1) + z_2 (\partial/\partial z_2) = q (\partial/\partial q) + p (\partial/\partial p). \quad (27.3.2)$$

Then, by Euler's relation, we have the result

$$\#\Sigma\#\mathcal{L}_{f^n} = \{\Sigma, \mathcal{L}_{f^n}\} = (n-1)\mathcal{L}_{f^n}. \quad (27.3.3)$$

See Exercises 1.5.1 and 7.6.7. (Remark: Sometimes  $\Sigma$  is called the *Euler* field because of its connection with the Euler relation.) We will say that  $\mathcal{L}_{f^n}$  is homogeneous of degree  $(n-1)$ . In view of (5.3.3) and (5.3.17), in the special case of a Hamiltonian vector field  $:f_n:$  there is the result

$$\#\Sigma\# :f_n: = \{\Sigma, :f_n:\} = (n-2) :f_n:. \quad (27.3.4)$$

Thus, the vector field  $:f_n:$  is homogeneous of degree  $(n-2)$ . Finally it is easily verified that there is a grading relation of the form

$$\{\mathcal{L}_{f^\ell}, \mathcal{L}_{g^m}\} = \mathcal{L}_{h^n} \text{ with } n = \ell + m - 1. \quad (27.3.5)$$

Let  $f_2$  be a quadratic polynomial in  $q, p$ . Then, using (3.5), we have the relation

$$\#f_2\#\mathcal{L}_{g^m} = \{ :f_2:, \mathcal{L}_{g^m}\} = \mathcal{L}_{h^m}. \quad (27.3.6)$$

We draw the important conclusion that the set of homogeneous vector fields  $\mathcal{L}_{g^m}$  transforms under and forms a representation of  $sp(2, \mathbb{R})$ .

What irreducible representations occur? Consider first the case of Hamiltonian vector fields. In this case

$$\#f_2\# :g_m: = \{ :f_2:, :g_m:\} = [f_2, g_m] := (:f_2: g_m):. \quad (27.3.7)$$

It follows from the previous section that the Hamiltonian vector fields  $:g_m:$  are transformed into each other under the action of  $sp(2, \mathbb{R})$  and carry the irreducible representation  $\Gamma(m)$ .

What about general vector fields? Note that

$$:q := \partial/\partial p \text{ and } :p := -\partial/\partial q. \quad (27.3.8)$$

It follows that any  $\mathcal{L}_{g^0}$  is a Hamiltonian vector field, and these fields carry the representation  $\Gamma(1)$ . Next consider the vector fields  $\mathcal{L}_{g^1}$ . They evidently form a 4-dimensional space spanned by the vector fields  $z_a(\partial/\partial z_b)$  with  $a = 1, 2$  and  $b = 1, 2$ .<sup>1</sup> We know that any  $:h_2:$  is such a vector field, and that these vector fields carry the representation  $\Gamma(2)$ , which is 3 dimensional. Also,  $\Sigma$  is of the form  $\mathcal{L}_{g^1}$ . From (3.4) we conclude that

$$\#f_2\#\Sigma = \{ :f_2:, \Sigma\} = -\{\Sigma, :f_2:\} = -\#\Sigma\# :f_2: = 0. \quad (27.3.9)$$

Consequently,  $\Sigma$  carries the representation  $\Gamma(0)$ . It follows that any  $\mathcal{L}_{g^1}$  can be written uniquely in the form

$$\begin{aligned} \mathcal{L}_{g^1} &= \lambda_1 :q^2: + \lambda_2 :qp: + \lambda_3 :p^2: + \lambda_4 \Sigma =: \lambda_1 q^2 + \lambda_2 qp + \lambda_3 p^2: + \lambda_4 \Sigma \\ &= :h_2: + \lambda_4 \Sigma. \end{aligned} \quad (27.3.10)$$

<sup>1</sup>Let  $z_1, z_2, \dots, z_m$  be  $m$  variables. Consider the  $m^2$  vector fields  $z_a(\partial/\partial z_b)$ . They can be shown to form a basis for the Lie algebra  $gl(m)$ . See Exercise 10.8.

The term  $: h_2 :$  is a member of the representation  $\Gamma(2)$ , and the term  $\lambda_4 \Sigma$  belongs to  $\Gamma(0)$ . That is, the vector fields  $\mathcal{L}_{\mathbf{g}^1}$  carry as a direct sum the representations  $\Gamma(2)$  and  $\Gamma(0)$ . Note that

$$\dim \mathcal{L}_{\mathbf{g}^1} = 4 = 3 + 1 = \dim \Gamma(2) + \dim \Gamma(0), \quad (27.3.11)$$

as required.

With this background in mind, let us consider the general case  $\mathcal{L}_{\mathbf{g}^m}$  with  $m \geq 1$ . Any such vector field can be written in the form

$$\mathcal{L}_{\mathbf{g}^m} = \sum_{a=1}^2 g_a^m (\partial/\partial z_a) \quad (27.3.12)$$

where  $g_1^m$  and  $g_2^m$  denote two homogeneous polynomials of degree  $m$ . We have just learned that the  $(\partial/\partial z_a)$  carry the representation  $\Gamma(1)$ , and we know from the previous section that the  $g_a^m$  carry the representation  $\Gamma(m)$ . It follows from the derivation property of  $\# f_2 \#$  that  $\mathcal{L}_{\mathbf{g}^m}$  must carry the direct product representation  $\Gamma(m) \otimes \Gamma(1)$ . See Exercise 3.2. Also, in the case of  $sp(2)$ , we have the Clebsch-Gordan series result

$$\Gamma(m) \otimes \Gamma(1) = \Gamma(m+1) \oplus \Gamma(m-1). \quad (27.3.13)$$

This is just the  $sp(2)$  analog of the familiar statement that spin  $m/2$  and spin  $1/2$  combine to make spin  $(m+1)/2$  and spin  $(m-1)/2$ . Recall (5.8.33) and remember that for the purposes of the present section and previous section we have made the definition  $n = 2j$ .

It follows that any  $\mathcal{L}_{\mathbf{g}^m}$  with  $m \geq 1$  has the unique decomposition

$$\mathcal{L}_{\mathbf{g}^m} = : h_{m+1} : + \mathcal{G}^{m-1}. \quad (27.3.14)$$

Here  $h_{m+1}$  is a unique homogeneous polynomial of degree  $m+1$  that is a member of the representation  $\Gamma(m+1)$  and  $\mathcal{G}^{m-1}$  is a unique vector field homogeneous of degree  $(m-1)$  that is a member of the representation  $\Gamma(m-1)$ . Let us introduce the notation

$$\mathcal{H}^{m+1} = : h_{m+1} : \quad (27.3.15)$$

to denote a Hamiltonian vector field that carries the representation  $\Gamma(m+1)$ . Then (3.13) can be written in the form

$$\mathcal{L}_{\mathbf{g}^m} = \mathcal{H}^{m+1} + \mathcal{G}^{m-1}. \quad (27.3.16)$$

We define  $\mathcal{G}^{m-1}$  to be the *non-Hamiltonian* part of  $\mathcal{L}_{\mathbf{g}^m}$ . What we have learned is that any homogeneous polynomial vector field in two variables can be uniquely decomposed into a Hamiltonian and a non-Hamiltonian part. We will learn subsequently that this result holds in any (even) number of variables.

In the case of two variables there is an additional step that can be made. Consider any vector field of the form  $f_{m-1} \Sigma$ . In view of (3.9) this vector field is a member of the representation  $\Gamma(m-1)$ . Thus, in the case of two variables we may write

$$\mathcal{L}_{\mathbf{g}^m} = : h_{m+1} : + f_{m-1} \Sigma \quad (27.3.17)$$

where both  $h_{m+1}$  and  $f_{m-1}$  are uniquely determined.

As a simple, but instructive, example of the decomposition just described, consider the case of the damped harmonic oscillator described by the equation of motion

$$\ddot{q} + 2\beta\dot{q} + q = 0. \quad (27.3.18)$$

This equation can be rewritten in the first-order form

$$\dot{q} = p, \quad (27.3.19)$$

$$\dot{p} = -(q + 2\beta p). \quad (27.3.20)$$

These equations can next be expressed in the Lie form

$$\dot{q} = \mathcal{L}q, \quad \dot{p} = \mathcal{L}p \quad (27.3.21)$$

where  $\mathcal{L}$  is the vector field

$$\mathcal{L} = p(\partial/\partial q) - (q + 2\beta p)(\partial/\partial p). \quad (27.3.22)$$

Evidently  $\mathcal{L}$  is of the form (3.10). By comparing coefficients we find the decomposition

$$\mathcal{L} = :-(p^2 + 2\beta pq + q^2)/2:-\beta\Sigma. \quad (27.3.23)$$

It is easily verified that the Hamiltonian  $(p^2 + 2\beta pq + q^2)/2$  produces simple harmonic motion with a frequency  $\omega_1$  given by the relation

$$\omega_1^2 = 1 - \beta^2. \quad (27.3.24)$$

Also, in this case, the vector fields  $\mathcal{H}^2$  and  $\mathcal{G}^0$  commute. Since the vector field  $-\beta\Sigma$  produces exponential decay, it follows that the general solution to (3.18) is of the form

$$q = Ae^{-\beta t} \sin(\omega_1 t + \phi). \quad (27.3.25)$$

What we have learned is that damping contributes both a non-Hamiltonian and a Hamiltonian part to the vector field. The non-Hamiltonian part produces exponential decay, and the Hamiltonian part shifts the frequency. For further detail, see Exercises 3.7 through 3.10.

## Exercises

**27.3.1.** A Lie algebra is called *simple* if it has no invariant subalgebras (ideals). See Section 8.9. Show that the Lie algebra  $su(2)$ , and hence also  $sp(2)$  and  $so(3)$ , is simple. Show that  $su(3)$  is simple. What are the ranks of  $sp(2)$ ,  $so(3)$ ,  $su(2)$ , and  $su(3)$ ? See Section 5.8.

**27.3.2.** Show that  $\#f_2\#$  has the derivation property

$$\#f_2\# \sum_a g_a (\partial/\partial z_a) = \sum_a (: f_2 : g_a) (\partial/\partial z_a) + \sum_a g_a \#f_2\# (\partial/\partial z_a). \quad (27.3.26)$$

**27.3.3.** Compare the dimensions of both sides of (3.13).

**27.3.4.** Use (5.3.26), and the discussion surrounding it, as well as (3.16) and (3.17) to show that  $\mathcal{G}^{m-1}$  is non-Hamiltonian.

**27.3.5.** Show that  $\Sigma$  can be written in the form

$$\Sigma = - \sum_{a,b} z_a J_{ab} : z_b : . \quad (27.3.27)$$

We know that both the  $z_a$  and the  $: z_b :$  transform according to  $\Gamma(1)$ . But  $\Sigma$  carries the representation  $\Gamma(0)$ . It follows that the numbers  $J_{ab}$  are the Clebsch-Gordan *coefficients* that couple  $\Gamma(1) \otimes \Gamma(1)'$  down to  $\Gamma(0)$ .

**27.3.6.** Consider the vector  $\mathcal{L}_{\mathbf{g}^2}$  given by

$$\mathcal{L}_{\mathbf{g}^2} = q^2 (\partial/\partial q). \quad (27.3.28)$$

Find  $h_3$  and  $f_1$  as in (3.15) for this vector field.

**27.3.7.** Verify (3.18) through (3.23). Verify that the vector field  $-\beta\Sigma$  produces exponential decay,

$$e^{-t\beta\Sigma} q = e^{-\beta t} q, \quad (27.3.29)$$

$$e^{-t\beta\Sigma} p = e^{-\beta t} p. \quad (27.3.30)$$

For the Hamiltonian

$$H = (p^2 + 2\beta pq + q^2)/2 \quad (27.3.31)$$

make the transformation of variables

$$q = \frac{1}{\sqrt{2}}(Q - P), \quad (27.3.32)$$

$$p = \frac{1}{\sqrt{2}}(Q + P). \quad (27.3.33)$$

Verify that this transformation is symplectic, and hence  $H$  is transformed to  $H'$  with

$$H' = (1/2)[(1 - \beta)P^2 + (1 + \beta)Q^2]. \quad (27.3.34)$$

Also, show that  $\Sigma$  is unchanged by this transformation,

$$q(\partial/\partial q) + p(\partial/\partial p) = Q(\partial/\partial Q) + P(\partial/\partial P). \quad (27.3.35)$$

Next make a second transformation of variables,

$$Q = [(1 - \beta)/(1 + \beta)]^{1/4} \bar{q}. \quad (27.3.36)$$

$$P = [(1 + \beta)/(1 - \beta)]^{1/4} \bar{p}. \quad (27.3.37)$$

Verify that this transformation is also symplectic, and hence  $H'$  is transformed to  $H''$  with

$$H'' = (1/2)(1 - \beta^2)^{1/2}(\bar{p}^2 + \bar{q}^2) = (\omega_1/2)(\bar{p}^2 + \bar{q}^2). \quad (27.3.38)$$

Again show that  $\Sigma$  is unchanged,

$$Q(\partial/\partial Q) + P(\partial/\partial P) = \bar{q}(\partial/\partial \bar{q}) + \bar{p}(\partial/\partial \bar{p}). \quad (27.3.39)$$

(This is the *second* time that  $\Sigma$  is unchanged. Why must this be? See Exercise 9.4.) Evidently, in accord with previous claims,  $H''$  produces simple harmonic motion with frequency  $\omega_1$ . And, since  $\Sigma$  is unchanged, the new variables still exhibit the same exponential decay as that in (3.29) and (3.30).

**27.3.8.** The oscillator described by (3.18) is underdamped when  $\beta < 1$ , critically damped when  $\beta = 1$ , and overdamped when  $\beta > 1$ . Exercise 3.6 deals with the underdamped case. Carry out a similar analysis for the critically and overdamped cases. Hint: For the overdamped case,  $H'$  as given by (3.34) produces *hyperbolic* motion. In this case, make the transformation of variables

$$Q = [(\beta - 1)/(\beta + 1)]^{1/4} \bar{q}, \quad (27.3.40)$$

$$P = [(\beta + 1)/(\beta - 1)]^{1/4} \bar{p}. \quad (27.3.41)$$

Verify that this transformation is symplectic and hence  $H'$  is transformed to  $H''$  with

$$H'' = (1/2)(\beta^2 - 1)^{1/2}(-\bar{p}^2 + \bar{q}^2). \quad (27.3.42)$$

Show that  $H''$  produces growth that goes like  $\exp[t(\beta^2 - 1)^{1/2}]$  as well as decay that goes as  $\exp[-t(\beta^2 - 1)^{1/2}]$ . For large  $\beta$  the growth rate of the growing term is almost as large as the decay rate in (3.29) and (3.30). They therefore nearly cancel. The net and well known result is that it takes a very long time for a highly overdamped oscillator to come to rest.

**27.3.9.** Find a pair of differential equations of the form

$$\dot{\bar{q}} = \dots,$$

$$\dot{\bar{p}} = \dots,$$

by expressing  $q, p$  in terms of  $\bar{q}, \bar{p}$  with the aid of (3.32) through (3.37) and using (3.19) and (3.20). Find the vector field for these differential equations and decompose it into Hamiltonian and non-Hamiltonian parts. Solve the differential equations.

**27.3.10.** Show that  $H$  as given by (3.31) and  $\mathcal{L}$  as given by (3.23) have the property

$$\mathcal{L}H = -\beta\Sigma H = -2\beta H. \quad (27.3.43)$$

Therefore,  $H$  must evolve according to the *nonoscillatory* rule

$$H = (\text{constant}) \times e^{-2\beta t}. \quad (27.3.44)$$

Verify directly from (3.19) and (3.25) that (3.44) is, in fact, correct.

**27.3.11.** Let  $G$  be the function

$$G = az_1^2 + bz_1z_2 + cz_2^2 = aq^2 + bqp + cp^2. \quad (27.3.45)$$

Find the associated gradient vector field  $\mathcal{L}_G$ . See Exercise 5.3.7. Decompose  $\mathcal{L}_G$  into Hamiltonian and non-Hamiltonian parts. Find  $G$  such that

$$\mathcal{L}_G = \Sigma. \quad (27.3.46)$$

Can a gradient vector field ever also be a Hamiltonian vector field?

**27.3.12.** The Van der Pol oscillator is described by the differential equation

$$\ddot{q} - 2\lambda(1 - q^2)\dot{q} + q = 0 \quad (27.3.47)$$

with  $\lambda > 0$ . Upon making the definition  $p = \dot{q}$ , show that (3.47) is produced by the vector field

$$\mathcal{L} = p(\partial/\partial q) - (q - 2\lambda p)(\partial/\partial p) - 2\lambda q^2 p(\partial/\partial p). \quad (27.3.48)$$

Evidently  $\mathcal{L}$  has the homogeneous decomposition

$$\mathcal{L} = \mathcal{L}_{\mathbf{g}^1} + \mathcal{L}_{\mathbf{g}^3} \quad (27.3.49)$$

where

$$\mathcal{L}_{\mathbf{g}^1} = p(\partial/\partial q) - (q - 2\lambda p)(\partial/\partial p), \quad (27.3.50)$$

$$\mathcal{L}_{\mathbf{g}^3} = -2\lambda q^2 p(\partial/\partial p). \quad (27.3.51)$$

Verify that these homogeneous vector fields in turn have the decompositions

$$\mathcal{L}_{\mathbf{g}^1} =: h_2 : + \mathcal{G}^0 \quad (27.3.52)$$

with

$$h_2 = -(p^2 - 2\lambda pq + q^2)/2, \quad (27.3.53)$$

$$\mathcal{G}^0 = \lambda[q(\partial/\partial q) + p(\partial/\partial p)] = \lambda\Sigma; \quad (27.3.54)$$

$$\mathcal{L}_{\mathbf{g}^3} =: h_4 : + \mathcal{G}^2, \quad (27.3.55)$$

with

$$h_4 = -(\lambda/2)q^3 p, \quad (27.3.56)$$

$$\mathcal{G}^2 = -(\lambda/2)q^2 \Sigma. \quad (27.3.57)$$

Show from (3.54) that the solution  $q = 0$  is unstable for  $\lambda > 0$ . Argue that for small  $\lambda$  the solutions to (3.47) should be nearly those for a simple harmonic oscillator, i.e., any circle in  $q, p$  phase space. Show from energy considerations that for small  $\lambda$  the Van der Pol oscillator should have a limit cycle that is nearly a circle in phase space (about the origin) of radius 2, which is indeed the case. Observe that

$$\mathcal{G}^0 + \mathcal{G}^2 = \lambda(1 - q^2/2)\Sigma. \quad (27.3.58)$$

Show that for a solution of the form  $q = A \sin(t + \phi)$  there is the relation

$$\langle (1 - q^2/2) \rangle = 1 - A^2/4 \quad (27.3.59)$$

where  $\langle \rangle$  denotes time averaging. Show that there is the general operator relation

$$\{ : f_2 : , (\mathcal{G}^0 + \mathcal{G}^2) \} = \lambda [ : f_2 : (1 - q^2/2)] \Sigma. \quad (27.3.60)$$

By considering an operator of the form

$$\exp(-t : f_2 :)(\mathcal{G}^0 + \mathcal{G}^2) \exp(t : f_2 :), \quad (27.3.61)$$

use (3.60) to find the relation

$$\langle \mathcal{G}^0 + \mathcal{G}^2 \rangle = \lambda \langle (1 - q^2/2) \rangle \Sigma. \quad (27.3.62)$$

Thus, on the limit cycle, the growth/damping due to  $(\mathcal{G}^0 + \mathcal{G}^2)$  averages to zero. Consider what appears to be a generalization of the Van der Pol oscillator described by the equation

$$\ddot{q} - 2\lambda \dot{q} + 2\tau q^2 \dot{q} + q = 0 \quad (27.3.63)$$

where  $\lambda$  and  $\tau$  are positive. Verify that (3.63) can be brought to the form (3.47) by a suitable scaling of  $q$ . Verify, for small  $\lambda$  and  $\tau$ , that (3.63) has a nearly circular limit cycle in phase space whose radius is given by the relation

$$A = 2\sqrt{(\lambda/\tau)}. \quad (27.3.64)$$

**27.3.13.** Suppose that  $f$  is an analytic function of the complex variable  $z = x + iy$ , and write the relations

$$w = f(z) \quad (27.3.65)$$

and

$$w = u(x, y) + iv(x, y). \quad (27.3.66)$$

Then, because  $f$  is assumed analytic,  $u$  and  $v$  satisfy the Cauchy-Riemann equations

$$\partial u / \partial x = \partial v / \partial y, \quad (27.3.67)$$

$$\partial u / \partial y = -\partial v / \partial x. \quad (27.3.68)$$

In terms of the phase-space variables  $\{q, p\}$ , consider the differential form

$$v(q, p) dq + u(q, p) dp. \quad (27.3.69)$$

According to Exercise 6.4.6, this form will be exact if there is the relation

$$\partial v / \partial p = \partial u / \partial q. \quad (27.3.70)$$

From the Cauchy-Riemann equation (3.67) we see that the relation (3.70) is in fact true, and therefore there is a function  $H$  such that

$$u(q, p) = \partial H / \partial p, \quad (27.3.71)$$

$$v(q, p) = \partial H / \partial q. \quad (27.3.72)$$

We know that any Hamiltonian gives rise to the Hamiltonian vector field  $: -H :$  given by the rule

$$: -H := (\partial H / \partial p)(\partial / \partial q) - (\partial H / \partial q)(\partial / \partial p). \quad (27.3.73)$$

In view of (3.71) and (3.72), we also have the relation

$$: -H := u(q, p)(\partial / \partial q) - v(q, p)(\partial / \partial p). \quad (27.3.74)$$

Thus, any analytic function  $f$  gives rise to a Hamiltonian vector field.

Consider the differential form

$$u(q, p)dq - v(q, p)dp. \quad (27.3.75)$$

Show, using the second Cauchy-Riemann equation (3.68), that this form is also exact so that there is a function  $K$  such that

$$u(q, p) = \partial K / \partial q, \quad (27.3.76)$$

$$v(q, p) = -\partial K / \partial p. \quad (27.3.77)$$

Show that the Hamiltonian vector field  $: -K :$  is given in terms of  $u$  and  $v$  by the relation

$$: -K := -v(q, p)(\partial / \partial q) - u(q, p)(\partial / \partial p). \quad (27.3.78)$$

Thus, any analytic function  $f$  also gives rise to a second Hamiltonian vector field. Show that (3.78) arises from (3.74) upon replacing  $f$  by  $if$ .

For the analytic function  $f$  given by

$$f(z) = z^2, \quad (27.3.79)$$

find the Hamiltonians  $H$  and  $K$ .

**27.3.14.** Review Exercise 3.13. For the analytic function  $f$  given by

$$f(z) = z^2, \quad (27.3.80)$$

consider the vector field  $\mathcal{L}$  given by

$$\mathcal{L} = u(q, p)(\partial / \partial q) + v(q, p)(\partial / \partial p). \quad (27.3.81)$$

Verify that this vector field is not Hamiltonian, and decompose it into Hamiltonian and non-Hamiltonian parts. Make analogous calculations for the vector field  $\mathcal{L}'$  given by

$$\mathcal{L}' = -v(q, p)(\partial / \partial q) + u(q, p)(\partial / \partial p). \quad (27.3.82)$$

**27.3.15.** Show that Duffing's equation (1.4.31) arises from the vector field  $\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_1 + \mathcal{L}_3$  where

$$\mathcal{L}_0 = (\epsilon \cos \omega \tau)(\partial / \partial p), \quad (27.3.83)$$

$$\mathcal{L}_1 = p(\partial / \partial q) - (q + 2\beta p)(\partial / \partial p), \quad (27.3.84)$$

$$\mathcal{L}_3 = -(q^3)(\partial / \partial p). \quad (27.3.85)$$

Verify that  $\mathcal{L}_0$  and  $\mathcal{L}_3$  are Hamiltonian,

$$\mathcal{L}_0 =: (\epsilon \cos \omega \tau)q : , \quad (27.3.86)$$

$$\mathcal{L}_3 = - : q^4 / 4 : . \quad (27.3.87)$$

Using (3.22) and (3.23), decompose  $\mathcal{L}_1$  into Hamiltonian and non-Hamiltonian parts.

## 27.4 Structure of $sp(4, \mathbb{R})$

The Lie algebra  $sp(4, \mathbb{R})$  is 10 dimensional, and its Cartan subalgebra is 2 dimensional. Therefore, in the Cartan basis, there should be 8 ladder operators. They are labelled by 8 two-component root vectors consisting of 4 vectors and their negatives. We will call these 4 vectors  $\alpha, \beta, \gamma$ , and  $\delta$ . They are given in terms of two orthogonal unit vectors  $e^1$  and  $e^2$  by the relations

$$\alpha = 2e^1, \quad (27.4.1)$$

$$\beta = e^1 + e^2, \quad (27.4.2)$$

$$\gamma = 2e^2, \quad (27.4.3)$$

$$\delta = -e^1 + e^2. \quad (27.4.4)$$

The eight  $sp(4)$  root vectors (the vectors  $\alpha, \beta, \gamma, \delta$  and their negatives) are shown in Figure 4.1. Note that all root vectors are of the form  $(\pm e^i \pm e^j)$  with  $i, j$  and the signs taken independently and the zero vector omitted. Thus, there are basically two kinds of root vectors: *short* root vectors with length  $\sqrt{2}$  and *long* root vectors with length 2. (And the angle between any two successive root vectors as one goes around the root diagram is 45 degrees.) They satisfy the normalization relations

$$\sum_{\mu} (\mathbf{e}^i \cdot \boldsymbol{\mu})(\boldsymbol{\mu} \cdot \mathbf{e}^j) = 12\delta_{ij}. \quad (27.4.5)$$

Again see Section 5.8 for an analogous treatment of  $su(3)$ .

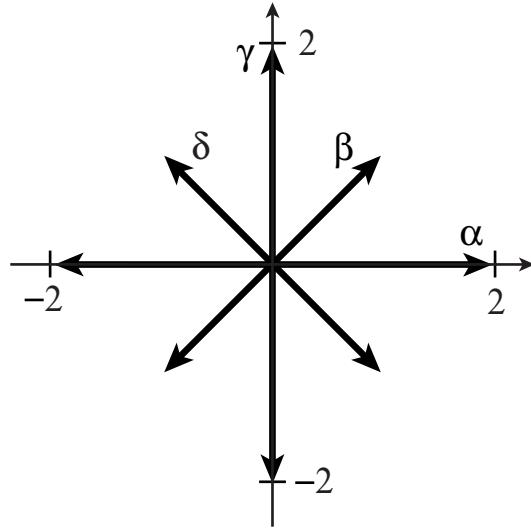


Figure 27.4.1: Root diagram showing the root vectors for  $sp(4)$ .

The Lie algebra  $sp(4, \mathbb{R})$  is generated by the monomials  $z_a z_b$  with  $a, b$  ranging from 1 to 4. See Section 5.7. For present purposes it is convenient to use as the basis for the Cartan subalgebra the polynomials

$$c^1 = -(i/2)(p_1^2 + q_1^2), \quad (27.4.6)$$

$$c^2 = -(i/2)(p_2^2 + q_2^2). \quad (27.4.7)$$

The Lie operators associated with  $c^1$  and  $c^2$  obviously commute. They are also Hermitian,

$$: c^j :^\dagger =: c^j :. \quad (27.4.8)$$

Finally, Lie transformations of the form

$$\begin{aligned} \mathcal{M}(\theta_1, \theta_2) &= \exp(-i\theta_1 : c^1 : -i\theta_2 : c^2 :) \\ &= \exp[-(\theta_1/2) : p_1^2 + q_1^2 : -(\theta_2/2) : p_2^2 + q_2^2 :] \end{aligned} \quad (27.4.9)$$

are real symplectic, unitary, and lie on a 2-torus which is a maximal torus in  $Sp(4, \mathbb{R})$ . See Sections 3.9, 5.9, and 7.2.

For the ladder operators in  $sp(4)$  we use the polynomials

$$r(\pm\alpha) = (\sqrt{2}/4)(q_1 \pm ip_1)^2, \quad (27.4.10)$$

$$r(\pm\beta) = (1/2)(q_1 \pm ip_1)(q_2 \pm ip_2), \quad (27.4.11)$$

$$r(\pm\gamma) = (\sqrt{2}/4)(q_2 \pm ip_2)^2, \quad (27.4.12)$$

$$r(\pm\delta) = (i/2)(q_1 \mp ip_1)(q_2 \pm ip_2). \quad (27.4.13)$$

Their associated Lie operators obey the conjugation relations

$$: r(\mu) :^\dagger =: r(-\mu) :. \quad (27.4.14)$$

It is easily verified that the Cartan subalgebra and ladder operators satisfy the Poisson bracket rules

$$[c^j, c^k] = 0, \quad (27.4.15)$$

$$[c^j, r(\mu)] = (\mathbf{e}^j \cdot \boldsymbol{\mu})r(\mu), \quad (27.4.16)$$

$$[r(\mu), r(-\mu)] = \sum_j (\mathbf{e}^j \cdot \boldsymbol{\mu})c^j, \quad (27.4.17)$$

as desired. There are also the relations

$$[r(\mu), r(\nu)] = N(\mu, \nu)r(\mu + \nu) \quad (27.4.18)$$

provided the sum  $(\mu + \nu)$  is again a root vector. All other brackets vanish. For the case of  $sp(4)$ , the  $N(\mu, \nu)$  have the values  $\pm\sqrt{2}$ . The positive  $N$ 's are  $N(\alpha, \delta)$ ,  $N(\beta, \delta)$ ,  $N(\beta, -\delta)$ ,  $N(\gamma, -\delta)$ ,  $N(\delta, -\beta)$ ,  $N(\delta, -\gamma)$ ,  $N(-\alpha, \beta)$ ,  $N(-\beta, *)$ ,  $N(-\beta, *)$ ,  $N(-\gamma, *)$ ,  $N(-\delta, *)$ ,  $N(-\delta, *)$ . We also note, as was true for  $sp(2)$ , that for the scalar product (7.3.12) the basis elements  $c^j$  and  $r(\mu)$  are orthonormal,

$$\langle c^j, c^j \rangle = \langle r(\mu), r(\mu) \rangle = 1, \quad (27.4.19)$$

$$\langle c^j, c^k \rangle = 0 \text{ for } j \neq k, \quad (27.4.20)$$

$$\langle c^j, r(\mu) \rangle = 0, \quad (27.4.21)$$

$$\langle r(\mu), r(\nu) \rangle = 0 \text{ for } \mu \neq \nu. \quad (27.4.22)$$

At this point, we remark that it is tempting to assume that the rank of a Lie algebra equals the maximum number of mutually commuting elements. (Some texts even make this claim!) This need not be the case. For  $sp(4)$ , which has rank 2, it is evident that the 3 elements  $:p_1^2:, :p_1 p_2:, :p_2^2:$  are mutually commuting. However, they are not Hermitian and, when exponentiated, do not even produce a torus (not to mention a maximal torus). Nor can other elements be found in the Lie algebra such that relations of the form (4.16) hold with the  $:p_k p_\ell:$  playing the role of the  $c$ 's. Instead, as will be evident in Section 27.5 [see (5.12), (5.14), and (5.16)], they are related by a symplectic unitary transformation to ladder operators. Therefore they do not meet the requirements to form a Cartan subalgebra.

We close this section by examining how  $sp(2)$ ,  $u(2)$ ,  $su(2)$ , and  $so(2)$  reside within  $sp(4)$ . The presence of  $sp(2)$  is evident. Comparison of Figures 1.1 and 4.1, and comparison of (1.1) and (4.10), shows that if we identify the coordinate pair  $q, p$  with the pair  $q_1, p_1$ , then the  $r(\pm\alpha)$  for  $sp(2)$  and  $sp(4)$  agree. Also, the  $c^1$  for  $sp(2)$  given by (1.11) agrees with the  $c^1$  given for  $sp(4)$  by (4.6). We also note that there is a second  $sp(2)$  within  $sp(4)$  generated by  $r(\pm\gamma)$  and  $c^2$ . It is identical to the  $sp(2)$  of Section 27.1 if we identify the pair  $q, p$  with  $q_2, p_2$ . Thus, there is an  $sp(2)$  within  $sp(4)$  for each equal and opposite pair of *long* root vectors. (See also Exercise 5.5.) These root vectors have length 2 as is required for an  $sp(2)$  root vector. See Section 27.1.

The presence of  $u(2)$  and  $su(2)$  within  $sp(4)$  is less evident. Comparison of (5.7.8) and (4.13) shows that there are the relations

$$r(\pm) = r(\mp\delta). \quad (27.4.23)$$

Also, comparison of (5.7.4), (5.7.9) and (4.6), (4.7) gives the relations

$$b^0 = i(c^1 + c^2), \quad (27.4.24)$$

$$c = (1/\sqrt{2})(c^1 - c^2). \quad (27.4.25)$$

Thus, the  $su(2)$  of Section 5.7 is associated with  $c$  and the  $r(\mp\delta)$  root vectors of  $sp(4)$ , and also including  $b^0$  yields  $u(2)$ . We also note that the spin 1 objects  $h^\pm$  and  $h^0$  of Section 5.7 given by (5.7.21) through (5.7.23) are proportional to the  $sp(4)$  generators  $r(\alpha)$ ,  $r(\beta)$ , and  $r(\gamma)$  given by (4.10) through (4.12). Reference to Figure 4.1 shows that they should indeed transform among each other under the action of  $r(\mp\delta)$  as given by (5.7.26) and (5.7.27). Moreover, we note that there is a second  $su(2)$  [and a corresponding  $u(2)$ ] within  $sp(4)$  associated with the root vectors  $r(\pm\beta)$ . Thus, there is an  $su(2)$  and a  $u(2)$  within  $sp(4)$  for each equal and opposite pair of *short* root vectors. These root vectors have length  $\sqrt{2}$  as required for  $su(2)$  root vectors. See Exercise 2.1.

Finally, there is an  $so(2)$  subalgebra within  $sp(4)$  whose presence is not at all obvious from looking at the  $sp(4)$  root diagram. Let  $J_z$  be the quadratic polynomial defined by the equation

$$J_z = q_1 p_2 - q_2 p_1. \quad (27.4.26)$$

It generates rotations in the  $q_1, q_2$  and  $p_1, p_2$  planes:

$$\exp : \theta J_z : q_1 = q_1 \cos \theta + q_2 \sin \theta, \quad (27.4.27)$$

$$\exp : \theta J_z : q_2 = -q_1 \sin \theta + q_2 \cos \theta; \quad (27.4.28)$$

$$\exp : \theta J_z p_1 = p_1 \cos \theta + p_2 \sin \theta, \quad (27.4.29)$$

$$\exp : \theta J_z : p_2 = -p_1 \sin \theta + p_2 \cos \theta. \quad (27.4.30)$$

From (4.13) we see that it is related to elements in the Cartan basis by the equation

$$J_z = -[r(\boldsymbol{\delta}) - r(-\boldsymbol{\delta})]. \quad (27.4.31)$$

## Exercises

**27.4.1.** Verify the Lie product rules (4.15) through (4.18).

**27.4.2.** Verify the relations (4.26) through (4.31). Show that  $J_z$  obeys the eigen relations

$$: J_z (q_1 \pm iq_2)^n = \mp in(q_1 \pm iq_2)^n, \quad (27.4.32)$$

$$: J_z : (p_1 \pm ip_2)^n = \mp in(p_1 \pm ip_2)^n. \quad (27.4.33)$$

**27.4.3.** Explore the properties of  $J'_z$  that is defined in analogy to  $J_z$  by the equation

$$J'_z = -[r(\boldsymbol{\beta}) - r(-\boldsymbol{\beta})]. \quad (27.4.34)$$

See Figure 4.1.

## 27.5 Representations of $sp(4, \mathbb{R})$

The description of representations of  $sp(4)$  follows the general Cartan procedure as described for  $su(3)$  in Section 5.8. For  $sp(4)$ , since it has rank 2, there are two fundamental weights  $\phi^1$  and  $\phi^2$ . They are given by the relations

$$\phi^1 = \mathbf{e}^1 = \boldsymbol{\alpha}/2, \quad (27.5.1)$$

$$\phi^2 = \mathbf{e}^1 + \mathbf{e}^2 = \boldsymbol{\beta}, \quad (27.5.2)$$

and are shown in Figure 5.1 along with the  $sp(4)$  root vectors. Thus, for  $sp(4)$ , every highest weight  $\mathbf{w}^h$  is of the form

$$\mathbf{w}^h = m\phi^1 + n\phi^2 = (m+n)\mathbf{e}^1 + n\mathbf{e}^2, \quad (27.5.3)$$

where  $m$  and  $n$  are arbitrary nonnegative integers. Correspondingly, for each  $m, n$  pair, there is an irreducible representation  $\Gamma(m, n)$  with highest weight  $\mathbf{w}^h$  given by (5.3). It can be shown that the dimension of  $\Gamma(m, n)$  is given by the relation

$$\dim \Gamma(m, n) = (1/6)(m+2n+3)(m+n+2)(m+1)(n+1). \quad (27.5.4)$$

See Exercise 5.1. It can also be shown that these representations are self conjugate,

$$\bar{\Gamma}(m, n) = \Gamma(m, n). \quad (27.5.5)$$

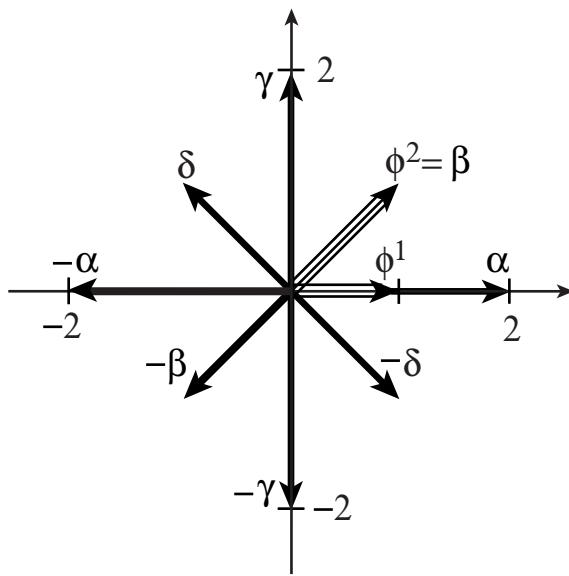


Figure 27.5.1: Fundamental weights  $\phi^1$  and  $\phi^2$  for  $sp(4)$ . The root vectors are also shown.

Table 27.5.1: Dimensions of Representations of  $sp(4)$ .

$m$	$n$	$\dim \Gamma(m, n)$	$m$	$n$	$\dim \Gamma(m, n)$
0	0	1	0	2	14
1	0	4	3	0	20
0	1	5	2	1	35
2	0	10	1	2	40
1	1	16	0	3	30

See Exercise 3.7.36. For quick reference the dimensions of the first few representations are listed in Table 5.1 above. Where there is no possibility of confusion, we will sometimes refer to a representation by its dimension.

From a knowledge of the root vectors and the highest weight it is a simple matter to construct weight diagrams for the low-dimensional representations. Figures 5.2 through 5.7 show weight diagrams for the first few representations. Inspection of these figures and reference to Table 5.1 shows that the weights must have unit multiplicities for the representations  $\Gamma(0, 0)$ ,  $\Gamma(1, 0)$ , and  $\Gamma(0, 1)$ . For  $\Gamma(2, 0)$ , which is the adjoint or regular representation, the weight vector at the origin has multiplicity 2. The representation  $\Gamma(1, 0)$  corresponds to the representation of  $sp(4)$  by  $4 \times 4$  matrices of the form  $JS$ . See (5.7.27) of Section 5.7. It happens that the Lie algebras for  $sp(4)$  and  $so(5)$  are equivalent over the complex field. The  $sp(4)$  representation  $\Gamma(0, 1)$ , which is 5 dimensional, is related to the obvious  $5 \times 5$  matrix representation of  $so(5)$ . See Exercise 5.4.

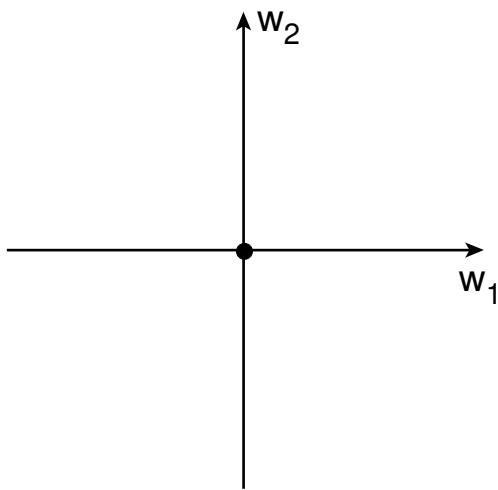


Figure 27.5.2: Weight diagram for the representation  $1 = \Gamma(0,0)$ .

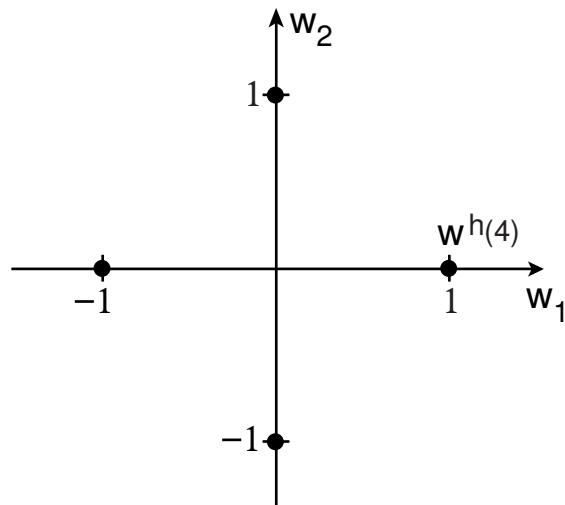
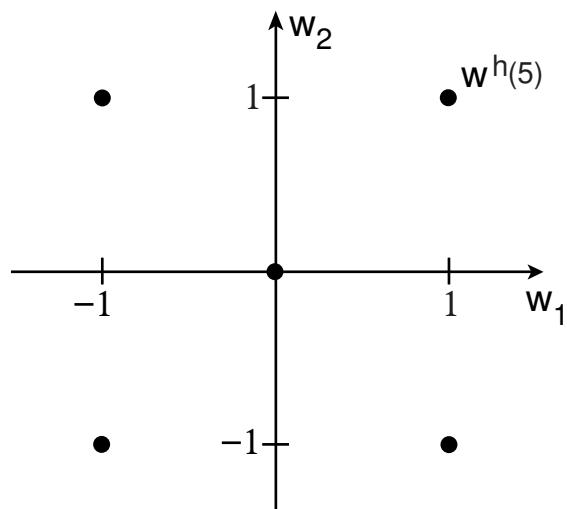


Figure 27.5.3: Weight diagram for the fundamental representation  $4 = \Gamma(1,0)$ .

Figure 27.5.4: Weight diagram for the representation  $5 = \Gamma(0, 1)$ .

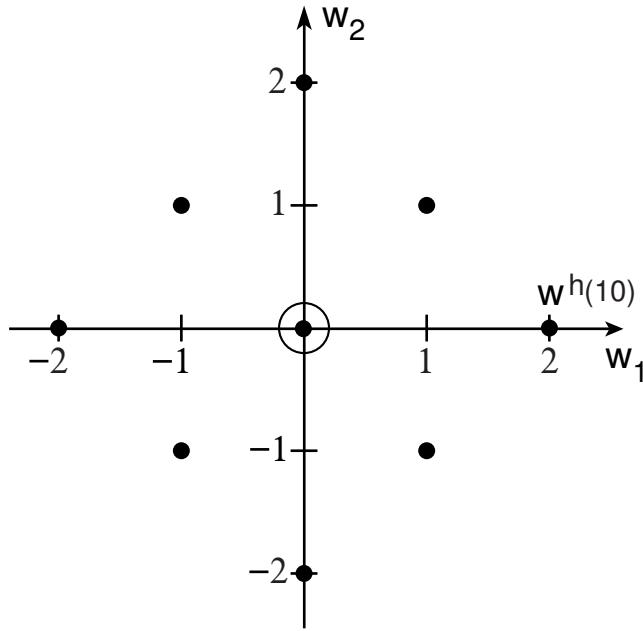


Figure 27.5.5: Weight diagram for the adjoint representation  $10 = \Gamma(2,0)$ . The circled weight at the origin has multiplicity 2. The other eight weights are located at the tips of the  $sp(4)$  root vectors.

Now, in mimicry of what was done before in Section 21.2, let  $\mathcal{P}_m$  denote the set of polynomials homogeneous of degree  $m$  in the variables  $q_1, p_1, q_2, p_2$ ; and let  $f_2$  be a quadratic polynomial in these variables. Then we have the relation

$$:f_2:\mathcal{P}_m \subseteq \mathcal{P}_m. \quad (27.5.6)$$

It follows that the set of homogeneous polynomials of degree  $m$  forms a representation of  $sp(4, \mathbb{R})$ .

What irreducible representations occur? The study of this question is again facilitated by a map  $\mathcal{A}(\pi/8)$  defined this time by the equation

$$\mathcal{A}(\pi/8) = \exp[-i(\pi/8) : p_1^2 - q_1^2 + p_2^2 - q_2^2 :]. \quad (27.5.7)$$

As before,  $\mathcal{A}$  is complex symplectic and unitary. Calculation gives the result

$$\begin{aligned} \mathcal{A}(\pi/8)(q_1^{r_1} p_1^{s_1} q_2^{r_2} p_2^{s_2}) = \\ (1/\sqrt{2})^{r_1+s_1+r_2+s_2} (i)^{s_1+s_2} (q_1 + ip_1)^{r_1} (q_1 - ip_1)^{s_1} (q_2 + ip_2)^{r_2} (q_2 - ip_2)^{s_2}. \end{aligned} \quad (27.5.8)$$

Evidently  $\mathcal{A}(\pm\pi/8)$  again transforms between what we will again call the Cartesian and resonance bases.

With the aid of  $\mathcal{A}$  we again define transformed Lie basis polynomials  $\tilde{c}^j$  and  $\tilde{r}(\mu)$  by the rule

$$\tilde{c}^1 = \mathcal{A}(-\pi/8)c^1 = -q_1 p_1, \quad (27.5.9)$$

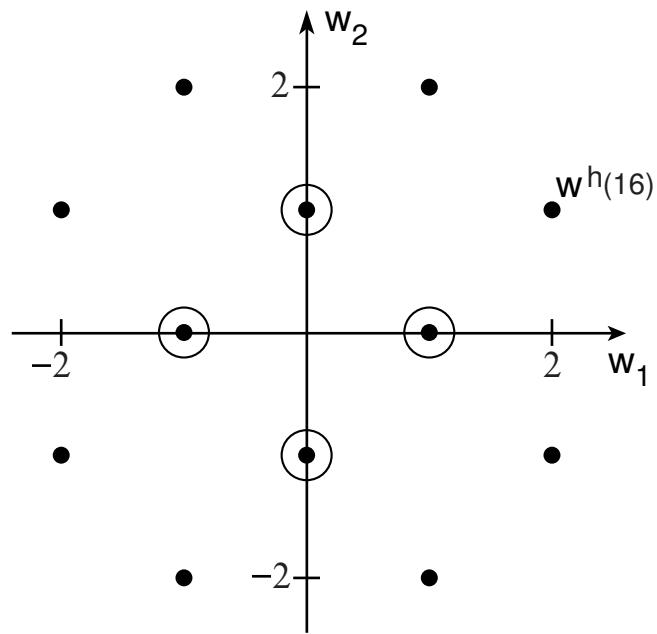


Figure 27.5.6: Weight diagram for the representation  $16 = \Gamma(1, 1)$ . The circled weights on the inner diamond have multiplicity 2.

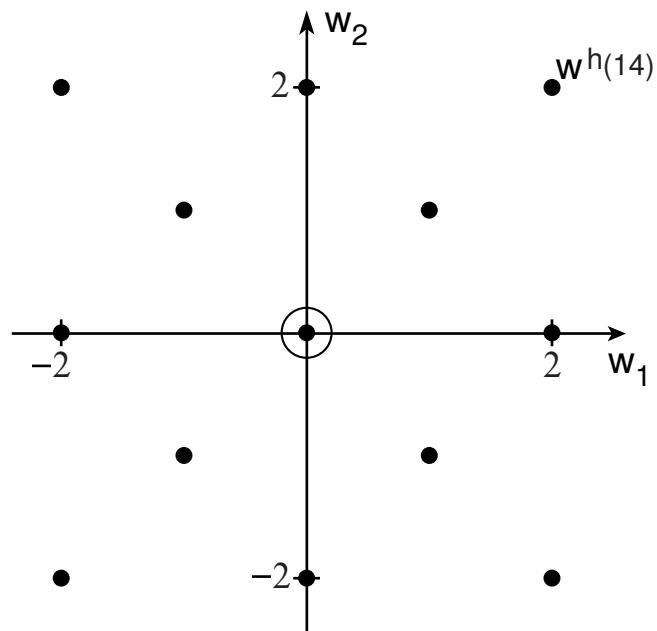


Figure 27.5.7: Weight diagram for the representation  $14 = \Gamma(0, 2)$ . The circled weight at the origin has multiplicity 2.

$$\tilde{c}^2 = \mathcal{A}(-\pi/8)c^2 = -q_2 p_2, \quad (27.5.10)$$

$$\tilde{r}(\boldsymbol{\alpha}) = \mathcal{A}(-\pi/8)r(\boldsymbol{\alpha}) = (1/\sqrt{2})q_1^2, \quad (27.5.11)$$

$$\tilde{r}(-\boldsymbol{\alpha}) = \mathcal{A}(-\pi/8)r(-\boldsymbol{\alpha}) = -(1/\sqrt{2})p_1^2, \quad (27.5.12)$$

$$\tilde{r}(\boldsymbol{\beta}) = \mathcal{A}(-\pi/8)r(\boldsymbol{\beta}) = q_1 q_2, \quad (27.5.13)$$

$$\tilde{r}(-\boldsymbol{\beta}) = \mathcal{A}(-\pi/8)r(-\boldsymbol{\beta}) = -p_1 p_2, \quad (27.5.14)$$

$$\tilde{r}(\boldsymbol{\gamma}) = \mathcal{A}(-\pi/8)r(\boldsymbol{\gamma}) = (1/\sqrt{2})q_2^2, \quad (27.5.15)$$

$$\tilde{r}(-\boldsymbol{\gamma}) = \mathcal{A}(-\pi/8)r(-\boldsymbol{\gamma}) = -(1/\sqrt{2})p_2^2, \quad (27.5.16)$$

$$\tilde{r}(\boldsymbol{\delta}) = \mathcal{A}(-\pi/8)r(\boldsymbol{\delta}) = p_1 q_2, \quad (27.5.17)$$

$$\tilde{r}(-\boldsymbol{\delta}) = \mathcal{A}(-\pi/8)r(-\boldsymbol{\delta}) = p_2 q_1. \quad (27.5.18)$$

Since  $\mathcal{A}$  is symplectic, the transformed basis polynomials obey the same Poisson bracket rules (4.15) through (4.18). Since  $\mathcal{A}$  is also unitary, the transformed basis polynomials also satisfy the orthonormality relations (4.19) through (4.22) and the conjugacy relations (4.8) and (4.14).

Now consider the actions of the Lie operators  $:\tilde{c}^j:$  and  $:\tilde{r}(\boldsymbol{\mu}):$  on the general monomials  $q_1^{r_1} p_1^{s_1} q_2^{r_2} p_2^{s_2}$ . Calculation gives the results

$$:\tilde{c}^1: q_1^{r_1} p_1^{s_1} q_2^{r_2} p_2^{s_2} = (r_1 - s_1) q_1^{r_1} p_1^{s_1} q_2^{r_2} p_2^{s_2}, \quad (27.5.19)$$

$$:\tilde{c}^2: q_1^{r_1} p_1^{s_1} q_2^{r_2} p_2^{s_2} = (r_2 - s_2) q_1^{r_1} p_1^{s_1} q_2^{r_2} p_2^{s_2}, \quad (27.5.20)$$

$$:\tilde{r}(\boldsymbol{\alpha}): q_1^{r_1} p_1^{s_1} q_2^{r_2} p_2^{s_2} = \sqrt{2} s_1 q_1^{r_1+1} p_1^{s_1-1} q_2^{r_2} p_2^{s_2}, \quad (27.5.21)$$

$$:\tilde{r}(-\boldsymbol{\alpha}): q_1^{r_1} p_1^{s_1} q_2^{r_2} p_2^{s_2} = \sqrt{2} r_1 q_1^{r_1-1} p_1^{s_1+1} q_2^{r_2} p_2^{s_2}, \quad (27.5.22)$$

$$:\tilde{r}(\boldsymbol{\beta}): q_1^{r_1} p_1^{s_1} q_2^{r_2} p_2^{s_2} = s_1 q_1^{r_1} p_1^{s_1-1} q_2^{r_2+1} p_2^{s_2} + s_2 q_1^{r_1+1} p_1^{s_1} q_2^{r_2} p_2^{s_2-1}, \quad (27.5.23)$$

$$:\tilde{r}(-\boldsymbol{\beta}): q_1^{r_1} p_1^{s_1} q_2^{r_2} p_2^{s_2} = r_1 q_1^{r_1-1} p_1^{s_1} q_2^{s_2} p_2^{s_2+1} + r_2 q_1^{r_1} p_1^{s_1+1} q_2^{r_2-1} p_2^{s_2}, \quad (27.5.24)$$

$$:\tilde{r}(\boldsymbol{\gamma}): q_1^{r_1} p_1^{s_1} q_2^{r_2} p_2^{s_2} = \sqrt{2} s_2 q_1^{r_1} p_1^{s_1} q_2^{r_2+1} p_2^{s_2-1}, \quad (27.5.25)$$

$$:\tilde{r}(-\boldsymbol{\gamma}): q_1^{r_1} p_1^{s_1} q_2^{r_2} p_2^{s_2} = \sqrt{2} r_2 q_1^{r_1} p_1^{s_1} q_2^{r_2-1} p_2^{s_2+1}, \quad (27.5.26)$$

$$:\tilde{r}(\boldsymbol{\delta}): q_1^{r_1} p_1^{s_1} q_2^{r_2} p_2^{s_2} = -r_1 q_1^{r_1-1} p_1^{s_1} q_2^{r_2+1} p_2^{s_2} + s_2 q_1^{r_1} p_1^{s_1+1} q_2^{r_2} p_2^{s_2-1}, \quad (27.5.27)$$

$$:\tilde{r}(-\boldsymbol{\delta}): q_1^{r_1} p_1^{s_1} q_2^{r_2} p_2^{s_2} = s_1 q_1^{r_1} p_1^{s_1-1} q_2^{r_2} p_2^{s_2+1} - r_2 q_1^{r_1+1} p_1^{s_1} q_2^{r_2-1} p_2^{s_2}. \quad (27.5.28)$$

Evidently any monomial of a given degree can be transformed into any other monomial of the same degree with the aid of the  $\tilde{r}(\boldsymbol{\mu})$ . Therefore  $sp(4)$  acts irreducibly on  $\mathcal{P}_m$ . Also,  $q_1^m$  is the vector of highest weight in  $\mathcal{P}_m$ , and has the weight  $\boldsymbol{\omega}^h = m\boldsymbol{\phi}^1$ ,

$$:\tilde{c}^1: q_1^m = mq_1^m = (\boldsymbol{e}^1 \cdot m\boldsymbol{\phi}^1)q_1^m, \quad (27.5.29)$$

$$:\tilde{c}^2: q_1^m = 0 = (\boldsymbol{e}^2 \cdot m\boldsymbol{\phi}^1)q_1^m.$$

Upon examination of (5.3) we conclude that, in the case of 4 variables,  $\mathcal{P}_m$  carries the representation  $\Gamma(m, 0)$  of  $sp(4)$ . From (7.3.36) and (5.4) we find, as expected, the result

$$\dim \mathcal{P}_m = N(m, 4) = (1/6)(m+3)(m+2)(m+1) = \dim \Gamma(m, 0). \quad (27.5.30)$$

Note that, unlike the case of 2 variables, the  $\mathcal{P}_m$  for various  $m$  do not carry *all* the representations of  $sp(4)$ , but only the representations  $\Gamma(m, 0)$ .

As before, we can let  $\mathcal{A}(\pi/8)$  act on both sides of (5.19) through (5.28). Doing so gives results analogous to those in (2.23) through (2.25). Consequently, as before, the  $:c^j:$  and  $:r(\boldsymbol{\mu}):$  act on the resonance basis in the same way that the  $:\tilde{c}^j:$  and  $:\tilde{r}(\boldsymbol{\mu}):$  act on the monomial basis.

## Exercises

**27.5.1.** Weyl discovered that for the simple Lie algebras the dimension of a representation  $\Gamma(\mathbf{w}^h)$  labeled by the highest weight  $\mathbf{w}^h$  is given the formula

$$\dim \Gamma(\mathbf{w}^h) = \prod_{\boldsymbol{\mu} > 0} [\boldsymbol{\mu} \cdot (\mathbf{w}^h + \boldsymbol{\mu}^+ / 2)] / [\boldsymbol{\mu} \cdot (\boldsymbol{\mu}^+ / 2)]. \quad (27.5.31)$$

Here the product is to be taken over all *positive* root vectors and  $\boldsymbol{\mu}^+$  is the sum of all positive roots as in (12.49). [As was the case for weights (see Section 5.8), we define a root  $\boldsymbol{\mu}$  to be positive (and write  $\boldsymbol{\mu} > 0$ ) if its first nonvanishing component is positive.] Show that the results (2.4), (5.4), and (5.8.21) for  $sp(2)$ ,  $sp(4)$ , and  $su(3)$  follow from Weyl's formula. If you are feeling algebraically rambunctious, verify (8.5) for  $sp(6)$ .

**27.5.2.** Verify (5.30).

**27.5.3.** Look at the  $sp(4)$  weight diagrams shown in Figures 5.2 through 5.7. Verify that the spacing between the dots in the directions of the long  $sp(4)$  root vectors is 2. These dots describe  $sp(2)$  representations within  $sp(4)$ . See Section 21.2. Verify that the spacing between the dots in the directions of the short  $sp(4)$  root vectors is  $\sqrt{2}$ . These dots describe  $su(2)$  representations within  $sp(4)$ . See Exercise 2.1.

**27.5.4.** In this exercise we will see that there are two obvious and mutually commuting  $sp(2)$  subalgebras residing within  $sp(4)$ . Then we will find one that is less obvious. This latter subalgebra is of use in the Lie algebraic treatment of (light) optical systems having axial symmetry. See Section 3.1 of Appendix X. Finally we will observe that there is a one-parameter family of mutually commuting  $sp(2)$  subalgebra pairs, and the two obvious subalgebras found initially belong to this family.

Consider the monomials

$$\tilde{r}(\boldsymbol{\alpha}) = (1/\sqrt{2})q_1^2, \quad (27.5.32)$$

$$\tilde{c}^1 = -q_1 p_1, \quad (27.5.33)$$

$$\tilde{r}(-\boldsymbol{\alpha}) = -(1/\sqrt{2})p_1^2; \quad (27.5.34)$$

$$\tilde{r}(\boldsymbol{\gamma}) = (1/\sqrt{2})q_2^2, \quad (27.5.35)$$

$$\tilde{c}^2 = -q_2 p_2, \quad (27.5.36)$$

$$\tilde{r}(-\boldsymbol{\gamma}) = -(1/\sqrt{2})p_2^2. \quad (27.5.37)$$

Note that, consistent with the symmetry of the  $sp(4)$  root vector diagram shown in Figure 4.1, the the ingredients of (5.35) through (5.37) are analogous to those of (5.32) through (5.34).

Verify that for the ingredients of (5.32) through (5.34) there are the Poisson bracket relations

$$[\tilde{c}^1, \tilde{r}(\pm\alpha)] = \pm 2\tilde{r}(\pm\alpha), \quad (27.5.38)$$

$$[\tilde{r}(\alpha), \tilde{r}(-\alpha)] = 2\tilde{c}^1. \quad (27.5.39)$$

As expected, reference to (1.13) and (1.14) reveals that these are the Cartan rules for  $sp(2)$ . Also, evidently the ingredients of (5.35) through (5.37) obey analogous rules. Finally, the ingredients of (5.32) through (5.34) are evidently in involution (commute) with the ingredients of (5.35) through (5.37).

To specify the the third and less obvious  $sp(2)$  residing within  $sp(4)$ , view  $q_1, q_2$  and  $p_1, p_2$  as components of vectors  $\mathbf{q}$  and  $\mathbf{p}$  by writing

$$\mathbf{q} = (q_1, q_2) \text{ and } \mathbf{p} = (p_1, p_2). \quad (27.5.40)$$

Also, make the definitions

$$q^2 = \mathbf{q} \cdot \mathbf{q} = (q_1)^2 + (q_2)^2, \quad (27.5.41)$$

$$\mathbf{q} \cdot \mathbf{p} = q_1 p_1 + q_2 p_2, \quad (27.5.42)$$

$$p^2 = \mathbf{p} \cdot \mathbf{p} = (p_1)^2 + (p_2)^2; \quad (27.5.43)$$

$$L_+ = (1/\sqrt{2})q^2, \quad (27.5.44)$$

$$L_0 = -\mathbf{q} \cdot \mathbf{p}, \quad (27.5.45)$$

$$L_- = -(1/\sqrt{2})p^2. \quad (27.5.46)$$

Verify that there are the Poisson bracket relations

$$[L_0, L_\pm] = \pm 2L_\pm, \quad (27.5.47)$$

$$[L_+, L_-] = 2L_0. \quad (27.5.48)$$

Evidently these relations are also the Cartan rules for  $sp(2)$ . Is this surprising? We see from the work leading up to this point that this result is to be expected.

Next, observe that the quantities in (5.44) through (5.46) are invariant under rotations in the 1, 2 plane: Review (4.26) through (4.30). See also Subsection 16.2.5.2 and (16.2.218) through (16.2.222) of Exercise 16.2.16. Consider the Lie operator  $:J_z:$  where  $J_z$  is defined by (4.26). Verify that there are the relations

$$:J_z: q^2 = :J_z: (\mathbf{q} \cdot \mathbf{p}) = :J_z: p^2 = 0. \quad (27.5.49)$$

Consider the Lie transformation  $\exp(\theta : J_z :)$ . Verify as expected that, for any  $\theta$ ,

$$\exp(\theta : J_z :) q^2 = q^2, \quad \exp(\theta : J_z :) (\mathbf{q} \cdot \mathbf{p}) = (\mathbf{q} \cdot \mathbf{p}), \quad \exp(\theta : J_z :) p^2 = p^2. \quad (27.5.50)$$

The last thing to do is for us/you to construct the promised one-parameter family of mutually commuting  $sp(2)$  subalgebra pairs. Looking at (5.32) through (5.37), define quadratic generators/polynomials  $\hat{r}(\boldsymbol{\alpha}, \theta) \dots$  by the rules

$$\hat{r}(\boldsymbol{\alpha}, \theta) = \exp(\theta : J_z :)\tilde{r}(\boldsymbol{\alpha}), \quad (27.5.51)$$

$$\hat{c}^1(\theta) = \exp(\theta : J_z :)\tilde{c}^1, \quad (27.5.52)$$

$$\hat{r}(-\boldsymbol{\alpha}, \theta) = \exp(\theta : J_z :)\tilde{r}(-\boldsymbol{\alpha}); \quad (27.5.53)$$

$$\hat{r}(\boldsymbol{\gamma}, \theta) = \exp(\theta : J_z :)\tilde{r}(\boldsymbol{\gamma}), \quad (27.5.54)$$

$$\hat{c}^2(\theta) = \exp(\theta : J_z :)\tilde{c}^2, \quad (27.5.55)$$

$$\hat{r}(-\boldsymbol{\gamma}, \theta) = \exp(\theta : J_z :)\tilde{r}(-\boldsymbol{\gamma}). \quad (27.5.56)$$

Verify from (4.27) through (4.30) that

$$\exp[(\pi/2) : J_z :]q_1 = q_2, \quad (27.5.57)$$

$$\exp[(\pi/2) : J_z :]q_2 = -q_1; \quad (27.5.58)$$

$$\exp[(\pi/2) : J_z :]p_1 = p_2, \quad (27.5.59)$$

$$\exp[(\pi/2) : J_z :]p_2 = -p_1; \quad (27.5.60)$$

$$\exp(\pi : J_z : )q_j = -q_j \text{ and } \exp(\pi : J_z : )p_j = -p_j. \quad (27.5.61)$$

Using (5.4.9) and (5.57) through (5.60), verify that

$$\begin{aligned} \exp[(\pi/2) : J_z :] \tilde{r}(\boldsymbol{\alpha}) &= (1/\sqrt{2}) \exp[(\pi/2) : J_z :] q_1^2 \\ &= (1/\sqrt{2}) \exp[(\pi/2) : J_z :](q_1 q_1) \\ &= (1/\sqrt{2})(\exp[(\pi/2) : J_z :] q_1)(\exp[(\pi/2) : J_z :] q_1) \\ &= (1/\sqrt{2})q_2 q_2 \\ &= \tilde{r}(\boldsymbol{\gamma}), \end{aligned} \quad (27.5.62)$$

$$\begin{aligned} \exp[(\pi/2) : J_z :] \tilde{c}^1 &= -\exp[(\pi/2) : J_z :](q_1 p_1) \\ &= -(\exp[(\pi/2) : J_z :] q_1)(\exp[(\pi/2) : J_z :] p_1) \\ &= -q_2 p_2 \\ &= \tilde{c}^2, \end{aligned} \quad (27.5.63)$$

$$\begin{aligned} \exp[(\pi/2) : J_z :] \tilde{r}(-\boldsymbol{\alpha}) &= -(1/\sqrt{2}) \exp[(\pi/2) : J_z :] p_1^2 \\ &= -(1/\sqrt{2}) \exp[(\pi/2) : J_z :](p_1 p_1) \\ &= -(1/\sqrt{2})(\exp[(\pi/2) : J_z :] p_1)(\exp[(\pi/2) : J_z :] p_1) \\ &= -(1/\sqrt{2})p_2 p_2 \\ &= \tilde{r}(-\boldsymbol{\gamma}). \end{aligned} \quad (27.5.64)$$

Evidently,  $\exp[(\pi/2) : J_z :]$  sends the left sides of (5.32) through (5.34) to the left sides of (5.35) through (5.37). It sends the elements of what we may call the first  $sp(2)$  to the elements of what we may call the second  $sp(2)$ .

What can be said about the effect of  $\exp(\pi : J_z :)$ ? Using (5.4.9) and (5.60), verify that it leaves all the generators in (5.32) through (5.37) *unchanged*. For example, there is the relation

$$\exp(\pi : J_z :) \tilde{r}(\boldsymbol{\alpha}) = \tilde{r}(\boldsymbol{\alpha}). \quad (27.5.65)$$

Show it follows that all elements on the left sides of (5.51) through (5.56) are *periodic* in  $\theta$  with period  $\pi$ . For example, there is the relation

$$\hat{r}(\boldsymbol{\alpha}, \theta \pm n\pi) = \hat{r}(\boldsymbol{\alpha}, \theta) \quad (27.5.66)$$

for any integer  $n$ . Indeed show, for example, that

$$\begin{aligned} \hat{r}(\boldsymbol{\alpha}, \theta) &= \exp(\theta : J_z :) \tilde{r}(\boldsymbol{\alpha}) = (1/\sqrt{2}) \exp(\theta : J_z :) q_1^2 \\ &= (1/\sqrt{2}) \exp(\theta : J_z :)(q_1 q_1) \\ &= (1/\sqrt{2}) [\exp(\theta : J_z :) q_1] [\exp(\theta : J_z :) q_1] \\ &= (1/\sqrt{2})(q_1 \cos \theta + q_2 \sin \theta)^2 \\ &= (1/\sqrt{2})(q_1^2 \cos^2 \theta + 2q_1 q_2 \cos \theta \sin \theta + q_2^2 \sin^2 \theta) \\ &= (1/\sqrt{8}) \{ [1 + \cos(2\theta)] q_1^2 + 2q_1 q_2 \sin(2\theta) + [1 - \cos(2\theta)] q_2^2 \} \\ &= (1/\sqrt{8}) \{ \sqrt{2}[1 + \cos(2\theta)] \tilde{r}(\boldsymbol{\alpha}) + 2q_1 q_2 \sin(2\theta) + \sqrt{2}[1 - \cos(2\theta)] \tilde{r}(\boldsymbol{\gamma}) \}. \end{aligned} \quad (27.5.67)$$

Recall the trigonometric identities

$$\sin(2\theta) = 2 \cos \theta \sin \theta, \quad (27.5.68)$$

$$\cos^2 \theta = (1/2)[1 + \cos(2\theta)], \quad (27.5.69)$$

$$\sin^2 \theta = (1/2)[1 - \cos(2\theta)]. \quad (27.5.70)$$

Check that the relation

$$\hat{r}(\boldsymbol{\alpha}, 0) = \tilde{r}(\boldsymbol{\alpha}) \quad (27.5.71)$$

and (5.61) and (5.65) and (5.66) are specific cases of (5.67).

The last task is to check that the elements on the left sides of (5.51) through (5.56) form, for any  $\theta$ , two commuting  $sp(2)$  Lie algebras. Here use will be made of (5.4.14). First verify that the elements in (5.51) through (5.53) commute with those in (5.54) through (5.56). Show, for example, that

$$\begin{aligned} [\hat{r}(\boldsymbol{\alpha}, \theta), \hat{r}(\boldsymbol{\gamma}, \theta)] &= [\exp(\theta : J_z :) \tilde{r}(\boldsymbol{\alpha}), \exp(\theta : J_z :) \tilde{r}(\boldsymbol{\gamma})] \\ &= \exp(\theta : J_z :) [\tilde{r}(\boldsymbol{\alpha}), \tilde{r}(\boldsymbol{\gamma})] \\ &= \exp(\theta : J_z :) 0 = 0. \end{aligned} \quad (27.5.72)$$

Finally, verify that the elements (5.51) through (5.53) and (5.54) through (5.56) both constitute  $sp(2)$  Lie algebras for any value of  $\theta$ . Show, for example, that

$$\begin{aligned} [\hat{c}^1(\theta), \hat{r}(\pm\alpha, \theta)] &= [\exp(\theta : J_z :)\tilde{c}^1, \exp(\theta : J_z :)\tilde{r}(\pm\alpha)] \\ &= \exp(\theta : J_z :)[\tilde{c}^1, \tilde{r}(\pm\alpha)] \\ &= \pm 2 \exp(\theta : J_z :)\tilde{r}(\pm\alpha) \\ &= \pm 2\hat{r}(\pm\alpha, \theta), \end{aligned} \tag{27.5.73}$$

as expected.

## 27.6 Symplectic Classification of Analytic Vector Fields in Four Variables

For the case of analytic vector fields, and in the spirit of Section 17.3, we need to consider in this section vector fields of the form  $\mathcal{L}_{\mathbf{g}^m}$  where the components of  $\mathbf{g}^m$  are homogeneous polynomials of degree  $m$  in the 4 variables  $z_1$  through  $z_4$ . Mutatis mutandis, many of the same results follow as before. With  $\Sigma$  defined by

$$\Sigma = \sum_a z_a (\partial/\partial z_a), \tag{27.6.1}$$

the relations (3.2) and (3.3) are still true. Also, (3.4) remains true, and (3.5) and (3.6) hold when  $f_2$  is quadratic in  $z_1$  through  $z_4$ . It follows that in the 4-variable case the Hamiltonian vector fields  $: h_m :$  are transformed into each other under the action of  $sp(4, \mathbb{R})$  and carry the irreducible representation  $\Gamma(m, 0)$ . Also, any  $\mathcal{L}_{\mathbf{g}^0}$  is a Hamiltonian vector field, and these fields carry the representation  $\Gamma(1, 0)$ .

Next consider the vector fields  $\mathcal{L}_{\mathbf{g}^1}$ . They form a 16-dimensional space spanned by the vector fields  $z_a (\partial/\partial z_b)$  with  $a, b = 1, 2, 3, 4$ . We know that any  $: h_2 :$  is such a vector field, and that these vector fields carry the representation  $\Gamma(2, 0)$ , which is 10 dimensional. Also,  $\Sigma$  is of the form  $\mathcal{L}_{\mathbf{g}^1}$  and, by (3.8), carries the 1-dimensional representation  $\Gamma(0, 0)$ . We will see that any  $\mathcal{L}_{\mathbf{g}^1}$  can be written uniquely in the form

$$\mathcal{L}_{\mathbf{g}^1} = \mathcal{H}^{2,0} + \mathcal{G}^{0,1} + \mathcal{G}^{0,0}. \tag{27.6.2}$$

Here  $\mathcal{H}^{2,0}$  denotes a Hamiltonian vector field of the form

$$\mathcal{H}^{2,0} =: h_2 :, \tag{27.6.3}$$

which therefore carries the representation  $\Gamma(2, 0)$ .  $\mathcal{G}^{0,0}$  is a non-Hamiltonian vector field that is a (constant) multiple of  $\Sigma$ , and therefore carries the representation  $\Gamma(0, 0)$ . Finally,  $\mathcal{G}^{0,1}$  is a non-Hamiltonian vector field that carries the representation  $\Gamma(0, 1)$ . Note from Table 5.1 that  $\Gamma(0, 1)$  has dimension 5 so that we have the completeness count  $16 = 10 + 5 + 1$ .

According to (6.3) finding a suitable basis for the vector fields in  $\mathcal{H}^{2,0}$  is equivalent to finding a suitable basis for the quadratic polynomials in 4-dimensional phase space. But

this has already been done: we may use the basis provided by the  $sp(4)$  generators given in the Cartesian Cartan basis by (5.9) through (5.18).

For the non-Hamiltonian parts we will prove the claims just made by exhibiting suitable bases for  $\mathcal{G}^{0,0}$  and  $\mathcal{G}^{0,1}$ . For our purposes, it is convenient to work in the Cartesian (monomial) basis and use the transformed generators defined in (5.9) through (5.18). Consider the 4 vector fields  $z_a(\partial/\partial z_a)$  with  $a = 1$  through 4. Evidently, they are mutually commuting. We also observe that the 3 vector fields  $:\tilde{c}^1: : \tilde{c}^2: : \tilde{\mathcal{G}}_{0,0}^{0,0} = \Sigma$  are made from the  $z_a(\partial/\partial z_a)$ ,

$$:\tilde{c}^1: = -:q_1 p_1: = -p_1(\partial/\partial p_1) + q_1(\partial/\partial q_1) = z_1(\partial/\partial z_1) - z_3(\partial/\partial z_3), \quad (27.6.4)$$

$$:\tilde{c}^2: = -:q_2 p_2: = -p_2(\partial/\partial p_2) + q_2(\partial/\partial q_2) = z_2(\partial/\partial z_2) - z_4(\partial/\partial z_4), \quad (27.6.5)$$

$$\tilde{\mathcal{G}}_{0,0}^{0,0} = \Sigma = z_1(\partial/\partial z_1) + z_2(\partial/\partial z_2) + z_3(\partial/\partial z_3) + z_4(\partial/\partial z_4). \quad (27.6.6)$$

As a fourth such linearly independent vector field we take the element  $\tilde{\mathcal{G}}_{0,0}^{0,1}$  defined by the equation

$$\tilde{\mathcal{G}}_{0,0}^{0,1} = z_1(\partial/\partial z_1) + z_3(\partial/\partial z_3) - z_2(\partial/\partial z_2) - z_4(\partial/\partial z_4). \quad (27.6.7)$$

See Exercise 6.1. In addition we define other elements  $\tilde{\mathcal{G}}_{k,\ell}^{0,1}$  by the equations

$$\begin{aligned} \tilde{\mathcal{G}}_{1,1}^{0,1} &= (1/2)\#\tilde{r}(\boldsymbol{\beta})\#\tilde{\mathcal{G}}_{0,0}^{0,1} = (1/2)\{:\tilde{r}(\boldsymbol{\beta}):, \tilde{\mathcal{G}}_{0,0}^{0,1}\} \\ &= -z_1(\partial/\partial z_4) + z_2(\partial/\partial z_3), \end{aligned} \quad (27.6.8)$$

$$\begin{aligned} \tilde{\mathcal{G}}_{-1,-1}^{0,1} &= (1/2)\#\tilde{r}(-\boldsymbol{\beta})\#\tilde{\mathcal{G}}_{0,0}^{0,1} = (1/2)\{:\tilde{r}(-\boldsymbol{\beta}):, \tilde{\mathcal{G}}_{0,0}^{0,1}\} \\ &= z_4(\partial/\partial z_1) - z_3(\partial/\partial z_2), \end{aligned} \quad (27.6.9)$$

$$\begin{aligned} \tilde{\mathcal{G}}_{-1,1}^{0,1} &= (1/2)\#\tilde{r}(\boldsymbol{\delta})\#\tilde{\mathcal{G}}_{0,0}^{0,1} = (1/2)\{:\tilde{r}(\boldsymbol{\delta}):, \tilde{\mathcal{G}}_{0,0}^{0,1}\} \\ &= -z_3(\partial/\partial z_4) - z_2(\partial/\partial z_1), \end{aligned} \quad (27.6.10)$$

$$\begin{aligned} \tilde{\mathcal{G}}_{1,-1}^{0,1} &= (1/2)\#\tilde{r}(-\boldsymbol{\delta})\#\tilde{\mathcal{G}}_{0,0}^{0,1} = (1/2)\{:\tilde{r}(-\boldsymbol{\delta}):, \tilde{\mathcal{G}}_{0,0}^{0,1}\} \\ &= z_4(\partial/\partial z_3) + z_1(\partial/\partial z_2). \end{aligned} \quad (27.6.11)$$

Since  $\Sigma$  commutes with all the  $sp(4)$  generators, we immediately have the results

$$\#\tilde{c}^j\#\tilde{\mathcal{G}}_{0,0}^{0,0} = \{:\tilde{c}^j:, \Sigma\} = -\{\Sigma, :\tilde{c}^j:\} = 0, \quad (27.6.12)$$

$$\#r(\boldsymbol{\mu})\#\tilde{\mathcal{G}}_{0,0}^{0,0} = \{:\tilde{r}(\boldsymbol{\mu}):, \Sigma\} = -\{\Sigma, :\tilde{r}(\boldsymbol{\mu}): \} = 0. \quad (27.6.13)$$

Thus, in keeping with its labels,  $\tilde{\mathcal{G}}_{0,0}^{0,0}$  carries the representation  $\Gamma(0,0)$ .

Direct computation shows that the five elements  $\tilde{\mathcal{G}}_{k,\ell}^{0,1}$  obey the rules

$$\#\tilde{c}^j\#\tilde{\mathcal{G}}_{k,\ell}^{0,1} = \{:\tilde{c}^j:, \tilde{\mathcal{G}}_{k,\ell}^{0,1}\} = \mathbf{e}^j \cdot (k\mathbf{e}^1 + \ell\mathbf{e}^2)\tilde{\mathcal{G}}_{k,\ell}^{0,1}. \quad (27.6.14)$$

That is why the subscripts are taken to have the  $k, \ell$  values shown. Reference to Figure 5.4 shows that the right sides of (6.14) are the components of the weights for the representation

$\Gamma(0, 1)$ . In particular, we see that  $\tilde{\mathcal{G}}_{1,1}^{0,1}$  occupies the highest weight site  $\mathbf{w}^h$  given by (5.3) for the representation  $\Gamma(0, 1)$ . Therefore there should be the ladder relations

$$\#\tilde{r}(\boldsymbol{\alpha})\#\tilde{\mathcal{G}}_{1,1}^{0,1} = \#\tilde{r}(\boldsymbol{\beta})\#\tilde{\mathcal{G}}_{1,1}^{0,1} = \#\tilde{r}(-\boldsymbol{\delta})\#\tilde{\mathcal{G}}_{1,1}^{0,1} = 0. \quad (27.6.15)$$

Direct calculation shows that these relations are true. Similarly, there are the ladder relations

$$\#\tilde{r}(\pm\boldsymbol{\alpha})\#\tilde{\mathcal{G}}_{0,0}^{0,1} = 0, \quad (27.6.16)$$

$$\#\tilde{r}(\pm\boldsymbol{\gamma})\#\tilde{\mathcal{G}}_{0,0}^{0,1} = 0, \quad (27.6.17)$$

because there are no weights in  $\Gamma(0, 1)$ , see Figure 5.4, at the sites  $\pm\boldsymbol{\alpha}, \pm\boldsymbol{\gamma}$ . Further calculation gives the relation

$$\#\tilde{r}(-\boldsymbol{\beta})\#\tilde{\mathcal{G}}_{1,1}^{0,1} = \tilde{\mathcal{G}}_{0,0}^{0,1}, \quad (27.6.18)$$

and all the other ladder relations one expects for the representation  $\Gamma(0, 1)$ .

For the sake of comparison, consider the Hamiltonian vector fields :  $\tilde{r}(\pm\boldsymbol{\beta})$  : and :  $\tilde{r}(\pm\boldsymbol{\delta})$  :. They belong to  $\mathcal{H}^{2,0}$  and occupy the same sites as  $\tilde{\mathcal{G}}_{\pm 1, \pm 1}^{0,1}$  and  $\tilde{\mathcal{G}}_{\mp 1, \pm 1}^{0,1}$ , respectively, in Figure 5.5. They have the form

$$:\tilde{r}(\boldsymbol{\beta}): = q_1 q_2 := z_1(\partial/\partial z_4) + z_2(\partial/\partial z_3), \quad (27.6.19)$$

$$:\tilde{r}(-\boldsymbol{\beta}): = -:p_1 p_2 := z_4(\partial/\partial z_1) + z_3(\partial/\partial z_2), \quad (27.6.20)$$

$$:\tilde{r}(\boldsymbol{\delta}): = p_1 q_2 := z_3(\partial/\partial z_4) - z_2(\partial/\partial z_1), \quad (27.6.21)$$

$$:\tilde{r}(-\boldsymbol{\delta}): = p_2 q_1 := z_4(\partial/\partial z_3) - z_1(\partial/\partial z_2). \quad (27.6.22)$$

Evidently the vector fields (6.7) through (6.11) and (6.19) through (6.22) are linearly independent. However, in contrast to (6.15), we have the nonzero result

$$\#\tilde{r}(-\boldsymbol{\delta})\# :\tilde{r}(\boldsymbol{\beta}) : = -(\sqrt{2}) :\tilde{r}(\boldsymbol{\alpha}) :. \quad (27.6.23)$$

Our proof is complete, and in the process of proof we have exhibited explicit expressions for the 5 vector fields  $\tilde{\mathcal{G}}_{0,0}^{0,1}$ ,  $\tilde{\mathcal{G}}_{\pm 1, \pm 1}^{0,1}$ , and  $\tilde{\mathcal{G}}_{\mp 1, \pm 1}^{0,1}$  that span  $\tilde{\mathcal{G}}^{0,1}$  in the monomial basis. If desired, these vector fields can be transformed to the resonance basis with the aid of the operator

$$\hat{\mathcal{A}}(\pi/8) = \exp[-i(\pi/8)\#p_1^2 - q_1^2 + p_2^2 - q_2^2\#]. \quad (27.6.24)$$

Of course, the results of such a transformation will again be linear combinations of  $\tilde{\mathcal{G}}_{0,0}^{0,1}$ ,  $\tilde{\mathcal{G}}_{\pm 1, \pm 1}^{0,1}$ , and  $\tilde{\mathcal{G}}_{\mp 1, \pm 1}^{0,1}$  because  $\hat{\mathcal{A}}(\pi/8)$  is generated by an  $\#f_2\#$  and we know that the set of  $\mathcal{G}^{0,1}$  is transformed into itself under such transformations. For example, we have the result

$$\mathcal{G}_{0,0}^{0,1} = \hat{\mathcal{A}}(\pi/8)\tilde{\mathcal{G}}_{0,0}^{0,1} = \quad (27.6.25)$$

At this juncture we point out that there is a one-to-one correspondence between the elements  $\tilde{\mathcal{G}}_{0,0}^{0,0}$  and  $\tilde{\mathcal{G}}_{k,\ell}^{0,1}$  and the matrices  $JA$  of Section 4.3. We first note that, according to (4.3.3), the matrices  $JA$  are transformed into themselves under the (commutator) action of  $sp(2n)$ , and therefore must form a representation of  $sp(2n)$ . Also, taken together, the matrices  $JS$  [which generate  $sp(2n)$ ] and the matrices  $JA$  generate  $g\ell(2n)$ . Similarly, the

vector fields  $\mathcal{H}^{2,0}$  [which generate  $sp(2n)$ ] and the vector fields  $\mathcal{G}^{0,1}$  and  $\mathcal{G}^{0,0}$  span  $\mathcal{L}_{\mathbf{g}^1}$ , and it is easily verified that  $\mathcal{L}_{\mathbf{g}^1}$  in turn generates  $g\ell(2n)$ . In analogy to (7.2.4), write the relations

$$\tilde{\mathcal{G}}_{0,0}^{0,0} z_c = [J\tilde{A}(0, 0; 0, 0)z]_c = \sum_d [J\tilde{A}(0, 0; 0, 0)]_{cd} z_d, \quad (27.6.26)$$

$$\tilde{\mathcal{G}}_{k,\ell}^{0,1} z_c = [J\tilde{A}(0, 1; k, \ell)z]_c = \sum_d [J\tilde{A}(0, 1; k, \ell)]_{cd} z_d. \quad (27.6.27)$$

Then, from the definitions (6.6) through (6.11), we find the results

$$\tilde{A}(0, 0; 0, 0) = -J = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad (27.6.28)$$

$$\tilde{A}(0, 1; 0, 0) = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \quad (27.6.29)$$

$$\tilde{A}(0, 1; 1, 1) = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (27.6.30)$$

$$\tilde{A}(0, 1; -1, -1) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \quad (27.6.31)$$

$$\tilde{A}(0, 1; -1, 1) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (27.6.32)$$

$$\tilde{A}(0, 1; 1, -1) = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \quad (27.6.33)$$

Note that, as expected, the matrices  $\tilde{A}$  are antisymmetric and span the space of  $4 \times 4$  antisymmetric matrices. Also, the matrix  $JA$ , when exponentiated and with  $A$  given by any multiple of the  $\tilde{A}$  in (6.28), produces a positive multiple of the identity matrix. The remaining  $JA$ , with  $A$  any linear combination of the five  $\tilde{A}$  given by (6.29) through (6.33), are traceless and therefore are in  $s\ell(4, \mathbb{R})$ .

We now turn to the general case  $\mathcal{L}_{\mathbf{g}^m}$  with  $m \geq 1$ . Any such vector field can be written in the form

$$\mathcal{L}_{\mathbf{g}^m} = \sum_{a=1}^4 g_a^m (\partial/\partial z_a). \quad (27.6.34)$$

We know that the  $g_a^m$  carry the representation  $\Gamma(m, 0)$  and the  $(\partial/\partial z_a)$  carry the representation  $\Gamma(1, 0)$ . It follows as before from the derivation property of  $\#f_2\#$  that  $\mathcal{L}_{\mathbf{g}^m}$  must carry the direct product representation  $\Gamma(m, 0) \otimes \Gamma(1, 0)$ . In the case of  $sp(4)$  there is the Clebsch-Gordan series result

$$\Gamma(m, 0) \otimes \Gamma(1, 0) = \Gamma(m+1, 0) \oplus \Gamma(m-1, 1) \oplus \Gamma(m-1, 0). \quad (27.6.35)$$

Consequently, any  $\mathcal{L}_{\mathbf{g}^m}$  with  $m \geq 1$  has the unique decomposition

$$\mathcal{L}_{\mathbf{g}^m} = \mathcal{H}^{m+1,0} + \mathcal{G}^{m-1,1} + \mathcal{G}^{m-1,0}. \quad (27.6.36)$$

Here  $\mathcal{H}^{m+1,0}$  is a Hamiltonian vector field that carries the representation  $\Gamma(m+1, 0)$ , and  $\mathcal{G}^{m-1,1}$  and  $\mathcal{G}^{m-1,0}$  are non-Hamiltonian vector fields that carry the representations  $\Gamma(m-1, 1)$  and  $\Gamma(m-1, 0)$ , respectively.

The vector field  $\mathcal{H}^{m+1,0}$  is of the form :  $h_{m+1}$  :. Finding a basis for  $\mathcal{G}^{m-1,0}$  is also easy. Since  $\Sigma$  carries the representation  $\Gamma(0, 0)$ , we have the result that  $\mathcal{G}^{m-1,0}$  is of the form

$$\mathcal{G}^{m-1,0} = f_{m-1}\Sigma, \quad (27.6.37)$$

where  $f_{m-1}$  is any homogeneous polynomial of degree  $(m-1)$ . Finding the basis elements for  $\mathcal{G}^{m-1,1}$  in general requires some work. One may, for example, follow a procedure similar to that used to find the basis for  $\mathcal{G}^{0,1}$ . A more systematic procedure would be to compute and tabulate the complete set of Clebsch-Gordan coefficients for the first several representations of  $sp(4)$ . These coefficients could then be used for modest  $m$  to find a basis for  $\Gamma(m-1, 1)$  in terms of the basis elements for  $\Gamma(m, 0) \times \Gamma(1, 0)$ .

As a specific example, let us consider vector fields of the form  $\mathcal{L}_{\mathbf{g}^2}$ . This set of fields has dimension  $4N(2, 4) = 40$ . The set of Hamiltonian vector fields  $\mathcal{H}^{3,0}$  has dimension  $N(3, 4) = 20$ . The set of non-Hamiltonian vector fields  $\mathcal{G}^{1,0}$  has dimension  $\dim \Gamma(1, 0) = 4$ . See Figure 5.3. They are spanned by the basis elements  $z_a\Sigma$ . The set of non-Hamiltonian vector fields  $\mathcal{G}^{1,1}$  has dimension  $\dim \Gamma(1, 1) = 16$ . See Figure 5.6. Finding a basis for them would require more work. Some further tools for this task are described in Sections 21.10 and 21.11.2.

## Exercises

**27.6.1.** The choice of  $\tilde{\mathcal{G}}_{0,0}^{0,1}$  as given in (6.7) must be made with care to assure that it is “pure”  $\Gamma(0, 1)$  and contains no  $\Gamma(0, 0)$  or  $\Gamma(2, 0)$  “contamination”. For example, one could add any amount of  $\Sigma$  and :  $\tilde{c}^j$  : to the selected  $\tilde{\mathcal{G}}_{0,0}^{0,1}$  and still satisfy (6.14) with  $k, \ell = 0$ . Show that the truth of (6.18) ensures that  $\tilde{\mathcal{G}}_{0,1}^{0,0}$  as defined by (6.7) has no  $\Gamma(0, 0)$  contamination. Show that the truth of (6.16) and (6.17) ensures that  $\tilde{\mathcal{G}}_{0,0}^{0,1}$  as defined by (6.7) has no  $\Gamma(2, 0)$  contamination. Verify the relations (6.8) through (6.23). The basis elements given by (6.7) through (6.11) have the weights displayed in Figure 5.4. Verify that any attempt to raise or lower an element to produce one with a weight different from those shown in Figure 5.4 leads to a null result as in (6.15) through (6.18).

**27.6.2.** Verify (6.28) through (6.33).

**27.6.3.** The relation (6.35) implies the relation

$$\begin{aligned} [\dim \Gamma(m, 0)][\dim \Gamma(1, 0)] = \\ \dim \Gamma(m+1, 0) + \dim \Gamma(m-1, 1) + \dim \Gamma(m-1, 0). \end{aligned} \quad (27.6.38)$$

Verify this relation using (5.4).

## 27.7 Structure of $sp(6, \mathbb{R})$

The Lie algebra  $sp(6, \mathbb{R})$  is 21 dimensional, and its Cartan subalgebra is 3 dimensional. Therefore, in the Cartan basis, there should be 18 ladder operators. They are labelled by 18 three-component root vectors consisting of 9 vectors and their negatives. For convenience, we will call these 9 vectors  $\alpha^j$ ,  $\beta^j$ , and  $\gamma^j$  where  $j$  ranges from 1 through 3. They are given in terms of three orthogonal unit vectors  $e^1$  through  $e^3$  by the relations

$$\alpha^1 = 2e^1, \quad (27.7.1)$$

$$\alpha^2 = e^1 + e^2, \quad (27.7.2)$$

$$\alpha^3 = e^1 - e^2, \quad (27.7.3)$$

$$\beta^1 = 2e^2, \quad (27.7.4)$$

$$\beta^2 = e^2 + e^3, \quad (27.7.5)$$

$$\beta^3 = e^2 - e^3, \quad (27.7.6)$$

$$\gamma^1 = 2e^3, \quad (27.7.7)$$

$$\gamma^2 = e^3 + e^1, \quad (27.7.8)$$

$$\gamma^3 = e^3 - e^1. \quad (27.7.9)$$

The 18  $sp(6)$  root vectors are shown in Figure 7.1. Note they are all of the form  $(\pm e^i \pm e^j)$  with the signs taken independently and the zero vector omitted. They satisfy the normalization relations

$$\sum_{\mu} (e^i \cdot \mu)(\mu \cdot e^j) = 16\delta_{ij}. \quad (27.7.10)$$

The Lie algebra  $sp(6, \mathbb{R})$  is generated by the monomials  $z_a z_b$  with  $a, b$  ranging from 1 to 6. In analogy with the case of  $sp(4, \mathbb{R})$ , it is convenient to use as the basis for the Cartan subalgebra the polynomials

$$c^1 = -(i/2)(p_1^2 + q_1^2), \quad (27.7.11)$$

$$c^2 = -(i/2)(p_2^2 + q_2^2), \quad (27.7.12)$$

$$c^3 = -(i/2)(p_3^2 + q_3^2). \quad (27.7.13)$$

Their associated Lie operators are Hermitian and, when exponentiated, generate a 3-torus which is a maximal torus in  $sp(6, \mathbb{R})$ . For the ladder operators in  $sp(6)$  we use the polynomials

$$r(\pm \alpha^1) = (\sqrt{2}/4)(q_1 \pm ip_1)^2, \quad (27.7.14)$$

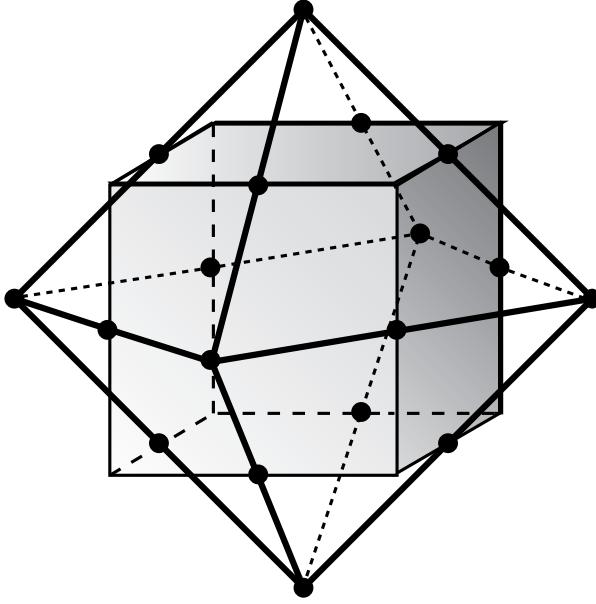


Figure 27.7.1: Root diagram showing the root vectors for  $sp(6)$ . The 6 tips of the long root vectors  $\pm\alpha^1, \pm\beta^1, \pm\gamma^1$  form the vertices of a regular octahedron. These root vectors have length 2. The remaining 12 short root vectors have length  $\sqrt{2}$ , and their tips lie at the midpoints of the 12 edges of the unit cube (the cube with edge 2).

$$r(\pm\alpha^2) = (1/2)(q_1 \pm ip_1)(q_2 \pm ip_2), \quad (27.7.15)$$

$$r(\pm\alpha^3) = (i/2)(q_1 \pm ip_1)(q_2 \mp ip_2), \quad (27.7.16)$$

$$r(\pm\beta^1) = (\sqrt{2}/4)(q_2 \pm ip_2)^2, \quad (27.7.17)$$

$$r(\pm\beta^2) = (1/2)(q_2 \pm ip_2)(q_3 \pm ip_3), \quad (27.7.18)$$

$$r(\pm\beta^3) = (i/2)(q_2 \pm ip_2)(q_3 \mp ip_3), \quad (27.7.19)$$

$$r(\pm\gamma^1) = (\sqrt{2}/4)(q_3 \pm ip_3)^2, \quad (27.7.20)$$

$$r(\pm\gamma^2) = (1/2)(q_3 \pm ip_3)(q_1 \pm ip_1), \quad (27.7.21)$$

$$r(\pm\gamma^3) = (i/2)(q_3 \pm ip_3)(q_1 \mp ip_1). \quad (27.7.22)$$

Their associated Lie operators obey the standard conjugation relations (4.14). Also, the Lie algebra generated by the  $c^j$  and the  $r(\mu)$  satisfy the standard rules (4.15) through (4.18). For the case of  $sp(6)$ , the  $N(\mu, \nu)$  have the values  $\pm\sqrt{2}$ . The positive  $N$ 's are  $*$ . As before, for the scalar product (7.3.12), the basis elements  $c^j$  and  $r(\mu)$  are orthonormal and therefore satisfy the relations (4.19) through (4.22).

We close this section by examining how  $sp(4)$ ,  $su(3)$ , and  $so(3)$  reside within  $sp(6)$ . The presence of  $sp(4)$  within  $sp(6)$  is obvious. Comparison of (4.6) and (4.7) with (7.11) and (7.12) shows that the  $c^j$  (with  $j = 1, 2$ ) are identical for  $sp(4)$  and  $sp(6)$ . Also, comparison of (4.1) through (4.4) with (7.1) through (7.4) indicates that, apart from labeling, the

root vectors of  $sp(4)$  are identical to a subset of those for  $sp(6)$ . Therefore, there is the correspondence

$$\boldsymbol{\alpha} \leftrightarrow \boldsymbol{\alpha}^1, \quad \boldsymbol{\beta} \leftrightarrow \boldsymbol{\alpha}^2, \quad \boldsymbol{\gamma} \leftrightarrow \boldsymbol{\beta}^1, \quad \boldsymbol{\delta} \leftrightarrow -\boldsymbol{\alpha}^3. \quad (27.7.23)$$

Figure 7.2 shows the  $sp(6)$  root vectors of Figure 7.1 viewed from above (looking against the  $\mathbf{e}^3$  axis). From this perspective, it is obvious that the root vectors (7.23) are arranged as required for  $sp(4)$ . See Figure 4.1. Finally, comparison of (4.10) through (4.13) with (7.14) through (7.17) gives the relations

$$r(\pm \boldsymbol{\alpha}) = r(\pm \boldsymbol{\alpha}^1), \quad r(\pm \boldsymbol{\beta}) = r(\pm \boldsymbol{\alpha}^2), \quad r(\pm \boldsymbol{\gamma}) = r(\pm \boldsymbol{\beta}^1), \quad r(\pm \boldsymbol{\delta}) = r(\mp \boldsymbol{\alpha}^3). \quad (27.7.24)$$

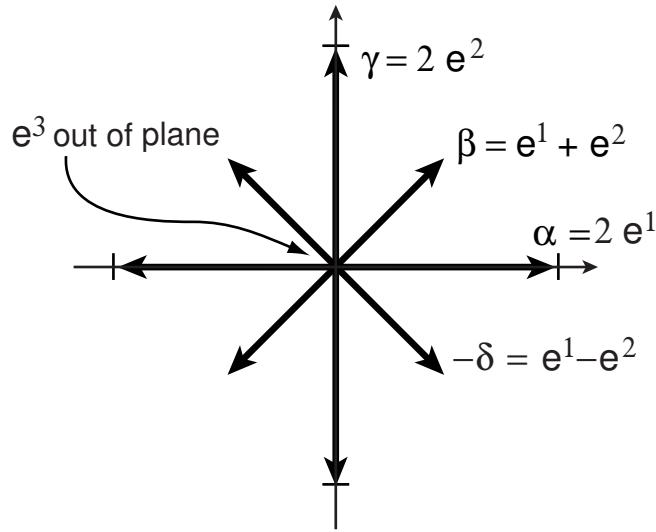


Figure 27.7.2: Top view of  $sp(6)$  root vectors of Figure 7.1 showing root vectors of an  $sp(4)$  subgroup. Only the  $sp(6)$  root vectors in the  $\mathbf{e}^1, \mathbf{e}^2$  plane are displayed. For clarity, all others are omitted. The vector  $\mathbf{e}^3$  is out of the plane of the paper.

The  $sp(4)$  just identified within  $sp(6)$  is the obvious one. Continued examination of the  $sp(6)$  root diagram of Figure 7.1 indicates that there are two more  $sp(4)$  subgroups within  $sp(6)$  gotten from the one just described by cyclically permuting the indices 1, 2, 3 on the variables  $q_1, q_2, q_3$  and  $p_1, p_2, p_3$ .

The presence of  $su(3)$  and  $u(3)$  within  $sp(6)$  is more subtle. Consider the  $sp(6)$  root vectors  $\boldsymbol{\alpha}^3, \boldsymbol{\beta}^3, \boldsymbol{\gamma}^3$  given by (7.3), (7.6), and (7.9). They all have length  $\sqrt{2}$ . They are also linearly dependent and therefore lie in a plane,

$$\boldsymbol{\alpha}^3 + \boldsymbol{\beta}^3 + \boldsymbol{\gamma}^3 = 0. \quad (27.7.25)$$

Within this plane they radiate from the origin like spokes equally “spaced” by angles of  $120^\circ$ . To verify this assertion, first note that the normal to this plane is given by the vector

$$\boldsymbol{\alpha}^3 \times \boldsymbol{\beta}^3 = \boldsymbol{\beta}^3 \times \boldsymbol{\gamma}^3 = \boldsymbol{\gamma}^3 \times \boldsymbol{\alpha}^3 = \mathbf{e}^1 + \mathbf{e}^2 + \mathbf{e}^3. \quad (27.7.26)$$

Let  $\mathbf{n}$  be the normal unit vector

$$\mathbf{n} = (\mathbf{e}^1 + \mathbf{e}^2 + \mathbf{e}^3)/\sqrt{3}. \quad (27.7.27)$$

Use of (\*) shows that there is the relation

$$\beta^3 = R(\mathbf{n}, 2\pi/3)\alpha^3, \quad (27.7.28)$$

$$\gamma^3 = R(\mathbf{n}, 2\pi/3)\beta^3, \quad (27.7.29)$$

$$\alpha^3 = R(\mathbf{n}, 2\pi/3)\gamma^3. \quad (27.7.30)$$

Figure 7.3 shows the  $sp(6)$  root vectors of Figure 7.1 viewed against the unit vector  $\mathbf{n}$ . From this perspective it is evident that the vectors  $\alpha^3$ ,  $\beta^3$ ,  $\gamma^3$  and their negatives are arranged as required for the root vectors of  $su(3)$ . Comparison of Figures 5.8.1 and 7.3 gives the correspondence

$$\pm \alpha \leftrightarrow \pm \alpha^3, \quad \pm \beta \leftrightarrow \mp \gamma^3, \quad \pm \gamma \leftrightarrow \pm \beta^3. \quad (27.7.31)$$

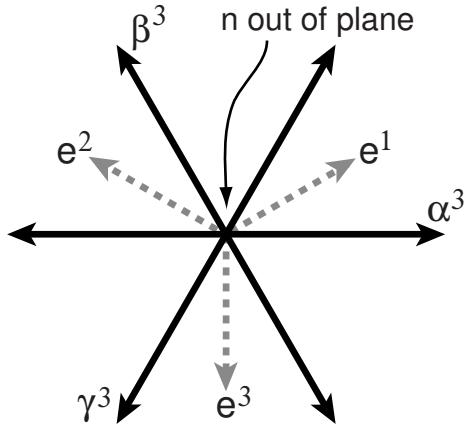


Figure 27.7.3: View against the unit vector  $\mathbf{n}$  of the  $sp(6)$  root vectors of Figure 7.1 showing root vectors of an  $su(3)$  subgroup. Only the  $sp(6)$  root vectors in the  $\alpha^3$ ,  $\beta^3$ ,  $\gamma^3$  plane and the  $e^1$ ,  $e^2$ ,  $e^3$  axes are displayed. For clarity, all others are omitted. The vector  $\mathbf{n}$  is out of the plane of the paper.

Correspondingly, comparison of (7.16), (7.19), and (7.22) with (5.8.35) gives the relations

$$r(\pm \alpha) = r(\pm \alpha^3), \quad r(\pm \beta) = r(\mp \gamma^3), \quad r(\pm \gamma) = r(\pm \beta^3). \quad (27.7.32)$$

Finally, comparison of (5.8.5) with (7.11) through (7.13) gives the relations

$$b^0 = i(c^1 + c^2 + c^3), \quad b^3 = i(c^1 - c^2), \quad b^8 = (i/\sqrt{3})(c^1 + c^2 - 2c^3). \quad (27.7.33)$$

The ladder elements  $r$  in (7.32) combined with  $b^3$  and  $b^8$  in (7.33) span  $su(3)$ ; and they all together along with  $b^0$  span  $u(3)$ .

The  $su(3)$  [and corresponding  $u(3)$ ] just identified within  $sp(6)$  is one of several such subgroups. Continued examination of Figure 7.1 indicates that there are three more. We know that the root vectors  $\pm \alpha^1$ ,  $\pm \beta^1$ ,  $\pm \gamma^1$  form the 6 vertices of a regular octahedron. An octahedron has 8 triangular faces consisting of 4 opposite pairs. There is an  $su(3)$  set of root vectors in each plane through the origin lying between and parallel to each pair of opposite faces.

Finally, there is an  $so(3)$  subalgebra within  $sp(6)$  whose presence is not obvious from looking at the  $sp(6)$  root diagram. The  $L_j$  defined by (5.8.89) generate simultaneous rotations in the  $q_1, q_2, q_3$  and  $p_1, p_2, p_3$  spaces. By (7.16), (7.19), and (7.22) they are related to elements in the Cartan basis by the equation

$$L_1 = q_2 p_3 - q_3 p_2 = r(\beta^3) - r(-\beta^3), \quad (27.7.34)$$

$$L_2 = q_3 p_1 - q_1 p_3 = r(\gamma^3) - r(-\gamma^3), \quad (27.7.35)$$

$$L_3 = q_1 p_2 - q_2 p_1 = r(\alpha^3) - r(-\alpha^3). \quad (27.7.36)$$

Note that these elements are all within  $su(3)$ .

## Exercises

**27.7.1.** Verify that, with the scalar product (7.3.12), the basis elements  $c^j$  and  $r(\mu)$  form an orthonormal set.

**27.7.2.** Verify the relations (7.28) through (7.30).

**27.7.3.** Verify that there are other  $so(3)$  subalgebras in  $sp(6)$  associated with the other  $su(3)$  subalgebras in  $sp(6)$ .

## 27.8 Representations of $sp(6, \mathbb{R})$

The description of representations of  $sp(6)$  follows the same general Cartan procedure as described for  $su(3)$  in Section 5.8 and  $sp(4)$  in Section 21.5. For  $sp(6)$ , since it has rank 3, there are three fundamental weights  $\phi^1, \phi^2$  and  $\phi^3$ . They are given by the relations

$$\phi^1 = e^1 = \alpha^1/2, \quad (27.8.1)$$

$$\phi^2 = e^1 + e^2 = \alpha^2, \quad (27.8.2)$$

$$\phi^3 = e^1 + e^2 + e^3, \quad (27.8.3)$$

and are shown in Figure 8.1 along with the  $sp(6)$  root vectors. Thus, for  $sp(6)$ , every highest weight  $w^h$  is of the form

$$w^h = \ell\phi^1 + m\phi^2 + n\phi^3 = (\ell + m + n)e^1 + (m + n)e^2 + ne^3, \quad (27.8.4)$$

where  $\ell, m$ , and  $n$  are arbitrary nonnegative integers. Correspondingly, for each  $\ell, m, n$  triplet, there is an irreducible representation  $\Gamma(\ell, m, n)$  with highest weight  $w^h$  given by (8.4). It can be shown that the dimension of  $\Gamma(\ell, m, n)$  is given by the relation

$$\begin{aligned} \dim \Gamma(\ell, m, n) &= \frac{1}{720}(\ell + 2m + 2n + 5)(\ell + m + 2n + 4)(\ell + m + n + 3) \\ &\quad \times (\ell + m + 2)(m + 2n + 3)(m + n + 2)(\ell + 1)(m + 1)(n + 1). \end{aligned} \quad (27.8.5)$$

Again see Exercise 5.1. The representations are also self conjugate,

$$\bar{\Gamma}(\ell, m, n) = \Gamma(\ell, m, n). \quad (27.8.6)$$

See Exercise 3.7.36. For quick reference the dimensions of the first few representations are listed in Table 8.1 below. Where there is no possibility of confusion, we will sometimes refer to a representation by its dimension. Note that  $\Gamma(0, 1, 0)$  and  $\Gamma(0, 0, 1)$  both have dimension 14.

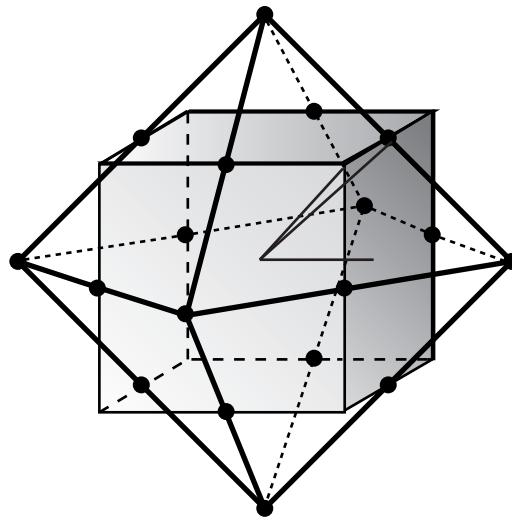


Figure 27.8.1: Fundamental weights  $\phi^1$ ,  $\phi^2$  and  $\phi^3$  for  $sp(6)$ . The root vectors are also shown.

Table 27.8.1: Dimensions of Representations of  $sp(6)$ .

$\ell$	$m$	$n$	$\dim \Gamma(\ell, m, n)$	$\ell$	$m$	$n$	$\dim \Gamma(\ell, m, n)$
0	0	0	1	3	0	0	56
1	0	0	6	2	1	0	189
0	1	0	14	2	0	1	216
0	0	1	14	1	2	0	350
2	0	0	21	1	1	1	512
1	1	0	64	1	0	2	378
1	0	1	70	0	3	0	385
0	2	0	90	0	2	1	616
0	1	1	126	0	1	2	594
0	0	2	84	0	0	3	330

From a knowledge of the root vectors and the highest weight it is a simple matter to construct weight diagrams for the various low-dimensional representations. Figures 8.2 through 8.5 show weight diagrams for the first few representations. Inspection of these

figures and reference to Table 8.1 shows that the weights must have unit multiplicities for the representations  $\Gamma(0, 0, 0)$  and  $\Gamma(1, 0, 0)$ . For  $\Gamma(0, 1, 0)$  the weight at the origin has multiplicity 2; and for  $\Gamma(2, 0, 0)$ , which is the adjoint or regular representation, the weight vector at the origin has multiplicity 3.

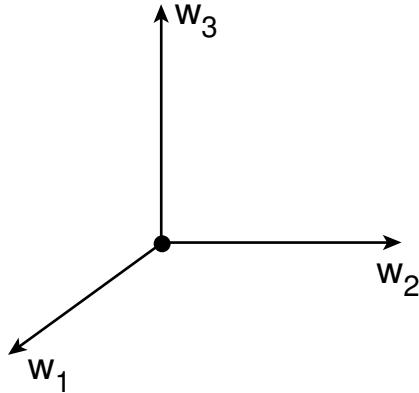


Figure 27.8.2: Weight diagram for the representation  $1 = \Gamma(0, 0, 0)$ .

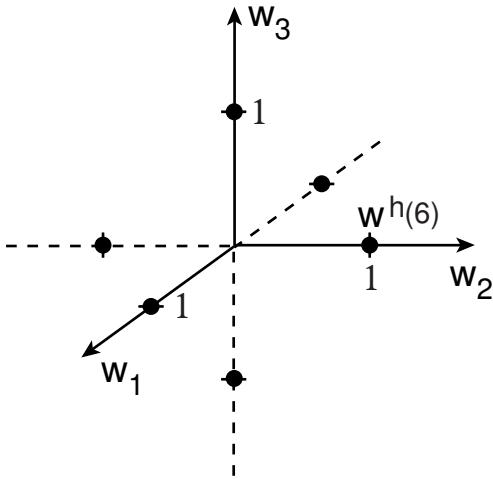


Figure 27.8.3: Weight diagram for the fundamental representation  $6 = \Gamma(1, 0, 0)$ .

Let  $\mathcal{P}_\ell$  be the space of homogeneous polynomials of degree  $\ell$  in the variables  $z_a$  with  $a = 1, 6$ . Then, by arguments that are now familiar, the space  $\mathcal{P}_\ell$  forms a representation of  $sp(6, \mathbb{R})$ . To see what representations occur we again employ a complex symplectic and unitary  $\mathcal{A}(\pi/8)$  now defined by the equation

$$\mathcal{A}(\pi/8) = \exp[-i(\pi/8) : p_1^2 - q_1^2 + p_2^2 - q_2^2 + p_3^2 - q_3^2 :]. \quad (27.8.7)$$

It transforms between the Cartesian and resonance bases by the rule

$$\begin{aligned} \mathcal{A}(\pi/8)(q_1^{r_1} p_1^{s_1} q_2^{r_2} p_2^{s_2} q_3^{r_3} p_3^{s_3}) &= (1/\sqrt{2})^{r_1+s_1+r_2+s_2+r_3+s_3} (i)^{s_1+s_2+s_3} \times \\ &(q_1 + ip_1)^{r_1} (q_1 - ip_1)^{s_1} (q_2 + ip_2)^{r_2} (q_2 - ip_2)^{s_2} (q_3 + ip_3)^{r_3} (q_3 - ip_3)^{s_3}. \end{aligned} \quad (27.8.8)$$

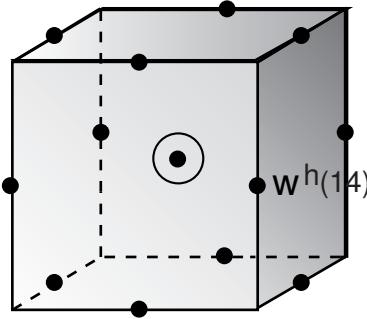


Figure 27.8.4: Weight diagram for the representation  $14 = \Gamma(0, 1, 0)$ . The circled weight at the origin has multiplicity 2. Observe from Figure 7.1 that the 12 other weights are located at the tips of the root vectors having length  $\sqrt{2}$ .

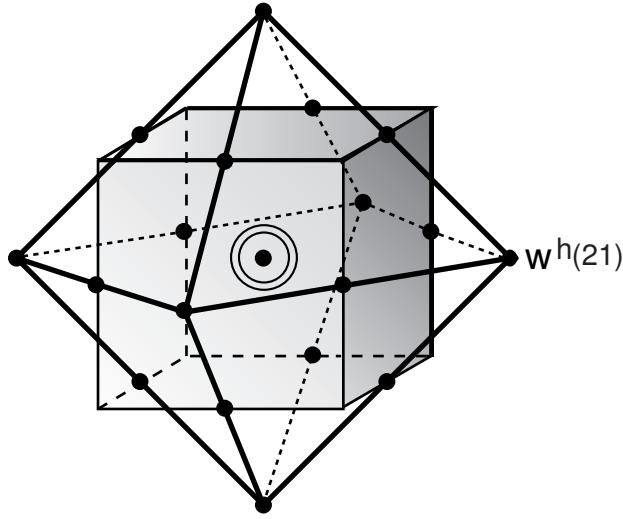


Figure 27.8.5: Weight diagram for the adjoint representation  $21 = \Gamma(2, 0, 0)$ . The doubly circled weight at the origin has multiplicity 3. The 18 other weights are located at the tips of the  $sp(6)$  root vectors.

Use of  $\mathcal{A}$  gives the transformed Lie basis polynomials  $\tilde{c}^j$  and  $\tilde{r}(\boldsymbol{\mu})$  listed below:

$$\tilde{c}^j = \mathcal{A}(-\pi/8)c^j = -q_j p_j, \quad (27.8.9)$$

$$\tilde{r}(\boldsymbol{\alpha}^1) = \mathcal{A}(-\pi/8)r(\boldsymbol{\alpha}^1) = (1/\sqrt{2})q_1^2, \quad (27.8.10)$$

$$\tilde{r}(-\boldsymbol{\alpha}^1) = \mathcal{A}(-\pi/8)r(-\boldsymbol{\alpha}^1) = -(1/\sqrt{2})p_1^2, \quad (27.8.11)$$

$$\tilde{r}(\boldsymbol{\alpha}^2) = \mathcal{A}(-\pi/8)r(\boldsymbol{\alpha}^2) = q_1 q_2, \quad (27.8.12)$$

$$\tilde{r}(-\boldsymbol{\alpha}^2) = \mathcal{A}(-\pi/8)r(-\boldsymbol{\alpha}^2) = -p_1 p_2, \quad (27.8.13)$$

$$\tilde{r}(\boldsymbol{\alpha}^3) = \mathcal{A}(-\pi/8)r(\boldsymbol{\alpha}^3) = q_1 p_2, \quad (27.8.14)$$

$$\tilde{r}(-\boldsymbol{\alpha}^3) = \mathcal{A}(-\pi/8)r(-\boldsymbol{\alpha}^3) = q_2 p_1, \quad (27.8.15)$$

$$\tilde{r}(\boldsymbol{\beta}^1) = \mathcal{A}(-\pi/8)r(\boldsymbol{\beta}^1) = (1/\sqrt{2})q_2^2, \quad (27.8.16)$$

$$\tilde{r}(-\beta^1) = \mathcal{A}(-\pi/8)r(-\beta^1) = -(1/\sqrt{2})p_2^2, \quad (27.8.17)$$

$$\tilde{r}(\beta^2) = \mathcal{A}(-\pi/8)r(\beta^2) = q_2q_3, \quad (27.8.18)$$

$$\tilde{r}(-\beta^2) = \mathcal{A}(-\pi/8)r(-\beta^2) = -p_2p_3, \quad (27.8.19)$$

$$\tilde{r}(\beta^3) = \mathcal{A}(-\pi/8)r(\beta^3) = q_2p_3, \quad (27.8.20)$$

$$\tilde{r}(-\beta^3) = \mathcal{A}(-\pi/8)r(-\beta^3) = q_3p_2, \quad (27.8.21)$$

$$\tilde{r}(\gamma^1) = \mathcal{A}(-\pi/8)r(\gamma^1) = (1/\sqrt{2})q_3^2, \quad (27.8.22)$$

$$\tilde{r}(-\gamma^1) = \mathcal{A}(-\pi/8)r(-\gamma^1) = -(1/\sqrt{2})p_3^2, \quad (27.8.23)$$

$$\tilde{r}(\gamma^2) = \mathcal{A}(-\pi/8)r(\gamma^2) = q_3q_1, \quad (27.8.24)$$

$$\tilde{r}(-\gamma^2) = \mathcal{A}(-\pi/8)r(-\gamma^2) = -p_3p_1, \quad (27.8.25)$$

$$\tilde{r}(\gamma^3) = \mathcal{A}(-\pi/8)r(\gamma^3) = q_3p_1, \quad (27.8.26)$$

$$\tilde{r}(-\gamma^3) = \mathcal{A}(-\pi/8)r(-\gamma^3) = q_1p_3. \quad (27.8.27)$$

Since  $\mathcal{A}$  is both symplectic and unitary, the transformed basis polynomials also obey the Poisson bracket rules (4.15) through (4.18), and also satisfy the orthonormality conditions (4.19) through (4.22) and the conjugation relations (4.8) and (4.14).

When the  $\tilde{c}^j$  and  $\tilde{r}(\mu)$  act on the monomials  $q_1^{r_1}p_1^{s_1}q_2^{r_2}p_2^{s_2}q_3^{r_3}p_3^{s_3}$ , there are relations analogous to those in (5.19) through (5.28), and it is evident that  $sp(6)$  acts irreducibly on  $\mathcal{P}_\ell$ . Also,  $q_1^\ell$  is the vector of highest weight in  $\mathcal{P}_\ell$  and has the weight  $\mathbf{w}^h = \ell\phi^1$ ,

$$:\tilde{c}^1:q_1^\ell = \ell q_1^\ell = (\mathbf{e}^1 \cdot \ell\phi^1)q_1^\ell = (\mathbf{e}^1 \cdot \mathbf{w}^h)q_1^\ell, \quad (27.8.28)$$

$$:\tilde{c}^2:q_1^\ell = 0 = (\mathbf{e}^2 \cdot \ell\phi^1)q_1^\ell = (\mathbf{e}^2 \cdot \mathbf{w}^h)q_1^\ell, \quad (27.8.29)$$

$$:\tilde{c}^3:q_1^\ell = 0 = (\mathbf{e}^3 \cdot \ell\phi^1)q_1^\ell = (\mathbf{e}^3 \cdot \mathbf{w}^h)q_1^\ell. \quad (27.8.30)$$

It follows that  $\mathcal{P}_\ell$  carries the representation  $\Gamma(\ell, 0, 0)$ . We also have, as expected, the result

$$\dim \mathcal{P}_\ell = N(\ell, 6) = (1/120)(\ell+5)(\ell+4)(\ell+3)(\ell+2)(\ell+1) = \dim \Gamma(\ell, 0, 0). \quad (27.8.31)$$

At this point we observe that the relations (5.8.27) and (5.8.31) can be written in the form

$$\begin{aligned} \Gamma(\ell, 0, 0) = & \sum_{m+n=\ell} \hat{\Gamma}(m, n) \oplus \sum_{m+n=\ell-2} \hat{\Gamma}(m, n) \oplus \sum_{m+n=\ell-4} \hat{\Gamma}(m, n) \oplus \cdots \\ & \oplus \hat{\Gamma}(0, 0), \text{ for } \ell \text{ even}; \end{aligned} \quad (27.8.32)$$

$$\begin{aligned} \Gamma(\ell, 0, 0) = & \sum_{m+n=\ell} \hat{\Gamma}(m, n) \oplus \sum_{m+n=\ell-2} \hat{\Gamma}(m, n) \oplus \sum_{m+n=\ell-4} \hat{\Gamma}(m, n) \oplus \cdots \\ & \oplus \hat{\Gamma}(1, 0) \oplus \hat{\Gamma}(0, 1), \text{ for } \ell \text{ odd}. \end{aligned} \quad (27.8.33)$$

[Here we have used the symbols  $\hat{\Gamma}(m, n)$  to denote representations of  $su(3)$  so as not to be confused with the symbols  $\Gamma(m, n)$  used in Section 21.5 to denote representations of  $sp(4)$ .] That is,  $sp(6)$  representations of the form  $\Gamma(\ell, 0, 0)$  can be decomposed into various  $\hat{\Gamma}(m, n)$

representations of its  $su(3)$  subgroup, and the representations listed occur once and only once.

Finally, as before, we can let  $\mathcal{A}(\pi/8)$  act on both sides of 6-variable relations analogous to (5.19) through (5.28). Doing so gives results analogous to those in (2.23) through (2.25). Consequently, as before, the  $:c^j:$  and  $:r(\boldsymbol{\mu}):$  act on the resonance basis in the same way that the  $:\tilde{c}^j:$  and  $:\tilde{r}(\boldsymbol{\mu}):$  act on the monomial basis.

## Exercises

**27.8.1.** From (7.14) verify the relation

$$r(\boldsymbol{\alpha}^1) - r(-\boldsymbol{\alpha}^1) = i\sqrt{2}q_1p_1. \quad (27.8.34)$$

Next verify that the transformation

$$\mathcal{U}(\theta) = \exp :i\theta q_1 p_1: \quad (27.8.35)$$

is symplectic and unitary, and satisfies the relations

$$\mathcal{U}(\theta)q_1 = \exp(-i\theta)q_1, \quad (27.8.36)$$

$$\mathcal{U}(\theta)p_1 = \exp(i\theta)p_1. \quad (27.8.37)$$

See Exercise 5.4.4. As a consequence verify the relations

$$\mathcal{U}(\theta)(p_1^2 + q_1^2) = p_1^2 \exp(2i\theta) + q_1^2 \exp(-2i\theta), \quad (27.8.38)$$

$$\mathcal{U}(\pi/2)(p_1^2 + q_1^2) = -(p_1^2 + q_1^2), \quad (27.8.39)$$

$$\mathcal{U}(\pi/2) : c^1 : \mathcal{U}^{-1}(\pi/2) = - : c^1 :. \quad (27.8.40)$$

Suppose that  $C^j$  and  $R(\boldsymbol{\mu})$  are any set of matrices that satisfies the commutation rules analogous to (4.15) through (4.18). That is, the  $C^j$  commute and the  $C^j$  and  $R(\boldsymbol{\mu})$  satisfy the rules (5.8.12) through (5.8.14). By this definition, they provide a matrix representation of  $sp(6)$ . Since the relation (8.40) is purely a consequence of Lie-algebraic rules, show that there must be the matrix relation

$$U(\pi/2)C^1[U(\pi/2)]^{-1} = -C^1, \quad (27.8.41)$$

where

$$U(\pi/2) = \exp\{(\pi/\sqrt{8})[R(\boldsymbol{\alpha}^1) - R(-\boldsymbol{\alpha}^1)]\}, \quad (27.8.42)$$

$$[U(\pi/2)]^{-1} = \exp\{-(\pi/\sqrt{8})[R(\boldsymbol{\alpha}^1) - R(-\boldsymbol{\alpha}^1)]\}. \quad (27.8.43)$$

Suppose  $|w_1 w_2 w_3\rangle$  is a vector in this representation with the property

$$C^j |w_1 w_2 w_3\rangle = w_j |w_1 w_2 w_3\rangle. \quad (27.8.44)$$

Show that the vector  $[U(\pi/2)]^{-1}|\boldsymbol{w}\rangle$  has the property

$$C^1[U^{-1}(\pi/2)]^{-1}|\boldsymbol{w}\rangle = -w_1[U(\pi/2)]^{-1}|\boldsymbol{w}\rangle. \quad (27.8.45)$$

Prove that if  $(w_1, w_2, w_3)$  is a weight vector, so is  $(-w_1, w_2, w_3)$ . Generalize this result to show that if  $(w_1, w_2, w_3)$  is a weight vector, so are the vectors  $(\pm w_1, \pm w_2, \pm w_3)$  where all  $\pm$  signs are taken independently. Verify similar results for  $sp(2)$  and  $sp(4)$ . Verify that the weight diagrams shown in Sections 21.2, 21.5, and 21.8 have this property.

**27.8.2.** Verify (8.31).

**27.8.3.** Verify by a dimension count that  $\Gamma(1, 0, 0)$  is the fundamental representation of  $sp(6)$  and  $\Gamma(2, 0, 0)$  is the adjoint representation. Repeat analogous calculations for the cases of  $sp(2)$  and  $sp(4)$ .

**27.8.4.** Work out the weight diagram for the  $sp(6)$  representation  $\Gamma(0, 0, 1)$ .

## 27.9 Symplectic Classification of Analytic Vector Fields in Six Variables

The symplectic classification of analytic vector fields in six variables is similar to the 4-variable case. As before, it suffices to consider homogeneous vector fields. The Hamiltonian vector fields :  $h_\ell$  : are transformed into each other under the action of  $sp(6, \mathbb{R})$ , and carry the representation  $\Gamma(\ell, 0, 0)$ . Any  $\mathcal{L}_{\mathbf{g}^0}$  is a Hamiltonian vector field, and these fields carry the representation  $\Gamma(1, 0, 0)$ . The vector field  $\Sigma$  defined by (6.1) with  $a$  ranging from 1 to 6 carries the representation  $\Gamma(0, 0, 0)$ .

The 6-dimensional analog of (6.34) shows that in this case the  $\mathcal{L}_{\mathbf{g}^\ell}$  carry the direct product representation  $\Gamma(\ell, 0, 0) \otimes \Gamma(1, 0, 0)$ . For  $sp(6)$  there is the Clebsch-Gordan series result

$$\Gamma(\ell, 0, 0) \otimes \Gamma(1, 0, 0) = \Gamma(\ell + 1, 0, 0) \oplus \Gamma(\ell - 1, 1, 0) \oplus \Gamma(\ell - 1, 0, 0). \quad (27.9.1)$$

Consequently, any  $\mathcal{L}_{\mathbf{g}^\ell}$  with  $\ell \geq 1$  has the unique decomposition

$$\mathcal{L}_{\mathbf{g}^\ell} = \mathcal{H}^{\ell+1,0,0} + \mathcal{G}^{\ell-1,1,0} + \mathcal{G}^{\ell-1,0,0}. \quad (27.9.2)$$

Here  $\mathcal{H}^{\ell+1,0,0}$  is a Hamiltonian vector field that carries the representation  $\Gamma(\ell + 1, 0, 0)$ , and is of the form :  $h_{\ell+1}$  :. The quantities  $\mathcal{G}^{\ell-1,1,0}$  and  $\mathcal{G}^{\ell-1,0,0}$  are non-Hamiltonian vector fields that carry the representations  $\Gamma(\ell - 1, 1, 0)$  and  $\Gamma(\ell - 1, 0, 0)$ , respectively. The vector fields  $\mathcal{G}^{\ell-1,0,0}$  are of the form

$$\mathcal{G}^{\ell-1,0,0} = f_{\ell-1} \Sigma \quad (27.9.3)$$

where  $f_{\ell-1}$  is any homogeneous polynomial of degree  $(\ell - 1)$ . The construction of the vector fields that span  $\mathcal{G}^{\ell-1,1,0}$  requires special effort.

As before we will work out the simplest case  $\ell = 1$  for which we have the result

$$\mathcal{L}_{\mathbf{g}^1} = \mathcal{H}^{2,0,0} + \mathcal{G}^{0,1,0} + \mathcal{G}^{0,0,0}. \quad (27.9.4)$$

Since  $\mathcal{L}_{\mathbf{g}^1}$  is spanned by the vector fields  $z_a(\partial/\partial z_b)$  with  $a, b = 1$  through 6, it has dimension 36. We know that  $\mathcal{H}^{2,0,0}$  has dimension  $N(2, 6) = 21$ , and  $\mathcal{G}^{0,0,0}$  has dimension 1. It follows that  $\mathcal{G}^{0,1,0}$  has dimension  $(36 - 21 - 1) = 14$ , which we know is the dimension of  $\Gamma(0, 1, 0)$ .

Similar to the case of  $sp(4)$  treated in Section 21.6, finding a suitable basis for the vector fields in  $\mathcal{H}^{2,0,0}$  is equivalent to finding a suitable basis for the quadratic polynomials in 6-dimensional phase space; and these basis polynomials may be taken to be the  $sp(6)$  generators given in the Cartesian Cartan basis by (8.9) through (8.27).

To find a basis for the non-Hamiltonian parts we will begin with the 6 mutually commuting vector fields  $z_a(\partial/\partial z_a)$  with  $a = 1$  through 6. The 4 vector fields  $:\tilde{c}^j:$  and  $\tilde{\mathcal{G}}_{0,0,0}^{0,0,0} = \Sigma$  are made from the  $z_a(\partial/\partial z_a)$ ,

$$:\tilde{c}^1 := z_1(\partial/\partial z_1) - z_4(\partial/\partial z_4), \quad (27.9.5)$$

$$:\tilde{c}^2 := z_2(\partial/\partial z_2) - z_5(\partial/\partial z_5), \quad (27.9.6)$$

$$:\tilde{c}^3 := z_3(\partial/\partial z_3) - z_6(\partial/\partial z_6), \quad (27.9.7)$$

$$\tilde{\mathcal{G}}_{0,0,0}^{0,0,0} = \Sigma = z_1(\partial/\partial z_1) + z_2(\partial/\partial z_2) + z_3(\partial/\partial z_3) + z_4(\partial/\partial z_4) + z_5(\partial/\partial z_5) + z_6(\partial/\partial z_6). \quad (27.9.8)$$

Let us define vector fields  $\tilde{\Sigma}^j$  by the equations

$$\tilde{\Sigma}^1 = q_1(\partial/\partial q_1) + p_1(\partial/\partial p_1) = z_1(\partial/\partial z_1) + z_4(\partial/\partial z_4), \quad (27.9.9)$$

$$\tilde{\Sigma}^2 = q_2(\partial/\partial q_2) + p_2(\partial/\partial p_2) = z_2(\partial/\partial z_2) + z_5(\partial/\partial z_5), \quad (27.9.10)$$

$$\tilde{\Sigma}^3 = q_3(\partial/\partial q_3) + p_3(\partial/\partial p_3) = z_3(\partial/\partial z_3) + z_6(\partial/\partial z_6). \quad (27.9.11)$$

They are obviously independent of the  $:\tilde{c}^j:$  and are also made from the  $z_a(\partial/\partial z_a)$ .

We already have the  $:\tilde{c}^j:$  and the combination

$$\Sigma = \tilde{\Sigma}^1 + \tilde{\Sigma}^2 + \tilde{\Sigma}^3. \quad (27.9.12)$$

As the 5th and 6th such linearly independent vectors we take the elements  ${}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0}$  and  ${}^8\tilde{\mathcal{G}}_{0,0,0}^{0,1,0}$  defined by the equations

$${}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = \tilde{\Sigma}^1 - \tilde{\Sigma}^2 = z_1(\partial/\partial z_1) + z_4(\partial/\partial z_4) - z_2(\partial/\partial z_2) - z_5(\partial/\partial z_5), \quad (27.9.13)$$

$$\begin{aligned} {}^8\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} &= \tilde{\Sigma}^1 + \tilde{\Sigma}^2 - 2\tilde{\Sigma}^3 \\ &= z_1(\partial/\partial z_1) + z_4(\partial/\partial z_4) + z_2(\partial/\partial z_2) \\ &\quad + z_5(\partial/\partial z_5) - 2z_3(\partial/\partial z_3) - 2z_6(\partial/\partial z_6). \end{aligned} \quad (27.9.14)$$

Here the superscripts “3” and “8” are used to refer to the analogous diagonal structure of the Gell-Mann matrices  $\lambda^3$  and  $\lambda^8$  of Section 5.8. It is easily verified that these 2 vector fields obey the relations

$$\#\tilde{c}^j \# {}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = 0, \quad (27.9.15)$$

$$\#\tilde{c}^j \# {}^8\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = 0, \quad (27.9.16)$$

and therefore are candidates for the center elements of Figure 8.4.

We define the remaining 12 vector fields that occupy the other sites of Figure 8.4 by the equations

$$\tilde{\mathcal{G}}_{1,1,0}^{0,1,0} = (1/2)\#\tilde{r}(\alpha^2)\# {}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = -z_1(\partial/\partial z_5) + z_2(\partial/\partial z_4), \quad (27.9.17)$$

$$\tilde{\mathcal{G}}_{-1,-1,0}^{0,1,0} = (1/2) \# \tilde{r}(-\boldsymbol{\alpha}^2) \# {}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = -z_4(\partial/\partial z_2) + z_5(\partial/\partial z_1), \quad (27.9.18)$$

$$\tilde{\mathcal{G}}_{1,-1,0}^{0,1,0} = (1/2) \# \tilde{r}(\boldsymbol{\alpha}^3) \# {}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = z_1(\partial/\partial z_2) + z_5(\partial/\partial z_4), \quad (27.9.19)$$

$$\tilde{\mathcal{G}}_{-1,1,0}^{0,1,0} = (1/2) \# \tilde{r}(-\boldsymbol{\alpha}^3) \# {}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = -z_2(\partial/\partial z_1) - z_4(\partial/\partial z_5), \quad (27.9.20)$$

$$\tilde{\mathcal{G}}_{0,1,1}^{0,1,0} = \# \tilde{r}(\boldsymbol{\beta}^2) \# {}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = -z_3(\partial/\partial z_5) + z_2(\partial/\partial z_6), \quad (27.9.21)$$

$$\tilde{\mathcal{G}}_{0,-1,-1}^{0,1,0} = \# \tilde{r}(-\boldsymbol{\beta}^2) \# {}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = z_5(\partial/\partial z_3) - z_6(\partial/\partial z_2), \quad (27.9.22)$$

$$\tilde{\mathcal{G}}_{0,1,-1}^{0,1,0} = \# \tilde{r}(\boldsymbol{\beta}^3) \# {}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = -z_2(\partial/\partial z_3) - z_6(\partial/\partial z_5), \quad (27.9.23)$$

$$\tilde{\mathcal{G}}_{0,-1,1}^{0,1,0} = \# \tilde{r}(-\boldsymbol{\beta}^3) \# {}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = z_3(\partial/\partial z_2) + z_5(\partial/\partial z_6), \quad (27.9.24)$$

$$\tilde{\mathcal{G}}_{1,0,1}^{0,1,0} = \# \tilde{r}(\boldsymbol{\gamma}^2) \# {}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = z_3(\partial/\partial z_4) - z_1(\partial/\partial z_6), \quad (27.9.25)$$

$$\tilde{\mathcal{G}}_{-1,0,-1}^{0,1,0} = \# \tilde{r}(-\boldsymbol{\gamma}^2) \# {}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = -z_4(\partial/\partial z_3) + z_6(\partial/\partial z_1), \quad (27.9.26)$$

$$\tilde{\mathcal{G}}_{-1,0,1}^{0,1,0} = \# \tilde{r}(\boldsymbol{\gamma}^3) \# {}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = -z_3(\partial/\partial z_1) - z_4(\partial/\partial z_6), \quad (27.9.27)$$

$$\tilde{\mathcal{G}}_{1,0,-1}^{0,1,0} = \# \tilde{r}(-\boldsymbol{\gamma}^3) \# {}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = z_1(\partial/\partial z_3) + z_6(\partial/\partial z_4). \quad (27.9.28)$$

They obey the relations

$$\# \tilde{c}^j \# \tilde{\mathcal{G}}_{k,\ell,m}^{0,1,0} = \mathbf{e}^j \cdot (k\mathbf{e}^1 + \ell\mathbf{e}^2 + m\mathbf{e}^3), \quad (27.9.29)$$

in keeping with the sites they occupy. We note that  $\tilde{\mathcal{G}}_{1,1,0}^{0,1,0}$  occupies the highest weight site  $\mathbf{w}^h$  given by (8.4) for the representation  $\Gamma(0, 1, 0)$ . Therefore there should be the ladder relation

$$\# \tilde{r}(\boldsymbol{\alpha}^3) \# \tilde{\mathcal{G}}_{1,1,0}^{0,1,0} = 0. \quad (27.9.30)$$

Direct calculation shows that this relation is true. By comparison, the vector field  $: \tilde{r}(\boldsymbol{\alpha}^2) :$  has the form

$$: \tilde{r}(\boldsymbol{\alpha}^2) : =: q_1 q_2 := z_1(\partial/\partial z_5) + z_2(\partial/\partial z_4). \quad (27.9.31)$$

It occupies the same 1,1,0 site in Figure 8.5,

$$\begin{aligned} \# \tilde{c}^j \# : \tilde{r}(\boldsymbol{\alpha}^2) : &= \{ : \tilde{c}^j : , : \tilde{r}(\boldsymbol{\alpha}^2) : \} =: [\tilde{c}^j, \tilde{r}(\boldsymbol{\alpha}^2)] : \\ &= (\mathbf{e}^j \cdot \boldsymbol{\alpha}^2) : \tilde{r}(\boldsymbol{\alpha}^2) := [\mathbf{e}^j \cdot (\mathbf{e}^1 + \mathbf{e}^2)] : \tilde{r}(\boldsymbol{\alpha}^2) :, \end{aligned} \quad (27.9.32)$$

and is evidently linearly independent of  $\tilde{\mathcal{G}}_{1,1,0}^{0,1,0}$ . It satisfies the relation

$$\# \tilde{r}(\boldsymbol{\alpha}^3) \# : \tilde{r}(\boldsymbol{\alpha}^2) := -(\sqrt{2}) : \tilde{r}(\boldsymbol{\alpha}^1) : . \quad (27.9.33)$$

It can be verified that the  ${}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0}$ ,  ${}^8\tilde{\mathcal{G}}_{0,0,0}^{0,1,0}$ , and  $\tilde{\mathcal{G}}_{k,\ell,m}^{0,1,0}$  satisfy all the ladder relations one expects for the representation  $\Gamma(0, 1, 0)$ . For example, there is the relation

$$(1/2) \# \tilde{r}(-\boldsymbol{\alpha}^2) \# \# \tilde{r}(\boldsymbol{\alpha}^2) \# {}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = {}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} \quad (27.9.34)$$

which shows that  ${}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0}$  has no  $\Gamma(0, 0, 0)$  contamination. (See Exercise 6.1.) There is also the relation

$$\# \tilde{r}(-\boldsymbol{\beta}^2) \# \# \tilde{r}(\boldsymbol{\beta}^2) \# {}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = (1/2) ({}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} - {}^8\tilde{\mathcal{G}}_{0,0,0}^{0,1,0}) \quad (27.9.35)$$

which shows that  ${}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0}$  and  ${}^8\tilde{\mathcal{G}}_{0,0,0}^{0,1,0}$  can be transformed into each other, and that  ${}^8\tilde{\mathcal{G}}_{0,0,0}^{0,1,0}$  as well has no  $\Gamma(0, 0, 0)$  contamination. There are also the relations

$$\#\tilde{r}(\boldsymbol{\nu})\#{}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = \#\tilde{r}(\boldsymbol{\nu})\#{}^8\tilde{\mathcal{G}}_{0,0,0}^{0,1,0} = 0 \quad (27.9.36)$$

for  $\boldsymbol{\nu} = \pm\boldsymbol{\alpha}^1, \pm\boldsymbol{\beta}^1, \pm\boldsymbol{\gamma}^1$ . Note that the sites  $\pm 2, 0, 0$  and  $0, \pm 2, 0$  and  $0, 0, \pm 2$  are empty in Figure 8.4 and occupied in Figure 8.5. Therefore (9.32) shows that  ${}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0}$  and  ${}^8\tilde{\mathcal{G}}_{0,0,0}^{0,1,0}$ , and hence the  $\tilde{\mathcal{G}}_{k,\ell,m}^{0,1,0}$ , have no  $\Gamma(2, 0, 0)$  contamination.

Finally, we observe that the results we have obtained in the monomial basis can be transformed if desired to the resonance basis with the aid of the operator

$$\hat{\mathcal{A}}(\pi/8) = \exp[-i(\pi/8)\#p_1^2 - q_1^2 + p_2^2 - q_2^2 + p_3^3 - q_3^2\#]. \quad (27.9.37)$$

## Exercises

**27.9.1.** Work out the analogs of the relations (9.17) through (9.28) with  ${}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0}$  replaced by  ${}^8\tilde{\mathcal{G}}_{0,0,0}^{0,1,0}$ .

**27.9.2.** Work out the  $6 \times 6$  matrices  $\tilde{A}$  corresponding to the vector fields  $\tilde{\mathcal{G}}_{0,0,0}^{0,0,0}$ ,  ${}^3\tilde{\mathcal{G}}_{0,0,0}^{0,1,0}$ ,  ${}^8\tilde{\mathcal{G}}_{0,0,0}^{0,1,0}$ , and  $\tilde{\mathcal{G}}_{k,\ell,m}^{0,1,0}$  in analogy to what was done for the  $4 \times 4$  case at the end of Section 21.6. Show that the matrix  $JA$ , when exponentiated and with  $A$  given by any multiple of the  $\tilde{A}$  associated with  $\tilde{\mathcal{G}}_{0,0,0}^{0,0,0}$ , produces a positive multiple of the identity matrix. Show that the remaining  $JA$ , with  $A$  any linear combination of the fourteen  $\tilde{A}$  associated with the remaining  $\tilde{\mathcal{G}}$  given by (9.13) and (9.14) and (9.17) through (9.28), are traceless and therefore are in  $sl(6, \mathbb{R})$ .

**27.9.3.** Write and verify the  $sp(6)$  analog of (6.38), as given in Exercise 6.3, using (9.1) and (8.5).

**27.9.4.** In Sections 21.3, 21.6, and 21.9 we learned that  $\Sigma$  always was invariant under the action of  $sp(2)$ ,  $sp(4)$ , and  $sp(6)$ . The purpose of this exercise is to show that this invariance is a consequence of a more general result. Consider  $m$ -dimensional Euclidean space with coordinates

$$x = (x_1, x_2, \dots, x_m). \quad (27.9.38)$$

Here  $m$  can be even or odd. Define a vector field  $\Sigma$  by the relation

$$\Sigma = \sum_{a=1}^m x_a \partial/\partial x_a. \quad (27.9.39)$$

Suppose that each  $x$  is sent to  $\bar{x}$  under the action of some linear, but invertible, transformation  $M$ ,

$$\bar{x} = Mx. \quad (27.9.40)$$

Define a transformed vector field  $\bar{\Sigma}$  by the rule

$$\bar{\Sigma} = \sum_{a=1}^m \bar{x}_a \partial/\partial \bar{x}_a. \quad (27.9.41)$$

You are to show that

$$\bar{\Sigma} = \Sigma. \quad (27.9.42)$$

That is,  $\Sigma$  is invariant under the action of  $M$ . Because  $M$  is any invertible matrix, we may say that  $\Sigma$  is invariant under the group  $GL(m, \mathbb{R})$ . Since  $Sp(2n, \mathbb{R})$  is a subgroup of  $GL(2n, \mathbb{R})$ , it follows that  $\Sigma$  is also invariant under the group  $Sp(2n, \mathbb{R})$ .

Begin by inverting (9.40),

$$x = M^{-1}\bar{x}, \quad (27.9.43)$$

and verify that this relation has the component form

$$x_a = \sum_b (M^{-1})_{ab} \bar{x}_b. \quad (27.9.44)$$

Show it follows that

$$\partial x_a / \partial \bar{x}_b = (M^{-1})_{ab}. \quad (27.9.45)$$

Verify by the chain rule that there is the relation

$$\partial / \partial \bar{x}_a = \sum_c (\partial x_c / \partial \bar{x}_a) \partial / \partial x_c = \sum_c (M^{-1})_{ca} \partial / \partial x_c. \quad (27.9.46)$$

Also, the relation (9.40) has the component form

$$\bar{x}_a = \sum_d M_{ad} x_d. \quad (27.9.47)$$

Verify, by employing (9.46) and (9.47) in (9.41), it follows that there is the relation

$$\begin{aligned} \bar{\Sigma} &= \sum_{acd} M_{ad} (M^{-1})_{ca} x_d \partial / \partial x_c = \sum_{acd} (M^{-1})_{ca} M_{ad} x_d \partial / \partial x_c \\ &= \sum_{cd} (M^{-1} M)_{cd} x_d \partial / \partial x_c = \sum_{cd} \delta_{cd} x_d \partial / \partial x_c \\ &= \sum_c x_c \partial / \partial x_c = \Sigma. \end{aligned} \quad (27.9.48)$$

For an infinitesimal (Lie-algebraic) version of (9.42), see Exercise 10.8.

**27.9.5.** Suppose an object of mass  $m$  is acted upon by a force  $\mathbf{F}$  arising from a potential  $V$  and a velocity dependent drag force,

$$\mathbf{F} = -\nabla V - 2\beta \mathbf{v}. \quad (27.9.49)$$

Define the particle's momentum  $\mathbf{p}$  in the usual way

$$\mathbf{p} = m\mathbf{v} \quad (27.9.50)$$

and show that Newton's equations of motion can be written in the form

$$\dot{\mathbf{q}} = \mathcal{L}\mathbf{q}, \quad \dot{\mathbf{p}} = \mathcal{L}\mathbf{p} \quad (27.9.51)$$

where  $\mathcal{L}$  is the vector field

$$\mathcal{L} = \sum_i (p_i/m)(\partial/\partial q_i) - (\partial V/\partial q_i)(\partial/\partial p_i) - (2\beta/m)p_i(\partial/\partial p_i). \quad (27.9.52)$$

Decompose  $\mathcal{L}$  into Hamiltonian and non-Hamiltonian parts to find the result

$$\mathcal{L} =: -[p^2/(2m) + \beta \mathbf{p} \cdot \mathbf{q} + V] : -\beta \Sigma. \quad (27.9.53)$$

Suppose  $V$  is quadratic in the components of  $\mathbf{q}$ . Show that in this case the Hamiltonian  $H$  defined by

$$H = p^2/(2m) + \beta \mathbf{p} \cdot \mathbf{q} + V \quad (27.9.54)$$

evolves according to the rule

$$H = (\text{constant}) \times e^{-2\beta t}. \quad (27.9.55)$$

**27.9.6.** Consider the case of a six-dimensional phase space and all Lie operators of the form  $:f_3:$ . What are the  $sp(6)$  transformation properties of the  $:f_3:\dagger$ ?

## 27.10 Scalar Product and Projection Operators for Vector Fields

Section 7.3 described a  $USp(2n)$  invariant scalar product for phase-space functions. Here we will see that there is a related scalar product for vector fields, and we will find that the use of this vector-field scalar product illuminates the discussion of previous sections.

For our present purposes it is convenient to employ a vector-field basis slightly different from that used in (5.3.17) and (5.3.18). Let  $s_\alpha$  be the various phase-space monomials indexed by  $\alpha$  in some convenient way as in Section 7.3 or Section 32.2. Take as vector-field basis elements the quantities  $\mathcal{L}_{\alpha a}$  defined by the equation

$$\mathcal{L}_{\alpha a} = s_\alpha : z_a :. \quad (27.10.1)$$

In view of the relations

$$:z_a := \sum_b J_{ab}(\partial/\partial z_b), \quad (27.10.2)$$

$$\partial/\partial z_a = - \sum_b J_{ab} :z_b :,$$

the  $\mathcal{L}_{\alpha a}$  manifestly form a satisfactory basis. Now define a scalar product for these basis elements (and hence, by linearity, for all vector fields) by the rule

$$\langle \mathcal{L}_{\alpha a}, \mathcal{L}_{\beta b} \rangle = \langle s_\alpha, s_\beta \rangle \langle z_a, z_b \rangle. \quad (27.10.3)$$

Here the scalar product on the left of (10.3) is the vector-field scalar product, and the scalar products on the right are the phase-space function scalar products of Section 7.3. Evidently the vector-field scalar product defined by (10.3) is positive definite.

Suppose two vector fields  $\mathcal{L}_g$  and  $\mathcal{L}_h$  are specified as in (5.3.17). Then, in view of (10.2), use of (10.3) gives the result

$$\langle \mathcal{L}_g, \mathcal{L}_h \rangle = \sum_a \langle g_a, h_a \rangle. \quad (27.10.4)$$

In the case that  $g$  and  $h$  are homogeneous, there is the immediate result

$$\langle \mathcal{L}_{g^m}, \mathcal{L}_{h^n} \rangle = 0 \text{ when } m \neq n. \quad (27.10.5)$$

As another interesting case, suppose  $g_m$  and  $h_m$  are homogeneous phase-space polynomials of degree  $m$ . Then use of (10.3) gives the result

$$\langle :g_m:, :h_m: \rangle = m \langle g_m, h_m \rangle. \quad (27.10.6)$$

Let  $:f_2:$  be the Hamiltonian vector field associated with any quadratic polynomial  $f_2$ , and consider the corresponding adjoint operator  $\# : f_2 : \#$ . For the action of  $\# : f_2 : \#$  on a general vector-field basis element we have the result

$$\# : f_2 : \# \mathcal{L}_{\alpha a} = (:f_2:s_\alpha) : z_a : + s_\alpha (:f_2:z_a) :. \quad (27.10.7)$$

See (3.25). Now watch closely! Take the scalar product of (10.7) with the general basis vector  $\mathcal{L}_{\alpha' a'}$  and manipulate the result to find the relation

$$\begin{aligned} \langle \# : f_2 : \# \mathcal{L}_{\alpha a}, \mathcal{L}_{\alpha' a'} \rangle &= \langle [(:f_2:s_\alpha) : z_a : + s_\alpha (:f_2:z_a) :], s_{\alpha'} : z_{a'} : \rangle \\ &= \langle (:f_2:s_\alpha) : z_a :, s_{\alpha'} : z_{a'} : \rangle + \langle s_\alpha : (:f_2:z_a) :, s_{\alpha'} : z_{a'} : \rangle \\ &= \langle :f_2:s_\alpha, s_{\alpha'} \rangle \langle z_a, z_{a'} \rangle + \langle s_\alpha, s_{\alpha'} \rangle \langle :f_2:z_a, z_{a'} \rangle \\ &= \langle s_\alpha, :f_2:^\dagger s_{\alpha'} \rangle \langle z_a, z_{a'} \rangle + \langle s_\alpha, s_{\alpha'} \rangle \langle z_a, :f_2:^\dagger z_{a'} \rangle \\ &= \langle s_\alpha : z_a :, (:f_2:^\dagger s_{\alpha'}) : z_{a'} : \rangle + \langle s_\alpha : z_a :, s_{\alpha'} : (:f_2:^\dagger z_{a'}) : \rangle \\ &= \langle s_\alpha : z_a :, [(:f_2:^\dagger s_{\alpha'}) : z_{a'} : + s_{\alpha'} : (:f_2:^\dagger z_{a'}) :] \rangle \\ &= \langle s_\alpha : z_a :, \# : f_2:^\dagger \# s_{\alpha'} : z_{a'} : \rangle = \langle \mathcal{L}_{\alpha a}, \# : f_2:^\dagger \# \mathcal{L}_{\alpha' a'} : \rangle. \end{aligned} \quad (27.10.8)$$

Here we have used (7.3.15). And, in view of (7.3.15) and (10.8), we have found the beautiful result

$$\# : f_2 : \#^\dagger = \# : f_2 :^\dagger \#. \quad (27.10.9)$$

In mimicry of (7.3.31), define the analogous operator  $\hat{\mathcal{M}}$ , which acts on vector fields, by the rule

$$\hat{\mathcal{M}} = \exp(\# : f_2^c : \#) \exp(i \# : f_2^a : \#). \quad (27.10.10)$$

As a consequence of (8.1.11), operators of the form (10.10) give a realization of the group  $USp(2n)$  acting on the space of vector fields. Moreover, in view of (7.3.26), (7.3.30), and (10.9), we have the result

$$\hat{\mathcal{M}}^\dagger = \hat{\mathcal{M}}^{-1}. \quad (27.10.11)$$

It follows that the vector-field scalar product defined by (10.3) is also  $USp(2n)$  invariant. Finally, as a special case, we see that the operators  $\hat{\mathcal{A}}$  defined by (6.24) and (9.37) are symplectic and unitary.

From Section 9 we know that the general homogeneous polynomial vector field (in 6 dimensions) has the decomposition (9.2). Since each term in the decomposition has different  $sp(6)$  [and therefore  $usp(6)$ ] transformation properties, we might expect that the different terms in the decomposition would be mutually orthogonal. This is indeed the case. That is, for the scalar product (10.3) and the decomposition (9.2), there are the relations

$$\begin{aligned}\langle \mathcal{H}^{\ell+1,0,0}, \mathcal{G}^{\ell-1,1,0} \rangle &= \langle \mathcal{H}^{\ell+1,0,0}, \mathcal{G}^{\ell-1,0,0} \rangle \\ &= \langle \mathcal{G}^{\ell-1,1,0}, \mathcal{G}^{\ell-1,0,0} \rangle = 0.\end{aligned}\quad (27.10.12)$$

Note that all the vector fields in (10.12) have the same degree of homogeneity. If the degrees are different, the vector fields are automatically orthogonal by (10.5).

The relations (10.12) will be proved in Subsection 11.2 by group-theoretic methods. Here we will begin to describe a complementary result. Section 9 showed that the decomposition (9.2) exists. However, given a specific vector field  $\mathcal{L}_{\mathbf{g}^\ell}$ , the only method proposed for finding  $\mathcal{H}^{\ell+1,0,0}$ ,  $\mathcal{G}^{\ell-1,1,0}$ , and  $\mathcal{G}^{\ell-1,0,0}$  was to construct in detail the bases for these spaces and then match coefficients. Fortunately, there is a more direct approach that accomplishes major aspects of this task. Part of this approach is described below, and the remainder will be described in Subsection 11.2.

Here we will show that there are linear projection operators  $\mathcal{P}^H$  and  $\mathcal{P}^G$  that can be described explicitly and that act on vector fields  $\mathcal{L}_{\mathbf{g}^\ell}$  to yield the results

$$\mathcal{P}^H \mathcal{L}_{\mathbf{g}^\ell} = \mathcal{H}^{\ell+1,0,0}, \quad (27.10.13)$$

$$\mathcal{P}^G \mathcal{L}_{\mathbf{g}^\ell} = \mathcal{G}^{\ell-1,1,0} + \mathcal{G}^{\ell-1,0,0} = \mathcal{G}_{\ell+1}. \quad (27.10.14)$$

They also have the properties

$$(\mathcal{P}^H)^2 = \mathcal{P}^H, \quad (\mathcal{P}^G)^2 = \mathcal{P}^G, \quad (27.10.15)$$

$$\mathcal{P}^H \mathcal{P}^G = \mathcal{P}^G \mathcal{P}^H = 0, \quad (27.10.16)$$

$$\mathcal{P}^H + \mathcal{P}^G = \mathcal{I}. \quad (27.10.17)$$

Here  $\mathcal{I}$  denotes the identity operator. Finally, the  $\mathcal{H}^{\ell+1,0,0}$  and  $\mathcal{G}_{\ell+1}$  defined by (10.13) and (10.14) satisfy

$$\langle \mathcal{H}^{\ell+1,0,0}, \mathcal{G}_{\ell+1} \rangle = 0. \quad (27.10.18)$$

We will also show directly that

$$\langle \mathcal{H}^{\ell+1,0,0}, \mathcal{G}^{\ell-1,1,0} \rangle = \langle \mathcal{H}^{\ell+1,0,0}, \mathcal{G}^{\ell-1,0,0} \rangle = 0. \quad (27.10.19)$$

We remark that the relations (10.13) and (10.14) are sufficient to carry out the decomposition required for factorizing general maps as will be done in Section 26.1.

We will first define the projection operators, and then show that they possess the advertised properties. Suppose we are given  $\mathcal{L}_{\mathbf{g}^\ell}$  and hence  $\mathbf{g}^\ell(z)$ . Then, in the spirit of (7.6.24), we *define* the homogeneous polynomial  $h_{\ell+1}$  by the rule

$$h_{\ell+1} = -[1/(\ell+1)] \sum_{ab} g_a^\ell(z) J_{ab} z_b. \quad (27.10.20)$$

Note that  $h_{\ell+1}$  depends *linearly* on the  $g_a^\ell$ . We now define  $\mathcal{P}^H$  by the rule

$$\mathcal{P}^H \mathcal{L}_{\mathbf{G}^\ell} = \mathcal{H}^{\ell+1,0,0} =: h_{\ell+1} : . \quad (27.10.21)$$

Let us compute the action of  $:h_{\ell+1}:$  on  $z_c$ . We find the intermediate result

$$\begin{aligned} :h_{\ell+1}: z_c &= [h_{\ell+1}, z_c] = -[1/(\ell+1)] \sum_{ab} J_{ab} [g_a^\ell z_b, z_c] \\ &= -[1/(\ell+1)] \sum_{ab} J_{ab} (g_a^\ell [z_b, z_c] + z_b [g_a^\ell, z_c]) \\ &= -[1/(\ell+1)] \{ \sum_{ab} g_a^\ell J_{ab} J_{bc} + \sum_{ab} J_{ab} z_b [g_a^\ell, z_c] \} \\ &= [1/(\ell+1)] \{ g_c^\ell + \sum_{ab} J_{ab} z_b [z_c, g_a^\ell] \}. \end{aligned} \quad (27.10.22)$$

Here we have used (1.7.10) and (3.1.3). Next write the tautology

$$[z_c, g_a^\ell] = [z_a, g_c^\ell] - A_{ac} \quad (27.10.23)$$

where

$$A_{ac} = [z_a, g_c^\ell] - [z_c, g_a^\ell]. \quad (27.10.24)$$

Note that  $A_{ac}$  is antisymmetric under the interchange of indices. Insertion of (10.23) into (10.22) gives the further result

$$:h_{\ell+1}: z_c = [1/(\ell+1)] \{ g_c^\ell + \sum_{ab} J_{ab} z_b [z_a, g_c^\ell] - \sum_{ab} J_{ab} z_b A_{ac} \}. \quad (27.10.25)$$

The center term on the right of (10.25) can be evaluated,

$$\begin{aligned} \sum_{ab} J_{ab} z_b [z_a, g_c^\ell] &= \sum_{ab} J_{ab} z_b :z_a :g_c^\ell = - \sum_{ab} J_{ba} z_b :z_a :g_c^\ell \\ &= \Sigma g_c^\ell = \ell g_c^\ell. \end{aligned} \quad (27.10.26)$$

Here (7.6.50), (3.1), and (3.26) have been used. Therefore (10.25) can be rewritten in the form

$$g_c^\ell =: h_{\ell+1} : z_c + [1/(\ell+1)] \sum_{ab} J_{ab} z_b A_{ac}. \quad (27.10.27)$$

The second term on the right can be manipulated further. Use the antisymmetry of  $J$  and the fact that  $a, b$  are dummy summation indices to write

$$[1/(\ell+1)] \sum_{ab} J_{ab} z_b A_{ac} = -[1/(\ell+1)] \sum_{ab} z_b J_{ba} A_{ac} = -[1/(\ell+1)] \sum_{ab} z_a J_{ab} A_{bc}. \quad (27.10.28)$$

As a result of this manipulation (10.27) can be rewritten in the form

$$g_c =: h_{\ell+1} : z_c - [1/(\ell+1)] \sum_{ab} z_a J_{ab} A_{bc}. \quad (27.10.29)$$

Correspondingly,  $\mathcal{L}_{\mathbf{g}^\ell}$  can be written in the form

$$\mathcal{L}_{\mathbf{g}^\ell} = : h_{\ell+1} : + \mathcal{L}_G \mathbf{g}^\ell \quad (27.10.30)$$

where

$${}^G g_c^\ell = -[1/(\ell+1)] \sum_{ab} z_a J_{ab} A_{bc}. \quad (27.10.31)$$

Note that  ${}^G \mathbf{g}^\ell$  is linear in  $\mathbf{g}^\ell$  since  $A$  is linear in  $\mathbf{g}^\ell$ . We now define the projection operator  $\mathcal{P}^G$  by the rule

$$\mathcal{P}^G \mathcal{L}_{\mathbf{g}^\ell} = \mathcal{G}^{\ell-1,1,0} + \mathcal{G}^{\ell-1,0,0} = \mathcal{G}_{\ell+1} = \mathcal{L}_G \mathbf{g}^\ell, \quad (27.10.32)$$

which is simply a rewriting of the relation

$$\mathcal{G}_{\ell+1} = \mathcal{L}_{\mathbf{g}^\ell} - : h_{\ell+1} :. \quad (27.10.33)$$

With the projection operator definitions (10.21) and (10.32), the relation (10.30) shows that (10.17) holds by construction.

It remains to be shown that the projection operators have the advertised properties. Suppose the  $\mathbf{g}^\ell$  corresponding to  $: h_{\ell+1} :$  is used in (10.31) to compute  ${}^G \mathbf{g}^\ell$ . According to (7.6.7) the matrix  $A$  given by (10.24) vanishes in this case. Consequently,  ${}^G \mathbf{g}^\ell$  is zero. This observation verifies the second assertion in (10.16). Conversely, suppose  ${}^G \mathbf{g}^\ell$  is used in (10.20) to compute  $h_{\ell+1}$ . We first observe that (10.31) can be rewritten in the form

$${}^G g_a^\ell = -[1/(\ell+1)] \sum_{cd} z_c J_{cd} A_{da}. \quad (27.10.34)$$

Consequently, we have the result

$$\begin{aligned} h_{\ell+1} &= -[1/(\ell+1)] \sum_{ab} {}^G g_a^\ell J_{ab} z_b \\ &= [1/(\ell+1)^2] \sum_{abcd} z_c J_{cd} A_{da} J_{ab} z_b \\ &= [1/(\ell+1)^2] \sum_{bc} z_c (JAJ)_{cb} z_b. \end{aligned} \quad (27.10.35)$$

However, since both  $J$  and  $A$  are antisymmetric, it follows that

$$(JAJ)^T = J^T (A)^T J^T = -JAJ. \quad (27.10.36)$$

Therefore the right side of (10.35) vanishes by antisymmetry and we find

$$h_{\ell+1} = 0. \quad (27.10.37)$$

This observation verifies the first assertion in (10.16). Finally, with the aid of (10.16) and (10.17), we find that

$$(\mathcal{P}^H)^2 = \mathcal{P}^H (\mathcal{I} - \mathcal{P}^G) = \mathcal{P}^H - \mathcal{P}^H \mathcal{P}^G = \mathcal{P}^H. \quad (27.10.38)$$

This calculation and its counterpart for  $\mathcal{P}^G$  verify (10.15).

To verify (10.18) we first calculate that

$$\begin{aligned} :h_{\ell+1}:z_c &= [h_{\ell+1}, z_c] = \sum_{de} (\partial h_{\ell+1}/\partial z_d) J_{de} (\partial z_c/\partial z_e) \\ &= \sum_{de} (\partial h_{\ell+1}/\partial z_d) J_{de} \delta_{ce} = - \sum_d J_{cd} (\partial h_{\ell+1}/\partial z_d). \end{aligned} \quad (27.10.39)$$

Consequently, we find using (10.4), (10.34), and (10.39) the intermediate result

$$\begin{aligned} \langle \mathcal{G}_{\ell+1}, \mathcal{H}^{\ell+1,0,0} \rangle &= \langle \mathcal{L}_G \mathbf{g}^\ell, :h_{\ell+1}: \rangle \\ &= [1/(\ell+1)] \sum_{abcd} J_{ab} J_{cd} \langle z_a A_{bc}, (\partial h_{\ell+1}/\partial z_d) \rangle \\ &= [1/(\ell+1)] \sum_{abcd} J_{ab} J_{cd} \langle A_{bc}, (\partial^2 h_{\ell+1}/\partial z_a \partial z_d) \rangle. \end{aligned} \quad (27.10.40)$$

Here, in the last line, (7.3.14) has also been used. Define tensors  $T^1$  and  $T^2$  by the rules

$$T^1_{abcd} = J_{ab} J_{cd}, \quad (27.10.41)$$

$$T^2_{abcd} = \langle A_{bc}, (\partial^2 h_{\ell+1}/\partial z_a \partial z_d) \rangle. \quad (27.10.42)$$

From the antisymmetry of  $J$  the tensor  $T^1$  has the symmetry property

$$T^1_{dcba} = J_{dc} J_{ba} = J_{ab} J_{cd} = T^1_{abcd}. \quad (27.10.43)$$

From the antisymmetry of  $A$  and the symmetry of  $(\partial^2 h_{\ell+1}/\partial z_a \partial z_d)$  the tensor  $T^2$  has the symmetry property

$$\begin{aligned} T^2_{dcba} &= \langle A_{cb}, (\partial^2 h_{\ell+1}/\partial z_d \partial z_a) \rangle \\ &= -\langle A_{bc}, (\partial^2 h_{\ell+1}/\partial z_a \partial z_d) \rangle = -T^2_{abcd}. \end{aligned} \quad (27.10.44)$$

It follows that

$$\sum_{abcd} J_{ab} J_{cd} \langle A_{bc}, (\partial^2 h_{\ell+1}/\partial z_a \partial z_d) \rangle = \sum_{abcd} T^1_{abcd} T^2_{abcd} = - \sum_{abcd} T^1_{dcba} T^2_{dcba} = 0, \quad (27.10.45)$$

and consequently

$$\langle \mathcal{G}_{\ell+1}, \mathcal{H}^{\ell+1,0,0} \rangle = 0. \quad (27.10.46)$$

To verify (10.19) suppose that  $\mathcal{L}_G \mathbf{g}^\ell$  is the vector field  $\mathcal{G}^{\ell-1,0,0}$  given in (9.3). Then we have the relation

$$g_c^\ell = f_{\ell-1} z_c. \quad (27.10.47)$$

Consequently, from (10.4), (10.39), and (10.47), we find the result

$$\begin{aligned} \langle \mathcal{G}^{\ell-1,0,0}, \mathcal{H}^{\ell+1,0,0} \rangle &= \langle \mathcal{L}_G \mathbf{g}^\ell, :h_{\ell+1}: \rangle = -[1/(\ell+1)] \sum_{cd} J_{cd} \langle z_c f_{\ell-1}, (\partial h_{\ell+1}/\partial z_d) \rangle \\ &= -[1/(\ell+1)] \sum_{cd} J_{cd} \langle f_{\ell+1}, (\partial^2 h_{\ell+1}/\partial z_c \partial z_d) \rangle = 0. \end{aligned} \quad (27.10.48)$$

Here we have again employed (7.3.14), and used the antisymmetry of  $J$  and the symmetry of  $(\partial^2 h_{\ell+1} / \partial z_c \partial z_d)$  to infer that the sum in (10.47) vanishes. Finally, in view of (10.46), (10.48), and the definition of  $\mathcal{G}_{\ell+1}$  as given in the second part of (10.14) or, equivalently, in (10.33), we conclude that both statements in (10.19) are correct.

Let us evaluate the scalar products between vector fields associated with linear transformations. Let  $F$  be any  $2n \times 2n$  matrix, possibly complex, and use it to define a vector field  $\mathcal{L}_{\mathbf{f}^1}$  by the rule

$$\mathcal{L}_{\mathbf{f}^1} = \sum_{ab} (JF)_{ab} z_b (\partial/\partial z_a). \quad (27.10.49)$$

Then, comparison of (5.3.17) and (10.49) gives the relation

$$f_a^1 = \sum_b (JF)_{ab} z_b. \quad (27.10.50)$$

From (10.49) we also find the result

$$\mathcal{L}_{\mathbf{f}^1} z_c = \sum_d (JF)_{cd} z_d, \quad (27.10.51)$$

which is analogous to (6.26) and (6.27).

Let  $G$  be a second  $2n \times 2n$  matrix, and use it to define the vector field  $\mathcal{L}_{\mathbf{g}^1}$ . Now compute the scalar product between  $\mathcal{L}_{\mathbf{f}^1}$  and  $\mathcal{L}_{\mathbf{g}^1}$ . Doing so gives the result

$$\begin{aligned} \langle \mathcal{L}_{\mathbf{f}^1}, \mathcal{L}_{\mathbf{g}^1} \rangle &= \sum_a \langle f_a^1, g_a^1 \rangle \\ &= \sum_{abc} \langle (JF)_{ab} z_b, (JG)_{ac} z_c \rangle \\ &= \sum_{abc} [(JF)_{ab}]^* (JG)_{ac} \langle z_b, z_c \rangle \\ &= \sum_{abc} [(JF)^\dagger]_{ba} (JG)_{ac} \delta_{bc} = \text{tr} [(JF)^\dagger JG] \\ &= \text{tr} [F^\dagger J^\dagger JG] = \text{tr} (F^\dagger G). \end{aligned} \quad (27.10.52)$$

Here a “\*” denotes complex conjugation, and use has been made of (3.1.6). Note that this scalar product is the same as that in (4.4.16).

Suppose  $S$  is a real symmetric matrix. Use it to define a quadratic polynomial  $h_2$  as in (7.2.3),

$$h_2 = -(1/2) \sum_{de} S_{de} z_d z_e. \quad (27.10.53)$$

Then, from (7.2.4), there is an associated vector field  $\mathcal{L}_{\mathbf{s}^1}$  given by the relation

$$\mathcal{L}_{\mathbf{s}^1} =: h_2 := \sum_{ab} (JS)_{ab} z_b (\partial/\partial z_a) = \sum_a s_a^1 (\partial/\partial z_a), \quad (27.10.54)$$

with

$$s_a^1 = \sum_b (JS)_{ab} z_b. \quad (27.10.55)$$

Define the Hamiltonian vector field  $\mathcal{H}^2$  by writing

$$\mathcal{H}^2 =: h_2 :, \quad (27.10.56)$$

and let  $\mathcal{G}_2$  be any non-Hamiltonian vector field of the form (10.49) with  $F$  being any real antisymmetric matrix  $A$ . Compare with (6.26) and (6.27). Then, using (10.52), we find the result

$$\langle \mathcal{G}_2, \mathcal{H}^2 \rangle = \text{tr } (A^T S) = -\text{tr } (AS) = 0. \quad (27.10.57)$$

Here we have used the easily proved fact that the trace of the product of an antisymmetric and a symmetric matrix is always zero. See (4.4.86). We observe that (10.57) is a special case of the general result (10.46).

We close this section with a further study of first-degree vector fields in  $2n$  variables. Consider again the relation (10.49) and decompose  $F$ , which we now assume to be real, into symmetric and antisymmetric parts by writing

$$F = S^F + A^F. \quad (27.10.58)$$

That is, we write

$$\mathcal{L}_{\mathbf{f}} = \sum_{ab} (JF)_{ab} z_b (\partial/\partial z_a) = \sum_{ab} [J(S^F + A^F)]_{ab} z_b (\partial/\partial z_a) = \mathcal{L}_{\mathbf{f}^S} + \mathcal{L}_{\mathbf{f}^A} \quad (27.10.59)$$

with

$$\mathcal{L}_{\mathbf{f}^S} = \sum_{ab} (JS^F)_{ab} z_b (\partial/\partial z_a) \quad (27.10.60)$$

and

$$\mathcal{L}_{\mathbf{f}^A} = \sum_{ab} (JA^F)_{ab} z_b (\partial/\partial z_a). \quad (27.10.61)$$

We will now verify directly, as expected, that  $\mathcal{L}_{\mathbf{f}^S}$  is a Hamiltonian vector field and  $\mathcal{L}_{\mathbf{f}^A}$  is a non-Hamiltonian vector field. In particular, we will show that there are the relations

$$\mathcal{P}^H \mathcal{L}_{\mathbf{f}^S} = \mathcal{L}_{\mathbf{f}^S}, \quad (27.10.62)$$

$$\mathcal{P}^G \mathcal{L}_{\mathbf{f}^A} = \mathcal{L}_{\mathbf{f}^A}. \quad (27.10.63)$$

## Exercises

**27.10.1.** Show that

$$: f : = \sum_a (\partial f / \partial z_a) : z_a : . \quad (27.10.64)$$

**27.10.2.** Verify (10.4).

**27.10.3.** Verify (10.6). Hint: You may use brute force or (10.49), (7.3.14), and (7.6.50).

**27.10.4.** Let  $\mathcal{L}_{f_{\ell-1}}$  be the vector field corresponding to (9.3), and let  $\mathcal{L}_{f'_{\ell-1}}$  be a second such field. Show that

$$\langle \mathcal{L}_{f_{\ell-1}}, \mathcal{L}_{f'_{\ell-1}} \rangle = (\ell - 1 + 2n) \langle f_{\ell-1}, f'_{\ell-1} \rangle \quad (27.10.65)$$

where  $(2n)$  is the phase-space dimension.

**27.10.5.** Let  $h_{\ell+1}$  be any homogeneous polynomial of degree  $(\ell + 1)$ . Find the  $\mathbf{g}^\ell$  in the  $\mathcal{L}_{\mathbf{g}^\ell}$  that equals :  $h_{\ell+1} ::$ . Insert this  $\mathbf{g}^\ell$  in (10.20) and verify that the  $h_{\ell+1}$  so produced agrees with the original  $h_{\ell+1}$ . You have again verified the first result in (10.15).

**27.10.6.** Consider vector fields that are in  $\mathcal{G}^{\ell-1,0,0}$  and therefore can be written in the form (10.47). Insert these  $\mathbf{g}^\ell$  into (10.20) and show that the  $h_{\ell+1}$  they produce vanish.

**27.10.7.** Verify that the vector fields  $\mathcal{H}^{2,0,0}$ ,  $\mathcal{G}^{0,1,0}$ , and  $\mathcal{G}^{0,0,0}$  found explicitly in Section 9 satisfy (10.12).

**27.10.8.** Let  $z_1, z_2, \dots, z_m$  be  $m$  variables. Consider the  $m^2$  vector fields  $\mathcal{L}_{ab}$  defined by the rule

$$\mathcal{L}_{ab} = z_a (\partial/\partial z_b). \quad (27.10.66)$$

Show that these vector fields obey the commutation rules

$$\{\mathcal{L}_{ab}, \mathcal{L}_{cd}\} = \delta_{bc} \mathcal{L}_{ad} - \delta_{ad} \mathcal{L}_{cb}. \quad (27.10.67)$$

Let  $A$  be an  $m \times m$  matrix. Associate with each such matrix the vector field  $\mathcal{L}^A$  defined by the rule

$$\mathcal{L}^A = \sum_{ab} A_{ab} \mathcal{L}_{ab}. \quad (27.10.68)$$

Show that there is the relation

$$\{\mathcal{L}^A, \mathcal{L}^B\} = \mathcal{L}^C \quad (27.10.69)$$

where

$$C = \{A, B\}. \quad (27.10.70)$$

That is, verify that the  $\mathcal{L}_{ab}$  yield a basis for the general linear group Lie algebra  $gl(m)$ .

Define a vector field  $\Sigma$  by the relation

$$\Sigma = \mathcal{L}^I = \sum_a \mathcal{L}_{aa} \quad (27.10.71)$$

where  $I$  is the identity matrix. Verify that there is the relation

$$\{\mathcal{L}_{ab}, \Sigma\} = 0. \quad (27.10.72)$$

You have shown that  $\Sigma$  is invariant under  $gl(m)$ .

**27.10.9.** Review Exercise 10.8 above. Show that the vector fields spanned by the elements

$$\mathcal{L}_{ab} = z_a (\partial/\partial z_b) - z_b (\partial/\partial z_a) \quad (27.10.73)$$

yield a basis for the Lie algebra  $so(m)$ . For the cases  $m = 2n = 2$ ,  $m = 2n = 4$ , and  $m = 2n = 6$  decompose these elements into Hamiltonian and non-Hamiltonian parts. Show that in each case the Hamiltonian parts span a Lie algebra, and identify these Lie algebras. What can be said about the non-Hamiltonian parts?

**27.10.10.** Use the machinery of this section to find the decompositions (3.51) through (3.56).

**27.10.11.** Consider the matrices  $\tilde{A}$  given by (6.28) through (6.33). Relate them to the matrices  $C^0$  through  $C^3$  and  $E^1$  and  $E^2$  given by (4.3.137) through (4.3.140) and (4.3.145) and (4.3.146). Also relate them to the matrices  $A^1$  through  $A^6$  given by (8.2.87) through (8.2.92). Verify directly that the trace of the product of any two different  $\tilde{A}$  matrices vanishes. Relate this fact to the assertion (10.12) and the relation (10.52). Show that all matrices of the form  $J\tilde{A}(0, 1; *, *)$  are traceless, and therefore all matrices of the form  $\exp[J\tilde{A}(0, 1; *, *)]$  have determinant +1.

## 27.11 Products and Casimir Operators

In this section we will explore the properties of products of entities when each entity taken by itself has well-defined properties under the action of the symplectic group. For example, if  $f_\ell$ ,  $g_m$ ,  $h_n$ ,  $\dots$  are homogeneous polynomials, we could ask about the transformation properties of the product  $[(f_\ell)(g_m)(h_n)\dots]$ . Or, we could ask about the transformation properties of the product of Lie operators  $[: f_\ell :: g_m :: h_n : \dots]$ . As a third example, we could ask about the properties of the product of adjoint operators  $[\#f_\ell\#\#g_m\#\#h_n\#\dots]$ . The first case, the transformation properties of the product  $[(f_\ell)(g_m)(h_n)\dots]$ , is simple because the polynomials  $f_\ell$ ,  $g_m$ ,  $h_n$ ,  $\dots$  can be multiplied together to yield some net polynomial, and the transformation properties of this polynomial are already known. The remaining two cases require more work.

### 27.11.1 The Quadratic Casimir Operator

We will find that a question of particular interest, and also the simplest, is to determine the transformation properties of the two-element products  $[: f_2 :: g_2 :]$  and  $[\#f_2\#\#g_2\#]$ . In the case of  $sp(2)$ , we know that  $: f_2 :$  and  $: g_2 :$  (and  $\#f_2\#$  and  $\#g_2\#$ ) each carry the representation  $\Gamma(2)$ , and therefore the product carries the representation  $\Gamma(2) \otimes \Gamma(2)$ . Also, there is the Clebsch-Gordan series result

$$\Gamma(2) \otimes \Gamma(2) = \Gamma(0) \oplus \Gamma(2) \oplus \Gamma(4). \quad (27.11.1)$$

(This is just the familiar statement for  $su(2)$  or  $sp(2)$  that spin 1 and spin 1 combine to make spin 0, 1, and 2.)

In the case of  $sp(4)$ , the corresponding representation for each factor is  $\Gamma(2, 0)$ ; and the corresponding Clebsch-Gordan series result is known from group theory to be

$$\Gamma(2, 0) \otimes \Gamma(2, 0) = \Gamma(0, 0) \oplus \Gamma(0, 1) \oplus \Gamma(0, 2) \oplus \Gamma(2, 0) \oplus \Gamma(2, 1) \oplus \Gamma(4, 0). \quad (27.11.2)$$

Finally, in the case of  $sp(6)$ , the representation for each factor is  $\Gamma(2, 0, 0)$ ; and the corresponding Clebsch-Gordan series result is

$$\begin{aligned} \Gamma(2, 0, 0) \otimes \Gamma(2, 0, 0) &= \\ \Gamma(0, 0, 0) \oplus \Gamma(0, 1, 0) \oplus \Gamma(0, 2, 0) \oplus \Gamma(2, 0, 0) \oplus \Gamma(2, 1, 0) \oplus \Gamma(4, 0, 0). \end{aligned} \quad (27.11.3)$$

Observe from (11.1) through (11.3) that in each case the *identity* representation [the representations  $\Gamma(0)$ ,  $\Gamma(0, 0)$ , and  $\Gamma(0, 0, 0)$ ] occurs once and only once. (Strictly speaking, we can only conclude that there is the potential for the identity representation to occur. The sought after quantity may in fact vanish. See Exercise 11.7.) Consequently, there must be some combination of quantities of the form  $[: f_2 :: g_2 :]$ , or of the form  $[\#f_2\#\#g_2\#]$ , that is *invariant* (commutes with all generators) under the action of the symplectic group. Moreover, this combination is unique up to an overall multiplicative constant. This combination is called the *Casimir* operator for the symplectic group or symplectic Lie algebra. More particularly, it is called the *quadratic* Casimir operator since it is composed of two factors.

Now that we know that a quadratic Casimir operator exists (and is unique), the problem is to find it explicitly. In effect, what we must do is find the Clebsch-Gordan *coefficients* that produce the identity representations in the series (11.1) through (11.3). We will work up to this task by stages.

Suppose  $L$  is a Lie algebra with basis elements  $B_1, B_2, \dots$ . Then, as in Section 3.7, the basis elements satisfy Lie product rules of the form

$$[B_\alpha, B_\beta] = \sum_\gamma c_{\alpha\beta}^\gamma B_\gamma. \quad (27.11.4)$$

Here  $[,]$  denotes the Lie product (however realized) and the quantities  $c_{\alpha\beta}^\gamma$  are the structure constants that specify  $L$ .

Next, suppose  $R$  is a *realization* of  $L$  in terms of  $m \times m$  matrices.<sup>2</sup> Then, for each basis element  $B_\alpha$ , there will be an associated matrix  $\hat{B}_\alpha$ , and these matrices will obey the commutation rules

$$\{\hat{B}_\alpha, \hat{B}_\beta\} = \sum_\gamma c_{\alpha\beta}^\gamma \hat{B}_\gamma \quad (27.11.5)$$

with the same structure constants as in (11.4). See Section 3.7.

Since a Lie algebra is a vector space, it is natural to consider the possibility of introducing some kind of scalar product among the elements of  $L$ . Suppose  $B$  and  $B'$  are any two elements in  $L$ , and let  $(B, B')$  denote their scalar product. Then, by linearity, there is the result

$$(B, B') = \sum_{\alpha\alpha'} b^\alpha (b')^{\alpha'} (B_\alpha, B_{\alpha'}) \quad (27.11.6)$$

where the  $b^\alpha$  and  $(b')^{\alpha'}$  are the components of  $B$  and  $B'$ ,

$$B = \sum_\alpha b^\alpha B_\alpha, \quad (27.11.7)$$

$$B' = \sum_{\alpha'} (b')^{\alpha'} B_{\alpha'}. \quad (27.11.8)$$

[Note that we have taken the scalar product to be *linear* (no complex conjugation), in both the components  $b^\alpha$  and  $(b')^{\alpha'}$  rather than antilinear (complex conjugation) in one and linear

---

<sup>2</sup>In this context we use the term *realization* rather than *representation* because in this chapter we wish, for the most part, to use the term *representation* only in the specific/technical sense of referring to some  $\Gamma(\dots)$ .

in the other as in (7.3.12).] The relation (11.6) can be rewritten in the form

$$(B, B') = \sum_{\alpha\alpha'} b^\alpha (b')^{\alpha'} g_{\alpha\alpha'} \quad (27.11.9)$$

where  $g_{\alpha\alpha'}$  is defined by writing

$$g_{\alpha\alpha'} = (B_\alpha, B_{\alpha'}). \quad (27.11.10)$$

In view of (11.9), the quantities  $g_{\alpha\alpha'}$  may be regarded as the entries in some kind of *metric tensor*, and the scalar product between any two elements in  $L$  is specified once the entries  $g_{\alpha\alpha'}$  are specified.

In principle, the entries  $g_{\alpha\alpha'}$  may be defined at will. However, it is advantageous to define  $g_{\alpha\alpha'}$  in a way that involves some properties of the Lie algebra  $L$  and has certain desired features. A way to do this is to define  $g_{\alpha\alpha'}$  with the aid of the realization  $R$  by writing

$$(B_\alpha, B_{\alpha'})_R = g_{\alpha\alpha'}^R = \text{tr} (\hat{B}_\alpha \hat{B}_{\alpha'}). \quad (27.11.11)$$

Here we have written the sub and superscript  $R$  to indicate that the realization  $R$  has been used. See Section 4.4, equation (4.4.39), for an analogous construction.

In the case of  $sp(2)$ , for example, suppose we use as a basis the matrices  $B^0$ ,  $F$ , and  $G$  associated with the quadratic phase-space polynomials  $b^0$ ,  $f$ , and  $g$  as described in Section 5.6. Then we find the result

$$g^F = \begin{array}{cccc} & b^0 & f & g \\ \begin{matrix} b^0 \\ f \\ g \end{matrix} & \begin{matrix} b^0 & -2 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{matrix} \end{array}. \quad (27.11.12)$$

Here we have employed the notation  $g^F$  to indicate that for the realization  $R$  we have used the  $2 \times 2$  *fundamental* or defining representation. Also, we have labeled the entries in  $g^F$  by the associated quadratic phase-space polynomials. Suppose instead of the basis  $B^0$ ,  $F$ , and  $G$  we use the Cartan basis of Section 21.1. The  $2 \times 2$  matrices associated with these basis elements are easily found. See Exercise 11.1. Using these matrices gives the result

$$g^F = \begin{array}{cccc} & c^1 & r(+\alpha) & r(-\alpha) \\ \begin{matrix} c^1 \\ r(+\alpha) \\ r(-\alpha) \end{matrix} & \begin{matrix} c^1 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 2 & 0 & 0 \end{matrix} \end{array}. \quad (27.11.13)$$

where we have now labeled the entries by the polynomials associated with the Cartan basis elements.

We observe that  $g^F$  is symmetric as is desired for a metric tensor. This symmetry property is true in general for any realization  $R$ ,

$$g_{\alpha\alpha'}^R = g_{\alpha'\alpha}^R, \quad (27.11.14)$$

because the trace operation has the permutation symmetry property (3.6.124). Indeed, by linearity and symmetry, we have the results

$$(B, B')_R = \text{tr}(\hat{B}\hat{B}') = \text{tr}(\hat{B}'\hat{B}) = (B', B)_R.$$

Analogous calculations can be carried out for the cases of  $sp(4)$  and  $sp(6)$  using the fundamental matrix representations of Sections 5.7 and 5.8. Here it is convenient to introduce additional notation. According to Section 5.5, associated with any quadratic polynomial  $f$  there is an associated Hamiltonian matrix  $JS^f$ . For any two quadratic polynomials  $f$  and  $g$ , let us make the definitions

$$\langle f, g \rangle_F = (f, g)_F = (JS^f, JS^g)_F = \text{tr}(JS^f JS^g). \quad (27.11.15)$$

Then, if we use the Cartan bases of Sections 21.4 and 21.7 and the fundamental representations we find, both for  $sp(4)$  and  $sp(6)$ , results analogous to (11.13) that can be written in the general form

$$\langle c^j, c^k \rangle_F = 2\delta_{jk}, \quad (27.11.16)$$

$$\langle c^j, r(\boldsymbol{\mu}) \rangle_F = 0, \quad (27.11.17)$$

$$\langle r(\boldsymbol{\mu}), r(\boldsymbol{\nu}) \rangle_F = 0, \text{ if } \boldsymbol{\mu} \neq -\boldsymbol{\nu}, \quad (27.11.18)$$

$$\langle r(\boldsymbol{\mu}), r(-\boldsymbol{\mu}) \rangle_F = 2. \quad (27.11.19)$$

Note that, as they stand, the relations (11.13) and (11.16) through (11.19) or, equivalently (11.11) with  $R = F$ , define a scalar product for the elements in the various  $sp(2n, \mathbb{R})$  Lie algebras. If we identify the Lie elements with their associated quadratic polynomials in the phase space variables  $z$  using relations of the form (5.5.1) and (5.5.3), then we have in effect also defined a scalar product  $\langle f, g \rangle_F$  among quadratic polynomials. However, unlike the scalar product of Section 7.3, this scalar product is only defined for *quadratic* polynomials. Exercises 11.1 and 11.3 examine the relation between these two scalar products.

The definition (11.11) has a further desirable property beyond symmetry that is less obvious. Let  $C$  be some element in  $L$ . Use it to *transform* any basis element  $B_\alpha$  into the element  $B_\alpha^{\text{tr}}$  by the rule

$$\begin{aligned} B_\alpha^{\text{tr}} &= \exp(\epsilon : C :) B_\alpha = B_\alpha + \epsilon : C : B_\alpha + (\epsilon^2/2!) : C :^2 B_\alpha + \dots \\ &= B_\alpha + \epsilon [C, B_\alpha] + (\epsilon^2/2!) [C, [C, B_\alpha]] + \dots \end{aligned} \quad (27.11.20)$$

Here  $: C :$  is a differential operator in the case that the Lie product is a Poisson bracket. Otherwise it is simply the *adjoint* operator defined by the property

$$: C : B_\alpha = [C, B_\alpha]. \quad (27.11.21)$$

See (3.7.31) and (5.3.2). The matrix analog of (11.20) in the realization  $R$  is the transformation

$$\begin{aligned} \hat{B}_\alpha^{\text{tr}} &= \exp(\epsilon \# \hat{C} \#) \hat{B}_\alpha = \hat{B}_\alpha + \epsilon \# \hat{C} \# \hat{B}_\alpha + (\epsilon^2/2!) \# C \#^2 B_\alpha + \dots \\ &= \hat{B}_\alpha + \epsilon \{ \hat{C}, \hat{B}_\alpha \} + (\epsilon^2/2!) \{ \hat{C}, \{ \hat{C}, \hat{B}_\alpha \} \} + \dots \end{aligned} \quad (27.11.22)$$

At this point we invoke the relation

$$\exp(\epsilon \# \hat{C} \#) \hat{B}_\alpha = \exp(\epsilon \hat{C}) \hat{B}_\alpha \exp(-\epsilon \hat{C}), \quad (27.11.23)$$

which is the matrix analog of (8.2.5) and derived in the same way. It follows that (11.22) can be rewritten in the form

$$\hat{B}_\alpha^{\text{tr}} = \exp(\epsilon \hat{C}) \hat{B}_\alpha \exp(-\epsilon \hat{C}). \quad (27.11.24)$$

Consequently, from (11.11) and (11.15), we have the result

$$\begin{aligned} (B_\alpha^{\text{tr}}, B_{\alpha'}^{\text{tr}})_R &= \text{tr}(\hat{B}_\alpha^{\text{tr}} \hat{B}_{\alpha'}^{\text{tr}}) \\ &= \text{tr}[\exp(\epsilon \hat{C}) \hat{B}_\alpha \exp(-\epsilon \hat{C}) \exp(\epsilon \hat{C}) \hat{B}_{\alpha'} \exp(-\epsilon \hat{C})] \\ &= \text{tr}[\exp(\epsilon \hat{C}) \hat{B}_\alpha \hat{B}_{\alpha'} \exp(-\epsilon \hat{C})] \\ &= \text{tr}[\exp(-\epsilon \hat{C}) \exp(\epsilon \hat{C}) \hat{B}_\alpha \hat{B}_{\alpha'}] = \text{tr}(\hat{B}_\alpha \hat{B}_{\alpha'}) \\ &= (B_\alpha, B_{\alpha'})_R. \end{aligned} \quad (27.11.25)$$

Here we have again used standard properties of the trace operation. See Exercise 3.6.7.<sup>3</sup> But we know that objects of the form  $[\exp(\epsilon : C :)]$  correspond to Lie group elements generated by the Lie algebra  $L$ . Therefore, (11.25) shows that the scalar product (11.11) has the remarkable property that it is *invariant* under the action of the group.

The relation (11.25) displays group invariance in *finite* (group) form. It is also instructive to view group invariance in *infinitesimal* (Lie-algebraic) form. This is easily done by retaining only the first two terms in (11.20) and equating powers of  $\epsilon$ . Doing so in (11.25) gives the result

$$([C, B_\alpha], B_{\alpha'})_R + (B_\alpha, [C, B_{\alpha'}])_R = 0. \quad (27.11.26)$$

Upon setting  $C = B_{\alpha''}$  and relabeling indices, (11.26) takes the beautifully symmetric, if less illuminating, forms

$$([B_\alpha, B_{\alpha'}], B_{\alpha''})_R = (B_\alpha, [B_{\alpha'}, B_{\alpha''}])_R, \quad (27.11.27)$$

$$(B_\alpha, [B_{\alpha'}, B_{\alpha''}])_R = (B_{\alpha'}, [B_{\alpha''}, B_\alpha])_R = (B_{\alpha''}, [B_\alpha, B_{\alpha'}])_R.$$

Have you ever encountered relations like (11.26) and (11.27) before? You have. See Exercise 11.8.

The invariance relation (11.25) has implications for the metric tensor  $g^R$ . Since the relation (11.20) involves only Lie products and sums, we know from (11.4) that there are transformation coefficients  $U_{\alpha\beta}$  such that (11.20) can be rewritten in the form

$$B_\alpha^{\text{tr}} = \sum_\beta U_{\alpha\beta} B_\beta. \quad (27.11.28)$$

Inserting this relation into (11.25) gives the results

$$\sum_{\beta\beta'} U_{\alpha\beta} U_{\alpha'\beta'} (B_\beta, B_{\beta'})_R = (B_\alpha, B_{\alpha'})_R, \quad (27.11.29)$$

---

<sup>3</sup>We are also in the uncomfortable position of using the symbols *tr* to stand both for *transformed* and *trace*. Some flexibility of mind is sometimes required.

or, with the aid of (11.10),

$$\sum_{\beta\beta'} U_{\alpha\beta} U_{\alpha'\beta'} g_{\beta\beta'}^R = g_{\alpha\alpha'}^R. \quad (27.11.30)$$

If we view the quantities  $U_{\alpha\beta}$  and  $g_{\alpha\alpha'}^R$  as entries in matrices, the relation (11.30) can be written in the compact form

$$Ug^R U^T = g^R. \quad (27.11.31)$$

Finally, we know that  $U$  is invertible. [Simply change the sign of  $\epsilon$  in (11.20).] Suppose that  $g^R$  is also invertible. We will soon see that it is for the symplectic Lie algebra. [Indeed,  $g^R$  can be shown to be invertible for all simple Lie algebras.] Then (11.31) can also be rewritten in the form

$$(U^T)^{-1}(g^R)^{-1}U^{-1} = (g^R)^{-1}, \quad (27.11.32)$$

or

$$U^T(g^R)^{-1}U = (g^R)^{-1}. \quad (27.11.33)$$

We are ready to construct the quadratic Casimir operator. Following the usual procedure, we define a metric tensor  $g_R^{\alpha\alpha'}$  with raised indices by the rule

$$g_R^{\alpha\alpha'} = [(g^R)^{-1}]_{\alpha\alpha'}. \quad (27.11.34)$$

Suppose  $\hat{B}_\alpha$  is any set of linear operators that obey (11.5), but do not necessarily belong to the realization  $R$  used to define  $g^R$ . They might, for example, be differential operators or matrices belonging to some other realization. We define the associated quadratic Casimir operator  $\mathcal{C}_2$  by the rule

$$\mathcal{C}_2 = \sum_{\alpha\alpha'} g_R^{\alpha\alpha'} \hat{B}_\alpha \hat{B}_{\alpha'}. \quad (27.11.35)$$

We must now show that  $\mathcal{C}_2$  has the desired properties. Suppose  $\exp(\epsilon \# \hat{C} \#)$  is applied to both sides of (11.35). Here  $\hat{C}$  is some linear combination of the  $\hat{B}_\alpha$ . Doing so, and making use of (11.22) and the isomorphism property (8.2.14), gives the result

$$\mathcal{C}_2^{\text{tr}} = \exp(\epsilon \# \hat{C} \#) \mathcal{C}_2 = \sum_{\alpha\alpha'} g_R^{\alpha\alpha'} \hat{B}_\alpha^{\text{tr}} \hat{B}_{\alpha'}^{\text{tr}}. \quad (27.11.36)$$

But we know that

$$\hat{B}_\alpha^{\text{tr}} = \sum_\beta U_{\alpha\beta} \hat{B}_\beta \quad (27.11.37)$$

since only the structure constants  $c_{\alpha\beta}^\gamma$  are involved in the computation of  $U$ . See (11.5) and (11.22). Therefore, (11.36) can be rewritten in the form

$$\mathcal{C}_2^{\text{tr}} = \sum_{\alpha\alpha'\beta\beta'} g_R^{\alpha\alpha'} U_{\alpha\beta} U_{\alpha'\beta'} \hat{B}_\beta \hat{B}_{\beta'}. \quad (27.11.38)$$

Also, when written in expanded form, (11.33) and (11.34) yield the relation

$$\sum_{\alpha\alpha'} g_R^{\alpha\alpha'} U_{\alpha\beta} U_{\alpha'\beta'} = g_R^{\beta\beta'}. \quad (27.11.39)$$

Consequently, we have result

$$\mathcal{C}_2^{\text{tr}} = \sum_{\beta\beta'} g_R^{\beta\beta'} \hat{B}_\beta \hat{B}_{\beta'} = \mathcal{C}_2. \quad (27.11.40)$$

That is,  $\mathcal{C}_2$  is invariant under group action.

As a special case of (11.40), set  $\hat{C} = \hat{B}_{\alpha''}$  and equate powers of  $\epsilon$  in (11.36) and (11.40). Doing so gives the infinitesimal result

$$\#\hat{B}_{\alpha''}\#\mathcal{C}_2 = \{\hat{B}_{\alpha''}, \mathcal{C}_2\} = 0. \quad (27.11.41)$$

That is, all Lie generators commute with  $\mathcal{C}_2$ . Put yet another way, the raised metric tensor entries  $g_R^{\alpha\alpha'}$  are the Clebsch-Gordan coefficients that couple together two copies of the representation associated with the  $B_\alpha$  (the adjoint representation) to form the identity representation.

There is still another way of looking at our result. Since the commutator is antisymmetric, the relation (11.41) also states that  $\mathcal{C}_2$  commutes with all Lie generators. And from this result, by the linearity and derivation properties of the commutator, we conclude that  $\mathcal{C}_2$  commutes with *all* products and sums of products of Lie generators.

## 27.11.2 Applications of the Quadratic Casimir Operator

Before continuing on to a discussion of higher-order Casimir operators, let us pause to make use of the quadratic Casimir operator for the symplectic group. For our discussion we will use the fundamental representation. Examination of (11.13) and (11.16) through (11.19) shows that in this case

$$(g^F)^2 = 4I \quad (27.11.42)$$

and hence

$$g_F^{\alpha\alpha'} = (1/4)g_{\alpha\alpha'}^F. \quad (27.11.43)$$

Therefore, in view of (11.16) through (11.19) and up to a normalization which we choose for convenience, the quadratic Casimir for the symplectic group is given by the relation

$$\mathcal{C}_2 = \sum_j (C^j)^2 + \sum_{\mu} R(-\mu)R(\mu). \quad (27.11.44)$$

Here the elements  $C^j$  and  $R(\mu)$  are some kind of linear operators or matrices that obey commutation rules analogous to (4.15) through (4.18).

We will soon apply  $\mathcal{C}_2$  to the highest weight state  $|\mathbf{w}^h\rangle$  in some representation  $\Gamma$ . Before doing so, it is useful to rewrite  $\mathcal{C}_2$  in a form that is convenient for this purpose. As was the case for weights (see Section 5.8), we define a root  $\mu$  to be *positive* if its first nonvanishing component is positive. For example, in the case of  $sp(4)$ , the roots  $\alpha, \beta, \gamma$ , and  $-\delta$  are positive. See Figure 4.1. Note that if  $\mu$  is positive, then  $-\mu$  is not positive. Conversely, if  $\mu$  is not positive, then  $-\mu$  is positive. Thus, half the root vectors are positive, and the other half (their negatives) are not. With this definition in mind, we may rewrite (11.44) in the form

$$\mathcal{C}_2 = \sum_j (C^j)^2 + \sum_{\mu>0} [R(\mu)R(-\mu) + R(-\mu)R(\mu)]. \quad (27.11.45)$$

Here the notation  $\boldsymbol{\mu} > 0$  indicates that  $\boldsymbol{\mu}$  is positive. Next write the simple identity

$$R(\boldsymbol{\mu})R(-\boldsymbol{\mu}) + R(-\boldsymbol{\mu})R(\boldsymbol{\mu}) = 2R(-\boldsymbol{\mu})R(\boldsymbol{\mu}) + \{R(\boldsymbol{\mu}), R(-\boldsymbol{\mu})\}. \quad (27.11.46)$$

But, by (4.17), we have the relation

$$\{R(\boldsymbol{\mu}), R(-\boldsymbol{\mu})\} = \sum_j (\mathbf{e}^j \cdot \boldsymbol{\mu}) C^j. \quad (27.11.47)$$

Consequently, (11.45) can be rewritten in the form

$$\mathcal{C}_2 = \sum_j (C^j)^2 + \sum_{\boldsymbol{\mu} > 0} \sum_j (\mathbf{e}^j \cdot \boldsymbol{\mu}) C^j + 2 \sum_{\boldsymbol{\mu} > 0} R(-\boldsymbol{\mu})R(\boldsymbol{\mu}). \quad (27.11.48)$$

There is one final simplification. Define  $\boldsymbol{\mu}^+$  to be the sum of all positive roots,

$$\boldsymbol{\mu}^+ = \sum_{\boldsymbol{\mu} > 0} \boldsymbol{\mu}. \quad (27.11.49)$$

With this definition,  $\mathcal{C}_2$  takes the form

$$\mathcal{C}_2 = \sum_j [(C^j)^2 + (\mathbf{e}^j \cdot \boldsymbol{\mu}^+) C^j] + 2 \sum_{\boldsymbol{\mu} > 0} R(-\boldsymbol{\mu})R(\boldsymbol{\mu}). \quad (27.11.50)$$

We are ready to apply  $\mathcal{C}_2$  to  $|\mathbf{w}^h\rangle$ . First observe that

$$R(\boldsymbol{\mu})|\mathbf{w}^h\rangle = 0 \text{ if } \boldsymbol{\mu} > 0. \quad (27.11.51)$$

Were this not so,  $|\mathbf{w}^h\rangle$  would not be an eigenvector of the  $C^j$  with highest weight. [See (5.8.16).] Now the virtue of writing  $\mathcal{C}_2$  in the form (11.50) is apparent. Also, we have the relations

$$\sum_j (C^j)^2 |\mathbf{w}^h\rangle = (\mathbf{w}^h \cdot \mathbf{w}^h) |\mathbf{w}^h\rangle, \quad (27.11.52)$$

$$\sum_j (\mathbf{e}^j \cdot \boldsymbol{\mu}^+) C^j |\mathbf{w}^h\rangle = (\boldsymbol{\mu}^+ \cdot \mathbf{w}^h) |\mathbf{w}^h\rangle. \quad (27.11.53)$$

It follows that  $|\mathbf{w}^h\rangle$  is an eigenvector of  $\mathcal{C}_2$  having eigenvalue  $\lambda(\mathbf{w}^h, \boldsymbol{\mu}^+)$ ,

$$\mathcal{C}_2 |\mathbf{w}^h\rangle = \lambda(\mathbf{w}^h, \boldsymbol{\mu}^+) |\mathbf{w}^h\rangle, \quad (27.11.54)$$

with

$$\lambda(\mathbf{w}^h, \boldsymbol{\mu}^+) = (\mathbf{w}^h \cdot \mathbf{w}^h) + (\boldsymbol{\mu}^+ \cdot \mathbf{w}^h). \quad (27.11.55)$$

There is one last observation: We know that every state in a representation can be obtained by suitable linear combinations of products of ladder operators and Cartan sub-algebra operators and constants applied to the highest-weight state. Also,  $\mathcal{C}_2$  commutes with all these operations. It follows that *all* the vectors in an irreducible representation are eigenvectors of  $\mathcal{C}_2$  with the *same* common eigenvalue  $\lambda(\mathbf{w}^h, \boldsymbol{\mu}^+)$ .

For future use, let us work out explicitly the eigenvalues of  $\mathcal{C}_2$  for general representations in the cases of  $sp(2)$ ,  $sp(4)$ , and  $sp(6)$ . Begin with  $sp(2)$ . In this case

$$\mathbf{w}^h = n\phi^1 = ne^1, \quad (27.11.56)$$

and

$$\boldsymbol{\mu}^+ = \boldsymbol{\alpha} = 2e^1. \quad (27.11.57)$$

Consequently, for the representation  $\Gamma(n)$ ,  $\mathcal{C}_2$  has the eigenvalue

$$\lambda(\mathbf{w}^h, \boldsymbol{\mu}^+) = n^2 + 2n = n(n+2) = 4j(j+1). \quad (27.11.58)$$

We also note that  $\mathcal{C}_2$  has the explicit form

$$\begin{aligned} \mathcal{C}_2 &= (C^1)^2 + R(-\boldsymbol{\alpha})R(\boldsymbol{\alpha}) + R(\boldsymbol{\alpha})R(-\boldsymbol{\alpha}) \\ &= 4\hat{J}_3^2 + 2\hat{J}_-\hat{J}_+ + 2\hat{J}_+\hat{J}_- \\ &= 4(\hat{J}_1^2 + \hat{J}_2^2 + \hat{J}_3^2). \end{aligned} \quad (27.11.59)$$

Here the quantities  $\hat{J}_{\pm}$  and  $\hat{J}_3$  (or  $\hat{J}_1$  to  $\hat{J}_3$ ) are some kind of linear operators or matrices that obey commutation rules analogous to (1.3), (1.4), or (1.21). The results (11.58) and (11.59) are those expected for  $su(2)$ . In particular, the quadratic Casimir operator is proportional to the square of the angular momentum, which is known to commute with the  $\hat{J}_k$ .

For the case of  $sp(4)$ ,

$$\mathbf{w}^h = m\phi^1 + n\phi^2 = (m+n)e^1 + ne^2, \quad (27.11.60)$$

and

$$\boldsymbol{\mu}^+ = \boldsymbol{\alpha} + \boldsymbol{\beta} + \boldsymbol{\gamma} + (-\boldsymbol{\delta}) = 4e^1 + 2e^2. \quad (27.11.61)$$

Consequently, for the representation  $\Gamma(m, n)$ ,  $\mathcal{C}_2$  has the eigenvalue

$$\lambda(\mathbf{w}^h, \boldsymbol{\mu}^+) = m^2 + 2mn + 2n^2 + 4m + 6n. \quad (27.11.62)$$

For the case of  $sp(6)$ ,

$$\mathbf{w}^h = \ell\phi^1 + m\phi^2 + n\phi^3 = (\ell+m+n)e^1 + (m+n)e^2 + ne^3, \quad (27.11.63)$$

and

$$\boldsymbol{\mu}^+ = \boldsymbol{\alpha}^1 + \boldsymbol{\alpha}^2 + \boldsymbol{\alpha}^3 + \boldsymbol{\beta}^1 + \boldsymbol{\beta}^2 + \boldsymbol{\beta}^3 + \boldsymbol{\gamma}^1 + \boldsymbol{\gamma}^2 + (-\boldsymbol{\gamma}^3) = 6e^1 + 4e^2 + 2e^3. \quad (27.11.64)$$

Consequently, for the representation  $\Gamma(\ell, m, n)$ ,  $\mathcal{C}_2$  has the eigenvalue

$$\lambda(\mathbf{w}^h, \boldsymbol{\mu}^+) = \ell^2 + 2m^2 + 3n^2 + 2\ell m + 2\ell n + 4mn + 6\ell + 10m + 12n. \quad (27.11.65)$$

With this background, we are prepared to use the quadratic Casimir operator to prove the orthogonality relations (10.12). Our main tool will be the adjoint Lie operator version of  $\mathcal{C}_2$  defined by writing

$$\mathcal{C}_2 = \sum_j (\# : c^j : \#)^2 + \sum_{\boldsymbol{\mu}} \# : r(-\boldsymbol{\mu}) : \# \# : r(\boldsymbol{\mu}) : \#. \quad (27.11.66)$$

From (4.8), (4.14), and (10.9) we find the conjugation relations

$$[(\# : c^j : \#)^2]^\dagger = [(\# : c^j : \#)^\dagger]^2 = (\# : c^j : \#)^\dagger \# = (\# : c^j : \#)^2, \quad (27.11.67)$$

$$\begin{aligned} [(\# : r(-\boldsymbol{\mu}) : \# \# : r(\boldsymbol{\mu}) : \#)^\dagger] &= [\# : r(\boldsymbol{\mu}) : \#]^\dagger [\# : r(-\boldsymbol{\mu}) : \#]^\dagger \\ &= [\# : r(\boldsymbol{\mu}) : \#]^\dagger [\# : r(-\boldsymbol{\mu}) : \#]^\dagger = [\# : r(-\boldsymbol{\mu}) : \# \# : r(\boldsymbol{\mu}) : \#]. \end{aligned} \quad (27.11.68)$$

It follows that  $\mathcal{C}_2$  is Hermitian,

$$\mathcal{C}_2^\dagger = \mathcal{C}_2. \quad (27.11.69)$$

Apply  $\mathcal{C}_2$  to any element of the vector fields  $\mathcal{H}^{\ell+1,0,0}$ ,  $\mathcal{G}^{\ell-1,1,0}$ , and  $\mathcal{G}^{\ell-1,0,0}$ . Based on (11.65), we find (writing in short-hand form) the results

$$\mathcal{C}_2 \mathcal{H}^{\ell+1,0,0} = [(\ell + 1)^2 + 6(\ell + 1)] \mathcal{H}^{\ell+1,0,0}, \quad (27.11.70)$$

$$\begin{aligned} \mathcal{C}_2 \mathcal{G}^{\ell-1,1,0} &= [(\ell - 1)^2 + 2 + 2(\ell - 1) + 6(\ell - 1) + 10] \mathcal{G}^{\ell-1,1,0} \\ &= [(\ell - 1)^2 + 8(\ell - 1) + 12] \mathcal{G}^{\ell-1,1,0}, \end{aligned} \quad (27.11.71)$$

$$\mathcal{C}_2 \mathcal{G}^{\ell-1,0,0} = [(\ell - 1)^2 + 6(\ell - 1)] \mathcal{G}^{\ell-1,0,0}. \quad (27.11.72)$$

Next consider matrix elements of the form  $\langle \mathcal{H}^{\ell+1,0,0}, \mathcal{C}_2 \mathcal{G}^{\ell-1,1,0} \rangle$ . From (11.71) we have the result

$$\langle \mathcal{H}^{\ell+1,0,0}, \mathcal{C}_2 \mathcal{G}^{\ell-1,1,0} \rangle = [(\ell - 1)^2 + 8(\ell - 1) + 12] \langle \mathcal{H}^{\ell+1,0,0}, \mathcal{G}^{\ell-1,1,0} \rangle. \quad (27.11.73)$$

However, using (11.69) and (11.70), these matrix elements also satisfy the relation

$$\begin{aligned} \langle \mathcal{H}^{\ell+1,0,0}, \mathcal{C}_2 \mathcal{G}^{\ell-1,1,0} \rangle &= \langle \mathcal{C}_2^\dagger \mathcal{H}^{\ell+1,0,0}, \mathcal{G}^{\ell-1,1,0} \rangle = \langle \mathcal{C}_2 \mathcal{H}^{\ell+1,0,0}, \mathcal{G}^{\ell-1,1,0} \rangle \\ &= [(\ell + 1)^2 + 6(\ell + 1)] \langle \mathcal{H}^{\ell+1,0,0}, \mathcal{G}^{\ell-1,1,0} \rangle. \end{aligned} \quad (27.11.74)$$

By combining (11.73) and (11.74) we find the result

$$(2\ell + 2) \langle \mathcal{H}^{\ell+1,0,0}, \mathcal{G}^{\ell-1,1,0} \rangle = 0. \quad (27.11.75)$$

In similar fashion we find the results

$$(4\ell + 12) \langle \mathcal{H}^{\ell+1,0,0}, \mathcal{G}^{\ell-1,0,0} \rangle = 0, \quad (27.11.76)$$

$$(2\ell + 10) \langle \mathcal{G}^{\ell-1,1,0}, \mathcal{G}^{\ell-1,0,0} \rangle = 0. \quad (27.11.77)$$

We observe that none of the quantities in parentheses on the left sides of (11.75) through (11.77) vanish for  $\ell \geq 0$ . Therefore all the scalar products of the form (10.12) vanish as advertised.

There is a related use of the quadratic Casimir operator that is also important. In Section 21.10 it was shown that there is an operator  $\mathcal{P}^G$  that projects out the non-Hamiltonian part of a general homogeneous vector field  $\mathcal{L}_{\mathbf{g}^\ell}$ ,

$$\mathcal{P}^G \mathcal{L}_{\mathbf{g}^\ell} = \mathcal{G}_{\ell+1}, \quad (27.11.78)$$

and an explicit procedure was given for finding  $\mathcal{G}_{\ell+1}$ . See (10.20) and (10.33). Now we will show how  $\mathcal{C}_2$  can be used to decompose  $\mathcal{G}_{\ell+1}$  into its separate parts  $\mathcal{G}^{\ell-1,1,0}$  and  $\mathcal{G}^{\ell-1,0,0}$ ,

$$\mathcal{G}_{\ell+1} = \mathcal{G}^{\ell-1,1,0} + \mathcal{G}^{\ell-1,0,0}. \quad (27.11.79)$$

Suppose  $\mathcal{C}_2$  is applied to both sides of (11.79). Then, from (11.71) and (11.72), we find the result

$$\begin{aligned} \mathcal{C}_2 \mathcal{G}_{\ell+1} &= \mathcal{C}_2 \mathcal{G}^{\ell-1,1,0} + \mathcal{C}_2 \mathcal{G}^{\ell-1,0,0} \\ &= [(\ell-1)^2 + 8(\ell-1) + 12] \mathcal{G}^{\ell-1,1,0} + [(\ell-1)^2 + 6(\ell-1)] \mathcal{G}^{\ell-1,0,0}. \end{aligned} \quad (27.11.80)$$

The two relations (11.79) and (11.80) can be solved for  $\mathcal{G}^{\ell-1,1,0}$  and  $\mathcal{G}^{\ell-1,0,0}$  to give the explicit results

$$\mathcal{G}^{\ell-1,1,0} = \{2\ell+10\}^{-1} \{ \mathcal{C}_2 - [(\ell-1)^2 + 6(\ell-1)] \} \mathcal{G}_{\ell+1}, \quad (27.11.81)$$

$$\mathcal{G}^{\ell-1,0,0} = -\{2\ell+10\}^{-1} \{ \mathcal{C}_2 - [(\ell-1)^2 + 8(\ell-1) + 12] \} \mathcal{G}_{\ell+1}. \quad (27.11.82)$$

Thus, given any homogeneous vector field  $\mathcal{L}_{\mathbf{g}^\ell}$ , we have an explicit procedure for finding its Hamiltonian part  $\mathcal{H}^{\ell+1,0,0}$  and its non-Hamiltonian parts  $\mathcal{G}^{\ell-1,1,0}$  and  $\mathcal{G}^{\ell-1,0,0}$ . Of course, this is not the full story. For some purposes we would like to have a complete set of basis vectors for the spaces  $\mathcal{H}^{\ell+1,0,0}$ ,  $\mathcal{G}^{\ell-1,1,0}$ , and  $\mathcal{G}^{\ell-1,0,0}$ . This was done for the case  $\ell = 1$  in Section 21.9, and we would like to have analogous results for all  $\ell$ , or for at least the first few values of  $\ell$  (say  $\ell = 2$  through 6 or so). The spaces  $\mathcal{H}^{\ell+1,0,0}$  and  $\mathcal{G}^{\ell-1,0,0}$  are relatively easy to handle because the elements of  $\mathcal{H}^{\ell+1,0,0}$  are of the form :  $h_{\ell+1}$  : and the elements of  $\mathcal{G}^{\ell-1,0,0}$  are of the form (9.3). In both cases one is working with homogeneous polynomials and must find suitable basis polynomials that correspond to the various weights in the weight diagrams for the representations  $\Gamma(\ell+1, 0, 0)$  and  $\Gamma(\ell-1, 0, 0)$ . This is relatively straightforward except for the problem of finding additional labels and associated properties when the weights have multiplicities higher than 1. Handling the space  $\mathcal{G}^{\ell-1,1,0}$  is more difficult. In this case it would be helpful to have explicit knowledge of the Clebsch-Gordan coefficients of  $sp(6)$  [and  $sp(4)$ ] for at least the relatively low-dimensional representations.

### 27.11.3 Higher-Order Casimir Operators

Before leaving the subject of Casimir operators, something should be said about cubic and higher-order Casimirs. For simplicity, only the cubic case will be considered, but the generalization to higher orders should be evident.

As before, we work with some realization  $R$ , and let  $\hat{B}_\alpha$  denote the basis elements of the Lie algebra in this realization. Then, in analogy to (11.11), we define a rank *three* tensor  ${}^3g_{\alpha\alpha'\alpha''}^R$  by writing

$${}^3g_{\alpha\alpha'\alpha''}^R = \text{tr} (\hat{B}_\alpha \hat{B}_{\alpha'} \hat{B}_{\alpha''}). \quad (27.11.83)$$

In view of (11.24) and the properties of the trace, we have the relation

$$\text{tr} (\hat{B}_\alpha^{\text{tr}} \hat{B}_{\alpha'}^{\text{tr}} \hat{B}_{\alpha''}^{\text{tr}}) = \text{tr} (\hat{B}_\alpha \hat{B}_{\alpha'} \hat{B}_{\alpha''}). \quad (27.11.84)$$

From this relation and (11.28) we deduce that  ${}_3g^R$  has the property

$$\sum_{\beta\beta'\beta''} U_{\alpha\beta}U_{\alpha'\beta'}U_{\alpha''\beta''} {}_3g_{\beta\beta'\beta''}^R = {}_3g_{\alpha\alpha'\alpha''}^R, \quad (27.11.85)$$

which is the analog of (11.30).

Next use  $g_R$  to raise the indices on  ${}_3g^R$  by the rule

$${}_3g_R^{\alpha\alpha'\alpha''} = g_R^{\alpha\gamma} g_R^{\alpha'\gamma'} g_R^{\alpha''\gamma''} {}_3g_{\gamma\gamma'\gamma''}^R, \quad (27.11.86)$$

Here, and in what follows, we have and will use the summation convention. The raised tensor  ${}_3g_R$  has the property

$${}_3g_R^{\alpha\alpha'\alpha''} U_{\alpha\beta}U_{\alpha'\beta'}U_{\alpha''\beta''} = {}_3g_R^{\beta\beta'\beta''}, \quad (27.11.87)$$

which is the analog of (11.33). [Note that in (11.85) the summation is over the second indices in the  $U$ 's, while in (11.87) it is over the first indices.] To verify (11.87), simply compute. From (11.86) we have

$${}_3g_R^{\alpha\alpha'\alpha''} U_{\alpha\beta}U_{\alpha'\beta'}U_{\alpha''\beta''} = g_R^{\alpha\gamma} U_{\alpha\beta} g^{\alpha'\gamma'} U_{\alpha'\beta'} g^{\alpha''\gamma''} U_{\alpha''\beta''} {}_3g_{\gamma\gamma'\gamma''}^R. \quad (27.11.88)$$

However, by changing indices, (11.85) can be written in the form

$${}_3g_{\gamma\gamma'\gamma''}^R = U_{\gamma\delta} U_{\gamma'\delta'} U_{\gamma''\delta''} {}_3g_{\delta\delta'\delta''}^R. \quad (27.11.89)$$

Now substitute (11.89) in (11.88) to get the result

$$\begin{aligned} {}_3g_R^{\alpha\alpha'\alpha''} U_{\alpha\beta}U_{\alpha'\beta'}U_{\alpha''\beta''} &= g_R^{\alpha\gamma} U_{\alpha\beta} U_{\gamma\delta} g_R^{\alpha'\gamma'} U_{\alpha'\beta'} U_{\gamma'\delta'} g_R^{\alpha''\gamma''} U_{\alpha''\beta''} U_{\gamma''\delta''} {}_3g_{\delta\delta'\delta''}^R \\ &= g_R^{\beta\delta} g_R^{\beta'\delta'} g_R^{\beta''\delta''} {}_3g_{\delta\delta'\delta''}^R = {}_3g_R^{\beta\beta'\beta''}, \end{aligned} \quad (27.11.90)$$

as claimed. Here repeated use has been made of (11.30).

We are now ready to construct the third-order Casimir operator. As before, suppose  $\hat{B}_\alpha$  is any set of linear operators that obey (11.5), but do not necessarily belong to the realization used to define  $g^R$ ,  ${}_3g^R$ , and hence  ${}_3g_R$ . We define the associated cubic Casimir operator  $\mathcal{C}_3$  by writing

$$\mathcal{C}_3 = {}_3g_R^{\alpha\alpha'\alpha''} \hat{B}_\alpha \hat{B}_{\alpha'} \hat{B}_{\alpha''}. \quad (27.11.91)$$

As a consequence of (11.87), this operator also has the invariance property

$$\mathcal{C}_3^{\text{tr}} = \exp(\epsilon \# C \#) \mathcal{C}_3 = \mathcal{C}_3, \quad (27.11.92)$$

and hence

$$\# \hat{B}_\delta \# \mathcal{C}_3 = \{ \hat{B}_\delta, \mathcal{C}_3 \} = 0 \text{ for all } \delta. \quad (27.11.93)$$

Indeed, using (11.37) and (11.87), we find the result

$$\begin{aligned} \mathcal{C}_3^{\text{tr}} &= {}_3g_R^{\alpha\alpha'\alpha''} \hat{B}_\alpha^{\text{tr}} \hat{B}_{\alpha'}^{\text{tr}} \hat{B}_{\alpha''}^{\text{tr}} \\ &= {}_3g_R^{\alpha\alpha'\alpha''} U_{\alpha\beta} U_{\alpha'\beta'} U_{\alpha''\beta''} \hat{B}_\beta \hat{B}_{\beta'} \hat{B}_{\beta''} \\ &= {}_3g_R^{\beta\beta'\beta''} \hat{B}_\beta \hat{B}_{\beta'} \hat{B}_{\beta''} = \mathcal{C}_3. \end{aligned} \quad (27.11.94)$$

Thus, as anticipated,  $\mathcal{C}_3$  also commutes with all Lie generators, and therefore also with all products and sums of products of generators. A special case of this result is that  $\mathcal{C}_2$  and  $\mathcal{C}_3$  commute.

Finally, we remark that it can be shown that a rank- $k$  simple Lie algebra has  $k$  functionally independent Casimir operators. Moreover, the eigenvalues of these Casimir operators, when acting on any vector in an irreducible representation, can be used to determine the representation. For example, in the case of  $sp(6)$ , the three Casimir operators  $\mathcal{C}_2$ ,  $\mathcal{C}_4$ , and  $\mathcal{C}_6$  are functionally independent and can be used to determine the values of  $\ell$ ,  $m$ ,  $n$  in  $\Gamma(\ell, m, n)$ .

## Exercises

**27.11.1.** Verify (11.12) using (11.11) and the matrices  $B^0$ ,  $F$ , and  $G$  given by (5.6.7), (5.6.13), and (5.6.14). Find the  $2 \times 2$  matrices associated with the elements associated with  $c^1$  and  $r(\pm\alpha)$  given by (1.11) and (1.12). Use these matrices to verify (11.13). Find  $g^F$  for  $sp(4)$  using the fundamental representation given by (5.7.42).

For quadratic polynomials, using (5.5.1) and (5.5.2) and the correspondences (5.5.3) and (5.5.4), we have made the definition  $\langle f, g \rangle_F = (JS^f, JS^g)_F = \text{tr}(JS^f JS^g)$ . Compare  $\langle f, g \rangle$  and  $\langle f, g \rangle_F$ . See Exercise 7.3.8. Verify the general result  $\langle f^a, f^c \rangle_F = (JS^{fa}, JS^{fc})_F = 0$ . Use (3.8.14), (3.8.22), (7.2.3), (7.2.4), and (7.3.53); and employ the notation  $S^{fa}$  and  $S^{fc}$  to denote the parts of  $S^f$  that anticommute and commute with  $J$ , respectively. See Exercise 7.3.10 for the analogous result  $\langle f_2^a, f_2^c \rangle = 0$ . For a pair of quadratic polynomials  $f$  and  $g$ , make the decompositions  $f = f^a + f^c$  and  $g = g^a + g^c$ . Verify the relation  $\langle f, g \rangle_F = 2\langle f^a, g^a \rangle - 2\langle f^c, g^c \rangle$ . As a special case there is the relation  $\langle f, f \rangle_F = 2\langle f^a, f^a \rangle - 2\langle f^c, f^c \rangle$ , which shows that the form  $\langle f, f \rangle_F$  is neither positive nor negative definite. This is to be expected because  $Sp(2n, \mathbb{R})$  is not compact. We also have the relation

$$\langle f^c, f^c \rangle_F = (JS^{fc}, JS^{fc})_F = -2\langle f^c, f^c \rangle < 0,$$

and we know that the  $JS^{fc}$  generate  $U(n)$ , the maximal compact subgroup of  $Sp(2n, \mathbb{R})$ . See the last comment in Exercise 11.3 below.

**27.11.2.** Verify the relations (11.16) through (11.19).

**27.11.3.** Review Exercise 11.1 above. This exercise further explores the relation between the Lie-algebraic metric for the Lie algebra  $usp(2n)$  and the  $Usp(2n)$  invariant scalar product. See Sections 5.10 and 7.3. Consider the  $sp(2n)$  Lie-algebraic metric given by (11.16) through (11.19). Define elements  $\Sigma$  and  $\Delta$  by the rules

$$\Sigma(\boldsymbol{\mu}) = (1/\sqrt{2})[r(\boldsymbol{\mu}) + r(-\boldsymbol{\mu})], \quad (27.11.95)$$

$$\Delta(\boldsymbol{\mu}) = (1/\sqrt{2})[r(\boldsymbol{\mu}) - r(-\boldsymbol{\mu})]. \quad (27.11.96)$$

Evidently, for  $\boldsymbol{\mu} > 0$ , the elements  $\Sigma(\boldsymbol{\mu})$  and  $\Delta(\boldsymbol{\mu})$  span the same space as the elements  $r(\boldsymbol{\mu})$  and  $r(-\boldsymbol{\mu})$ . Using (11.16) through (11.19) verify (with  $\boldsymbol{\mu}, \boldsymbol{\nu} > 0$ ) the relations

$$(c^j, \Sigma(\boldsymbol{\mu}))_F = (c^j, \Delta(\boldsymbol{\mu}))_F = 0, \quad (27.11.97)$$

$$(\Sigma(\boldsymbol{\mu}), \Delta(\boldsymbol{\nu}))_F = 0, \quad (27.11.98)$$

$$(\Sigma(\boldsymbol{\mu}), \Sigma(\boldsymbol{\nu}))_F = 2\delta_{\boldsymbol{\mu}\boldsymbol{\nu}}, \quad (27.11.99)$$

$$(\Delta(\boldsymbol{\mu}), \Delta(\boldsymbol{\nu}))_F = -2\delta_{\boldsymbol{\mu}\boldsymbol{\nu}}. \quad (27.11.100)$$

Show that : $\Sigma$ : and : $\Delta$ : obey the conjugacy relations

$$:\Sigma(\boldsymbol{\mu}) :^\dagger = :\Sigma(\boldsymbol{\mu}) :, \quad (27.11.101)$$

$$:\Delta(\boldsymbol{\mu}) :^\dagger = - :\Delta(\boldsymbol{\mu}) :. \quad (27.11.102)$$

Consider as a basis set the elements  $i\mathcal{C}^j$ ,  $i\Sigma(\boldsymbol{\mu})$ , and  $\Delta(\boldsymbol{\mu})$  with  $\boldsymbol{\mu} > 0$ . Call these elements  $b_\alpha$ . Show that their associated Lie operators are all anti-Hermitian,

$$:b_\alpha :^\dagger = - :b_\alpha :. \quad (27.11.103)$$

Consequently, they form a basis for  $usp(2n)$  and, when exponentiated, generate  $Usp(2n)$ . See Section 7.3. Verify that these elements satisfy the relation

$$(b_\alpha, b_\beta)_F = -2\delta_{\alpha\beta} = -2\langle b_\alpha, b_\beta \rangle. \quad (27.11.104)$$

Here the Lie-algebraic scalar product on the left is that given by (11.16) through (11.19), and the scalar product on the right is that given by (4.19) through (4.22) and arises from the construction of Section 7.3. Since we have been working over the complex field, we know that the Lie-algebraic scalar product is invariant under  $Sp(2n, C)$ . See (11.25). It is therefore also invariant under  $Usp(2n)$  because  $Usp(2n)$  is a subgroup of  $Sp(2n, C)$ . The relation (11.104) is consistent with this invariance because we already know from Section 7.3 that the scalar product on the right is invariant under  $Usp(2n)$ .

One last comment: From the discussion of Section 5.10 we know that  $Usp(2n)$  is compact. Inspection of (11.104) shows that the Lie-algebraic metric for  $usp(2n)$  is negative definite. It can be shown that for any simple Lie algebra the Lie-algebraic metric is negative definite if and only if the corresponding Lie group is compact. See also (11.109) and (11.110) for the cases of the compact groups  $SU(2)$  and  $SO(3, \mathbb{R})$ .

**27.11.4.** The relation (11.30) displays the invariance of the metric tensor under finite group action. Show that (11.27) describes this same invariance in infinitesimal form. In particular, use (11.27) to produce the relation

$$c_{\alpha\alpha'}^\gamma g_{\gamma\alpha''}^R = c_{\alpha'\alpha''}^\gamma g_{\gamma\alpha}^R. \quad (27.11.105)$$

Use the metric tensor  $g^R$  to lower the upper index on the structure constants by the rule

$$c_{\alpha\alpha'\alpha''} = c_{\alpha\alpha'\alpha''}^\gamma g_{\gamma\alpha}^R. \quad (27.11.106)$$

Show that the lowered structure constants are completely antisymmetric (antisymmetric under the interchange of any pair of adjacent indices).

**27.11.5.** Find the analog of the formulas (11.81) and (11.82) for the case of 4-dimensional phase space.

**27.11.6.** In the case of a 6-dimensional phase space, consider the vector field  $\mathcal{L}_{\mathbf{g}^2}$  given by the relation

$$\mathcal{L}_{\mathbf{g}^2} = (q_1)^2 \partial / \partial q_1. \quad (27.11.107)$$

Using the methods of Sections 21.10 and 21.11.2, decompose  $\mathcal{L}_{\mathbf{g}^2}$  into Hamiltonian and non-Hamiltonian parts  $\mathcal{H}^{3,0,0}$ ,  $\mathcal{G}^{1,1,0}$ , and  $\mathcal{G}^{1,0,0}$ .

**27.11.7.** For the case of a 2-dimensional phase space we know that quadratic functions  $f_2$  and  $g_2$  carry the representation  $\Gamma(2)$ . Therefore, from (11.1), we might naively expect that the product  $f_2 g_2$  might contain the identity representation  $\Gamma(0)$ . In analogy with (11.59), if the identity representation does occur, it should be a multiple of the polynomial  $(J_1^2 + J_2^2 + J_3^2)$  where the  $J_j$  are given by (1.1), (1.2), and (1.20). Show that in fact there is the relation

$$J_1^2 + J_2^2 + J_3^2 = 0. \quad (27.11.108)$$

Thus, in this case, the sought after quantity actually vanishes. In retrospect, this is to be expected because we know from Section 21.2 that quartic polynomials in two variables, of which all polynomials of the form  $f_2 g_2$  are examples, carry only the representation  $\Gamma(4)$ . Note that  $\Gamma(4)$  also occurs in the Clebsch-Gordan series (11.1).

**27.11.8.** The purpose of this exercise is to explore the consequences of the relations (11.27) in the case of  $su(2)$ , or equivalently,  $so(3, \mathbb{R})$ . In the case of  $su(2)$ , suppose we employ the realization provided by the  $K^\alpha$  matrices of Exercise 3.7.30. Verify the scalar product results

$$(K^\alpha, K^\beta)_F = \text{tr}(K^\alpha K^\beta) = (-i/2)^2 \text{tr}(\sigma^\alpha \sigma^\beta) = -(1/2)\delta_{\alpha\beta} = -(1/2)\mathbf{e}_\alpha \cdot \mathbf{e}_\beta. \quad (27.11.109)$$

Show that in the cases of  $su(2)$  and  $so(3, \mathbb{R})$  there are the related results

$$(K^\alpha, K^\beta)_K = (L^\alpha, L^\beta)_F = \text{tr}(L^\alpha L^\beta) = -2\delta_{\alpha\beta} = -2\mathbf{e}_\alpha \cdot \mathbf{e}_\beta. \quad (27.11.110)$$

Here the subscript  $K$  stands for *Killing* in anticipation of the next section. Next show, using the notation of Exercise 3.2.27, that there are the relations

$$(\mathbf{a} \cdot \mathbf{K}, \mathbf{b} \cdot \mathbf{K})_F = -(1/2)\mathbf{a} \cdot \mathbf{b}, \quad (27.11.111)$$

$$(\mathbf{a} \cdot \mathbf{K}, \mathbf{b} \cdot \mathbf{K})_K = (\mathbf{a} \cdot \mathbf{L}, \mathbf{b} \cdot \mathbf{L})_F = -2\mathbf{a} \cdot \mathbf{b}. \quad (27.11.112)$$

Now examine the first relation in (11.27). In the case of  $su(2)$ , and using the fundamental representation, it reads

$$(\{K_\alpha, K_\beta\}, K_\gamma)_F = (K_\alpha, \{K_\beta, K_\gamma\})_F. \quad (27.11.113)$$

Show that multiplying the left side of (11.113) by the quantity  $a_\alpha b_\beta c_\gamma$ , and summing over  $\alpha$ ,  $\beta$ , and  $\gamma$ , yield the result

$$\sum_{\alpha\beta\gamma} a_\alpha b_\beta c_\gamma (\{K_\alpha, K_\beta\}, K_\gamma)_F = (\{\mathbf{a} \cdot \mathbf{K}, \mathbf{b} \cdot \mathbf{K}\}, \mathbf{c} \cdot \mathbf{K})_F = -(1/2)(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}. \quad (27.11.114)$$

Show that multiplying the right side of (11.113) by the quantity  $a_\alpha b_\beta c_\gamma$ , and summing over  $\alpha$ ,  $\beta$ , and  $\gamma$ , yield the result

$$\sum_{\alpha\beta\gamma} a_\alpha b_\beta c_\gamma (K_\alpha, \{K_\beta, K_\gamma\})_F = (\mathbf{a} \cdot \mathbf{K}, \{\mathbf{b} \cdot \mathbf{K}, \mathbf{c} \cdot \mathbf{K}\})_F = -(1/2)\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}). \quad (27.11.115)$$

You have shown, in the case of  $su(2)$ , that (11.113) is equivalent to the familiar statement about the interchange of the dot and the cross in three-dimensional vector algebra,

$$(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}). \quad (27.11.116)$$

Carry out the analogous calculation for both the cases of  $su(2)$  and  $so(3, \mathbb{R})$  using (11.110) to again arrive at the conclusion (11.116). Finally show, in the cases of  $su(2)$  and  $so(3, R)$ , that the second relation in (11.27) is equivalent to the relation

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}). \quad (27.11.117)$$

**27.11.9.** Show that the first relation in (11.27) can be rewritten in the form

$$(- : B_{\alpha'} : B_\alpha, B_{\alpha''})_R = (B_\alpha, : B_{\alpha'} : B_{\alpha''})_R. \quad (27.11.118)$$

Here we have used the more compact notation

$$: B_{\alpha'} := \text{ad } B_{\alpha'}. \quad (27.11.119)$$

Compare (3.7.71), (5.3.2), and (11.21). Show that (11.118) implies the relation

$$: B_{\alpha'} :^\dagger = - : B_{\alpha'} : \quad (27.11.120)$$

with respect to the inner product  $( , )_R$ . That is,  $: B_{\alpha'} :$  is anti-Hermitian with respect to this inner product.

**27.11.10.** Exercise on the Casimir operator for  $SO(4, \mathbb{R})$ .

## 27.12 The Killing Form

Section 21.11.1 introduced the concept of a scalar product for the elements of a Lie algebra and defined a metric tensor with the aid of a realization  $R$ . An important special case of this construction is the Killing form. The *Killing form* is simply the metric tensor  $g^R$  in the case that the realization  $R$  is the *adjoint* representation. See the end of Section 3.7 to review the definition of the adjoint representation. Let us call this tensor  $g^K$  in honor of Killing. Then, using (11.11) and (3.7.56), we find the result

$$g^K_{\alpha\alpha'} = \text{tr } (\hat{B}_\alpha \hat{B}_{\alpha'}) = \sum_{\mu\nu} (\hat{B}_\alpha)_{\mu\nu} (\hat{B}_{\alpha'})_{\nu\mu} = \sum_{\mu\nu} c_{\alpha\nu}^\mu c_{\alpha'\mu}^\nu. \quad (27.12.1)$$

As (12.1) shows, the Killing form (metric tensor) has the advantage that it is constructed directly in terms of the structure constants. It is therefore directly available without further study of the Lie algebra.<sup>4</sup> By contrast, the  $g^F$  that we have been using for  $sp(2n)$

---

<sup>4</sup>Suppose a Lie algebra  $L$  is specified by presenting its structure constants. Then  $g^K$  is computable using (12.1). It can be shown that  $g^K$  is invertible if and only if  $L$  is semisimple.

is constructed from a knowledge of the fundamental  $2n \times 2n$  matrix representation. For most groups these matrices are usually much smaller than the matrices for the adjoint representation, and therefore, *if known*, far easier to use. For example, in the case of  $sp(6)$ , the fundamental representation involves  $6 \times 6$  matrices, and the adjoint representation involves  $21 \times 21$  matrices. [However, it turns out that for  $E_8(248)$  the lowest dimensional representation *is* the adjoint representation, and  $248 \times 248$  matrices are required.]

It can be shown in general for a *simple* Lie algebra that  $g^K$  and  $g^F$  are proportional. [For example, the identity representation occurs once and only once in the  $sp(2n)$  Clebsch-Gordan series (11.1) through (11.3).] From (11.16) we know that for the fundamental representation of  $sp(2n)$  there is the relation

$$(\hat{C}^1, \hat{C}^1)_F = 2. \quad (27.12.2)$$

And for the adjoint representation of  $sp(2n)$  there is the relation

$$(\hat{C}^1, \hat{C}^1)_K = \sum_{\mu} (\mathbf{e}^1 \cdot \boldsymbol{\mu})(\boldsymbol{\mu} \cdot \mathbf{e}^1) = 4n + 4. \quad (27.12.3)$$

See Exercise 12.1. It follows that  $g^K$  and  $g^F$  are related by the equation

$$g_{\alpha\alpha'}^K = (2n + 2)g_{\alpha\alpha'}^F. \quad (27.12.4)$$

Still a bit more can be said. As before, let  $g^R$  be the metric tensor obtained using the realization  $R$  as in (11.11). Then there is a relation of the form

$$g_{\alpha\alpha'}^R = \tau(R)g_{\alpha\alpha'}^F. \quad (27.12.5)$$

where  $\tau(R)$  is a *positive* proportionality constant that depends on the realization.<sup>5</sup> According to (12.4),  $\tau$  has the value  $(2n + 2)$  for the adjoint representation and the Lie algebra  $sp(2n)$ .

## Exercises

**27.12.1.** Verify (12.2). Using (3.7.56), show that in the adjoint representation the matrix  $\hat{C}^1$  is diagonal and has as its diagonal entries  $\ell$  zeroes (where  $\ell$  is the rank of the Lie algebra) and the numbers  $(\mathbf{e} \cdot \boldsymbol{\mu})$  where  $\boldsymbol{\mu}$  ranges over all the root vectors. Next show that  $(\hat{C}^1, \hat{C}^1)_K$  has the value

$$(\hat{C}^1, \hat{C}^1)_K = \text{tr}[(\hat{C}^1)^2] = \sum_{\mu} (\mathbf{e}^1 \cdot \boldsymbol{\mu})(\boldsymbol{\mu} \cdot \mathbf{e}^1). \quad (27.12.6)$$

Finally, given that the root vectors for  $sp(2n)$  are all combinations of the form  $\pm \mathbf{e}^j \pm \mathbf{e}^k$  with the signs taken independently, verify (12.3).

**27.12.2.** Review Exercise 3.7.30. The  $2 \times 2$  matrices  $K^\alpha$  and the  $3 \times 3$  matrices  $L^\alpha$  displayed there provide the fundamental and adjoint representations of  $su(2)$ , respectively. Use these matrices to construct metric tensors for  $su(2)$ . Show that

$$g_{\alpha\alpha'}^F = -(1/2)\delta_{\alpha\alpha'} \quad (27.12.7)$$

---

<sup>5</sup>We remark that a relation of the form (12.5) holds among the *irreducible* representations of any *simple* Lie algebra. It need not hold in general. See Exercise 12.2.

and

$$g_{\alpha\alpha'}^K = -2\delta_{\alpha\alpha'}. \quad (27.12.8)$$

Note that  $g^F$  and  $g^K$  are proportional with a positive proportionality constant, as expected because  $su(2)$  is simple. Note also that they are negative definite because  $SU(2)$  is compact. (Recall that it has the topology of  $S^3$ .) Finally, since the Lie algebras  $so(3, R)$  and  $su(2)$  are the same, these  $g^F$  and  $g^K$  also provide metric tensors for  $so(3, \mathbb{R})$ .

**27.12.3.** Use the matrices (4.4.31) through (4.4.34) as a basis for  $gl(2, \mathbb{R})$  and show that in this case  $g^F$  is given by (4.4.40). Find the adjoint representation for  $gl(2, \mathbb{R})$ , which will be a set of  $4 \times 4$  matrices, compute the Killing form  $g^K$ , and show that it is *singular*. Thus, (12.5) does not hold in this case. Show that  $gl(2, \mathbb{R})$  is not *simple* and also not *semisimple*. See also the discussion at the end of Section 3.7.

**27.12.4.** Using the  $sp(4)$  matrices (5.7.42) and matrices of the form  $JA$  with the antisymmetric matrices  $A$  given by (6.28) through (6.33), find  $g^F$  for  $gl(4, \mathbb{R})$ .

## 27.13 Enveloping Algebra

So far we have been exploring the properties of products of entities when each entity by itself has well-defined properties under the action of some group (in our case, the symplectic group). These entities were either polynomials in some variables on which the group acted, or Lie operators, or adjoint Lie operators. (They could also be matrices or other linear operators. As shown in Section 26.\*), they could also be moments of a particle distribution.) They were not necessarily in the Lie algebra of the group, but they had the two properties that they could be multiplied together (multiplication was defined) and they transformed in some systematic way under the action of the Lie algebra.

For some (perhaps mathematical) purposes it is useful to explore what can be done by working directly and abstractly with only the Lie algebra itself rather than various concrete entities such as polynomials, Lie operators, adjoint Lie operators, etc. But now there is a problem because, for an abstract Lie algebra, there is no meaning for the “ordinary” product of any two elements in the Lie algebra. That is, if  $A$  and  $B$  are elements in a Lie algebra  $L$ , the Lie product  $[A, B]$  is defined, but there is no meaning to the product  $AB$ . In particular, there is no meaning to an associative product  $ABC$  such that  $(AB)C = A(BC)$ . This apparent obstacle can be overcome by a clever construction. We will see that in many ways the *tensor* product can be used to play the role of an ordinary product.

Since  $L$  is a vector space, it is meaningful to talk about tensor products of the space with itself. For example, if the elements  $B_\alpha$  are a basis for  $L$ , we may consider the space of all linear combinations of tensor products of the form  $B_\alpha \otimes B_{\alpha'}$ . We will call this vector space  $L^2$ , and write

$$L^2 = L \otimes L. \quad (27.13.1)$$

Similarly, we may consider the space of all linear combinations of tensor products of the form  $B_\alpha \otimes B_{\alpha'} \otimes B_{\alpha''}$ . Note that for a tensor product there is the associative property

$$B_\alpha \otimes (B_{\alpha'} \otimes B_{\alpha''}) = (B_\alpha \otimes B_{\alpha'}) \otimes B_{\alpha''}. \quad (27.13.2)$$

We will call this vector space  $L^3$ , and write

$$L^3 = L \otimes L \otimes L. \quad (27.13.3)$$

It is now obvious how to define still higher-order tensor product spaces  $L^4, L^5, \dots$

We would also like to define  $L^1$  and  $L^0$ . For  $L^1$  we take the Lie algebra itself. That is,  $L^1$  is the vector space consisting of all linear combinations of the  $B_\alpha$ . What about  $L^0$ ? Since  $L$  is a vector space, there must be some associated field of scalars (say the complex numbers) with *unit* element 1. Let  $L^0$  be the vector space of all linear combinations (scalar multiples) of 1. Evidently  $L^0$  is a one-dimensional vector space, and consists of the field of scalars associated with  $L$ .

Now watch closely! Having defined the vector spaces  $L^n$ , we define the vector space  $\mathcal{T}$  to be the direct sum of all these vector spaces,

$$\mathcal{T} = L^0 \oplus L^1 \oplus L^2 \oplus L^3 + \dots \quad (27.13.4)$$

Suppose  $\mathcal{A}$  and  $\mathcal{B}$  are any two elements in  $\mathcal{T}$ . As a simple example, suppose they are of the form

$$\mathcal{A} = a + bB_\alpha = a1 + bB_\alpha, \quad (27.13.5)$$

$$\mathcal{B} = c + dB_\beta = c1 + dB_\beta. \quad (27.13.6)$$

Let us compute their tensor product. Doing so gives the result

$$\mathcal{A} \otimes \mathcal{B} = (a1 + bB_\alpha) \otimes (c1 + dB_\beta) = a1 \otimes c1 + a1 \otimes dB_\beta + bB_\alpha \otimes c1 + bB_\alpha \otimes dB_\beta. \quad (27.13.7)$$

Let us make the obvious rules

$$1 \otimes 1 = 1, \quad 1 \otimes B_\beta = B_\beta, \quad B_\alpha \otimes 1 = B_\alpha, \quad \text{etc.}, \quad (27.13.8)$$

which can be summarized abstractly by writing

$$L^0 \otimes L^0 = L^0, \quad L^0 \otimes L^n = L^n, \quad L^n \otimes L^0 = L^n. \quad (27.13.9)$$

Then we find the result

$$\mathcal{A} \otimes \mathcal{B} = ac + adB_\beta + bcB_\alpha + bdB_\alpha \otimes B_\beta. \quad (27.13.10)$$

We know that, by construction,  $\mathcal{T}$  is a vector space. We now see that it can be also be viewed as an *associative* algebra with the operation of multiplication taken to be the tensor product. The standard nomenclature is to call  $\mathcal{T}$  the *tensor algebra* of  $L$ .

Our next step is to give  $\mathcal{T}$  a Lie-algebraic structure. Suppose again that  $\mathcal{A}$  and  $\mathcal{B}$  are any two elements in  $\mathcal{T}$ . We define their commutator by the rule

$$\{\mathcal{A}, \mathcal{B}\} = \mathcal{A} \otimes \mathcal{B} - \mathcal{B} \otimes \mathcal{A}. \quad (27.13.11)$$

It is obvious that this commutator has the desired antisymmetry property (3.7.41). Let us check the Jacobi condition. We find the results

$$\{\mathcal{A}, \{\mathcal{B}, \mathcal{C}\}\} = \mathcal{A} \otimes \mathcal{B} \otimes \mathcal{C} - \mathcal{B} \otimes \mathcal{C} \otimes \mathcal{A} - \mathcal{A} \otimes \mathcal{C} \otimes \mathcal{B} + \mathcal{C} \otimes \mathcal{B} \otimes \mathcal{A}. \quad (27.13.12)$$

$$\{\mathcal{B}, \{\mathcal{C}, \mathcal{A}\}\} = \mathcal{B} \otimes \mathcal{C} \otimes \mathcal{A} - \mathcal{C} \otimes \mathcal{A} \otimes \mathcal{B} - \mathcal{B} \otimes \mathcal{A} \otimes \mathcal{C} + \mathcal{A} \otimes \mathcal{C} \otimes \mathcal{B}. \quad (27.13.13)$$

$$\{\mathcal{C}, \{\mathcal{A}, \mathcal{B}\}\} = \mathcal{C} \otimes \mathcal{A} \otimes \mathcal{B} - \mathcal{A} \otimes \mathcal{B} \otimes \mathcal{C} - \mathcal{C} \otimes \mathcal{B} \otimes \mathcal{A} + \mathcal{B} \otimes \mathcal{A} \otimes \mathcal{C}. \quad (27.13.14)$$

Inspection shows that if (13.12) through (13.14) are summed, all the terms on the right cancel in pairs to give the desired result

$$\{\mathcal{A}, \{\mathcal{B}, \mathcal{C}\}\} + \{\mathcal{B}, \{\mathcal{C}, \mathcal{A}\}\} + \{\mathcal{C}, \{\mathcal{A}, \mathcal{B}\}\} = 0. \quad (27.13.15)$$

Thus,  $\mathcal{T}$  has been made into a Lie algebra with the Lie product taken to be the tensor product commutator (13.11).

We continue our exploration by defining adjoint operators in the standard way. Suppose  $\mathcal{C}$  is some element in  $\mathcal{T}$  and let  $\mathcal{A}$  be any element in  $\mathcal{T}$ . Then we define the adjoint operator  $\#\mathcal{C}\#$ , which maps  $\mathcal{T}$  into itself, by the rule

$$\#\mathcal{C}\#\mathcal{A} = \{\mathcal{C}, \mathcal{A}\}. \quad (27.13.16)$$

We claim that  $\#\mathcal{C}\#$  is a *derivation*. To see this, compute  $\#\mathcal{C}\#(\mathcal{A} \otimes \mathcal{B})$  to find the result

$$\begin{aligned} \#\mathcal{C}\#(\mathcal{A} \otimes \mathcal{B}) &= \{\mathcal{C}, \mathcal{A} \otimes \mathcal{B}\} = \mathcal{C} \otimes \mathcal{A} \otimes \mathcal{B} - \mathcal{A} \otimes \mathcal{B} \otimes \mathcal{C} \\ &= \mathcal{C} \otimes \mathcal{A} \otimes \mathcal{B} - \mathcal{A} \otimes \mathcal{C} \otimes \mathcal{B} + \mathcal{A} \otimes \mathcal{C} \otimes \mathcal{B} \\ &\quad - \mathcal{A} \otimes \mathcal{B} \otimes \mathcal{C} + \mathcal{A} \otimes \mathcal{C} \otimes \mathcal{B} - \mathcal{A} \otimes \mathcal{C} \otimes \mathcal{B} \\ &= \{\mathcal{C}, \mathcal{A}\} \otimes \mathcal{B} + \mathcal{A} \otimes \{\mathcal{C}, \mathcal{B}\} \\ &= (\#\mathcal{C}\#\mathcal{A}) \otimes \mathcal{B} + \mathcal{A} \otimes (\#\mathcal{C}\#\mathcal{B}). \end{aligned} \quad (27.13.17)$$

Note that (13.17) may be viewed as a rule that tells one how to compute the commutator of  $\mathcal{C}$  with a product if one already knows how to compute the commutator of  $\mathcal{C}$  with the individual elements that form the product. As a further example of such a rule, consider a commutator of the form  $\{\mathcal{A} \otimes \mathcal{B}, \mathcal{C} \otimes \mathcal{D}\}$ , which is a commutator of two products. By using adjoint operator notation and the result (13.17) we find the relation

$$\begin{aligned} \{\mathcal{A} \otimes \mathcal{B}, \mathcal{C} \otimes \mathcal{D}\} &= \#\mathcal{A} \otimes \mathcal{B}\#(\mathcal{C} \otimes \mathcal{D}) \\ &= (\#\mathcal{A} \otimes \mathcal{B}\#\mathcal{C}) \otimes \mathcal{D} + \mathcal{C} \otimes (\#\mathcal{A} \otimes \mathcal{B}\#\mathcal{D}) \\ &= \{\mathcal{A} \otimes \mathcal{B}, \mathcal{C}\} \otimes \mathcal{D} + \mathcal{C} \otimes \{\mathcal{A} \otimes \mathcal{B}, \mathcal{D}\} \\ &= -\{\mathcal{C}, \mathcal{A} \otimes \mathcal{B}\} \otimes \mathcal{D} - \mathcal{C} \otimes \{\mathcal{D}, \mathcal{A} \otimes \mathcal{B}\} \\ &= -[\#\mathcal{C}\#(\mathcal{A} \otimes \mathcal{B})] \otimes \mathcal{D} - \mathcal{C} \otimes [\#\mathcal{D}\#(\mathcal{A} \otimes \mathcal{B})] \\ &= -(\#\mathcal{C}\#\mathcal{A}) \otimes \mathcal{B} \otimes \mathcal{D} - \mathcal{A} \otimes (\#\mathcal{C}\#\mathcal{B}) \otimes \mathcal{D} \\ &\quad - \mathcal{C} \otimes (\#\mathcal{D}\#\mathcal{A}) \otimes \mathcal{B} - \mathcal{C} \otimes \mathcal{A} \otimes (\#\mathcal{D}\#\mathcal{B}). \end{aligned} \quad (27.13.18)$$

We see that (13.18) gives a rule for finding the commutator of two products if we know how to compute the commutators of the constituents.

So far we have only made use of the vector-space structure of  $L$ . Let us now also employ its Lie product structure. With the results of the last paragraph still fresh in mind, we observe that the constituents of any element in  $\mathcal{T}$  are ultimately the  $B_\alpha$ . Therefore, any commutator in  $\mathcal{T}$  can ultimately be reduced to commutators among the  $B_\alpha$ . At this point

we would like to use the Lie product structure of  $L$  to stipulate these commutators by the rule

$$\{B_\alpha, B_\beta\} = B_\alpha \otimes B_\beta - B_\beta \otimes B_\alpha = [B_\alpha, B_\beta] = \sum_\gamma c_{\alpha\beta}^\gamma B_\gamma. \quad (27.13.19)$$

The motivation for this move is that  $\mathcal{T}$ , because it contains  $L^1 = L$  as a subspace, would then also contain a copy of  $L$  as a *Lie subalgebra*. Evidently (13.19) (which may be viewed as a *reduction* rule that reduces higher-order tensor products resulting from commutators to lower-order tensor products) is compatible with the antisymmetry and Jacobi properties of the commutator because of (3.7.44) and (3.7.45). There is therefore some hope that it can be enforced *consistently*. [Here is an example question of consistency: We can enforce the condition (13.19) and then multiply or commute. Or, we may first multiply or commute and then enforce the condition. Do these two procedures give the same result?] But hope is not enough when proof is required.

The standard way to show that (13.19) can be invoked consistently is to construct the *enveloping* algebra. Let  $O_{\alpha\beta}$  denote the element

$$O_{\alpha\beta} = \{B_\alpha, B_\beta\} - [B_\alpha, B_\beta]. \quad (27.13.20)$$

Next, let  $O$  denote the set of all linear combinations of the  $O_{\alpha\beta}$ ,

$$O = \text{set of all elements in } \mathcal{T} \text{ of the form } \sum_{\alpha\alpha'} b_\alpha b'_{\alpha'} O_{\alpha\alpha'}. \quad (27.13.21)$$

Finally, let  $\mathcal{O}$  be the set of all elements formed by tensor multiplying all of  $O$  on both the left and right by *all* elements in  $\mathcal{T}$  and forming all linear combinations of such elements,

$$\begin{aligned} \mathcal{O} = & \text{set of all linear combinations of elements in } \mathcal{T} \\ & \text{of the form } \mathcal{A} \otimes O \otimes \mathcal{B} \text{ for all } \mathcal{A}, \mathcal{B} \in \mathcal{T}. \end{aligned} \quad (27.13.22)$$

Evidently  $\mathcal{O}$  is a linear vector space since, by construction, all linear combinations of elements in  $\mathcal{O}$  are again in  $\mathcal{O}$ . The set of  $\mathcal{O}$  is also an associative algebra that is *invariant* under tensor multiplication on either the left or right side by any element in  $\mathcal{T}$ ,

$$\mathcal{A}' \otimes \mathcal{O} \in \mathcal{O}, \mathcal{O} \otimes \mathcal{B}' \in \mathcal{O} \text{ for all } \mathcal{A}', \mathcal{B}' \in \mathcal{T}. \quad (27.13.23)$$

That is so because all multiplications have also already occurred in the definition (13.22). For this reason,  $\mathcal{O}$  is called the *two-sided ideal* generated by the  $O_{\alpha\beta}$ . (Recall that in Section 8.9 an ideal was defined in the Lie-algebraic context to be the set of all elements invariant under the Lie product. Here the concept is the same except that the product is tensor multiplication from the left and the right.)

Suppose we use the set  $\mathcal{O}$  to set up an equivalence relation among all elements in  $\mathcal{T}$ . We will say that two elements  $\mathcal{A}$  and  $\mathcal{A}'$  in  $\mathcal{T}$  are equivalent if their difference (recall that both  $\mathcal{A}$  and  $\mathcal{A}'$  are vectors, and therefore can be added and subtracted) is in the set  $\mathcal{O}$ ,

$$\mathcal{A} \sim \mathcal{A}' \Leftrightarrow (\mathcal{A} - \mathcal{A}') \in \mathcal{O}. \quad (27.13.24)$$

It is easily verified, by a discussion analogous to that in Section 8.9, that (13.24) does indeed define an equivalence relation. Next, this equivalence relation can be used to set up

equivalence classes. By the standard arguments, when this is done, any element in  $\mathcal{O}$  will be in the equivalence class  $\{0\}$  that contains the zero vector in  $\mathcal{T}$ ,

$$\mathcal{B} \in \mathcal{O} \Leftrightarrow \{\mathcal{B}\} = \{0\}. \quad (27.13.25)$$

[For the analogous Lie-algebraic case, see (8.9.3).] Finally, let  $\mathcal{E}$  be the quotient space of  $\mathcal{T}$  with respect to  $\mathcal{O}$ ,

$$\mathcal{E} = \mathcal{T}/\mathcal{O}. \quad (27.13.26)$$

Since  $\mathcal{T}$  is an associative algebra and  $\mathcal{O}$  is an ideal in  $\mathcal{T}$ , the quotient space  $\mathcal{E}$  will also be an associative algebra. It is called the *enveloping algebra* of  $L$ . (It is also often called the *universal* enveloping algebra because it can be shown to be unique up to an isomorphism.)

We note that because of (13.25), all elements in  $\mathcal{O}$ , (that is, all elements in  $\mathcal{T}$  that contain  $O_{\alpha\beta}$ ) are automatically replaced by 0 in  $\mathcal{E}$ . This is equivalent to enforcing the condition (13.19) in  $\mathcal{E}$ . And, because  $\mathcal{E}$  is an associative algebra, we have verified that this condition can be enforced consistently.

Soon we will use the enveloping algebra to construct Casimir operators. To do so, it is useful to first explore further the property of adjoint operators  $\#\mathcal{C}\#$ . Since  $\#\mathcal{C}\#$  is a derivation, recall (13.17), the operator  $\exp(\epsilon\#\mathcal{C}\#)$  is an isomorphism for tensor multiplication,

$$[\exp(\epsilon\#\mathcal{C}\#)](\mathcal{A} \otimes \mathcal{B}) = \{[\exp(\epsilon\#\mathcal{C}\#)]\mathcal{A}\} \otimes \{[\exp(\epsilon\#\mathcal{C}\#)]\mathcal{B}\}. \quad (27.13.27)$$

Indeed, if  $\mathcal{F}$  is any element of  $\mathcal{T}$  composed of tensor products of the  $B_\alpha$ , we have the result

$$\begin{aligned} & [\exp(\epsilon\#\mathcal{C}\#)]\mathcal{F}(B_{\alpha_1}, B_{\alpha_2}, B_{\alpha_3} \dots) = \\ & \mathcal{F}([\exp(\epsilon\#\mathcal{C}\#)]B_{\alpha_1}, [\exp(\epsilon\#\mathcal{C}\#)]B_{\alpha_2}, [\exp(\epsilon\#\mathcal{C}\#)]B_{\alpha_3} \dots). \end{aligned} \quad (27.13.28)$$

See Section 8.2 for the standard arguments justifying this result.

Now suppose  $\mathcal{C} = C$  where  $C$  is some element in  $L$  as in (11.20). Then we have the results

$$\#C\#B_\alpha = \{C, B_\alpha\} = [C, B_\alpha] =: C : B_\alpha. \quad (27.13.29)$$

Here we have used (13.19). It follows from (11.20), (11.28), and (13.29) that in this case there is the relation

$$[\exp(\epsilon\#\mathcal{C}\#)]B_\alpha = \exp(\epsilon : C : )B_\alpha = U_{\alpha\beta}B_\beta, \quad (27.13.30)$$

where we have again used the summation convention. Correspondingly, in this case (13.28) can be rewritten in the form

$$[\exp(\epsilon\#\mathcal{C}\#)]\mathcal{F}(B_{\alpha_1}, B_{\alpha_2}, B_{\alpha_3} \dots) = \mathcal{F}(U_{\alpha_1\beta_1}B_{\beta_1}, U_{\alpha_2\beta_2}B_{\beta_2}, U_{\alpha_3\beta_3}B_{\beta_3} \dots). \quad (27.13.31)$$

We are now ready to discuss Casimir operators. In analogy with (11.35) we now define the quadratic Casimir operator  $\mathcal{C}_2$  to be the quantity

$$\mathcal{C}_2 = \sum_{\alpha\alpha'} g_R^{\alpha\alpha'} B_\alpha \otimes B_{\alpha'}. \quad (27.13.32)$$

Note that in this context  $\mathcal{C}_2$  is not first of all an “operator”, but rather is an element in the enveloping algebra  $\mathcal{E}$ . In analogy with (11.38), let us see how it transforms. We find the result

$$\begin{aligned}\mathcal{C}_2^{\text{tr}} &= \exp(\epsilon \# \mathcal{C} \#) \mathcal{C}_2 = \sum_{\alpha\alpha'} g_R^{\alpha\alpha'} (U_{\alpha\beta} B_\beta) \otimes (U_{\alpha'\beta'} B_{\beta'}) \\ &= \sum_{\alpha\alpha'} g_R^{\alpha\alpha'} U_{\alpha\beta} U_{\alpha'\beta'} B_\beta \otimes B_{\beta'} \\ &= \sum_{\beta\beta'} (\sum_{\alpha\alpha'} g_R^{\alpha\alpha'} U_{\alpha\beta} U_{\alpha'\beta'}) B_\beta \otimes B_{\beta'} \\ &= \sum_{\beta\beta'} g_R^{\beta\beta'} B_\beta \otimes B_{\beta'} = \mathcal{C}_2.\end{aligned}\tag{27.13.33}$$

Here we have used (13.31) and (11.39). We see that  $\mathcal{C}_2$  is again invariant. Also, the infinitesimal version of (13.33) with  $\mathcal{C} = C = B_{\alpha''}$  gives the result

$$\# B_{\alpha''} \# \mathcal{C}_2 = \{B_{\alpha''}, \mathcal{C}_2\} = 0.\tag{27.13.34}$$

We see that  $\mathcal{C}_2$  commutes with all the elements in  $L$ . Moreover, since everything in the enveloping algebra is constructed from elements in  $L$ , it follows that  $\mathcal{C}_2$  commutes with all the elements of the enveloping algebra,

$$\{\mathcal{C}_2, \mathcal{E}\} = 0.\tag{27.13.35}$$

At this point we pause to note that we might use (13.19) to rearrange [in analogy to (11.46) and (11.47)] the terms in  $\mathcal{C}_2$  as given by (13.32) to get an expression for  $\mathcal{C}_2$  analogous to (11.50). What would happen if we then compute  $\{B_{\alpha''}, \mathcal{C}_2\}$  using the rearranged  $\mathcal{C}_2$ ? According to our previous discussion about consistency,  $\mathcal{O}$  is an ideal thereby guaranteeing that the quotient space  $\mathcal{T}/\mathcal{O} = \mathcal{E}$  is an associative algebra. Therefore the result should be (and is indeed) the same.

The construction of higher-order Casimir operators proceeds in a similar fashion. For example, the analog of (11.86) is

$$\mathcal{C}_3 =_3 g_R^{\alpha\alpha'\alpha''} B_\alpha \otimes B_{\alpha'} \otimes B_{\alpha''}.\tag{27.13.36}$$

Again, it is an element in the enveloping algebra. It too commutes with all the elements in  $L$ , and therefore also with all of  $\mathcal{E}$ . In general, in the context of the present discussion, we may define a Casimir operator to be any element in the enveloping algebra that commutes with *all* the elements in the enveloping algebra. Put another way, the Casimir operators form the *center* of the enveloping algebra.

## Exercises

**27.13.1.** From (13.4) it follows that  $\mathcal{T}$ , the tensor algebra of  $L$ , can be expressed in the form

$$\mathcal{T} = \mathcal{T}^0 \oplus \mathcal{T}^1 \oplus \mathcal{T}^2 \oplus \mathcal{T}^3 \oplus \cdots,\tag{27.13.37}$$

where

$$\mathcal{T}^n = L^n. \quad (27.13.38)$$

Suppose that  $L$  has dimension  $k$ . Show that each subspace  $\mathcal{T}^n$  then has dimension

$$\dim \mathcal{T}^n = k^n. \quad (27.13.39)$$

The algebra  $\mathcal{E}$ , the enveloping algebra of  $L$ , can also be decomposed in the form

$$\mathcal{E} = \mathcal{E}^0 \oplus \mathcal{E}^1 \oplus \mathcal{E}^2 \oplus \mathcal{E}^3 \oplus \dots, \quad (27.13.40)$$

where each subspace  $\mathcal{E}^n$  of “degree”  $n$  is spanned by the tensor products of  $n$  basis elements  $(B_{\alpha_1} \otimes B_{\alpha_2} \otimes B_{\alpha_3} \otimes \dots \otimes B_{\alpha_n})$ . However, in the case of  $\mathcal{E}$ , the relation (13.19) can be used to rearrange the basis elements in each  $\mathcal{E}^n$  so that the subscripts have a definite standard ordering. For example, we may arrange them in ascending order,

$$\begin{aligned} & \text{rearranged } (B_{\alpha_1} \otimes B_{\alpha_2} \otimes B_{\alpha_3} \otimes \dots \otimes B_{\alpha_n}) = (B_{\beta_1} \otimes B_{\beta_2} \otimes B_{\beta_3} \otimes \dots \otimes B_{\beta_n}) \\ & \text{with } \beta_1 \leq \beta_2 \leq \beta_3 \leq \dots \leq \beta_n. \end{aligned} \quad (27.13.41)$$

In the rearrangement process various terms of lower degree may be generated, but they simply feed down to  $\mathcal{E}^{n-1}$ , etc. It follows that the various terms in some standard ordering, say that shown on the right side of (13.41), span  $\mathcal{E}^n$ . It can be shown that they are linearly independent as well, and therefore form a *basis* for  $\mathcal{E}^n$ . This basis is called the *Poincaré-Birkhoff-Witt* basis. Show that the dimension of  $\mathcal{E}^n$  is given by the relation

$$\dim \mathcal{E}^n = N(n, k) \quad (27.13.42)$$

with  $N(n, k)$  given by (7.3.40). Hint: Once a standard ordering has been established, the counting of basis elements is the same as counting monomials.

**27.13.2.** Suppose  $R$  is a realization of some Lie algebra  $L$ . Thus, if the  $B_\alpha$  form a basis of  $L$ , there are associated matrices  $\hat{B}_\alpha$  in the realization  $R$ . Let  $\mathcal{R}$  be the *linear* map that sends the  $B_\alpha$  to the  $\hat{B}_\alpha$ ,

$$\mathcal{R}(B_\alpha) = \hat{B}_\alpha. \quad (27.13.43)$$

(Note that since both  $L$  and the set of  $m \times m$  matrices are vector spaces, it makes sense to talk about a linear map that sends one into the other.) Then, by the definition of a realization, we have the relation

$$\begin{aligned} \mathcal{R}([B_\alpha, B_\beta]) &= \mathcal{R}\left(\sum_\gamma c_{\alpha\beta}^\gamma B_\gamma\right) = \sum_\gamma c_{\alpha\beta}^\gamma \hat{B}_\gamma = \{\hat{B}_\alpha, \hat{B}_\beta\} \\ &= \{\mathcal{R}(B_\alpha), \mathcal{R}(B_\beta)\}. \end{aligned} \quad (27.13.44)$$

Next, let us extend the definition of  $\mathcal{R}$  to have it act on any basis element in the tensor algebra  $\mathcal{T}$  by the rule

$$\begin{aligned} \mathcal{R}(B_{\alpha_1} \otimes B_{\alpha_2} \otimes B_{\alpha_3} \dots) &= \mathcal{R}(B_{\alpha_1})\mathcal{R}(B_{\alpha_2})\mathcal{R}(B_{\alpha_3})\dots \\ &= \hat{B}_{\alpha_1}\hat{B}_{\alpha_2}\hat{B}_{\alpha_3}\dots. \end{aligned} \quad (27.13.45)$$

Show that  $\mathcal{R}$  sends any element of the two-sided ideal  $\mathcal{O}$  to the zero matrix,

$$\mathcal{R}(O_{\alpha\beta}) = 0, \quad \mathcal{R}(O) = 0, \quad \mathcal{R}(\mathcal{O}) = 0. \quad (27.13.46)$$

Thus, the image of any element in  $\mathcal{T}$  under the action of  $\mathcal{R}$  depends only on the equivalence class to which the element belongs, and we may equally well view  $\mathcal{R}$  as acting on  $\mathcal{T}/\mathcal{O} = \mathcal{E}$ . Show that  $\mathcal{R}$  sends the Casimir operator (13.32) defined in the enveloping algebra context to the Casimir operator (11.35) defined in the representation context. Show that in general anything that it discovered about Casimir operators in the enveloping algebra context is immediately transferable to the realization context, and vice versa.

## 27.14 The Symplectic Lie Algebras $sp(8)$ and Beyond

The previous sections in this chapter have treated the cases of  $sp(2)$ ,  $sp(4)$ , and  $sp(6)$ . The Lie algebraic structure of all the  $sp(2n)$ , for example root vectors and fundamental weight vectors, is also known. In particular, for  $sp(2n)$ , a representation is characterized by  $n$  non-negative integers  $k_1, k_2, \dots, k_n$  and may be denoted by the symbols  $\Gamma(k_1, k_2, \dots, k_n)$ . Homogeneous polynomials of degree  $\ell$  in the  $2n$  components of  $z$  again carry representations of  $sp(2n)$ , and for these representations there is the result

$$k_1 = \ell, \quad (27.14.1)$$

$$k_j = 0 \text{ for } j = 2, 3, \dots, n. \quad (27.14.2)$$

There is also an analogous Clebsch-Gordon series result of the form (9.1) where all entries in  $\Gamma(k_1, k_2, \dots, k_n)$  are zero save for the first two,

$$\begin{aligned} & \Gamma(\ell, 0, 0, \dots) \otimes \Gamma(1, 0, 0, \dots) \\ &= \Gamma(\ell + 1, 0, 0, \dots) \oplus \Gamma(\ell - 1, 1, 0, \dots) \\ & \quad \oplus \Gamma(\ell - 1, 0, 0, \dots). \end{aligned} \quad (27.14.3)$$

Thus, the symplectic classification of all analytic vector fields in any (even) dimension is in principle known. Moreover, any  $\mathcal{L}_{\mathbf{g}^\ell}$  with  $\ell \geq 1$  has the unique decomposition

$$\mathcal{L}_{\mathbf{g}^\ell} = \mathcal{H}^{\ell+1,0,0,\dots} + \mathcal{G}^{\ell-1,1,0,\dots} + \mathcal{G}^{\ell-1,0,0,\dots}. \quad (27.14.4)$$

Here  $\mathcal{H}^{\ell+1,0,0,\dots}$  is a Hamiltonian vector field that carries the representation  $\Gamma(\ell + 1, 0, 0, \dots)$ , and is of the form :  $h_{\ell+1} \cdot$ . The quantities  $\mathcal{G}^{\ell-1,1,0,\dots}$  and  $\mathcal{G}^{\ell-1,0,0,\dots}$  are non-Hamiltonian vector fields that carry the representations  $\Gamma(\ell - 1, 1, 0, \dots)$  and  $\Gamma(\ell - 1, 0, 0, \dots)$ , respectively.

## Exercises

**27.14.1.** Show that any  $2n \times 2n$  matrix that commutes with all  $sp(2n)$  matrices (in the fundamental representation) must be a multiple of  $I$ . Show that any  $2n \times 2n$  matrix that commutes with all  $Sp(2n)$  matrices (in the fundamental representation) must be a multiple of  $I$ .

## 27.15 Momentum Maps, Noether's Theorem, and Casimirs

The previous sections, among other things, have shown how to decompose an analytic vector field  $\mathcal{L}_g$  into Hamiltonian and non-Hamiltonian parts. Here we address a somewhat more restricted question. Suppose we are given a vector field  $\mathcal{L}_g$ , and also know that it came from some Hamiltonian  $h$  so that there is in principle the relation

$$\mathcal{L}_g =: h : . \quad (27.15.1)$$

That is,  $\mathcal{L}_g$  is a Hamiltonian vector field. We then say that there is a *momentum map*  $\mu$  that sends  $\mathcal{L}_g$  to  $h$ ,

$$\mu(\mathcal{L}_g) = h. \quad (27.15.2)$$

Here the name *momentum* is associated with the fact that in some simple examples the resulting  $h$  is some kind of momentum such as linear or angular momentum.

In this section we will develop/review what is required for  $\mathcal{L}_g$  to be Hamiltonian, and then see how to determine  $h$  in terms of  $g$ . We will also see how momentum maps are related to integrals of motion and, when there are several integrals of motion, how to construct from them integrals of motion that are in involution.

### 27.15.1 Momentum Maps, Noether's Theorem, and Conservation Laws

Why might one be interested in momentum maps? Given a vector field  $\mathcal{L}_g$ , we may define a family of maps  $\mathcal{M}(\tau)$ , not to be confused with momentum maps, by the rule

$$\mathcal{M}(\tau) = \exp(\tau \mathcal{L}_g). \quad (27.15.3)$$

[For a discussion of some of the properties of general Lie operators (general vector fields) and their associated Lie transformations, see Exercises 5.3.10 and 5.4.14.] The maps  $\mathcal{M}(\tau)$  send phase space into itself according to the relation

$$\bar{z}(\tau) = \mathcal{M}(\tau)z, \quad (27.15.4)$$

and they evidently form a one-parameter group. Now suppose the motion of some system is governed by some Hamiltonian  $H(z, t)$  and suppose that this Hamiltonian is invariant under the action of  $\mathcal{M}(\tau)$ ,

$$\mathcal{M}(\tau)H(z, t) = H(z, t). \quad (27.15.5)$$

From this invariance/symmetry relation and (15.3) we conclude that

$$\mathcal{L}_g H(z, t) = 0. \quad (27.15.6)$$

But, if (15.1) holds, (15.6) can be rewritten as

$$\mathcal{L}_g H(z, t) =: h : H = [h, H] = 0, \quad (27.15.7)$$

and we see that  $h$  is an integral of motion. Thus, the existence of symmetry and a momentum map implies the existence of an integral of motion (a *conserved quantity* or *conservation law*) and vice versa.<sup>6</sup>

How can one test a vector field to see if it is Hamiltonian and therefore there is a momentum map? We have already seen that the maps  $\mathcal{M}(\tau)$  form a one-parameter group. Suppose we now require that these maps be symplectic for all  $\tau$ . That is, we require that

$$[\bar{z}_a(\tau), \bar{z}_b(\tau)] = J_{ab} \text{ for all } \tau. \quad (27.15.8)$$

Put another way, we require that each transformation  $\mathcal{M}(\tau)$  *preserve* the symplectic structure of phase space. For small  $\tau$  we have the result

$$\mathcal{M}(\tau) = \mathcal{I} + \tau \mathcal{L}_{\mathbf{g}} + O(\tau^2). \quad (27.15.9)$$

It follows that

$$\bar{z}(\tau)_a = \mathcal{M}(\tau)z_a = z_a + \tau \mathcal{L}_{\mathbf{g}} z_a + O(\tau)^2 = z_a + \tau g_a + O(\tau)^2. \quad (27.15.10)$$

Here we have used the result

$$\mathcal{L}_{\mathbf{g}} z_a = g_a. \quad (27.15.11)$$

Upon employing (15.10) we find the result

$$\begin{aligned} [\bar{z}_a(\tau), \bar{z}_b(\tau)] &= [z_a, z_b] + \tau \{[z_a, g_b] + [g_a, z_b]\} + O(\tau^2) \\ &= J_{ab} + \tau \{[z_a, g_b] + [g_a, z_b]\} + O(\tau^2). \end{aligned} \quad (27.15.12)$$

Enforcing (15.8) and equating powers of  $\tau$  give the result

$$[z_a, g_b] + [g_a, z_b] = 0. \quad (27.15.13)$$

We have seen this condition before in Lemma 6.2 of Section 7.6. There we learned that (15.13) implies and is implied by the relation

$$g_a =: h : z_a, \quad (27.15.14)$$

which, in view of (15.11), is equivalent to the relation (15.1). Thus, (15.13) is a necessary and sufficient condition for  $\mathcal{L}_{\mathbf{g}}$  to be a Hamiltonian vector field.

We also learned how to construct  $h$ . It is given, up to an additive constant, by the relation

$$h(z) = - \int_P^z \sum_{cd} g_c(z') J_{cd} dz'_d \quad (27.15.15)$$

where  $P$  is *any* path ending at the point  $z$ . A convenient path is that which connects the origin and  $z$  by a straight line,

$$z'(\lambda) = \lambda z \text{ with } \lambda \in [0, 1]. \quad (27.15.16)$$

---

<sup>6</sup>Observe that the assumed symmetry described by (15.5) is a *continuous* symmetry. The condition (15.5) is supposed to hold over a continuous range of  $\tau$ . We also remark that *Emmy Noether* (1882-1935) was the first to explore in detail the connection between continuous symmetries and conservation laws. For this reason results of this subsection, or analogous results, are often called *Noether's theorem*.

For this path we find the explicit result

$$h(z) = - \sum_{cd} z_d J_{cd} \int_0^1 d\lambda g_c(\lambda z). \quad (27.15.17)$$

This relation specifies  $h$  in terms of the  $g_c$ , and hence in terms of  $\mathcal{L}\mathbf{g}$ . It therefore provides the map  $\mu$  described in (15.2). Note that this specification is equivalent to the *definition* (10.20) of the Hamiltonian part of a general homogeneous vector field.

At this point we remark that in some circumstances one initially has a transformation or a group of transformations whose action is only on the position coordinates  $q$ . Such transformations can always be extended to symplectic actions on phase space, and thus in this case the existence of Hamiltonian vector fields is guaranteed. Recall Exercise 6.5.2.

As a first example of the momentum map process, consider (for a 6-dimensional phase space) the case where

$$\begin{aligned} g_1(z) &= 1, \\ g_a(z) &= 0, \quad a \neq 1, \end{aligned} \quad (27.15.18)$$

so that

$$\mathcal{L}\mathbf{g} = \partial/\partial z_1 = \partial/\partial q_1. \quad (27.15.19)$$

It is easily verified that

$$(\mathcal{L}\mathbf{g})^n z = 0 \text{ for } n \geq 2, \quad (27.15.20)$$

from which it follows that

$$\bar{z}_a(\tau) = \exp(\tau \mathcal{L}\mathbf{g}) z_a = z_a + \tau \delta_{a,1}. \quad (27.15.21)$$

That is,  $\mathcal{L}\mathbf{g}$  generates translations in phase space along the  $z_1 = q_1$  axis. Evidently, the  $g_a$  specified by (15.18) satisfy (15.13) so that  $\mathcal{L}\mathbf{g}$  is a Hamiltonian vector field. Finally, the formula (15.17) for  $h$  is easily evaluated to give the result

$$h(z) = -z_4 = -p_1, \quad (27.15.22)$$

the negative of the first component of the linear momentum. Thus, invariance under translation implies the conservation of linear momentum, and vice versa.

As a second example, suppose that

$$\begin{aligned} \mathcal{L}\mathbf{g} &= z_1 \partial/\partial z_2 - z_2 \partial/\partial z_1 + z_4 \partial/\partial z_5 - z_5 \partial/\partial z_4 \\ &= (q_1 \partial/\partial q_2 - q_2 \partial/\partial q_1) + (p_1 \partial/\partial p_2 - p_2 \partial/\partial p_1). \end{aligned} \quad (27.15.23)$$

For this example the nonzero  $g_a$  are given by the relations

$$\begin{aligned} g_1(z) &= -z_2, \\ g_2(z) &= z_1, \\ g_4(z) &= -z_5, \\ g_5(z) &= z_4. \end{aligned} \quad (27.15.24)$$

With the  $g_a(z)$  in view, it is easily checked that (15.13) holds. It therefore makes sense to continue on to compute  $h$ . The integrals appearing on the right side of (15.17) are easy to evaluate because the  $h_a(z)$  are homogeneous of degree one. Doing so gives the results

$$\int_0^1 d\lambda g_c(\lambda z) = \int_0^1 d\lambda \lambda g_c(z) = (1/2)g_c(z). \quad (27.15.25)$$

Finally, employing (15.25) in (15.17) gives the result

$$h(z) = -(z_1 z_5 - z_2 z_4) = -(q_1 p_2 - q_2 p_1). \quad (27.15.26)$$

From the second line of (15.23) we recognize  $\mathcal{L}_g$  as the generator of simultaneous rotations in the  $q_1, q_2$  and  $p_1, p_2$  planes, and from (15.26) we see that  $h$  is the negative of the third component of the angular momentum. Thus, invariance under rotation implies the conservation of angular momentum, and vice versa.

### 27.15.2 Use of Casimirs

In general, if a Hamiltonian is invariant under the action of some  $n$ -dimensional group that preserves symplectic structure, there will be  $n$  associated integrals of motion. These integrals, however, need not be in mutual involution. Think, for example, of the components of angular momentum. Their Poisson bracket Lie algebra provides a realization of  $su(2)$  [or, equivalently,  $so(3, \mathbb{R})$ ], and they are therefore not in involution. Generally integrals will be in involution if, and only if, the corresponding Lie operators commute. Recall (5.3.14). In this subsection we will explore briefly how Casimirs can sometimes be used to construct integrals that are in involution.

Suppose that a Hamiltonian  $H$  is indeed invariant under the action of some  $n$ -dimensional group that preserves symplectic structure, and therefore there are  $n$  associated integrals of motion. Call these integrals  $h^\alpha$ . Let  $C$  be any function of these integrals,

$$C = C(h^1, h^2, \dots, h^n). \quad (27.15.27)$$

Then we know from Exercise 5.2.4 that  $C$  will also be an integral of motion. Our goal will be to construct a  $C$  such that it is functionally independent of any one of the  $h^\alpha$ , but is also in involution with any of them.

We know that the  $h^\alpha$  will form a Lie algebra with the Poisson bracket serving as a Lie product. That is, there will be relations of the form

$$[h^\alpha, h^\beta] = \sum_\gamma c_{\alpha\beta}^\gamma h^\gamma. \quad (27.15.28)$$

(Note that these relations are consistent with Poisson's theorem that states that the Poisson bracket of two integrals of motion is again an integral of motion. See Exercise 5.2.3.) The structure constants  $c_{\alpha\beta}^\gamma$  can be used to construct a Killing metric tensor  $g_{\alpha\alpha'}^K$ , and from  $g_{\alpha\alpha'}^K$ , assuming it is invertible, we can construct  $g_K^{\alpha\alpha'}$ . With  $g_K^{\alpha\alpha'}$  in hand, we can define the function  $C_2$  by the rule

$$C_2 = \sum_{\alpha\alpha'} g_K^{\alpha\alpha'} h^\alpha h^{\alpha'}. \quad (27.15.29)$$

In analogy to the calculations carried out for the quadratic Casimir operator  $C_2$  in Section 21.11.1, it is easily verified that  $C_2$  is in involution with the  $h^\alpha$ . It can happen that  $C_2$  vanishes identically. See Exercise 21.11.7. But, if  $C_2$  does not vanish, we have found two functionally independent integrals in involution, namely  $C_2$  and any one of the  $h^\alpha$ .

At this point we might go on to find additional integrals  $C_3$ , etc. constructed in analogy to the higher-order Casimir operators. They will also be in involution and, if these integrals are nonvanishing and functionally independent, we will have found additional nontrivial integrals. In general, assuming the Lie algebra in question is simple, we may hope to find as many integrals in involution as there are labels necessary to specify a representation for the Lie algebra and to specify vectors within a representation.

Let us apply this construction to the rotation group example of the previous subsection. Suppose that  $H$  has the three integrals

$$h^1 = q_2 p_3 - q_3 p_2, \quad (27.15.30)$$

$$h^2 = q_3 p_1 - q_1 p_3, \quad (27.15.31)$$

$$h^3 = q_1 p_2 - q_2 p_1. \quad (27.15.32)$$

They form an  $su(2)$  Lie algebra,

$$[h^1, h^2] = h^3, \text{ etc.} \quad (27.15.33)$$

The metric tensor for  $su(2)$  is given in Exercise 21.12.2. It follows, after a convenient renormalization, that we may take for  $C_2$  the quantity

$$C_2 = (h^1)^2 + (h^2)^2 + (h^3)^2, \quad (27.15.34)$$

which is the square of the angular momentum.

## Exercises

**27.15.1.** Verify that (15.5) implies (15.6), and conversely.

**27.15.2.** Suppose (15.13) holds so that  $\mathcal{L}_g$  is a Hamiltonian vector field. Suppose also that  $\mathcal{L}_g$  can be decomposed into a sum of homogeneous parts as in (3.1). (This will certainly be possible if  $\mathcal{L}_g$  is analytic.) Show that then (10.20) and (15.17) are equivalent.

**27.15.3.** Verify that use of (15.18) in (15.17) does yield (15.22). Verify that the  $g_a$  given by (15.18) and the  $h$  given by (15.22) do indeed satisfy (15.1).

**27.15.4.** Verify, by evaluating the effect of  $\exp(\tau \mathcal{L}_g)$  on phase space, that  $\mathcal{L}_g$  as given by (15.23) does indeed generate simultaneous rotations in the  $q_1, q_2$  and  $p_1, p_2$  planes. Verify (15.24) through (15.26). Verify that the  $g_a$  given by (15.24) and the  $h$  given by (15.26) do indeed satisfy (15.1).

**27.15.5.** Verify the relations (15.33). Verify that  $C_2$  and any one of the  $h^\alpha$  are in involution.



# Bibliography

## Group and Lie Algebra Theory

See also the Lie Group Theory sections of the Bibliographies for Chapters 3 and 5.

- [1] A.O. Barut and R. Raczka, *Theory of Group Representations and Applications*, World Scientific (1986).
- [2] N. Bourbaki, *Lie Groups and Lie Algebras, Elements of Mathematics, Chapters 1-3*, Springer-Verlag (1989).
- [3] T. Brocker and T.T. Dieck, *Representations of Compact Lie Groups*, Springer-Verlag (1985).
- [4] R. Cahn, *Semi-Simple Lie Algebras and their Representations*, Dover (2006).
- [5] J-Q. Chen, *Group Representation Theory for Physicists*, World Scientific (1989).
- [6] A.R. Edmonds, *Angular Momentum in Quantum Mechanics*, Princeton University Press(1957).
- [7] M.E. Rose, *Elementary Theory of Angular Momentum*, John Wiley and Sons (1957).
- [8] L. Biedenharn and J. Louck, *Angular Momentum in Quantum Physics: Theory and Application*, Cambridge University Press (2009).
- [9] W. Fulton and J. Harris, *Representation Theory, A First Course*, Corrected third printing, Springer-Verlag (1996).
- [10] H. Georgi, *Lie Algebras in Particle Physics*, Perseus Books (1999).
- [11] R. Goodman and N.R. Wallach, *Representations and Invariants of the Classical Groups*, Cambridge University Press (1998).
- [12] R. Goodman and N.R. Wallach, *Symmetry, Representations, and Invariants*, Springer (2009).
- [13] J.E. Humphreys, *Introduction to Lie Algebras and Representation Theory*, Springer-Verlag (1972).
- [14] J.E. Humphreys, *Representations of Semisimple Lie Algebras in the BBG Category  $\mathcal{O}$* , American Mathematical Society (2008).

- [15] N. Jacobson, *Lie Algebras*, Interscience Publishers (1962).
- [16] A.W. Knapp, *Lie Groups Beyond an Introduction*, Second Edition, Birkhäuser (2005).
- [17] A.W. Knapp, *Representation Theory of Semisimple Groups, An Overview Based on Examples*, Princeton (1986).
- [18] D.H. Sattinger and O.L. Weaver, *Lie Groups and Algebras with Applications to Physics, Geometry, and Mechanics*, Springer-Verlag (1986).
- [19] V.S. Varadarajan, *Lie Groups, Lie Algebras, and Their Representations*, Springer-Verlag (1984).
- [20] J.E. Campbell, *Introductory Treatise on Lie's Theory of Finite Continuous Transformation Groups*, Chelsea Publishing (1903 and 1966).
- [21] H. Weyl, *The Classical Groups: Their Invariants and Representations*, Princeton University Press (1946).
- [22] B.G. Wybourne, *Classical Groups for Physicists*, John Wiley and Sons (1974).
- [23] W. Rossmann, *Lie Groups, An Introduction Through Linear Groups*, Oxford (2002).
- [24] A. Baker, *Matrix Groups: An Introduction to Lie Group Theory*, Springer (2006).
- [25] 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).
- [26] P. Di Francesco, P. Mathieu, and D. Sénéchal, *Conformal Field Theory*, Springer (1997). See Chapter 13.
- [27] D. Montgomery and L. Zippin, *Topological Transformation Groups*, Interscience (1955).
- [28] P. Tondeur, *Introduction to Lie Groups and Transformation Groups*, Lecture Notes in Mathematics 7, Springer-Verlag (1965).
- [29] P. Cvitanović, *Group Theory, Birdtracks, Lie's, and Exceptional Groups*, Princeton (2008).
- [30] J.-P. Serre, *Lie Algebras and Lie Groups: 1964 Lectures Given at Harvard University*, Springer (2005).
- [31] J.-P. Serre and G. Jones, *Complex Semisimple Lie Algebras*, Springer (2001).
- [32] J. Dieudonné, *Special Functions and Linear Representations of Lie Groups*, American Mathematical Society (1980).
- [33] J. Dieudonné, *Treatise on Analysis*, Volumes 10-IV and 10-V in the series Pure and Applied Mathematics, Academic Press (1974 and 1977).

- [34] Zhong-Qi Ma, *Group Theory for Physicists*, World Scientific (2007).
- [35] F. Iachello, *Lie Algebras and Applications*, 2nd edition, Springer (2015).
- [36] A. Zee, *Group Theory in a Nutshell for Physicists*, Princeton University Press (2016).
- [37] B. Hall, *Lie Groups, Lie Algebras, and Representations: an Elementary Introduction*, 2nd edition, Springer (2015).
- [38] A. Baker, *Matrix Groups: An Introduction to Lie Group Theory*, Springer (2002).
- [39] K. Erdmann and M. Wildon, *Introduction to Lie Algebras*, Springer (2011).
- [40] H. Pollatsek, *Lie Groups: A Problem-Oriented Introduction via Matrix Groups*, Mathematical Association of America (2009).
- [41] Robert Hermann has published several books on various applications of Lie theory as well as others that provide commentary on various papers of Lie and related work of other mathematicians of his time period. For a list of some of them, see the Wikipedia Web site [http://en.wikipedia.org/wiki/Robert\\_Hermann\\_\(mathematician\)](http://en.wikipedia.org/wiki/Robert_Hermann_(mathematician)).

#### Overview and History of the Theory of Lie Algebras and Lie Groups

- [42] R. Howe, “A Century of Lie Theory,” *American Mathematical Society Centennial Proceedings, Vol. 2: Mathematics into the Twenty-first Century*, Providence, R.I., American Mathematical Society, pp. 101-320 (1992).
- [43] T. Hawkins, *Emergence of the Theory of Lie Groups: An Essay in the History of Mathematics 1869-1926*, Springer (2000). See also the review of Hawkins’ book by D. Rowe published in *Notices of the American Mathematical Society* **50**, p. 668 (June/July 2003).
- [44] A. Borel, *Essays in the History of Lie Groups and Algebraic Groups*, American Mathematical Society (2001).
- [45] A. Stubhaug, *The Mathematician Sophus Lie*, Springer-Verlag (2002).
- [46] W. Schmid, “Poincaré and Lie Groups”, *Bulletin Amer. Math. Soc. (N.S.)* **6** pp. 175-186 (1982). This article can also be found in *The Mathematical Heritage of Henri Poincaré, Proceedings of Symposia in Pure Mathematics of the American Mathematical Society* **39**, Parts 1 and 2, F. Browder, Edit., American Mathematical Society (1983).

#### Decomposing Vector Fields

- [47] D. Lewis and J.E. Marsden, “The Hamiltonian-Dissipative Decomposition of Normal Forms of Vector Fields”, *Proc. of the Conf. on Bifurcation Theory and its Num. An.*, pp. 51-78, Xi’an Jiaotong Univ. Press (1989).

#### Enveloping Algebras and Casimir Operators

See also the references above on Group and Lie Algebra Theory.

- [48] J. Dixmier, *Enveloping Algebras*, North-Holland Publishing (1977).
- [49] R. Campoamor-Stursberg, “A new matrix method for the Casimir operators of the Lie algebras  $wsp(N, R)$  and  $Isp(2N, R)$ ”, *J. Phys. A: Math. Gen.* **38**, p. 4187 (2005).
- [50] K. Maurischat, “Casimir operators for symplectic groups”, arXiv:1011.4777v1 [math.NT] (2010).

#### Combinatorics of $Sp(2n, \mathbb{C})$ Representations

- [51] Sheila Sundaram, *On the Combinatorics of Representations of  $Sp(2n, C)$* , Ph. D. thesis, Massachusetts Institute of Technology (1986).

#### Computer Programs

The programs listed below cover all the simple Lie groups, and are extremely useful. They were employed, for example, to check Table 8.1, to compute weights and multiplicities for various representations, and to produce Clebsch-Gordan series such as (9.1).

- [52] W.G. McKay, J. Patera, and D.W. Rand, *SimpLie User’s Manual*, Macintosh Software for Representations of Simple Lie Algebras, Centre de recherches mathématiques, Université de Montréal, C.P. 6128-A, Montréal, QC H3C 3J7, Canada (1990).
- [53] M.A.A. van Leeuwen, A.M. Cohen, and B. Lisser, *Lie Manual describing LIE Version 2.1*. LIE is a software package for Lie group theoretical computations developed by the Computer Algebra Group of CWI (Centrum voor Wiskunde en Informatica, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands). For further information see the Web site <http://www-math.univ-poitiers.fr/~maavl/LiE/>.
- [54] The Computational Algebra Group within the University of Sydney School of Mathematics and Statistics has produced the program CAYLEY and its replacement MAGMA. They provide a mathematically rigorous environment for computing with algebraic, number-theoretic, combinatoric, and geometric objects. For further information, see the Web site <http://magma.maths.usyd.edu.au/magma/>
- [55] There is a program called *Schur* for calculating properties of Lie groups and symmetric functions. For further information, see the Web site <http://schur.sourceforge.net>

#### Momentum Maps (sometimes called Moment Maps)

- [56] V. Guillemin and S. Sternberg, *Symplectic Techniques in Physics*, Cambridge (1984).
- [57] J.-M. Souriau, *Structure of Dynamical Systems: a Symplectic View of Physics*, Birkhäuser (1997).
- [58] J.E. Marsden and T. S. Ratiu, *Introduction to Mechanics and Symmetry*, Second Edition, Springer (1999).

- [59] D. D. Holm, *Geometric Mechanics, Part I: Dynamics and Symmetry*, Imperial College Press, World Scientific (2008).
- [60] D. D. Holm, T. Schmah, C. Stoica, and D. C. P. Ellis, *Geometric Mechanics, from Finite to Infinite Dimensions*, Oxford (2009).
- [61] A. Cannas da Silva, *Lectures on Symplectic Geometry*, Corrected second printing, Springer-Verlag (2008).
- [62] D. Neuenschwander, *Emmy Noether's Wonderful Theorem*, Johns Hopkins University Press (2011).
- [63] Y. Kosmann-Schwarzbach, *The Noether Theorems, Invariance and Conservation Laws in the Twentieth Century*, B. Schwarzbach, Translator, Springer (2011).
- [64] P. J. Olver, Review of Y. Kosmann-Schwarzbach's Springer (2011) book "The Noether Theorems, Invariance and Conservation Laws in the Twentieth Century", *Bulletin of the American Mathematical Society* **50** 161 (2012).
- [65] P. J. Olver, *Applications of Lie Groups to Differential Equations*, Springer-Verlag (1993).
- [66] J. Butterfield, "On Symplectic Reduction in Classical Mechanics", See the Web site <https://arxiv.org/pdf/physics/0507194.pdf>.



# Chapter 28

## More About Various Groups and their Interrelations and other Miscellanea

Relations between  $SO(3)$ ,  $SU(2)$ ,  $Sp(2, \mathbb{C})$ , the Lorentz group,  $Sp(4)$ ,  $SO(5)$ ,  $SU(4)$ , and  $SO(6)$

### Exercises

#### Geodesics, Affine Geodesics, and Lorentz Orbits in Minkowski Space

**28.0.6.** 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  $\tau_n$  were *equally* spaced. See (6.272) through (6.274). What happens if we relax this assumption? Suppose we assume only that

$$\tau_{n+1} > \tau_n \quad (28.0.1)$$

so that

$$\tau^{in} < \tau_1 < \tau_2 \cdots \tau_{N-1} < \tau^{fin}. \quad (28.0.2)$$

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

$$x_{n+1} - x_n = x_{ag}(\tau_{n+1}) - x_{ag}(\tau_n) = b(\tau_{n+1} - \tau_n) = [(\tau_{n+1} - \tau_n)/(\tau^{fin} - \tau^{in})](z - y). \quad (28.0.3)$$

Consequently, find that

$$\begin{aligned} I(x_n, x_{n+1}) &= \sum_{ij} g_{ij}(x_{n+1} - x_n)^i (x_{n+1} - x_n)^j \\ &= [(\tau_{n+1} - \tau_n)/(\tau^{fin} - \tau^{in})]^2 \sum_{ij} g_{ij}(z - y)^i (z - y)^j \\ &= [(\tau_{n+1} - \tau_n)/(\tau^{fin} - \tau^{in})]^2 I(y, z). \end{aligned} \quad (28.0.4)$$

Correspondingly, verify that

$$(\tau_{n+1} - \tau_n)^{-1} I(x_n, x_{n+1}) = (\tau_{n+1} - \tau_n)(\tau^{fin} - \tau^{in})^{-2} I(y, z). \quad (28.0.5)$$

With the aid of (6.296), show that the sum rule (6.277) still holds even when the  $\tau_n$  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

$$E_{\tau_n, \tau_{n+1}} = \int_{\tau_n}^{\tau_{n+1}} L_E(\tau) d\tau. \quad (28.0.6)$$

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

$$L_E = (1/2)(\tau^{fin} - \tau^{in})^{-2} I(y, z). \quad (28.0.7)$$

It follows that

$$E_{\tau_n, \tau_{n+1}} = \int_{\tau_n}^{\tau_{n+1}} L_E(\tau) d\tau = L_E \int_{\tau_n}^{\tau_{n+1}} d\tau = (1/2)(\tau_{n+1} - \tau_n)(\tau^{fin} - \tau^{in})^{-2} I(y, z). \quad (28.0.8)$$

Verify that combining (6.296) and (6.299) yields the result

$$2E_{\tau_n, \tau_{n+1}} = (\tau_{n+1} - \tau_n)^{-1} I(x_n, x_{n+1}). \quad (28.0.9)$$

Finally, verify that combining (6.264) and (6.300) converts the sum rule (6.277) into the relation

$$E_{\tau^{in}, \tau_1} + E_{\tau_1, \tau_2} + \cdots + E_{\tau_{N-1}, \tau^{fin}} = E_{\tau^{in}, \tau^{fin}} = E. \quad (28.0.10)$$

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.

**28.0.7.** This exercise examines how a Lorentz transformation acts on the electromagnetic field tensor  $F^{\alpha\beta}$  as given by (6.56), which we repeat below:

$$F^{\mu\nu} = \begin{pmatrix} 0 & -B_z & B_y & E_x/c \\ B_z & 0 & -B_x & E_y/c \\ -B_y & B_x & 0 & E_z/c \\ -E_x/c & -E_y/c & -E_z/c & 0 \end{pmatrix}, \quad (28.0.11)$$

so that  $F^{12} = -B_z$ , etc. Let us review some background information: A Lorentz transformation, when acting on space-time, is a linear transformation described by a matrix which we will call  $\Lambda$ . See (6.2.49). Its action on the four-vector (6.42) is given by the relation

$$\bar{x}^\alpha = \sum_\mu \Lambda^{\alpha\mu} x^\mu. \quad (28.0.12)$$

Its action on a tensor  $F^{\alpha\beta}$  is given by the relation

$$\bar{F}^{\alpha\beta} = \sum_{\mu\nu} \Lambda^{\alpha\mu} \Lambda^{\beta\nu} F^{\mu\nu}. \quad (28.0.13)$$

See Exercise 6.2.6. The matrix  $\Lambda$  satisfies the relation

$$\Lambda g \Lambda^T = g. \quad (28.0.14)$$

See (6.2.51). Verify, if we view  $F$  as a matrix, then (6.311) can be written in the form

$$\bar{F} = \Lambda F \Lambda^T. \quad (28.0.15)$$

[Note that the left side of (6.312) and the right side of (6.313) have an identical structure. That is because both  $g$  and  $F$  are rank two tensors, and therefore are acted upon by a Lorentz transformation  $\Lambda$  in the same fashion. But the difference between  $g$  and  $F$  is that  $g$  is invariant under this action, and  $F$  is not.]

As a brief interlude, we comment that there is a somewhat different way of viewing the transformation properties of  $g$ , which is also an opportunity to introduce a further concept and associated nomenclature. If we view  $g^{\mu\nu}$  as a second-rank tensor, we may define a transformed tensor  $\bar{g}^{\alpha\beta}$  by the rule

$$\bar{g}^{\alpha\beta} = \sum_{\mu\nu} \Lambda^{\alpha\mu} \Lambda^{\beta\nu} g^{\mu\nu}. \quad (28.0.16)$$

According to (0.14) there is the result

$$\bar{g}^{\alpha\beta} = g^{\alpha\beta}. \quad (28.0.17)$$

We may say that  $g^{\alpha\beta}$  is an *invariant* tensor. Alternatively, we may say that Lorentz transformations preserve the Minkowski metric. Transformations that preserve a metric are called *isometries*. Thus, Lorentz transformations are isometries for Minkowski space.

Let us return to the main discussion. Watch as we now perform some trickery: Define a matrix  $G$  by the rule

$$G = F - \lambda g \quad (28.0.18)$$

where  $\lambda$  is a parameter. Define a transformed matrix  $\bar{G}$  by the rule

$$\bar{G} = \Lambda G \Lambda^T. \quad (28.0.19)$$

Show that there is the result

$$\bar{G} = \bar{F} - \lambda g. \quad (28.0.20)$$

Define polynomials  $P(F, \lambda)$  and  $P(\bar{F}, \lambda)$  by the rules

$$P(F, \lambda) = \det(G), \quad (28.0.21)$$

$$P(\bar{F}, \lambda) = \det(\bar{G}). \quad (28.0.22)$$

For the identity component of the Lorentz group the Lorentz transformation matrix  $\Lambda$  has the property

$$\det(\Lambda) = 1. \quad (28.0.23)$$

See Exercise 7.3.27. Show, using (6.315), that

$$\det(\bar{G}) = \det(G) \quad (28.0.24)$$

and consequently

$$P(\bar{F}, \lambda) = P(F, \lambda). \quad (28.0.25)$$

It therefore behooves us to compute  $P(F, \lambda)$ .

Show, because  $F$  is antisymmetric, only even powers of  $\lambda$  can occur in  $P(F, \lambda)$ . Indeed, show that

$$P(F, \lambda) = -\lambda^4 + \lambda^2[\mathbf{B} \cdot \mathbf{B} - (1/c^2)\mathbf{E} \cdot \mathbf{E}] + (1/c^2)(\mathbf{E} \cdot \mathbf{B})^2. \quad (28.0.26)$$

Define functions  $I_1$  and  $I_2$  of  $F$  (the electromagnetic fields  $\mathbf{E}$  and  $\mathbf{B}$ ) by the rules

$$I_1(F) = \mathbf{B} \cdot \mathbf{B} - (1/c^2)\mathbf{E} \cdot \mathbf{E}, \quad (28.0.27)$$

$$I_2(F) = (1/c)(\mathbf{E} \cdot \mathbf{B}). \quad (28.0.28)$$

Then, by (6.321) and upon equating powers of  $\lambda$ , we find using (6.322) that

$$I_1(\bar{F}) = I_1(F), \quad (28.0.29)$$

and

$$[I_2(\bar{F})]^2 = [I_2(F)]^2 \quad (28.0.30)$$

where  $\bar{F}$  is the field tensor composed of  $\bar{\mathbf{E}}$  and  $\bar{\mathbf{B}}$ , the fields resulting from applying the Lorentz transformation  $\Lambda$  to the fields  $\mathbf{E}$  and  $\mathbf{B}$ . That is, the field functions  $I_1$  and  $[I_2]^2$  are Lorentz invariant.

In fact, the function  $I_2$  itself, and not just its square, is Lorentz invariant,

$$I_2(\bar{F}) = I_2(F). \quad (28.0.31)$$

Evidently, (6.326) follows from (6.327). But can (6.327) be proved from (6.326)? It can, by *continuity*: First, suppose  $F$  is such that  $I_2(F) = 0$ . Then we have the chain of reasoning

$$I_2(F) = 0 \Rightarrow [I_2(F)]^2 = 0 \Rightarrow [I_2(\bar{F})]^2 = 0 \Rightarrow I_2(\bar{F}) = 0, \quad (28.0.32)$$

which establishes (6.327) in this case. But what about the case  $I_2(F) \neq 0$ ? In this case, without further reasoning, we may only conclude from (6.326) that

$$I_2(\bar{F}) = I_2(F) \text{ or } I_2(\bar{F}) = -I_2(F). \quad (28.0.33)$$

We wish to rule out the second possibility.

Let  $\hat{\Lambda}(\tau)$  be a continuous *path* in the identity component of the Lorentz group that connects the identity element  $I$  to the element  $\Lambda$ . Such a path is easily specified using, in Lie form, a *polar decomposition* for elements in the identity component of the Lorentz group. For details, see Exercise 7.3.27. There it is shown that Lorentz group elements  $\Lambda$  in the identity component can be written in the form

$$\Lambda(\lambda, \mathbf{m}; \theta, \mathbf{n}) = \exp(\lambda \mathbf{m} \cdot \mathbf{N}) \exp(\theta \mathbf{n} \cdot \mathbf{L}) \quad (28.0.34)$$

where  $\mathbf{N}$  and  $\mathbf{L}$  are Lie generators for *boosts* and *rotations*, respectively.<sup>1</sup> We now define a

---

<sup>1</sup>Note that the parameters  $\lambda$  appearing in (6.314) and (6.330) are not the same.

path  $\hat{\Lambda}(\tau)$  with the desired properties by the rule

$$\hat{\Lambda}(\tau) = \exp(\tau\lambda\mathbf{m} \cdot \mathbf{N}) \exp(\tau\theta\mathbf{n} \cdot \mathbf{L}). \quad (28.0.35)$$

By construction, this path satisfies the relations

$$\hat{\Lambda}(0) = I \text{ and } \hat{\Lambda}(1) = \Lambda. \quad (28.0.36)$$

Next define a sequence of field tensors  $\hat{F}(\tau)$  by the rule

$$\hat{F}(\tau) = \hat{\Lambda}(\tau) F \hat{\Lambda}^T(\tau) \quad (28.0.37)$$

with the results that

$$\hat{F}(0) = F \text{ and } \hat{F}(1) = \bar{F}. \quad (28.0.38)$$

Finally, define a function  $\hat{I}_2(\tau)$  by the rule

$$\hat{I}_2(\tau) = I_2[\hat{F}(\tau)]. \quad (28.0.39)$$

Verify that  $\hat{I}_2(\tau)$  is a continuous function, and has the properties

$$\hat{I}_2(0) = I_2(F) \text{ and } \hat{I}_2(1) = I_2(\bar{F}). \quad (28.0.40)$$

We now have all the required ingredients to complete our argument. We have already assumed  $I_2(F) \neq 0$ . Next assume that the second option in (6.329) holds,

$$I_2(\bar{F}) = -I_2(F). \quad (28.0.41)$$

Show that then

$$\hat{I}_2(1) = I_2(\bar{F}) = -I_2(F) = -\hat{I}_2(0). \quad (28.0.42)$$

That is, the function  $\hat{I}_2(\tau)$  *changes sign* as  $\tau$  goes from  $\tau = 0$  to  $\tau = 1$ . At this point *Bernard Bolzano* (1781-1848) would exclaim, based on his *intermediate value theorem*, that  $\hat{I}_2(\tau)$  must vanish for some  $\tau$  value somewhere in the interval  $\tau \in (0, 1)$ . Let  $\tau_0 \in (0, 1)$  be a/the  $\tau$  value for which  $\hat{I}_2(\tau)$  takes on the intermediate value 0,

$$\hat{I}_2(\tau_0) = 0. \quad (28.0.43)$$

Now we may make the reasoning chain

$$\hat{I}_2(\tau_0) = 0 \Rightarrow [\hat{I}_2(\tau_0)]^2 = 0 \Rightarrow [\hat{I}_2(\tau_0)]^2 \neq [I_2(F)]^2, \quad (28.0.44)$$

which shows that  $[I_2]^2$  is not Lorentz invariant. We have reached a contradiction, and therefore the first option in (6.329) must hold. That is,  $I_2$  itself must be Lorentz invariant, as claimed.

There is also an alternate proof that  $I_1$  and  $I_2$  are Lorentz invariant and, in fact are manifestly Lorentz invariant. Begin by recalling some standard definitions: The tensor  $F_{\mu\nu}$ , the lower-index/covariant version of  $F^{\alpha\beta}$ , is defined by the rule

$$F_{\mu\nu} = \sum_{\alpha\beta} g_{\mu\alpha} g_{\nu\beta} F^{\alpha\beta}. \quad (28.0.45)$$

Verify that

$$F_{\mu\nu} = \begin{pmatrix} 0 & -B_z & B_y & -E_x/c \\ B_z & 0 & -B_x & -E_y/c \\ -B_y & B_x & 0 & -E_z/c \\ E_x/c & E_y/c & E_z/c & 0 \end{pmatrix}, \quad (28.0.46)$$

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  $\mathbf{E} \rightarrow -\mathbf{E}$ . The tensor  ${}^*F^{\mu\nu}$ , the tensor *dual* to  $F^{\alpha\beta}$ , is defined by the rule

$${}^*F^{\mu\nu} = (1/2) \sum_{\alpha\beta} \epsilon^{\mu\nu\alpha\beta} F_{\alpha\beta}. \quad (28.0.47)$$

Here  $\epsilon^{\alpha\beta\gamma\delta}$  is the completely antisymmetric tensor with  $\epsilon^{1234} = 1$ . Verify that a particular application of the rule (6.343) yields the relation  ${}^*F^{12} = F_{34}$ . Verify the general result

$${}^*F^{\mu\nu} = \begin{pmatrix} 0 & -E_z/c & E_y/c & -B_x \\ E_z/c & 0 & -E_x/c & -B_y \\ -E_y/c & E_x/c & 0 & -B_z \\ B_x & B_y & B_z & 0 \end{pmatrix}, \quad (28.0.48)$$

so that  ${}^*F^{12} = -E_z/c$ , etc. Evidently the elements of  ${}^*F^{\mu\nu}$  are obtained from those of  $F^{\mu\nu}$  by making the substitutions  $\mathbf{E}/c \rightarrow -\mathbf{B}$  and  $\mathbf{B} \rightarrow \mathbf{E}/c$ .

With these definitions in hand, employ them to show that there are the manifestly Lorentz invariant relations

$$I_1(F) = (1/2) \sum_{\alpha\beta} F^{\alpha\beta} F_{\alpha\beta}, \quad (28.0.49)$$

$$I_2(F) = (1/8) \sum_{\alpha\beta\gamma\delta} \epsilon^{\alpha\beta\gamma\delta} F_{\alpha\beta} F_{\gamma\delta} = (1/4) \sum_{\alpha\beta} {}^*F^{\alpha\beta} F_{\alpha\beta}. \quad (28.0.50)$$

Finally, for future use in Exercise 8.2.12, note from (6.322) that there is the result

$$\det(F) = P(F, 0) = (1/c^2)(\mathbf{E} \cdot \mathbf{B})^2 = [I_2(F)]^2. \quad (28.0.51)$$

You have shown that (6.325) and (6.327) comprise a *necessary* condition for a field pair  $\bar{\mathbf{E}}, \bar{\mathbf{B}}$  and  $\mathbf{E}, \mathbf{B}$  to be related by a Lorentz transformation. It can be shown that (6.325) and (6.327) also comprise a *sufficient* condition. See Exercise 6.2.12. That is, if for a given pair  $\bar{F}, F$  (6.325) and (6.327) hold, then there must a Lorentz transformation  $\Lambda$  such that (6.313) holds.

Since it has been established that (6.325) and (6.327) comprise a necessary and sufficient condition, it follows that *any* Lorentz invariant electromagnetic field function must be expressible as some function of  $I_1$  and  $I_2$ .

**28.0.8.** The purpose of this exercise is to explore the relation between the Lie algebras  $sp(2, \mathbb{R})$ ,  $sl(2, \mathbb{R})$ ,  $su(2)$ , and  $usp(2)$ . From the work of Exercise 3.7.29 we know that  $sl(n, \mathbb{R})$  and  $su(n)$  are equivalent over the complex field and therefore, as a particular case,  $sl(2, \mathbb{R})$  and  $su(2)$  are equivalent over the complex field. Also, from Section 5.10.1, we know that  $sp(2n, \mathbb{R})$  and  $usp(2n)$  are equivalent over the complex field and therefore, as a particular case,  $sp(2, \mathbb{R})$  and  $usp(2)$  are equivalent over the complex field. In fact,  $sp(2, \mathbb{R})$

and  $usp(2)$  are the same. See Exercise 5.10.17. There remains the case of  $sp(2, \mathbb{R})$  and  $su(2)$ . Your task in this exercise is to verify that they are equivalent over the complex field and to explore various features of this equivalence. In Exercise 3.25 you will have the privilege of studying how  $sp(2, \mathbb{R})$  and  $su(2)$  are related to  $sl(2, \mathbb{C})$ .

Suppose that the quantities  $B_\alpha$ , for  $\alpha = 1, 2$ , or  $3$ , are *any* set of matrices or operators that obey the  $sp(2, \mathbb{R})$  commutation rules given by (3.7.69) through (3.7.71). Define associated matrices/operators  $B'_\alpha$  by the (change of basis) rules

$$B'_1 = -iB_1, \quad (28.0.52)$$

$$B'_2 = -B_2, \quad (28.0.53)$$

$$B'_3 = -iB_3. \quad (28.0.54)$$

Verify that the  $B'_\alpha$  obey the commutation rules

$$\{B'_1, B'_2\} = B'_3, \quad (28.0.55)$$

$$\{B'_2, B'_3\} = B'_1, \quad (28.0.56)$$

$$\{B'_3, B'_1\} = B'_2, \quad (28.0.57)$$

iff the  $B_\alpha$  obey the  $sp(2, \mathbb{R})$  commutation rules given by (3.7.69) through (3.7.71). But, according to Equations (3.7.173) and (3.7.174) of Exercise 3.7.31, the commutation rules (3.86) through (3.88) are those for  $su(2)$ . Thus  $sp(2, \mathbb{R})$  and  $su(2)$  are indeed equivalent over the complex field. [Note that the relations (3.83) through (3.85) do in fact involve complex coefficients.]

Suppose the quantities  $B_\alpha$  are, in fact, the  $2 \times 2$  matrices that appear on the right sides of (3.7.66) through (3.7.68). Show that in this case the matrices  $B'_\alpha$  are the matrices given by the relations

$$B'_\alpha = K^\alpha. \quad (28.0.58)$$

Recall the definitions (3.7.169) through (3.7.171).

What are the polynomials  $b_\alpha$  and  $b'_\alpha$  associated with the  $B_\alpha$  and the  $B'_\alpha$ , and what are the properties of their associated Lie operators? Based on (5.6.6), (5.6.11), and (5.6.12), and (3.7.66) through (3.7.68), verify that

$$b_1 = (1/2)f = (1/4)(p^2 - q^2), \quad (28.0.59)$$

$$b_2 = (1/2)b^0 = (1/4)(p^2 + q^2), \quad (28.0.60)$$

$$b_3 = (1/2)g = (1/2)qp. \quad (28.0.61)$$

Based on (3.83) through (3.85) and (3.89) through (3.91), verify that

$$b'_1 = -ib_1 = -i(1/2)f = -i(1/4)(p^2 - q^2), \quad (28.0.62)$$

$$b'_2 = -b_2 = -(1/2)b^0 = -(1/4)(p^2 + q^2), \quad (28.0.63)$$

$$b'_3 = -ib_3 = -i(1/2)g = -i(1/2)qp. \quad (28.0.64)$$

Verify the Poisson bracket relations

$$[b_1, b_2] = -b_3, \quad (28.0.65)$$

$$[b_2, b_3] = -b_1, \quad (28.0.66)$$

$$[b_3, b_1] = b_2. \quad (28.0.67)$$

They are to be expected from (3.7.69) through (3.7.71) which, as we have already seen, are a variant of the commutation rules for  $sp(2, \mathbb{R})$ . Verify, as expected from (3.86) through (3.88), that there are the Poisson bracket relations

$$[b'_1, b'_2] = b'_3, \quad (28.0.68)$$

$$[b'_2, b'_3] = b'_1, \quad (28.0.69)$$

$$[b'_3, b'_1] = b'_2, \quad (28.0.70)$$

which are a variant of the commutation rules for  $su(2)$ . Recall (3.16) through (3.18). Verify the conjugacy relations

$$:b_1:^\dagger =:b_1:, \quad (28.0.71)$$

$$:b_2:^\dagger = -:b_2:, \quad (28.0.72)$$

$$:b_3:^\dagger =:b_3:; \quad (28.0.73)$$

$$:b'_\alpha:^\dagger = -:b'_\alpha:. \quad (28.0.74)$$

Thus  $:b_2:$  is anti-Hermitian,  $:b_1:$  and  $:b_3:$  are Hermitian, and the  $:b'_\alpha:$  are anti-Hermitian.

It is also useful to introduce another basis for  $sp(2, \mathbb{R})$ . Let  $\ell_\pm$  and  $\ell_0$  be the monomials

$$\ell_+ = -(1/2)(f + b^0) = -(1/2)p^2, \quad (28.0.75)$$

$$\ell_- = -(1/2)(f - b^0) = (1/2)q^2, \quad (28.0.76)$$

$$\ell_0 = (1/2)g = (1/2)qp. \quad (28.0.77)$$

(Note that the  $\ell_\pm$  are linear combinations of  $b^0$  and  $f$  with *real* coefficients.) Define associated Lie operators by the rules

$$\mathcal{L}_+ =: \ell_+ :, \quad (28.0.78)$$

$$\mathcal{L}_0 =: \ell_0 :, \quad (28.0.79)$$

$$\mathcal{L}_- =: \ell_- :. \quad (28.0.80)$$

Using (7.3.16) through (7.3.18) verify that

$$(\mathcal{L}_+)^{\dagger} =: -p^2/2 :^\dagger =: q^2/2 := \mathcal{L}_-, \quad (28.0.81)$$

$$(\mathcal{L}_0)^{\dagger} =: (1/2)qp :^\dagger =: (1/2) :qp := \mathcal{L}_0, \quad (28.0.82)$$

$$(\mathcal{L}_-)^{\dagger} =: (1/2)q^2/2 :^\dagger =: -(1/2)p^2/2 := \mathcal{L}_+. \quad (28.0.83)$$

Also, let  $L_{\pm}$  and  $L_0$  be the matrices associated with  $\mathcal{L}_{\pm}$  and  $\mathcal{L}_0$ . Verify that they are given by the relations

$$L_+ = -(1/2)(F + B^0) = - \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = -(1/2)(\sigma^1 + i\sigma^2), \quad (28.0.84)$$

$$L_0 = (1/2)G = (1/2) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = (1/2)\sigma^3, \quad (28.0.85)$$

$$L_- = -(1/2)(F - B^0) = - \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = -(1/2)(\sigma^1 - i\sigma^2). \quad (28.0.86)$$

Verify that Lie operators (3.108) through (3.110) obey the commutation rules

$$\{\mathcal{L}_+, \mathcal{L}_-\} = 2\mathcal{L}_0, \quad (28.0.87)$$

$$\{\mathcal{L}_0, \mathcal{L}_+\} = \mathcal{L}_+, \quad (28.0.88)$$

$$\{\mathcal{L}_0, \mathcal{L}_-\} = -\mathcal{L}_-. \quad (28.0.89)$$

Since the  $\ell_{\pm}$  and  $\ell_0$  are real polynomials and the  $\mathcal{L}_{\pm}$  and  $\mathcal{L}_0$  are the Lie operators associated with these polynomials, the commutation rules (3.117) through (3.119) are a variant of those for  $sp(2, \mathbb{R})$ . Verify, as expected by construction, that  $L_{\pm}$  and  $L_0$  obey the same commutation rules. Finally, we observe that they are also the commutation rules for  $su(2)$  in its raising and lowering operator form. See Sections 27.1 and 27.2. That is, by a change of basis involving complex coefficients, the commutation rules for the usual form of  $su(2)$  can be brought to the  $sp(2, \mathbb{R})$  commutation rules (3.117) through (3.119). This change in form is another instance of the equivalence of  $sp(2, \mathbb{R})$  and  $su(2)$  under a change of basis involving complex coefficients.

**28.0.9.** Review Exercise 3.24. The purpose of this exercise is to show that  $sp(2, \mathbb{R})$  and  $su(2)$  are subalgebras of  $sl(2, \mathbb{C})$ , to see how they fit within  $sl(2, \mathbb{C})$ , and to make analogous statements about the relations between the corresponding groups  $Sp(2, \mathbb{R})$ ,  $SU(2)$ , and  $SL(2, \mathbb{C})$ .

The group  $SL(n, \mathbb{C})$  is the set of  $n \times n$  matrices with entries drawn from the field  $\mathbb{C}$  and having determinant +1. Let  $B_1$  through  $B_3$  be the matrices given by (3.7.66) through (3.7.68) and let  $\gamma$  be a three-component vector with entries drawn from  $\mathbb{C}$ ,

$$\gamma = (\gamma_1, \gamma_2, \gamma_3). \quad (28.0.90)$$

According to Exercise 3.1.3, a  $2 \times 2$  matrix is symplectic iff its determinant is +1. And, according to Exercise 3.7.10, the exponent for the exponential form of any matrix (which always exists in some neighborhood of the identity) has trace 0 iff the the matrix has determinant +1. Thus, in agreement with Exercise 3.7.26,  $sl(n, \mathbb{C})$ , the Lie algebra of  $SL(n, \mathbb{C})$ , consists of all  $n \times n$  matrices with entries drawn from the field  $\mathbb{C}$  and having trace 0. In particular,  $sl(2, \mathbb{C})$  consists of all  $2 \times 2$  matrices with entries drawn from the field  $\mathbb{C}$  and having trace 0. Verify, therefore, that  $sl(2, \mathbb{C})$  consists of matrices of the form

$$(1/2) \begin{pmatrix} \gamma_3 & \gamma_1 + \gamma_2 \\ \gamma_1 - \gamma_2 & -\gamma_3 \end{pmatrix} = \gamma \cdot \mathbf{B} \quad (28.0.91)$$

where

$$\boldsymbol{\gamma} \cdot \mathbf{B} = \sum_j \gamma_j B_j. \quad (28.0.92)$$

Let  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  be three-component vectors with entries drawn from  $\mathbb{R}$ ,

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3), \quad (28.0.93)$$

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3). \quad (28.0.94)$$

Decompose  $\boldsymbol{\gamma}$  into real and imaginary parts by writing

$$\boldsymbol{\gamma} = \boldsymbol{\alpha} + i\boldsymbol{\beta}. \quad (28.0.95)$$

Verify that  $sp(2, \mathbb{R})$  consists of the matrices (3.121) evaluated with

$$\boldsymbol{\beta} = 0 \text{ and } \boldsymbol{\alpha} \text{ unrestricted.} \quad (28.0.96)$$

Consequently, in the  $sp(2, \mathbb{R})$  case, the  $sl(2, \mathbb{C})$  matrix (3.121) takes the form

$$(1/2) \begin{pmatrix} \alpha_3 & \alpha_1 + \alpha_2 \\ \alpha_1 - \alpha_2 & -\alpha_3 \end{pmatrix} = \boldsymbol{\alpha} \cdot \mathbf{B}. \quad (28.0.97)$$

What is required for the matrix in (3.121) to be in  $su(2)$ ? Show that it must be anti-Hermitian. What does this property require of the vector  $\boldsymbol{\gamma}$ ? Verify that

$$(1/2) \begin{pmatrix} \gamma_3 & \gamma_1 + \gamma_2 \\ \gamma_1 - \gamma_2 & -\gamma_3 \end{pmatrix}^\dagger = \sum_j \bar{\gamma}_j B_j^\dagger = \bar{\gamma}_1 B_1 - \bar{\gamma}_2 B_2 + \bar{\gamma}_3 B_3. \quad (28.0.98)$$

[Recall that  $B_1$  and  $B_3$  are Hermitian and  $B_2$  is anti-Hermitian. See (3.7.66) through (3.7.68).] Show that the matrix in (3.121) is anti-Hermitian iff

$$\bar{\gamma}_1 = -\gamma_1, \quad (28.0.99)$$

$$\bar{\gamma}_2 = \gamma_2, \quad (28.0.100)$$

$$\bar{\gamma}_3 = -\gamma_3. \quad (28.0.101)$$

Verify it follows that  $su(2)$  consists of the matrices (3.121) evaluated with

$$\alpha_1 = 0 \text{ and } \beta_1 \text{ is unrestricted,} \quad (28.0.102)$$

$$\beta_2 = 0 \text{ and } \alpha_2 \text{ is unrestricted,} \quad (28.0.103)$$

$$\alpha_3 = 0 \text{ and } \beta_3 \text{ is unrestricted.} \quad (28.0.104)$$

Consequently, show that in the  $su(2)$  case, the  $sl(2, \mathbb{C})$  matrix (3.121) takes the form

$$\begin{aligned} (1/2) \begin{pmatrix} i\beta_3 & i\beta_1 + \alpha_2 \\ i\beta_1 - \alpha_2 & -i\beta_3 \end{pmatrix} &= \gamma_1 B_1 + \gamma_2 B_2 + \gamma_3 B_3 \\ &= i\beta_1 B_1 + \alpha_2 B_2 + i\beta_3 B_3 \\ &= -\beta_1 B'_1 - \alpha_2 B'_2 - \beta_3 B'_3 \\ &= -\beta_1 K^1 - \alpha_2 K^2 - \beta_3 K^3. \end{aligned} \quad (28.0.105)$$

Recall the relations (3.7.169) through (3.7.171) and (7.3.86) through (7.3.89).

Is it possible for the matrix (3.122) to be in *both*  $sp(2, \mathbb{R})$  and  $su(2)$ ? Upon comparing the conditions (3.127) and the conditions (3.133) through (3.135), show that the answer is *yes* provided that

$$\beta = 0, \alpha_1 = \alpha_3 = 0, \text{ and } \alpha_2 \text{ is unrestricted.} \quad (28.0.106)$$

Thus  $sp(2, \mathbb{R})$  and  $su(2)$  have in common the one-dimensional Lie algebra (over the real field  $\mathbb{R}$ ) spanned by  $B_2$ .

You have shown that  $sp(2, \mathbb{R})$  and  $su(2)$  are both subalgebras of  $sl(2, \mathbb{C})$ . Moreover, you have shown that these two subalgebras are equivalent under the (complex) change of basis given by (7.3.83) through (7.3.85).

So far you have been treating Lie algebras. What can be said about the associated Lie groups  $Sp(2, \mathbb{R})$ ,  $SU(2)$ , and  $SL(2, \mathbb{C})$ ?

We begin with the case corresponding to the one-dimensional Lie algebra spanned by  $B_2$ , which is easy. The matrix  $B_2$  is real and antisymmetric. See (3.7.67). Show, therefore, that elements of the form  $\exp(\theta B_2)$ , with  $\theta$  real, comprise an  $SO(2, \mathbb{R})$  group. See (3.7.93) or (5.9.12). Note that this  $SO(2, \mathbb{R})$  group is in both  $Sp(2, \mathbb{R})$  and  $SU(2)$ .

To continue, we know much about  $Sp(2, \mathbb{R})$  and, in particular, that it has the topology  $E^2 \times T^1$ . See (5.9.10) in Section 5.9.1.<sup>2</sup> Similarly, we know much about  $SU(2)$  and, in particular, that it has the topology  $S^3$ . See Exercise 5.10.13. What remains is to examine the case of  $SL(2, \mathbb{C})$ .

Let  $\mathbf{a}$  and  $\mathbf{b}$  be three-vectors with real entries. Employ the notation

$$\mathbf{a} \cdot \boldsymbol{\sigma} = \sum_j a_j \sigma^j, \text{ etc.} \quad (28.0.107)$$

Verify that the matrix  $\mathbf{a} \cdot \boldsymbol{\sigma}$  is Hermitian and the matrix  $i\mathbf{b} \cdot \boldsymbol{\sigma}$  is anti-Hermitian. Show that any matrix of the form  $\exp[(1/2)\mathbf{a} \cdot \boldsymbol{\sigma}]$  is Hermitian and positive definite. Show that any matrix of the form  $\exp[(1/2)i\mathbf{b} \cdot \boldsymbol{\sigma}]$  is in  $SU(2)$ . Show that any element  $M$  in  $SL(2, \mathbb{C})$  can be written in the form

$$M = \exp[(1/2)\mathbf{a} \cdot \boldsymbol{\sigma}] \exp[(1/2)i\mathbf{b} \cdot \boldsymbol{\sigma}] \quad (28.0.108)$$

where both factors are unique. (This result is the complex analog of orthogonal polar decomposition. See Section 4.2.) What is the topology of  $SL(2, \mathbb{C})$ ? We know that the second factor in (3.139), because it is in  $SU(2)$ , has the topology  $S^3$ . Show that the first factor has the topology of  $E^3$ . It follows that  $SL(2, \mathbb{C})$  has the topology  $E^3 \times S^3$ .

Using the parameterization (3.139), how does one obtain the two subgroups  $SU(2)$  and  $Sp(2, \mathbb{R})$ ? Show that  $SU(2)$  consists of the matrices (3.139) evaluated with

$$\mathbf{a} = 0 \text{ and } \mathbf{b} \text{ unrestricted.} \quad (28.0.109)$$

Show that  $Sp(2, \mathbb{R})$  consists of the matrices (3.139) evaluated with

$$a_1 \text{ and } a_3 \text{ unrestricted, } a_2 = 0; b_1 = b_3 = 0, b_2 \text{ unrestricted.} \quad (28.0.110)$$

In summary,  $SL(2, \mathbb{C})$  has both  $SU(2)$  and  $Sp(2, \mathbb{R})$  subgroups.

---

<sup>2</sup>Observe that the  $T^1$  in  $Sp(2, \mathbb{R})$  is the  $SO(2, \mathbb{R})$  just described.

Finally, show that the  $SO(2, \mathbb{R})$  that is in both  $SU(2)$  and  $Sp(2, \mathbb{R})$  consists of the matrices (3.139) evaluated with

$$\mathbf{a} = 0; b_1 = b_3 = 0, b_2 \text{ unrestricted.} \quad (28.0.111)$$

**28.0.10.** Review Exercise 1.6.17. Recall that it treated the net interval  $I(y, z)$  for the affine geodesic joining the points  $y$  and  $z$  in Minkowski space and the infinitesimal interval  $ds^2$ , and showed that they were connected by the integral and differential relations

$$\int_{\tau^{in}}^{\tau^{fin}} d\tau [ds^2/(d\tau)^2]|_{x=x_{ag}} = (\tau^{fin} - \tau^{in})^{-1} I(y, z)$$

and

$$I(x, x + dx) = \sum_{ij} g_{ij}(x + dx - x)^i (x + dx - x)^j = \sum_{ij} g_{ij}(dx)^i (dx)^j = ds^2. \quad (28.0.112)$$

See (1.6.256) and (1.6.286).

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,

$$y' = f(y). \quad (28.0.113)$$

Suppose that  $f$  has the property that

$$I(y', z') = I(y, z) \quad (28.0.114)$$

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

$$y' = f(y) = d + \Lambda y \quad (28.0.115)$$

where  $d$  is some fixed four-component vector and  $\Lambda$  is some fixed  $4 \times 4$  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,

$$d = f(0), \quad (28.0.116)$$

and define a new transformation  $h(y)$  by the rule

$$h(y) = f(y) - d. \quad (28.0.117)$$

Verify that  $h$  maps the origin into itself,

$$h(0) = 0, \quad (28.0.118)$$

and also satisfies

$$I\{h(y), h(z)\} = I(y, z). \quad (28.0.119)$$

Verify that explicit evaluation of both sides of (3.150) gives the result

$$h(y) \cdot h(y) - 2h(y) \cdot h(z) + h(z) \cdot h(z) = y \cdot y - 2y \cdot z + z \cdot z. \quad (28.0.120)$$

Show that setting  $z = 0$  in (3.151) gives the result

$$h(y) \cdot h(y) = y \cdot y. \quad (28.0.121)$$

Show that combining (3.151) and (3.152) yields the result

$$h(y) \cdot h(z) = y \cdot z. \quad (28.0.122)$$

Let  $e^1$  to  $e^4$  be the points/vectors

$$e^1 = (1000), e^2 = (0100), \text{ etc.} \quad (28.0.123)$$

Define points/vectors  $c^j$  by the rule

$$c^j = h(e^j). \quad (28.0.124)$$

Show that using (3.153) and (3.155) gives the result

$$c^i \cdot c^j = h(e^i) \cdot h(e^j) = e^i \cdot e^j = g^{ij}. \quad (28.0.125)$$

Prove, therefore, that the vectors  $c^j$  are linearly independent and can be used as a basis set. Now define a matrix  $\Lambda$  by the rule

$$\Lambda e^j = c^j, \quad (28.0.126)$$

and show that this definition results in the explicit relation

$$\Lambda^{ij} = (e^i, \Lambda e^j) = (e^i, c^j) \quad (28.0.127)$$

where  $(*, *)$  denotes the usual/ordinary scalar product.

Let  $y$  be an arbitrary point having the expansion

$$y = \sum_j \eta^j e^j, \quad (28.0.128)$$

and set

$$y'' = h(y). \quad (28.0.129)$$

Show, since the  $c^j$  form a basis, that one may write

$$y'' = \sum_j g^{jj} \{c^j \cdot y''\} c^j. \quad (28.0.130)$$

Using (3.153), (3.155), (3.159), and (3.160), verify that

$$c^j \cdot y'' = h(e^j) \cdot h(y) = e^j \cdot y = g^{jj} \eta^j. \quad (28.0.131)$$

Show that combining this information with (3.157), (3.160), and (3.161) yields the result

$$h(y) = y'' = \sum_j (g^{jj})^2 \eta^j c^j = \sum_j \eta^j \Lambda e^j = \Lambda y. \quad (28.0.132)$$

Finally, verify that going back to (3.148) gives the advertised result

$$f(y) = h(y) + d = \Lambda y + d. \quad (28.0.133)$$

**28.0.11.** The purpose of this exercise is to study the Lie algebra of the Lorentz group and the Lie/exponential representation of group elements. But, before doing so, we digress to observe that the Lorentz group has four *separate* components, only one of which contains the identity element.

To see that the Lorentz group has four separate components, begin by verifying that the matrices  $g$  (spatial inversion),  $-g$  (temporal inversion), and  $-I$  (total inversion), as well as  $I$ , are all Lorentz transformations. Next, starting with the relation (6.2.50), verify the line of reasoning

$$\begin{aligned} \Lambda^T g \Lambda = g \Rightarrow \det(\Lambda^T g \Lambda) = \det(g) \Rightarrow \det(\Lambda^T) \det(g) \det(\Lambda) = \det(g) \Rightarrow \\ \det(\Lambda^T) \det(\Lambda) = 1 \Rightarrow [\det(\Lambda)]^2 = 1 \Rightarrow \det(\Lambda) = \pm 1. \end{aligned} \quad (28.0.134)$$

Since the determinant of a matrix is a continuous function of its entries, the last relation in (3.165) shows that the Lorentz transformations with determinant  $+1$  are separated in matrix space from those with determinant  $-1$ . Also we know that the matrix  $g$  is a Lorentz transformation, and is easily verified to have  $\det(g) = -1$ . Therefore, if any Lorentz transformation matrix with determinant  $+1$  is multiplied by  $g$ , show that the result will be a Lorentz transformation matrix with determinant  $-1$ . Finally, verify the line of reasoning

$$\begin{aligned} \Lambda^T g \Lambda = g \Rightarrow (\Lambda^T g \Lambda)^{44} = g^{44} \Rightarrow (\Lambda^{44})^2 - \sum_{\mu=1}^3 (\Lambda^{\mu 4})^2 = 1 \Rightarrow \\ |\Lambda^{44}| \geq 1 \Leftrightarrow \Lambda^{44} \geq 1 \text{ or } \Lambda^{44} \leq -1. \end{aligned} \quad (28.0.135)$$

Evidently, among the four possibilities embraced by the last relations in (3.165) and (3.166), only the component with

$$\det(\Lambda) = 1 \text{ and } \Lambda^{44} \geq 1 \quad (28.0.136)$$

can contain the identity element. Verify that this component is a subgroup, and the other three are not. Verify that all the elements in the other components can be obtained by multiplying the elements in the identity component by  $g$  or  $-g$  or  $-I$ .

With this digression behind us, we turn to studying the Lie structure of the identity component of the Lorentz group. Suppose  $\Lambda$  is sufficiently near the identity matrix  $I$  so that it can be written in the form

$$\Lambda = \exp(\epsilon S) = I + \epsilon S + O(\epsilon^2) \quad (28.0.137)$$

where  $\epsilon$  is a small parameter and  $S$  is a matrix to be determined. Show that inserting (3.168) into (6.2.50) and equating powers of  $\epsilon$  yields the condition

$$S^T g + gS = 0 \Leftrightarrow \quad (28.0.138)$$

$$S^T = -gSg. \quad (28.0.139)$$

Verify that the condition (3.170) is also sufficient for the  $\Lambda$  given by (3.168) to satisfy (6.2.50) exactly. Verify that matrices  $S$  that satisfy (3.170) form a Lie algebra.

Let us pause at this point to see how the tilde Lie algebraic conjugacy operator defined by  $\tilde{\mathcal{C}}$  in Exercise 3.7.36 applies to elements in the Lorentz group Lie algebra. Show from the definition (3.7.219) and (3.170) that

$$\tilde{\mathcal{C}}(S) = -S^T = gSg = gSg^{-1}. \quad (28.0.140)$$

(Here we have used the fact that  $g = g^{-1}$ ). Evidently, for the Lorentz group Lie algebra, this conjugate representation is equivalent to the original representation. Note that Lorentz transformation matrices  $\Lambda$  act on four-vectors, and four-vectors carry the representation  $\Gamma(1/2, 1/2)$ . See Exercise 7.3.29. You have shown that this representation is self conjugate under the tilde operation.

What happens if we instead use the conjugacy operators  $\check{\mathcal{C}}$  and  $\hat{\mathcal{C}}$ ? Verify that

$$\check{\mathcal{C}}(S) = \bar{S} = S \quad (28.0.141)$$

and

$$\hat{\mathcal{C}}(S) = -S^\dagger = -S^T = gSg = gSg^{-1} \quad (28.0.142)$$

because  $S$  is a real matrix. It follows that, for the Lorentz group Lie algebra, the conjugate representation is equivalent to the original representation no matter what conjugacy operator is used.

Now let us work out the consequences of (3.170) in detail. Begin by computing some matrix elements. Verify the following line of reasoning for diagonal elements:

$$\begin{aligned} S^T = -gSg \Rightarrow (S^T)^{jj} &= -\sum_{k\ell} g^{jk} S^{k\ell} g^{\ell j} \Rightarrow S^{jj} = -S^{jj} \\ &\Rightarrow \text{all diagonal elements of } S \text{ vanish.} \end{aligned} \quad (28.0.143)$$

Verify the following line of reasoning for  $j4$  and  $4j$  elements:

$$S^T = -gSg \Rightarrow (S^T)^{j4} = -\sum_{k\ell} g^{jk} S^{k\ell} g^{\ell 4} \Rightarrow S^{4j} = -g^{jj} S^{j4} \Rightarrow S^{4j} = S^{j4} \text{ for } j \neq 4. \quad (28.0.144)$$

Verify, consequently, that  $S$  can be written in the form

$$S = \begin{pmatrix} A & \mathbf{a} \\ \mathbf{a}^T & 0 \end{pmatrix} \quad (28.0.145)$$

where  $A$  is a  $3 \times 3$  matrix and  $\mathbf{a}$  is a three-component vector. Now take (3.170) into account once again. Show that it implies the matrix relation

$$\begin{pmatrix} A^T & \mathbf{a} \\ \mathbf{a}^T & 0 \end{pmatrix} = -\begin{pmatrix} -I & \mathbf{o} \\ \mathbf{o} & 1 \end{pmatrix} \begin{pmatrix} A & \mathbf{a} \\ \mathbf{a}^T & 0 \end{pmatrix} \begin{pmatrix} -I & \mathbf{o} \\ \mathbf{o} & 1 \end{pmatrix} \quad (28.0.146)$$

where  $\mathbf{o}$  is a three-component vector all of whose entries are zero. Verify that carrying out the matrix multiplications appearing on the right side of (3.177) yields the final result

$$\begin{pmatrix} A^T & \mathbf{a} \\ \mathbf{a}^T & 0 \end{pmatrix} = \begin{pmatrix} -A & \mathbf{a} \\ \mathbf{a}^T & 0 \end{pmatrix}. \quad (28.0.147)$$

Consequently  $A$  must be antisymmetric,

$$A^T = -A, \quad (28.0.148)$$

and  $\mathbf{a}$  can be any three-component vector.

We are ready to set up a convenient (and pleasing) basis for the matrices  $S$ . In analogy to (3.7.178) through (3.7.180) define in the present context matrices  $L^1$  through  $L^3$  by the rules

$$L^1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (28.0.149)$$

$$L^2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (28.0.150)$$

$$L^3 = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (28.0.151)$$

Also, define matrices  $N^1$  through  $N^3$  by the rules

$$N^1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad (28.0.152)$$

$$N^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad (28.0.153)$$

$$N^3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (28.0.154)$$

Verify that these matrices form a basis for the vector space of matrices  $S$  that satisfy (3.176) with the condition (3.179). Thus, the Lie algebra of such matrices is six dimensional. Indeed, verify that there are the commutation rules

$$\{L^j, L^k\} = \sum_{\ell} \epsilon_{jk\ell} L^{\ell}, \quad (28.0.155)$$

$$\{L^j, N^k\} = \sum_{\ell} \epsilon_{jk\ell} N^{\ell}, \quad (28.0.156)$$

$$\{N^j, N^k\} = - \sum_{\ell} \epsilon_{jk\ell} L^{\ell}. \quad (28.0.157)$$

Finally, observe that the  $L^j$  are sent into themselves under the tilde operation (3.171), and the  $N^j$  are sent into their negatives. Glancing at (3.186) through (3.188), we see that the tilded basis elements satisfy the same commutation relations as the original elements, as expected.

Consider matrices  $\Lambda$  of the form

$$\Lambda(\lambda, \mathbf{m}; \theta, \mathbf{n}) = \exp(\lambda \mathbf{m} \cdot \mathbf{N}) \exp(\theta \mathbf{n} \cdot \mathbf{L}) \quad (28.0.158)$$

where  $\mathbf{m}$  and  $\mathbf{n}$  are unit vectors and

$$\mathbf{m} \cdot \mathbf{N} = \sum_j m_j N^j, \text{ etc.} \quad (28.0.159)$$

Verify that  $\mathbf{m} \cdot \mathbf{N}$  is Hermitian and  $\mathbf{n} \cdot \mathbf{L}$  is anti-Hermitian, and therefore (3.189) is a polar decomposition. Show that every  $\Lambda$  in the identity component of the Lorentz group can be uniquely written in this form. Show that all  $\Lambda$  in all four components of the Lorentz group can be uniquely written in the form.

$$\Lambda(\lambda, \mathbf{m}; \theta, \mathbf{n}; r, s) = g^r (-g)^s \exp(\lambda \mathbf{m} \cdot \mathbf{N}) \exp(\theta \mathbf{n} \cdot \mathbf{L}) \text{ with } r = 0, 1 \text{ and } s = 0, 1. \quad (28.0.160)$$

To simplify nomenclature, from here on we will refer to the identity component of the Lorentz group simply as the Lorentz group.

Evidently, as follows from the work of Exercise 3.7.31, the factor  $\exp(\theta \mathbf{n} \cdot \mathbf{L})$  in (3.189) produces spatial rotations by angle  $\theta$  about the axis  $\mathbf{n}$ . What can be said about the factor  $\exp(\lambda \mathbf{m} \cdot \mathbf{N})$ ? Your next task is to verify that it produces velocity transformations along the  $\mathbf{m}$  axis and to find the relation between  $\lambda$  and the magnitude of the velocity.

Begin with the case  $\mathbf{m} = \mathbf{e}_3$ , in which case we are interested in the effect of  $\exp(\lambda N^3)$ . Verify that

$$(N^3)^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (28.0.161)$$

and

$$(N^3)^3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = N^3. \quad (28.0.162)$$

Show, therefore, that there is the intermediate result

$$\begin{aligned} \exp(\lambda N^3) &= I + \lambda N^3 + \lambda^2 (N^3)^2 / 2! + \lambda^3 (N^3)^3 / 3! + \dots \\ &= I + N^3 \sinh(\lambda) + (N^3)^2 [\cosh(\lambda) - 1]. \end{aligned} \quad (28.0.163)$$

Verify that employing (3.185) and (3.192) in (3.194) gives the final matrix result

$$\exp(\lambda N^3) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cosh(\lambda) & \sinh(\lambda) \\ 0 & 0 & \sinh(\lambda) & \cosh(\lambda) \end{pmatrix}. \quad (28.0.164)$$

For extra credit, evaluate  $\mathbf{m} \cdot \mathbf{N}$  and its second and third powers. Show that

$$(\mathbf{m} \cdot \mathbf{N})^3 = \mathbf{m} \cdot \mathbf{N}. \quad (28.0.165)$$

Use the results you find to generalize (3.194) and to evaluate  $\exp(\lambda \mathbf{m} \cdot \mathbf{N})$  for arbitrary unit vector  $\mathbf{m}$ .

To continue, review Exercise 6.2.7 and, in particular, the result (6.2.53). How are (6.2.53) and (3.195) related? Let  $\tau$  be a parameter and consider the world line

$$x^1 = x^2 = x^3 = 0, \quad t = \tau \quad (28.0.166)$$

so that

$$x(\tau) = \{0, 0, 0, c\tau\}. \quad (28.0.167)$$

This is the world line for a particle at rest at the spatial origin.<sup>3</sup> Suppose  $\bar{x}$  is the result of applying  $\exp(\lambda N^3)$  to  $x$ . Verify that according to (3.195) there is the result

$$\bar{x}(\tau) = \exp(\lambda N^3)x(\tau) = \{0, 0, c \sinh(\lambda)\tau, c \cosh(\lambda)\tau\} \quad (28.0.168)$$

so that

$$\bar{x}^3 = c \sinh(\lambda)\tau \quad (28.0.169)$$

and

$$c\bar{t} = c \cosh(\lambda)\tau \Leftrightarrow \bar{t} = \cosh(\lambda)\tau. \quad (28.0.170)$$

Using (3.200) and (3.201) verify that, after the transformation/boost (3.195), the particle will be moving along the  $+e_3$  axis with velocity  $v$  given by

$$v = d\bar{x}_3/d\bar{t} = c \tanh(\lambda). \quad (28.0.171)$$

Using the definition

$$v = (v/c)c = \beta c \Leftrightarrow \beta = v/c, \quad (28.0.172)$$

verify that

$$\beta = \tanh(\lambda). \quad (28.0.173)$$

The quantity

$$\lambda = \tanh^{-1}(\beta) \quad (28.0.174)$$

is called the *rapidity*. At this point we may make a remark about sign choices. Observe that had we replaced the matrices defining the  $N^j$  by their negatives in (3.183) through (3.185), the commutation rules (3.186) through (3.188) would be unchanged. However, a minus sign

---

<sup>3</sup>Note that for a particle to be possibly at rest in some inertial frame it must have finite mass.

would then appear in (3.202). Since it seems desirable for a positive rapidity to result in a positive velocity, the sign choice we have made seems to be the more natural.

Here is an occasion for two brief interludes: For the first, review Exercise 3.7.37 and suppose the tilde group element conjugacy relation defined by the operator  $\tilde{\mathcal{D}}$  is applied to the Lorentz group element  $\Lambda$  given by (3.189). Show that

$$\tilde{\mathcal{D}}[\Lambda(\lambda, \mathbf{m}; \theta, \mathbf{n})] = \exp(-\lambda \mathbf{m} \cdot \mathbf{N}) \exp(\theta \mathbf{n} \cdot \mathbf{L}) = \Lambda(-\lambda, \mathbf{m}; \theta, \mathbf{n}). \quad (28.0.175)$$

Evidently the effect of  $\tilde{\mathcal{D}}$  is to change the sign of the rapidity and therefore the sign of the boost velocity.

For the second interlude, suppose two velocity transformations with rapidities  $\lambda_1$  and  $\lambda_2$  are made successively in the same direction. Then, from the group property

$$\exp(\lambda_1 \mathbf{m} \cdot \mathbf{N}) \exp(\lambda_2 \mathbf{m} \cdot \mathbf{N}) = \exp[(\lambda_1 + \lambda_2) \mathbf{m} \cdot \mathbf{N}], \quad (28.0.176)$$

we see that rapidities add. Show, as a result, that there is the relation

$$\beta_3 = \tanh(\lambda_3) = \tanh(\lambda_1 + \lambda_2). \quad (28.0.177)$$

Verify the chain of reasoning

$$\begin{aligned} \beta_3 &= \tanh(\lambda_1 + \lambda_2) = [\tanh(\lambda_1) + \tanh(\lambda_2)]/[1 + \tanh(\lambda_1) \tanh(\lambda_2)] \\ &= (\beta_1 + \beta_2)/(1 + \beta_1 \beta_2). \end{aligned} \quad (28.0.178)$$

This is the relativistic law for the addition of parallel velocities.

We also take this opportunity to remark that in general velocity transformations do not commute because the right side of (3.188) does not vanish, but rather contains rotation generators. Consequently a sequence of velocity transformations can produce a rotation, called a *Wigner rotation*. This failure to commute, which is a relativistic phenomena, is the origin of *Thomas precession*.

To return to the main theme, verify using (3.204) that

$$\begin{aligned} \gamma &= 1/(1 - \beta^2)^{1/2} = 1/[1 - \tanh^2(\lambda)]^{1/2} = \cosh(\lambda)/[\cosh^2(\lambda) - \sinh^2(\lambda)]^{1/2} \\ &= \cosh(\lambda). \end{aligned} \quad (28.0.179)$$

Also verify that

$$\beta\gamma = \tanh(\lambda) \cosh(\lambda) = \sinh(\lambda). \quad (28.0.180)$$

Consequently, (3.195) can be rewritten in the form

$$\Lambda(\lambda, \mathbf{e}_3; 0, \mathbf{n}) = \exp(\lambda N^3) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cosh(\lambda) & \sinh(\lambda) \\ 0 & 0 & \sinh(\lambda) & \cosh(\lambda) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \gamma & \beta\gamma \\ 0 & 0 & \beta\gamma & \gamma \end{pmatrix} \quad (28.0.181)$$

in agreement with (6.2.53).

**28.0.12.** Review Exercise 3.27. We have seen that the commutation rules (3.186) are those for  $so(3, \mathbb{R})$  and they generate spatial rotations,  $SO(3, \mathbb{R})$  transformations, described by Lorentz transformation group elements of the form  $\exp(\theta \mathbf{n} \cdot \mathbf{L})$ . Are there other subgroups of the Lorentz group?

Curiously, and generally not discussed in the literature, the Lorentz group also has  $Sp(2, \mathbb{R})$  subgroups. Consider, for example, the generators  $N^3$ ,  $N^1$ , and  $L^2$ . Verify from (3.187) and (3.188) that they obey the commutation rules

$$\{L^2, N^3\} = N^1, \quad (28.0.182)$$

$$\{L^2, N^1\} = -N^3, \quad (28.0.183)$$

$$\{N^3, N^1\} = -L^2, \quad (28.0.184)$$

and therefore generate a subalgebra. Next look at the  $sp(2, \mathbb{R})$  commutation rules (3.7.69) through (3.7.71). Verify, under the correspondences

$$B_1 \leftrightarrow -N^1, \quad (28.0.185)$$

$$B_2 \leftrightarrow -L^2, \quad (28.0.186)$$

$$B_3 \leftrightarrow -N^3, \quad (28.0.187)$$

which taken together behave like a change of basis using *real* coefficients, that the commutation relations (3.213) through (3.215) are the same as the  $sp(2, \mathbb{R})$  commutation rules (3.7.69) through (3.7.71).<sup>4</sup> Verify, therefore, that Lorentz transformations of the form

$$\Lambda(\lambda_3, \lambda_1, \theta) = \exp(\lambda_3 N^3 + \lambda_1 N^1) \exp(\theta L^2) \quad (28.0.188)$$

form a group that is isomorphic to the group  $Sp(2, \mathbb{R})$ .

How could one have guessed that the Lorentz group would have both  $SO(3, \mathbb{R})$  and  $Sp(2, \mathbb{R})$  subgroups? According to Exercise 3.25 the group  $SL(2, \mathbb{C})$  has both  $SU(2)$  and  $Sp(2, \mathbb{R})$  subgroups. In Exercise 8.2.14 you will learn that the Lorentz group is homomorphic to  $SL(2, \mathbb{C})$ . In fact,  $SL(2, \mathbb{C})$  is the covering group for the Lorentz group. Armed with this knowledge, one would expect that the Lorentz group would have both  $SO(3, \mathbb{R})$  and  $Sp(2, \mathbb{R})$  subgroups.

Finally, let  $\Lambda^f$  be any *fixed* Lorentz group element. Verify that all Lorentz group elements  $\Lambda'$  of the form

$$\Lambda' = \Lambda^f \exp(\lambda_1 N^3 + \lambda_2 N^1) \exp(\theta L^2) (\Lambda^f)^{-1} \quad (28.0.189)$$

also comprise an  $Sp(2, \mathbb{R})$  subgroup. Thus the Lorentz group has many  $Sp(2, \mathbb{R})$  subgroups depending on the choice of  $\Lambda^f$ . Make a similar argument for the case of  $SO(3, \mathbb{R})$ .

**28.0.13.** The purpose of this exercise is to describe some representations of the Lorentz group. We begin with the observation that the Lorentz group is *not* compact. Indeed, for example, looking at (3.208) we see that  $\lambda$  can be arbitrarily large thereby yielding matrices  $\Lambda$  that are arbitrarily far from the origin in matrix space. It follows that the

---

<sup>4</sup>Making the correspondences (3.216) through (3.218) was facilitated by the common appearance of Pauli matrices in (3.7.66) through (3.7.68), (5.6.7), (5.6.13), (5.6.14), and (7.3.236) through (7.3.241).

Lorentz group does not have any *finite* dimensional *unitary* representations because unitary matrices are bounded in matrix space.<sup>5</sup> However, the Lorentz group does have nonunitary finite dimensional representations. They are useful for constructing classical and quantum fields including the Higgs fields, Dirac fields for leptons (neutrinos, electrons ···) and quarks, vector boson fields (gluons, photons,  $W^\pm$ ,  $Z^0$ ), the graviton field, and more. Some facts about these finite dimensional representations are the subject of this exercise.

Let  $\check{L}^j$  and  $\check{N}^j$  be a set of matrices/operators that obey the Lorentz group Lie algebra commutation rules (3.186) through (3.188). Define related matrices/operators  $A^j$  and  $B^j$  by the rules

$$A^j = (\check{L}^j + i\check{N}^j)/2, \quad (28.0.190)$$

$$B^j = (\check{L}^j - i\check{N}^j)/2, \quad (28.0.191)$$

Note that these definitions are essentially a change of basis with coefficients drawn from the complex field  $\mathbb{C}$ . Verify, from (3.186) through (3.188), that the  $A^j$  and  $B^k$  obey the commutation rules

$$\{A^j, B^k\} = 0, \quad (28.0.192)$$

$$\{A^j, A^k\} = \sum_{\ell} \epsilon_{jkl} A^{\ell}, \quad (28.0.193)$$

$$\{B^j, B^k\} = \sum_{\ell} \epsilon_{jkl} B^{\ell}. \quad (28.0.194)$$

That is, the  $A^j$  and  $B^k$  commute, and separately obey the commutation rules for  $su(2)$ . You have shown that, over the complex field, the Lie algebra of the Lorentz group is equivalent to  $su(2) \oplus su(2)$ , the direct sum of two commuting  $su(2)$  Lie algebras. It follows that the Lorentz group Lie algebra, unlike the classical and exceptional Lie algebras listed in Table 3.7.2, is *not* a simple Lie algebra.<sup>6</sup> In retrospect, we should have already known this. Recall Exercise 3.7.40.

As is familiar from their occurrence in Quantum Mechanics, the representations of  $su(2)$  are labelled by a quantity  $j$  that can take on the values

$$j = 0, 1/2, 1, 3/2, 2 \dots . \quad (28.0.195)$$

We also recall that the dimension of an  $su(2)$  representation is given by the quantity  $(2j+1)$ . Since the Lie algebra of the Lorentz group is equivalent to  $su(2) \oplus su(2)$ , we will label a representation of the Lorentz group Lie algebra by the symbol  $\Gamma(j_1, j_2)$  where  $j_1$  and  $j_2$  are the  $j$  values associated with the  $A_k$  and  $B_k$  Lie algebras, respectively.<sup>7</sup> Verify that the dimension of  $\Gamma(j_1, j_2)$  is given by the relation

$$\dim \Gamma(j_1, j_2) = (2j_1 + 1)(2j_2 + 1). \quad (28.0.196)$$

---

<sup>5</sup>The Lorentz group does have *infinite* dimensional unitary representations. Their discussion is beyond the scope of this book.

<sup>6</sup>It is, however, semisimple since it is the direct sum of two commuting  $su(2)$  Lie algebras and  $su(2)$  is simple.

<sup>7</sup>The symbol  $D(j_1, j_2)$  is also frequently used to denote a representation of the Lorentz group or its Lie algebra.

What is the representation in the case that the matrices  $\check{L}^j$  and  $\check{N}^j$  are the matrices given by (3.180) through (3.185)? This will be the representation carried by the matrices  $\Lambda$  given by (3.189) acting on four-vectors. To answer this question we may compute the  $su(2)$  Casimir operators given by  $\mathbf{A} \cdot \mathbf{A}$  and  $\mathbf{B} \cdot \mathbf{B}$  where

$$\mathbf{A} \cdot \mathbf{A} = \sum_j (A^j)^2, \text{ etc.} \quad (28.0.197)$$

(For a discussion of Casimir operators see Exercise 3.7.31 and Section 27.11.) Verify from (3.221) that

$$\sum_j (A^j)^2 = (1/4) \sum_j [(L^j)^2 + i(L^j N^j + N^j L^j) - (N^j)^2]. \quad (28.0.198)$$

Show that the following results hold:

$$\sum_j (L^j)^2 = -2 \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (28.0.199)$$

[see (3.7.215)],

$$L^j N^j = N^j L^j = 0, \quad (28.0.200)$$

$$\sum_j (N^j)^2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}. \quad (28.0.201)$$

Show, therefore, that

$$\sum_j (A^j)^2 = (1/4) \sum_j [(L^j)^2 - (N^j)^2] = -(3/4)I, \quad (28.0.202)$$

and that the same result holds for  $\mathbf{B} \cdot \mathbf{B}$ . Since the  $su(2)$  quadratic Casimir operator has the value  $-j(j+1)$  it follows that in the four-vector case there are the relations

$$j_1 = j_2 = 1/2. \quad (28.0.203)$$

You have shown that four-vectors carry the representation  $\Gamma(1/2, 1/2)$ . Note, from the work of Exercise 3.27, we know that  $\Gamma(1/2, 1/2)$  is self conjugate under any of the conjugacy operations. This fact is customarily expressed by writing

$$\bar{\Gamma}(1/2, 1/2) = \Gamma(1/2, 1/2) \quad (28.0.204)$$

where here, to be precise, the bar represents any of the conjugacy operations. We also observe from (3.189) that, since the generators given by (3.180) through (3.185) are real, the matrices  $\Lambda$  are real. Therefore, we say that the representation  $\Gamma(1/2, 1/2)$  is real.

It can be shown that Dirac 4-spinors (to be defined subsequently) carry the representation  $\Gamma(0, 1/2) \oplus \Gamma(1/2, 0)$ , and antisymmetric tensors such as the electromagnetic field tensor  $F^{\mu\nu}$  carry the representation  $\Gamma(0, 1) \oplus \Gamma(1, 0)$ . (See Exercises 8.2.17 and 3.33.) It can be shown that these representations are also real. Verify that four-vectors, Dirac 4-spinors, and antisymmetric tensors have the expected dimensions of 4, 4, and 6, respectively. Verify, using (6.2.51), that the metric tensor  $g^{\mu\nu}$  carries the representation  $\Gamma(0, 0)$ . Show that the same is true of the completely antisymmetric (*Levi-Civita*) tensor/symbol  $\epsilon^{\alpha\beta\gamma\delta}$ .

**28.0.14.** Review Exercise 3.7.26. One of its purposes was to show that  $SL(n, \mathbb{C})$ , the set of all  $n \times n$  complex matrices with determinant +1, forms a group; and to verify that  $sl(n, \mathbb{C})$ , the set of all  $n \times n$  complex matrices with trace 0, is its Lie algebra. The purpose of this exercise is to study in some detail (the simplest case)  $SL(2, \mathbb{C})$  and its Lie algebra  $sl(2, \mathbb{C})$ .

To begin, define matrices  $\hat{L}^j$  and  $\hat{N}^j$  by the rules

$$\hat{L}^1 = K^1 = (-i/2)\sigma^1 = (-i/2) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (28.0.205)$$

$$\hat{L}^2 = K^2 = (-i/2)\sigma^2 = (-i/2) \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = (-1/2) \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad (28.0.206)$$

$$\hat{L}^3 = K^3 = (-i/2)\sigma^3 = (-i/2) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}; \quad (28.0.207)$$

$$\hat{N}^1 = (1/2)\sigma^1 = (1/2) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (28.0.208)$$

$$\hat{N}^2 = (1/2)\sigma^2 = (1/2) \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad (28.0.209)$$

$$\hat{N}^3 = (1/2)\sigma^3 = (1/2) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (28.0.210)$$

See Exercise 3.7.31 and (3.7.169) through (3.7.171). Note that

$$\hat{N}^j = i\hat{L}^j. \quad (28.0.211)$$

Verify that the  $\hat{L}^j$  and  $\hat{N}^j$  form a basis for  $sl(2, \mathbb{C})$  and obey the commutation rules

$$\{\hat{L}^j, \hat{L}^k\} = \sum_{\ell} \epsilon_{jk\ell} \hat{L}^{\ell}, \quad (28.0.212)$$

$$\{\hat{L}^j, \hat{N}^k\} = \sum_{\ell} \epsilon_{jk\ell} \hat{N}^{\ell}, \quad (28.0.213)$$

$$\{\hat{N}^j, \hat{N}^k\} = - \sum_{\ell} \epsilon_{jk\ell} \hat{L}^{\ell}. \quad (28.0.214)$$

Observe that the structure constants in (3.243) through (3.245) are the same as those in (3.183) through (3.185). Therefore the Lie algebra  $sl(2, \mathbb{C})$  is the *same* as the Lie algebra of the Lorentz group.

Consider  $SL(2, \mathbb{C})$  matrices  $\hat{\Lambda}$  of the form

$$\hat{\Lambda}(\lambda, \mathbf{m}; \theta, \mathbf{n}) = \exp(\lambda \mathbf{m} \cdot \hat{\mathbf{N}}) \exp(\theta \mathbf{n} \cdot \hat{\mathbf{L}}) \quad (28.0.215)$$

were  $\mathbf{m}$  and  $\mathbf{n}$  are unit vectors and

$$\mathbf{m} \cdot \hat{\mathbf{N}} = \sum_j m_j \hat{N}^j, \text{ etc.} \quad (28.0.216)$$

Verify that  $\mathbf{m} \cdot \hat{\mathbf{N}}$  is Hermitian and  $\mathbf{n} \cdot \hat{\mathbf{L}}$  is anti-Hermitian, and therefore (3.246) is a polar decomposition. Show that every matrix in  $SL(2, \mathbb{C})$  can be uniquely written in this form. See Exercise 4.2.5.

You have shown that the Lie algebra  $sl(2, \mathbb{C})$  is the *same* as the Lie algebra of the Lorentz group. Verify, moreover, that there are the analogous polar decompositions (3.246) and (3.189). We therefore expect that there is an intimate connection between the Lorentz group and the group  $SL(2, \mathbb{C})$ . This connection is explored in Exercise 8.2.14 where it is shown that  $SL(2, \mathbb{C})$  is the covering group of the Lorentz group.

At this point it is possible to make three remarks: For the first, review Exercise 3.7.36. Suppose the grave Lie algebraic conjugacy operator defined by  $\hat{\mathcal{C}}$  is applied to the elements  $\hat{L}^j$  and  $\hat{N}^j$  that comprise a basis for the  $sl(2, \mathbb{C})$  Lie algebra. Show from the definition (3.7.225) and (3.236) through (3.241) that there are the results

$$\hat{\mathcal{C}}(\hat{L}^j) = \hat{L}^j = -(\hat{L}^j)^\dagger = \hat{L}^j, \quad (28.0.217)$$

$$\hat{\mathcal{C}}(\hat{N}^j) = \hat{N}^j = -(\hat{N}^j)^\dagger = -\hat{N}^j. \quad (28.0.218)$$

Evidently the  $\hat{L}^j$  are left in peace and the  $\hat{N}^j$  change sign. Verify, as expected, that these transformed elements obey the same commutation rules (3.243) through (3.245) as the original elements.

Can the  $\hat{L}^j, \hat{N}^j$  be related to the  $\hat{L}^j, \hat{N}^j$  by a similarity transformation as in (3.7.218)? Suppose we assume so. Then there will be the relations

$$\hat{L}^j = E \hat{L}^j E^{-1}, \quad (28.0.219)$$

$$\hat{N}^j = E \hat{N}^j E^{-1}. \quad (28.0.220)$$

Verify that combining (3.248) through (3.251) yields the relations

$$\hat{L}^j = E \hat{L}^j E^{-1}, \quad (28.0.221)$$

$$\hat{N}^j = -E \hat{N}^j E^{-1}. \quad (28.0.222)$$

But, from (3.242) and (3.253), we conclude that

$$i \hat{L}^j = -i E \hat{L}^j E^{-1} \Rightarrow \hat{L}^j = -E \hat{L}^j E^{-1}. \quad (28.0.223)$$

Observe that (3.252) and the far right side of (3.254) disagree! It follows there is *no*  $E$  for which (3.250) and (3.251) hold. Therefore the representations of  $sl(2, \mathbb{C})$  provided by the  $\hat{L}^j, \hat{N}^j$  and the  $\hat{L}^j, \hat{N}^j$  are *not* equivalent.

For the second remark, review Exercise 3.7.37. Suppose the grave group element conjugacy relation defined by the operator  $\hat{\mathcal{D}}$  is applied to the  $SL(2, \mathbb{C})$  group elements  $\hat{\Lambda}$  given by (3.246). Show that

$$\hat{\mathcal{D}}[\hat{\Lambda}(\lambda, \mathbf{m}; \theta, \mathbf{n})] = \exp(-\lambda \mathbf{m} \cdot \hat{\mathbf{N}}) \exp(\theta \mathbf{n} \cdot \hat{\mathbf{L}}) = \hat{\Lambda}(-\lambda, \mathbf{m}; \theta, \mathbf{n}). \quad (28.0.224)$$

Evidently the effect of  $\hat{\mathcal{D}}$  is to change the sign of the  $SL(2, \mathbb{C})$  analog of the rapidity and therefore the sign of the  $SL(2, \mathbb{C})$  analog of the boost velocity.

For the third remark we again comment on sign choices. In accord with our earlier finding, observe that had we replaced the matrices defining the  $\hat{N}^j$  by their negatives in (3.239) through (3.241), the commutation rules (3.243) through (3.245) would be unchanged. However, as we will later see at the end of Exercise 8.2.14, the sign choice we have made is necessary for the construction of a natural map between the group  $SL(2, \mathbb{C})$  and the Lorentz group.

In analogy to the study in Exercise 3.29 of the representations of the Lorentz group carried by various entities such as four-vectors, the last task of this exercise is to examine what representations are involved in the case of  $SL(2, \mathbb{C})$ . Following (3.221) and (3.222), form the matrices

$$\hat{A}^j = (\hat{L}^j + i\hat{N}^j)/2, \quad (28.0.225)$$

$$\hat{B}^j = (\hat{L}^j - i\hat{N}^j)/2, \quad (28.0.226)$$

using for the  $\hat{L}^j$  and  $\hat{N}^j$  the matrices (3.236) through (3.241). Verify that so doing yields the results

$$\hat{A}^j = 0, \quad (28.0.227)$$

$$\hat{B}^j = (-i/2)\sigma^j. \quad (28.0.228)$$

Continue on to show that

$$\sum_j (\hat{A}^j)^2 = 0, \quad (28.0.229)$$

$$\sum_j (\hat{B}^j)^2 = -(3/4)I. \quad (28.0.230)$$

Verify it follows from (3.260) and (3.261) that there are the results

$$j_1 = 0 \quad (28.0.231)$$

and

$$j_2 = 1/2. \quad (28.0.232)$$

You have shown that the use of  $SL(2, \mathbb{C})$  produces the  $\Gamma(0, 1/2)$  representation of the Lorentz group.

We have seen that the representations of  $sl(2, \mathbb{C})$  provided by the  $\hat{L}^j, \hat{N}^j$  and the  $\hat{\tilde{L}}^j, \hat{\tilde{N}}^j$  are not equivalent and that the representation provided by the  $\hat{L}^j, \hat{N}^j$  is the  $\Gamma(0, 1/2)$  representation. What representation is provided by the  $\hat{\tilde{L}}^j, \hat{\tilde{N}}^j$ ? Again following (3.221) and (3.222), form the matrices

$$\hat{\tilde{A}}^j = (\hat{\tilde{L}}^j + i\hat{\tilde{N}}^j)/2, \quad (28.0.233)$$

$$\hat{\hat{B}}^j = (\hat{\hat{L}}^j - i\hat{\hat{N}}^j)/2, \quad (28.0.234)$$

using for the  $\hat{\hat{L}}^j$  and  $\hat{\hat{N}}^j$  the matrices (3.248) and (3.249). Verify that so doing yields the results

$$\hat{\hat{A}}^j = (\hat{\hat{L}}^j + i\hat{\hat{N}}^j)/2 = (\hat{L}^j - i\hat{N}^j)/2 = (-i/2)\sigma^j, \quad (28.0.235)$$

$$\hat{\hat{B}}^j = (\hat{\hat{L}}^j - i\hat{\hat{N}}^j)/2 = (\hat{L}^j + i\hat{N}^j)/2 = 0. \quad (28.0.236)$$

Continue on to show that

$$\sum_j (\hat{\hat{A}}^j)^2 = -(3/4)I, \quad (28.0.237)$$

$$\sum_j (\hat{\hat{B}}^j)^2 = 0. \quad (28.0.238)$$

Verify it follows from (3.268) and (3.269) that there are the results

$$j_1 = 1/2 \quad (28.0.239)$$

and

$$j_2 = 0. \quad (28.0.240)$$

You have shown that the use of the  $\hat{\hat{L}}^j, \hat{\hat{N}}^j$  produces the  $\Gamma(1/2, 0)$  representation of the Lorentz group. Correspondingly, we may write

$$\bar{\Gamma}(0, 1/2) = \Gamma(1/2, 0) \quad (28.0.241)$$

where the bar denotes the result of using the grave conjugation operation  $\bar{\mathcal{C}}$ . It can be shown that there is the general Lorentz group/Lie-algebra representation conjugacy relation

$$\bar{\Gamma}(j_1, j_2) = \Gamma(j_2, j_1) \quad (28.0.242)$$

of which (3.235) and (3.272) are particular cases.

**28.0.15.** Review Exercise 7.3.30. There you learned that  $SL(2, \mathbb{C})$  and the Lorentz group have the same Lie algebra. The purpose of this exercise is to show that a *subgroup* of  $SL(3, \mathbb{C})$  also has the Lorentz group Lie algebra, and to discover what representation of the Lorentz group is provided by this subgroup. In Exercise 3.7.26 you learned that  $sl(3, \mathbb{C})$  consists of  $3 \times 3$  complex matrices with trace 0. A  $3 \times 3$  complex matrix requires 9 complex and hence 18 real numbers for its specification. Requiring that the trace vanish imposes one complex and hence two real conditions among these numbers with the result that  $sl(3, \mathbb{C})$  has  $18 - 2 = 16$  real dimensions. For comparison, we know that  $sl(2, \mathbb{C})$  and the Lorentz group Lie algebra have 6 real dimensions. Therefore, if this exercise is to succeed, we must find a suitable six-dimensional *subalgebra* of  $sl(3, \mathbb{C})$ .

Recall the  $so(3, \mathbb{C})$  matrices  $L^j$  defined by (3.7.178) through (3.7.180). Note that they are traceless. Use them to define  $sl(3, \mathbb{C})$  matrices  $\check{L}^j$  and  $\check{N}^j$  by the rules

$$\check{L}^1 = L^1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad (28.0.243)$$

$$\check{L}^2 = L^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad (28.0.244)$$

$$\check{L}^3 = L^3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad (28.0.245)$$

$$\check{N}^1 = iL^1 = i \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad (28.0.246)$$

$$\check{N}^2 = iL^2 = i \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad (28.0.247)$$

$$\check{N}^3 = iL^3 = i \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (28.0.248)$$

Note that, in analogy with (3.242),

$$\check{N}^j = i\check{L}^j. \quad (28.0.249)$$

Verify that the  $\check{L}^j$  and  $\check{N}^j$  obey the commutation rules

$$\{\check{L}^j, \check{L}^k\} = \sum_{\ell} \epsilon_{jk\ell} \check{L}^{\ell}, \quad (28.0.250)$$

$$\{\check{L}^j, \check{N}^k\} = \sum_{\ell} \epsilon_{jk\ell} \check{N}^{\ell}, \quad (28.0.251)$$

$$\{\check{N}^j, \check{N}^k\} = - \sum_{\ell} \epsilon_{jk\ell} \check{L}^{\ell}. \quad (28.0.252)$$

Evidently they span a *subalgebra* of  $sl(3, \mathbb{C})$ . Moreover, the structure constants in (3.281) through (3.283) are the same as those in (3.186) through (3.188). Therefore this subalgebra is the *same* as the Lie algebra of the Lorentz group.

What representation of the Lorentz group Lie algebra is provided by the  $\check{L}^j$  and  $\check{N}^j$ ? Review Exercise 3.29. Employ in (3.221) and (3.222) the  $\check{L}^j$  and  $\check{N}^j$  given by (3.274) through (3.279) to find the results

$$\check{A}^j = (\check{L}^j + i\check{N}^j)/2 = (L^j - L^j)/2 = 0, \quad (28.0.253)$$

$$\check{B}^j = (\check{L}^j - i\check{N}^j)/2 = (L^j + L^j)/2 = L^j. \quad (28.0.254)$$

Continue on to show that

$$\sum_j (\check{A}^j)^2 = 0, \quad (28.0.255)$$

$$\sum_j (\check{B}^j)^2 = \sum_j (L^j)^2 = -(2)I. \quad (28.0.256)$$

See (3.7.215). Verify it follows from (3.286) and (3.287) that there are the results

$$j_1 = 0 \quad (28.0.257)$$

and

$$j_2 = 1. \quad (28.0.258)$$

You have shown that the use of the  $sl(3, \mathbb{C})$  subalgebra produces the  $\Gamma(0, 1)$  representation of the Lorentz group Lie algebra.

Suppose the grave Lie algebraic conjugacy operator defined by  $\check{\mathcal{C}}$  is applied to the elements  $\check{L}^j$  and  $\check{N}^j$  that comprise a basis for the  $sl(3, \mathbb{C})$  Lie subalgebra. Show from the definition (3.7.225) and (3.274) through (3.279) that there are the results

$$\check{\mathcal{C}}(\check{L}^j) = \check{\check{L}}^j = -(\check{L}^j)^\dagger = \check{L}^j, \quad (28.0.259)$$

$$\check{\mathcal{C}}(\check{N}^j) = \check{\check{N}}^j = -(\check{N}^j)^\dagger = -\check{N}^j. \quad (28.0.260)$$

Evidently the  $\check{L}^j$  are left in peace and the  $\check{N}^j$  change sign. Verify, as expected, that the  $\check{\check{L}}^j$ ,  $\check{\check{N}}^j$  also provide a representation of the Lorentz group Lie algebra.

Suppose instead the breve Lie algebraic conjugacy operator defined by  $\check{\check{\mathcal{C}}}$  is applied to the elements  $\check{L}^j$  and  $\check{N}^j$  that comprise a basis for the  $sl(3, \mathbb{C})$  Lie subalgebra. Show from the definition (3.7.224) and (3.274) through (3.279) that there are the results

$$\check{\check{\mathcal{C}}}(\check{L}^j) = \check{\check{L}}^j = \bar{\check{L}}^j = \check{L}^j, \quad (28.0.261)$$

$$\check{\check{\mathcal{C}}}(\check{N}^j) = \check{\check{N}}^j = \bar{\check{N}}^j = -\check{N}^j. \quad (28.0.262)$$

Evidently, for the  $sl(3, \mathbb{C})$  Lie subalgebra, the grave and breve operations have the *same* effect.

Read again the part of Exercise 3.30 that showed the representations of the Lorentz group Lie algebra provided by  $\hat{L}^j, \hat{N}^j$  and  $\check{L}^j, \check{N}^j$  are not equivalent. Construct a similar proof that the representations of the Lorentz group Lie algebra provided by  $\check{L}^j, \check{N}^j$  and  $\check{\check{L}}^j, \check{\check{N}}^j$  are not equivalent. Also, state and prove an analog of (3.255). Finally, show that the  $\check{\check{L}}^j, \check{\check{N}}^j$  produce the  $\Gamma(1, 0)$  representation of the Lorentz group Lie algebra so that

$$\bar{\Gamma}(0, 1) = \Gamma(1, 0), \quad (28.0.263)$$

which is again a particular case of (3.273).

**28.0.16.** Recall that under the action of a Lorentz transformation  $\Lambda$  the electromagnetic field tensor  $F^{\mu\nu}$  defined by

$$F^{\mu\nu} = \begin{pmatrix} 0 & -B_z & B_y & E_x/c \\ B_z & 0 & -B_x & E_y/c \\ -B_y & B_x & 0 & E_z/c \\ -E_x/c & -E_y/c & -E_z/c & 0 \end{pmatrix} \quad (28.0.264)$$

transforms according to the rule

$$\hat{F}^{\alpha\beta} = \sum_{\mu\nu} \Lambda^{\alpha\mu} \Lambda^{\beta\nu} F^{\mu\nu} \Leftrightarrow \hat{F} = \Lambda F \Lambda^T. \quad (28.0.265)$$

Review Exercise 1.6.17. (Here we use a hat  $\hat{\cdot}$  rather than a bar  $\bar{\cdot}$  as a distinguishing mark because later we will want to use a bar to indicate complex conjugation.) This exercise is the first of two exercises whose purpose is to relate the transformation rule (3.296) to the Lorentz Lie algebra/group representations  $\Gamma(0, 1)$  and  $\Gamma(1, 0)$  found in Exercise 3.31 above. It will be limited to Lorentz transformations that are near the identity transformation. It will be followed by a subsequent exercise that extends the results found here to all Lorentz transformations.

Before beginning our exploration it is convenient to introduce some new notation. Define a tensor-valued function  $\mathcal{F}^{\mu\nu}$  of  $\mathbf{E}$  and  $\mathbf{B}$  by the rule

$$\mathcal{F}^{\mu\nu}(\mathbf{E}, \mathbf{B}) = \begin{pmatrix} 0 & -B_z & B_y & E_x/c \\ B_z & 0 & -B_x & E_y/c \\ -B_y & B_x & 0 & E_z/c \\ -E_x/c & -E_y/c & -E_z/c & 0 \end{pmatrix}. \quad (28.0.266)$$

With this definition we may rewrite (3.295) in the form

$$F^{\mu\nu} = \mathcal{F}^{\mu\nu}(\mathbf{E}, \mathbf{B}), \quad (28.0.267)$$

and we may also write

$$\hat{F}^{\mu\nu} = \mathcal{F}^{\mu\nu}(\hat{\mathbf{E}}, \hat{\mathbf{B}}) = \begin{pmatrix} 0 & -\hat{B}_z & \hat{B}_y & \hat{E}_x/c \\ \hat{B}_z & 0 & -\hat{B}_x & \hat{E}_y/c \\ -\hat{B}_y & \hat{B}_x & 0 & \hat{E}_z/c \\ -\hat{E}_x/c & -\hat{E}_y/c & -\hat{E}_z/c & 0 \end{pmatrix} \quad (28.0.268)$$

where  $\hat{\mathbf{E}}$  and  $\hat{\mathbf{B}}$  are the transformed fields associated with  $\hat{F}$ . Finally, for compactness of notation, we will sometimes omit the tensor indices to simply write

$$F = \mathcal{F}(\mathbf{E}, \mathbf{B}) \quad (28.0.269)$$

and

$$\hat{F} = \mathcal{F}(\hat{\mathbf{E}}, \hat{\mathbf{B}}). \quad (28.0.270)$$

Now let us begin our exploration by considering some particular cases. Suppose  $\Lambda$  is the Lorentz transformation for a *small* rotation  $\theta$  about the  $z$  axis,

$$\Lambda = \exp(\theta L^3) = I + \theta L^3 + O(\theta)^2. \quad (28.0.271)$$

Show that in this case (3.296) becomes

$$\hat{F} = \Lambda F \Lambda^T = (I + \theta L^3)F(I - \theta L^3) + O(\theta)^2 = F + \theta \{L^3, F\} + O(\theta)^2. \quad (28.0.272)$$

(Recall that  $L^3$  is antisymmetric.) Verify that

$$\begin{aligned} FL^3 &= \begin{pmatrix} 0 & -B_z & B_y & E_x/c \\ B_z & 0 & -B_x & E_y/c \\ -B_y & B_x & 0 & E_z/c \\ -E_x/c & -E_y/c & -E_z/c & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} -B_z & 0 & 0 & 0 \\ 0 & -B_z & 0 & 0 \\ B_x & B_y & 0 & 0 \\ -E_y/c & E_x/c & 0 & 0 \end{pmatrix}, \end{aligned} \quad (28.0.273)$$

$$L^3 F = [F^T (L^3)^T]^T = (FL^3)^T = \begin{pmatrix} -B_z & 0 & B_x & -E_y/c \\ 0 & -B_z & B_y & E_x/c \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (28.0.274)$$

and therefore

$$\{L^3, F\} = \begin{pmatrix} 0 & 0 & B_x & -E_y/c \\ 0 & 0 & B_y & E_x/c \\ -B_x & -B_y & 0 & 0 \\ E_y/c & -E_x/c & 0 & 0 \end{pmatrix}. \quad (28.0.275)$$

Verify that use of (3.297) through (3.306) yields the relations

$$\hat{E}_x = E_x - \theta E_y + O(\theta^2), \quad (28.0.276)$$

$$\hat{E}_y = E_y + \theta E_x + O(\theta^2), \quad (28.0.277)$$

$$\hat{E}_z = E_z + O(\theta^2); \quad (28.0.278)$$

$$\hat{B}_x = B_x - \theta B_y + O(\theta^2), \quad (28.0.279)$$

$$\hat{B}_y = B_y + \theta B_x + O(\theta^2), \quad (28.0.280)$$

$$\hat{B}_z = B_z + O(\theta^2). \quad (28.0.281)$$

What are we to make of the relations (3.307) through (3.312)? Verify that (3.307) through (3.309) can be rewritten in the form

$$\begin{pmatrix} \hat{E}_x \\ \hat{E}_y \\ \hat{E}_z \end{pmatrix} = \begin{pmatrix} 1 & -\theta & 0 \\ \theta & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} + O(\theta^2) \quad (28.0.282)$$

or, in matrix/vector notation,

$$\hat{\mathbf{E}} = \exp(\theta \check{L}^3) \mathbf{E} + O(\theta^2). \quad (28.0.283)$$

See (3.276). Verify that (3.310) through (3.312) can be rewritten analogously.

At this point, in anticipation of further results, it is convenient to define two three-dimensional complex vectors  $\mathbf{F}^\pm$  (sometimes called *Faraday* vectors) by the rules

$$\mathbf{F}^\pm = \mathbf{E} \pm i\mathbf{c}\mathbf{B}. \quad (28.0.284)$$

Verify that the results (3.307) through (3.312) can be rewritten in the form

$$\begin{pmatrix} \hat{F}_x^+ \\ \hat{F}_y^+ \\ \hat{F}_z^+ \end{pmatrix} = \begin{pmatrix} 1 & -\theta & 0 \\ \theta & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} F_x^+ \\ F_y^+ \\ F_z^+ \end{pmatrix} + O(\theta^2) \quad (28.0.285)$$

or, in matrix/vector notation,

$$\hat{\mathbf{F}}^+ = \exp(\theta \check{L}^3) \mathbf{F}^+ + O(\theta^2). \quad (28.0.286)$$

Let us extend our notation a bit further. Evidently a knowledge of  $\mathbf{F}^+$  is equivalent to a knowledge of  $\mathbf{E}$  and  $\mathbf{B}$ , and vice versa. Indeed, from (3.315) we see that

$$\mathbf{E} = \Re(\mathbf{F}^+) \quad (28.0.287)$$

and

$$\mathbf{B} = (1/c)\Im(\mathbf{F}^+). \quad (28.0.288)$$

We may therefore view  $\mathbf{F}^+$  as being an argument for  $\mathcal{F}$  so that (3.300) and (3.301) can also be written as

$$F = \mathcal{F}(\mathbf{F}^+) \quad (28.0.289)$$

and

$$\hat{F} = \mathcal{F}(\hat{\mathbf{F}}^+). \quad (28.0.290)$$

Verify, using this extended notation, that the results (3.301) through (3.303) and (3.317) can be rewritten in the form

$$\exp(\theta L^3) \mathcal{F}(\mathbf{F}^+) [\exp(\theta L^3)]^T = \mathcal{F}[\exp(\theta \check{L}^3) \mathbf{F}^+] + O(\theta^2). \quad (28.0.291)$$

We have studied the case of a small *rotation* about the  $z$  axis. As a second particular case, suppose  $\Lambda$  is the Lorentz transformation for a small *boost*  $\lambda$  along the  $z$  axis,

$$\Lambda = \exp(\lambda N^3) = I + \lambda N^3 + O(\lambda)^2. \quad (28.0.292)$$

Show that in this case (3.296) becomes

$$\hat{F} = \Lambda F \Lambda^T = (I + \lambda N^3) F (I + \lambda N^3) + O(\lambda^2) = F + \lambda \{N^3, F\}_+ + O(\lambda^2). \quad (28.0.293)$$

(Here  $\{*, *\}_+$  denotes an anticommutator, and we have used the fact that  $N^3$  is symmetric.) Verify that

$$\begin{aligned} FN^3 &= \begin{pmatrix} 0 & -B_z & B_y & E_x/c \\ B_z & 0 & -B_x & E_y/c \\ -B_y & B_x & 0 & E_z/c \\ -E_x/c & -E_y/c & -E_z/c & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & E_x/c & B_y \\ 0 & 0 & E_y/c & -B_x \\ 0 & 0 & E_z/c & 0 \\ 0 & 0 & 0 & -E_z/c \end{pmatrix}, \end{aligned} \quad (28.0.294)$$

$$N^3 F = [F^T (N^3)^T]^T = -(FN^3)^T = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -E_x/c & -E_y/c & -E_z/c & 0 \\ -B_y & B_x & 0 & E_z/c \end{pmatrix}, \quad (28.0.295)$$

and therefore

$$\{N^3, F\}_+ = \begin{pmatrix} 0 & 0 & E_x/c & B_y \\ 0 & 0 & E_y/c & -B_x \\ -E_x/c & -E_y/c & 0 & 0 \\ -B_y & B_x & 0 & 0 \end{pmatrix}. \quad (28.0.296)$$

Verify that use of (3.297) through (3.299), and (3.324) through (3.327) yields the relations

$$\hat{E}_x = E_x + \lambda c B_y + O(\lambda^2), \quad (28.0.297)$$

$$\hat{E}_y = E_y - \lambda c B_x + O(\lambda^2), \quad (28.0.298)$$

$$\hat{E}_z = E_z + O(\lambda^2); \quad (28.0.299)$$

$$\hat{B}_x = B_x - \lambda E_y/c + O(\lambda^2), \quad (28.0.300)$$

$$\hat{B}_y = B_y + \lambda E_x/c + O(\lambda^2), \quad (28.0.301)$$

$$\hat{B}_z = B_z + O(\lambda^2). \quad (28.0.302)$$

What are we to make of the transformation results (3.328) through (3.333)? Verify, in terms of the Faraday vector  $\mathbf{F}^+$ , that

$$\begin{aligned} \hat{F}_x^+ &= \hat{E}_x + i c \hat{B}_x = E_x + i c B_x + \lambda c B_y - i \lambda E_y + O(\lambda^2) \\ &= E_x + i c B_x - i \lambda (E_y + i c B_y) + O(\lambda^2) \\ &= F_x^+ - i \lambda F_y^+ + O(\lambda^2), \end{aligned} \quad (28.0.303)$$

$$\begin{aligned} \hat{F}_y^+ &= \hat{E}_y + i c \hat{B}_y = E_y + i c B_y - \lambda c B_x + i \lambda E_x + O(\lambda^2) \\ &= E_y + i c B_y + i \lambda (E_x + i c B_x) + O(\lambda^2) \\ &= F_y^+ + i \lambda F_x^+ + O(\lambda^2), \end{aligned} \quad (28.0.304)$$

$$\hat{F}_z^+ = \bar{E}_z + i c \hat{B}_z = E_z + i c B_z + O(\lambda^2) = F_z^+ + O(\lambda^2). \quad (28.0.305)$$

Consequently the results (3.328) through (3.336) can be rewritten in the form

$$\begin{pmatrix} \hat{F}_x^+ \\ \hat{F}_y^+ \\ \hat{F}_z^+ \end{pmatrix} = \begin{pmatrix} 1 & -i\lambda & 0 \\ i\lambda & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} F_x^+ \\ F_y^+ \\ F_z^+ \end{pmatrix} + O(\lambda^2) \quad (28.0.306)$$

or, in matrix/vector notation,

$$\hat{\mathbf{F}}^+ = \exp(\lambda \check{N}^3) \mathbf{F}^+ + O(\lambda^2). \quad (28.0.307)$$

See (3.279). Finally, using the extended notation (3.320) and (3.321), verify that the results (3.324) through (3.338) can be rewritten in the form

$$\exp(\lambda N^3) \mathcal{F}(\mathbf{F}^+) [\exp(\lambda N^3)]^T = \mathcal{F}[\exp(\lambda \check{N}^3) \mathbf{F}^+] + O(\lambda^2). \quad (28.0.308)$$

**28.0.17.** This exercise is a sequel to Exercise 3.32. It established, for *small*  $\theta$  and *small*  $\lambda$ , the key results (3.322) and (3.339). Associated with these results are the correspondences

$$\Lambda = \exp(\theta L^3) \Leftrightarrow \hat{\mathbf{F}}^+ = \exp(\theta \check{L}^3) \mathbf{F}^+ + O(\theta^2), \quad (28.0.309)$$

$$\Lambda = \exp(\lambda N^3) \Leftrightarrow \hat{\mathbf{F}}^+ = \exp(\lambda \check{N}^3) \mathbf{F}^+ + O(\lambda^2). \quad (28.0.310)$$

The purpose of this exercise is to extend these results and to relate them to the Lorentz Lie algebra/group representations  $\Gamma(0, 1)$  and  $\Gamma(1, 0)$ .

The first extension is to observe that the  $O(\theta^2)$  error terms in (3.322) and (3.340) and the  $O(\lambda^2)$  error terms in (3.339) and (3.341) are in fact *identically zero!* To see this, in the case of the  $O(\theta^2)$  error terms, observe that there is the *group* relation

$$\exp[\theta L^3] = \exp[(1/2)\theta L^3] \exp[(1/2)\theta L^3]. \quad (28.0.311)$$

Verify that employing this group relation in (3.322) yields the result

$$\begin{aligned} \exp(\theta L^3) \mathcal{F}(\mathbf{F}^+) [\exp(\theta L^3)]^T &= \\ \exp[(1/2)\theta L^3] \exp[(1/2)\theta L^3] \mathcal{F}(\mathbf{F}^+) \{ \exp[(1/2)\theta L^3] \}^T \{ \exp[(1/2)\theta L^3] \}^T &= \\ \exp[(1/2)\theta L^3] \mathcal{F} \{ \exp[(1/2)\theta \check{L}^3] \mathbf{F}^+ \} \{ \exp[(1/2)\theta L^3] \}^T + O[(\theta/2)^2] &= \\ \mathcal{F} \{ \exp[(1/2)\theta \check{L}^3] \exp[(1/2)\theta \check{L}^3] \mathbf{F}^+ \} + 2O[(\theta/2)^2] &= \\ \mathcal{F}[\exp(\theta \check{L}^3) \mathbf{F}^+] + (1/2)O(\theta^2). & \end{aligned} \quad (28.0.312)$$

That is, the possible  $O(\theta^2)$  error term in (3.322) has been replaced by the possible  $(1/2)O(\theta^2)$  error term in (3.343). Similarly, using in (3.322) the group relation

$$\exp[\theta L^3] = \{ \exp[(1/n)\theta L^3] \}^n \quad (28.0.313)$$

will yield the result

$$\exp(\theta L^3) \mathcal{F}(\mathbf{F}^+) [\exp(\theta L^3)]^T = \mathcal{F}[\exp(\theta \check{L}^3) \mathbf{F}^+] + (1/n)O(\theta^2). \quad (28.0.314)$$

Therefore, upon letting  $n \rightarrow \infty$ , we find the result

$$\exp(\theta L^3) \mathcal{F}(\mathbf{F}^+) [\exp(\theta L^3)]^T = \mathcal{F}[\exp(\theta \check{L}^3) \mathbf{F}^+] \quad (28.0.315)$$

as claimed.

In an analogous way, *mutatis mutandis* and again using group properties, (3.339) becomes the relation

$$\exp(\lambda N^3) \mathcal{F}(\mathbf{F}^+) [\exp(\lambda N^3)]^T = \mathcal{F}[\exp(\lambda \check{N}^3) \mathbf{F}^+]. \quad (28.0.316)$$

At this point let us make a sanity check. Verify that

$$\exp(\lambda \check{N}^3) = \begin{pmatrix} \cosh(\lambda) & -i \sinh(\lambda) & 0 \\ i \sinh(\lambda) & \cosh(\lambda) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (28.0.317)$$

and therefore in this case it follows that

$$\hat{F}_x^+ = \cosh(\lambda) F_x^+ - i \sinh(\lambda) F_y^+, \quad (28.0.318)$$

$$\hat{F}_y^+ = i \sinh(\lambda) F_x^+ + \cosh(\lambda) F_y^+, \quad (28.0.319)$$

$$\hat{F}_z^+ = F_z^+, \quad (28.0.320)$$

so that

$$\hat{E}_x = \Re(\hat{F}_x^+) = \cosh(\lambda) E_x + \sinh(\lambda) c B_y = \gamma E_x + \beta \gamma c B_y, \quad (28.0.321)$$

$$\hat{E}_y = \Re(\hat{F}_y^+) = \cosh(\lambda) E_y - \sinh(\lambda) c B_x = \gamma E_y - \beta \gamma c B_x, \quad (28.0.322)$$

$$\hat{E}_z = \Re(\hat{F}_z^+) = E_z; \quad (28.0.323)$$

$$\hat{B}_x = (1/c) \Im(\hat{F}_x^+) = \cosh(\lambda) B_x - \sinh(\lambda)(1/c) E_y = \gamma B_x - \beta \gamma(1/c) E_y, \quad (28.0.324)$$

$$\hat{B}_y = (1/c) \Im(\hat{F}_y^+) = \cosh(\lambda) B_y + \sinh(\lambda)(1/c) E_x = \gamma B_y + \beta \gamma(1/c) E_x, \quad (28.0.325)$$

$$\hat{B}_z = (1/c) \Im(\hat{F}_z^+) = B_z. \quad (28.0.326)$$

[See (3.210) and (3.211).] The relations (3.352) through (3.357) are the expected ones for a boost along the  $z$  axis, and therefore the sanity check has been passed.

So far we have considered rotations about the  $z$  axis and boosts along the  $z$  axis. The second extension is to consider other axes. Verify that, since we have already allowed  $\mathbf{E}$  and  $\mathbf{B}$  to be completely arbitrary, there must be the obvious generalizations: Suppose  $\Lambda$  is the Lorentz transformation for a general rotation,

$$\Lambda = \exp(\theta \mathbf{n} \cdot \mathbf{L}). \quad (28.0.327)$$

In this case (3.346) has the generalization

$$\exp(\theta \mathbf{n} \cdot \mathbf{L}) \mathcal{F}(\mathbf{F}^+) [\exp(\theta \mathbf{n} \cdot \mathbf{L})]^T = \mathcal{F}[\exp(\theta \mathbf{n} \cdot \check{\mathbf{L}}) \mathbf{F}^+]. \quad (28.0.328)$$

Or suppose  $\Lambda$  is the Lorentz transformation for a general boost,

$$\Lambda = \exp(\lambda \mathbf{m} \cdot \mathbf{N}). \quad (28.0.329)$$

In this case (3.347) has the generalization

$$\exp(\lambda \mathbf{m} \cdot \mathbf{N}) \mathcal{F}(\mathbf{F}^+) [\exp(\lambda \mathbf{m} \cdot \mathbf{N})]^T = \mathcal{F}[\exp(\lambda \mathbf{m} \cdot \check{\mathbf{N}}) \mathbf{F}^+]. \quad (28.0.330)$$

The final extension is to consider general Lorentz transformations

$$\Lambda(\lambda, \mathbf{m}; \theta, \mathbf{n}) = \exp(\lambda \mathbf{m} \cdot \mathbf{N}) \exp(\theta \mathbf{n} \cdot \mathbf{L}). \quad (28.0.331)$$

Recall (3.189). Verify that in this case there is the result

$$\begin{aligned} \Lambda \mathcal{F}(\mathbf{F}^+) \Lambda^T &= \exp(\lambda \mathbf{m} \cdot \mathbf{N}) \exp(\theta \mathbf{n} \cdot \mathbf{L}) \mathcal{F}(\mathbf{F}^+) [\exp(\theta \mathbf{n} \cdot \mathbf{L})]^T [\exp(\lambda \mathbf{m} \cdot \mathbf{N})]^T = \\ &= \exp(\lambda \mathbf{m} \cdot \mathbf{N}) \mathcal{F}[\exp(\theta \mathbf{n} \cdot \check{\mathbf{L}}) \mathbf{F}^+] [\exp(\lambda \mathbf{m} \cdot \mathbf{N})]^T = \\ &= \mathcal{F}[\exp(\lambda \mathbf{m} \cdot \check{\mathbf{N}}) \exp(\theta \mathbf{n} \cdot \check{\mathbf{L}}) \mathbf{F}^+]. \end{aligned} \quad (28.0.332)$$

Verify, consequently, that there is the general correspondence

$$\Lambda = \exp(\lambda \mathbf{m} \cdot \mathbf{N}) \exp(\theta \mathbf{n} \cdot \mathbf{L}) \Leftrightarrow \hat{\mathbf{F}}^+ = \exp(\lambda \mathbf{m} \cdot \check{\mathbf{N}}) \exp(\theta \mathbf{n} \cdot \check{\mathbf{L}}) \mathbf{F}^+. \quad (28.0.333)$$

Your last task is to explore how these results relate to the Lorentz Lie algebra/group representations  $\Gamma(0, 1)$  and  $\Gamma(0, 1)$ . You already found in Exercise 3.31 that the matrices  $\check{\mathbf{L}}$  and  $\check{\mathbf{N}}$  constitute a basis for the  $\Gamma(0, 1)$  representation of the Lorentz group Lie algebra. Consequently, according to the right side equation in (3.364), the vectors  $\mathbf{F}^+$  carry the  $\Gamma(0, 1)$  representation of the Lorentz Lie algebra/group. Now form the complex conjugate of the right side equation in (3.364). Verify that so doing yields the relation

$$\hat{\bar{\mathbf{F}}}^+ = \exp(\lambda \mathbf{m} \cdot \check{\mathbf{N}}) \exp(\theta \mathbf{n} \cdot \check{\bar{\mathbf{L}}}) \bar{\mathbf{F}}^+ \quad (28.0.334)$$

where a bar  $\bar{\phantom{A}}$  denotes complex conjugation. [Verify from (3.7.1) that  $\overline{\exp(B)} = \exp(\bar{B})$  for any matrix  $B$ .] Next verify that

$$\hat{\bar{\mathbf{F}}}^+ = \hat{\mathbf{F}}^- \text{ and } \bar{\mathbf{F}}^+ = \mathbf{F}^-. \quad (28.0.335)$$

Also show from (3.290) through (3.293) that for the matrices  $\check{\mathbf{L}}$  and  $\check{\mathbf{N}}$  there are the relations

$$\check{\bar{\mathbf{L}}} = \check{\mathbf{L}} \text{ and } \check{\bar{\mathbf{N}}} = \check{\mathbf{N}}. \quad (28.0.336)$$

Verify, therefore, that (3.365) can be rewritten in the form

$$\hat{\mathbf{F}}^- = \exp(\lambda \mathbf{m} \cdot \check{\mathbf{N}}) \exp(\theta \mathbf{n} \cdot \check{\bar{\mathbf{L}}}) \mathbf{F}^-. \quad (28.0.337)$$

Verify, accordingly, that the correspondence (3.364) implies the correspondence

$$\Lambda = \exp(\lambda \mathbf{m} \cdot \mathbf{N}) \exp(\theta \mathbf{n} \cdot \mathbf{L}) \Leftrightarrow \hat{\mathbf{F}}^- = \exp(\lambda \mathbf{m} \cdot \check{\mathbf{N}}) \exp(\theta \mathbf{n} \cdot \check{\bar{\mathbf{L}}}) \mathbf{F}^-, \quad (28.0.338)$$

and vice versa. You already found in Exercise 3.31 that the matrices  $\check{\mathbf{L}}$  and  $\check{\mathbf{N}}$  constitute a basis for the  $\Gamma(1, 0)$  representation of the Lorentz group Lie algebra. Consequently, according to (3.368), the vectors  $\mathbf{F}^-$  carry the  $\Gamma(1, 0)$  representation of the Lorentz Lie algebra/group. Put another way, in transforming according to the rule (3.296), the electromagnetic field tensor  $F^{\mu\nu}$  carries both the Lie algebra/group representations  $\Gamma(0, 1)$  and  $\Gamma(1, 0)$ . Note since (3.296) involves only real quantities when acting on  $F$  and hence on  $\mathbf{E}$  and  $\mathbf{B}$ , this net representation is real. See, for example, (3.352) through (3.357). We would now like to see in more detail how this net representation carries both the  $\Gamma(0, 1)$  and  $\Gamma(1, 0)$  representations. For this, see Exercise 7.3.35.

**28.0.18.** The purpose of this exercise is to develop some general purpose matrix machinery for working with complex matrices that will be of subsequent use.<sup>8</sup> Suppose  $k_1$  and  $k_2$  are two  $n \times n$  possibly complex matrices. Decompose each  $k_j$  into real and imaginary parts by writing

$$k_j = \Re k_j + i \Im k_j \quad (28.0.339)$$

so that

$$\begin{aligned} k_1 k_2 &= (\Re k_1 + i \Im k_1)(\Re k_2 + i \Im k_2) = \\ &(\Re k_1 \Re k_2 - \Im k_1 \Im k_2) + i(\Re k_1 \Im k_2 + \Im k_1 \Re k_2). \end{aligned} \quad (28.0.340)$$

---

<sup>8</sup>This machinery is analogous to some of the machinery in Section 3.9.

Next suppose we define  $k_3$  by the rule

$$k_3 = k_1 k_2 \quad (28.0.341)$$

and also make the decomposition (3.370) for  $k_3$ . Verify it follows from (3.370) through (3.372) that

$$\Re k_3 = \Re k_1 \Re k_2 - \Im k_1 \Im k_2 \quad (28.0.342)$$

and

$$\Im k_3 = \Re k_1 \Im k_2 + \Im k_1 \Re k_2. \quad (28.0.343)$$

Now comes an interesting construction: Given any  $n \times n$  matrix  $k$  define, in terms of  $k$ , a  $2n \times 2n$  real matrix  $K(k)$  by the rule

$$K(k) = \begin{pmatrix} \Re k & -\Im k \\ \Im k & \Re k \end{pmatrix}. \quad (28.0.344)$$

Here each entry on the right side of (3.375) is an  $n \times n$  block.<sup>9</sup> Verify that for *real* scalars  $\lambda$  there is the scalar multiplication result

$$K(\lambda k) = \lambda K(k). \quad (28.0.345)$$

Verify that there is the additive isomorphism

$$K(k_1 + k_2) = K(k_1) + K(k_2). \quad (28.0.346)$$

More remarkably, verify using (3.370) through (3.375) that there is the multiplicative isomorphism

$$K(k_1 k_2) = K(k_1) K(k_2). \quad (28.0.347)$$

That is, show that there is the matrix relation

$$\begin{pmatrix} \Re k_3 & -\Im k_3 \\ \Im k_3 & \Re k_3 \end{pmatrix} = \begin{pmatrix} \Re k_1 & -\Im k_1 \\ \Im k_1 & \Re k_1 \end{pmatrix} \begin{pmatrix} \Re k_2 & -\Im k_2 \\ \Im k_2 & \Re k_2 \end{pmatrix}. \quad (28.0.348)$$

Let  $I^{[n]}$  and  $I^{[2n]}$  be the  $n \times n$  and  $2n \times 2n$  identity matrices, respectively. Verify that

$$K(I^{[n]}) = I^{[2n]}. \quad (28.0.349)$$

Suppose that  $k$  is invertible. Verify that

$$K(k^{-1}) = [K(k)]^{-1}. \quad (28.0.350)$$

For another remarkable result, let  $W$  be the matrix defined by (3.9.12), which we write more precisely as

$$W = \frac{1}{\sqrt{2}} \begin{pmatrix} I^{[n]} & iI^{[n]} \\ iI^{[n]} & I^{[n]} \end{pmatrix}. \quad (28.0.351)$$

---

<sup>9</sup>Note that if  $K$  is known, then  $k$  can be found from a knowledge of the upper and lower left blocks of  $K$ . Thus the mapping between  $k$  and  $K$  is invertible.

Verify the similarity transformation relation

$$\begin{aligned}
 WK(k)W^{-1} &= (1/2) \begin{pmatrix} I^{[n]} & iI^{[n]} \\ iI^{[n]} & I^{[n]} \end{pmatrix} \begin{pmatrix} \Re k & -\Im k \\ \Im k & \Re k \end{pmatrix} \begin{pmatrix} I^{[n]} & -iI^{[n]} \\ -iI^{[n]} & I^{[n]} \end{pmatrix} \\
 &= (1/2) \begin{pmatrix} I^{[n]} & iI^{[n]} \\ iI^{[n]} & I^{[n]} \end{pmatrix} \begin{pmatrix} \Re k + i\Im k & -i\Re k - \Im k \\ \Im k - i\Re k & -i\Im k + \Re k \end{pmatrix} \\
 &= \begin{pmatrix} \Re k + i\Im k & 0 \\ 0 & \Re k - i\Im k \end{pmatrix} = \begin{pmatrix} k & 0 \\ 0 & \bar{k} \end{pmatrix}. \tag{28.0.352}
 \end{aligned}$$

Again, each block occurring in (3.383) is  $n \times n$ . For bonus points show from (3.383) that

$$\det[K(k)] = |\det(k)|^2. \tag{28.0.353}$$

(Recall Exercise 3.3.2.) Finally, associated with the additive and multiplicative isomorphisms (3.376) through (3.378), there are various a Lie algebraic/group properties. Suppose  $b_1$  and  $b_2$  are two possibly complex  $n \times n$  matrices. Using (3.376) through (3.378), verify that

$$\begin{aligned}
 K(\{b_1, b_2\}) &= K(b_1 b_2 - b_2 b_1) = K(b_1 b_2) - K(b_2 b_1) \\
 &= K(b_1)K(b_2) - K(b_2)K(b_1) \\
 &= \{K(b_1), K(b_2)\}.
 \end{aligned} \tag{28.0.354}$$

Thus  $K$  is a Lie product (commutator) isomorphism. Suppose  $b$  is a possibly complex  $n \times n$  matrix. Let us try to evaluate  $K[\exp(b)]$ . Verify that there is the result

$$K[\exp(b)] = K[\sum_{\ell} (1/\ell!) b^{\ell}] = \sum_{\ell} (1/\ell!) K(b^{\ell}) = \sum_{\ell} (1/\ell!) [K(b)]^{\ell} = \exp[K(b)], \tag{28.0.355}$$

which provides a relation between group elements  $\exp(b)$  and group elements  $\exp[K(b)]$ .

**28.0.19.** Review Exercise 7.3.34. The purpose of this exercise is to apply the matrix machinery developed there to the Lorentz group. Make the Ansatz

$$k = \exp(\lambda \mathbf{m} \cdot \check{\mathbf{N}}) \exp(\theta \mathbf{n} \cdot \check{\mathbf{L}}) \tag{28.0.356}$$

so that the right side of (3.364) can be rewritten in the form

$$\hat{\mathbf{F}}^+ = k \mathbf{F}^+. \tag{28.0.357}$$

Decompose  $k$  into real and imaginary parts by writing

$$k = \Re k + i\Im k. \tag{28.0.358}$$

Verify that by so doing (3.388) can be rewritten in the form

$$\hat{\mathbf{E}} + ic\hat{\mathbf{B}} = (\Re k + i\Im k)(\mathbf{E} + ic\mathbf{B}) = (\Re k \mathbf{E} - \Im k c \mathbf{B}) + i(\Im k \mathbf{E} + \Re k c \mathbf{B}). \tag{28.0.359}$$

Now equate real and imaginary parts of (3.390) to obtain, because  $\mathbf{E}$  and  $c\mathbf{B}$  are real, the relations

$$\hat{\mathbf{E}} = \Re k\mathbf{E} - \Im kc\mathbf{B}, \quad (28.0.360)$$

$$c\hat{\mathbf{B}} = \Im k\mathbf{E} + \Re kc\mathbf{B}. \quad (28.0.361)$$

Introduce a *real* six-component vector  $u$  by the rule

$$u = (E_x, E_y, E_z; cB_x, cB_y, cB_z). \quad (28.0.362)$$

Verify that the relations (3.391) and (3.392) can be summarized in the form

$$\hat{u} = K(k)u. \quad (28.0.363)$$

Note that all quantities appearing in (3.394) are real. Verify from (3.384) that if (3.387) holds, then  $\det[K(k)] = 1$ . Moreover, in view of (3.378), the real  $6 \times 6$  matrices  $K(k)$  provide a representation of the Lorentz group; and evidently this is the representation carried by the  $F^{\mu\nu}$ .

Finally, look at (3.383) whose far right side involves  $k$  and  $\bar{k}$ . Verify from (3.367) and (3.387) that

$$\bar{k} = \exp(\lambda \mathbf{m} \cdot \bar{\check{\mathbf{N}}}) \exp(\theta \mathbf{n} \cdot \bar{\check{\mathbf{L}}}) = \exp(\lambda \mathbf{m} \cdot \check{\mathbf{N}}) \exp(\theta \mathbf{n} \cdot \check{\mathbf{L}}). \quad (28.0.364)$$

We already know, from the work of Exercise 3.31 and the earlier discussion in this exercise, that  $\check{\mathbf{L}}$  and  $\check{\mathbf{N}}$  and hence  $k$  carry the  $\Gamma(0, 1)$  representation of the Lorentz Lie algebra/group, and  $\check{\mathbf{L}}$  and  $\check{\mathbf{N}}$  and hence  $\bar{k}$  carry the  $\Gamma(1, 0)$  representation of the Lorentz Lie algebra/group. Note that the right side of (3.383) is *block diagonal* for all  $k$  and that  $W$  as given by (3.382) is *fixed* and thus *independent* of  $k$ . Therefore the representation carried by the  $K(k)$  is *reducible*. It follows that the representation carried by the  $K(k)$  and hence by the  $F^{\mu\nu}$  is *equivalent* to the *direct sum* representation

$$\Gamma(0, 1) \oplus \Gamma(1, 0) = \Gamma(0, 1) \oplus \bar{\Gamma}(0, 1) = \bar{\Gamma}(1, 0) \oplus \Gamma(1, 0). \quad (28.0.365)$$

Here use has also been made of (3.273). Moreover, upon combining (3.383), (3.387), and (3.395), we find for Lorentz group elements the relation

$$WK[\exp(\lambda \mathbf{m} \cdot \check{\mathbf{N}}) \exp(\theta \mathbf{n} \cdot \check{\mathbf{L}})]W^{-1} = \\ \begin{pmatrix} \exp(\lambda \mathbf{m} \cdot \check{\mathbf{N}}) \exp(\theta \mathbf{n} \cdot \check{\mathbf{L}}) & 0 \\ 0 & \exp(\lambda \mathbf{m} \cdot \bar{\check{\mathbf{N}}}) \exp(\theta \mathbf{n} \cdot \bar{\check{\mathbf{L}}}) \end{pmatrix}. \quad (28.0.366)$$

So far we have been employing the construction/definition (3.375) to set up the *group* isomorphism (3.378) between the complex  $3 \times 3$  matrices  $k$  given by (3.387) that provide the  $\Gamma(0, 1)$  representation of the Lorentz group and the real  $6 \times 6$  matrices  $K(k)$  that provide a representation that is equivalent to the representation (3.396). To complete the story, we would like to have a corresponding *Lie algebraic* isomorphism. That is what (3.385) does. Verify that from (3.281) through (3.283) and (3.385) it follows that

$$\{K(\check{L}^j), K(\check{L}^k)\} = \sum_{\ell} \epsilon_{jkl} K(\check{L}^{\ell}), \quad (28.0.367)$$

$$\{K(\check{L}^j), K(\check{N}^k)\} = \sum_{\ell} \epsilon_{jk\ell} K(\check{N}^\ell), \quad (28.0.368)$$

$$\{K(\check{N}^j), K(\check{N}^k)\} = - \sum_{\ell} \epsilon_{jk\ell} K(\check{L}^\ell). \quad (28.0.369)$$

Moreover, the matrices  $K(\check{L}^j)$  and  $K(\check{N}^j)$  are all real: Verify, since the  $\check{L}^j$  are real, it follows that

$$K(\check{L}^j) = \begin{pmatrix} \check{L}^j & 0 \\ 0 & \check{L}^j \end{pmatrix}. \quad (28.0.370)$$

Since the  $\check{N}^j$  are pure imaginary, see (3.280), verify that

$$K(\check{N}^j) = \begin{pmatrix} 0 & -\check{L}^j \\ \check{L}^j & 0 \end{pmatrix}. \quad (28.0.371)$$

Consequently, the  $K(\check{L}^j)$  and  $K(\check{N}^j)$  provide a representation of the Lorentz group Lie algebra by  $6 \times 6$  real matrices. (Is this representation related to the adjoint representation?) Finally, for any  $3 \times 3$  matrix  $b$ , there is the relation

$$WK(b)W^{-1} = \begin{pmatrix} b & 0 \\ 0 & \bar{b} \end{pmatrix}. \quad (28.0.372)$$

See (3.383). It follows that the  $K(\check{L}^j)$  and  $K(\check{N}^j)$  provide, for the Lorentz group Lie algebra, a representation that is equivalent to the representation (3.396).

**28.0.20.** Review Exercise 3.7.41. It examines the relation between the group commutator and the associated Lie-algebraic commutator. In this exercise you will evaluate the Lorentz group commutator for *boosts* along the 3 and 1 axes.<sup>10</sup> Recall from (3.185) that the generators for these boosts satisfy the commutation rule

$$\{N^3, N^1\} = -L^2, \quad (28.0.373)$$

and the fact that this commutator involves a rotation generator is the source of Wigner (1902-1995) rotations and Thomas (1903-1992) precession. According to (3.184) the remaining commutation rules among the generators  $N^3, N^1, L^2$  are

$$\{L^2, N^3\} = N^1, \quad (28.0.374)$$

$$\{L^2, N^1\} = -N^3. \quad (28.0.375)$$

Your task is to evaluate the Lorentz *group* commutator

$$h(s) = \exp(-sN^1) \exp(-sN^3) \exp(sN^1) \exp(sN^3). \quad (28.0.376)$$

That is, your task is to find the net effect of a boost along the 3 axis with rapidity  $s$  followed by a boost along the 1 axis with the same rapidity followed by a boost along the 3 axis with

---

<sup>10</sup>Why not choose the 1 and 2 axes which, from a notational perspective, would be more natural? Although physically there should be no difference, we will see that our choice is computationally more convenient.

rapidity  $-s$  finally followed by a boost along the 1 axis with rapidity  $-s$ . Roughly speaking, we may refer to this operation as the *concatenation* of four boosts along the sides of a square. Note that the rapidities along the 3 and 1 axes add to zero, and therefore we might intuit that there is no net boost. However, in view of (3.251), we might expect some net rotation. Finally note that, in view of the work of Exercise 3.28, you will *equivalently* be exploring some properties of  $Sp(2, \mathbb{R})$ . That is, (3.254) is also an  $Sp(2, \mathbb{R})$  group commutator.

As an initial exploratory step, your first sub-task is to find  $h(s)$  through terms of  $O(s^3)$  making use of the BCH formula (3.7.41). This may be done in steps: Verify that

$$\begin{aligned} \exp(sN^1)\exp(sN^3) &= \exp[sN^1 + sN^3 + (s^2/2)\{N^1, N^3\} \\ &\quad + (s^3/12)\{N^1, \{N^1, N^3\}\} + (s^3/12)\{N^3, \{N^3, N^1\}\} + O(s^4)] = \\ &\exp[sN^1 + sN^3 + (s^2/2)L^2 \\ &\quad + (s^3/12)\{N^1, L^2\} - (s^3/12)\{N^3, L^2\} + O(s^4)] = \\ &\exp[sN^1 + sN^3 + (s^2/2)L^2 + (s^3/12)N^3 + (s^3/12)N^1 + O(s^4)]. \end{aligned} \quad (28.0.377)$$

Similarly, verify that

$$\begin{aligned} \exp(-sN^1)\exp(-sN^3) &= \exp[-sN^1 - sN^3 + (s^2/2)\{N^1, N^3\} \\ &\quad - (s^3/12)\{N^1, \{N^1, N^3\}\} - (s^3/12)\{N^3, \{N^3, N^1\}\} + O(s^4)] = \\ &\exp[-sN^1 - sN^3 + (s^2/2)L^2 \\ &\quad - (s^3/12)\{N^1, L^2\} + (s^3/12)\{N^3, L^2\} + O(s^4)] = \\ &\exp[-sN^1 - sN^3 + (s^2/2)L^2 - (s^3/12)N^3 - (s^3/12)N^1 + O(s^4)]. \end{aligned} \quad (28.0.378)$$

Next show that (3.255) and (3.256) may be combined to give the result

$$\begin{aligned} h(s) &= \exp(-sN^1)\exp(-sN^3)\exp(sN^1)\exp(sN^3) = \\ &\exp[-sN^1 - sN^3 + (s^2/2)L^2 - (s^3/12)N^3 - (s^3/12)N^1 + O(s^4)] \times \\ &\exp[sN^1 + sN^3 + (s^2/2)L^2 + (s^3/12)N^3 + (s^3/12)N^1 + O(s^4)] = \\ &\exp[-sN^1 - sN^3 + (s^2/2)L^2 + O(s^4)] \exp[sN^1 + sN^3 + (s^2/2)L^2 + O(s^4)] = \\ &\exp[s^2L^2 + (1/2)\{-sN^1 - sN^3 + (s^2/2)L^2, sN^1 + sN^3 + (s^2/2)L^2\} + O(s^4)] = \\ &\exp[s^2L^2 + (s^3/2)\{L^2, N^1 + N^3\} + O(s^4)] = \\ &\exp[-(s^3/2)N^3 + (s^3/2)N^1 + O(s^4)] \exp[s^2L^2 + O(s^4)]. \end{aligned} \quad (28.0.379)$$

[Note that, in passing from lines 2 and 3 in (3.257) to subsequent lines, the terms proportional to  $s^3/12$  cancel through  $O(s^3)$ .] Observe that the far right side of (3.257) is written in polar form. [See Subsection 4.2.2, Exercise 4.2.5, Exercise 7.3.30, (3.186), and (3.241).] Therefore, through terms of  $O(s^3)$ , the grand result of the four boosts appearing in the Lorentz group commutator is a net rotation about the 2 axis proportional to  $s^2$  and a boost in the  $-\mathbf{e}_3 + \mathbf{e}_1$  direction proportional to  $s^3$ . [Verify that the same result can be obtained using (3.7.241).] Our intuition about there being no net boost is wrong, but not entirely wrong since there is no boost term proportional to  $s$ . That there is a net boost at all is a relativistic,  $O[(v/c)^3]$ , effect.

The composition of non-collinear boosts appears to be a complicated matter. What can be said about higher order terms? As it stands, the evaluation of (3.254) requires working

with  $4 \times 4$  matrices. However, since what really matters in this situation are the commutation rules which we know are the same for the Lorentz group and  $SL(2, \mathbb{C})$ , we could equally well evaluate the associated  $SL(2, \mathbb{C})$  group commutator

$$\hat{h}(s) = \exp(-s\hat{N}^1) \exp(-s\hat{N}^3) \exp(s\hat{N}^1) \exp(s\hat{N}^3). \quad (28.0.380)$$

This task is simpler because it involves only the exponentiation and multiplication of  $2 \times 2$  matrices, which will make it easier to work to higher order in  $s$ . Recall (3.232) through (3.237) and the results of Exercise 5.7.7 that summarizes various properties of the Pauli matrices. Moreover, in view of (3.233), (3.235), and (2.327), we will be able to work with *real* matrices, which simplifies numerical computation. This fact is the actual reason for choosing the boosts to be along the 3 and 1 axes. Our choice was dictated by the way in which the Pauli matrices have been defined.

Still, there is work to be done. Verify the following preliminary results:

$$\begin{aligned} \exp(s\hat{N}^3) &= \exp[(s/2)\sigma^3] = \cosh[(s/2)\sigma^3] + \sinh[(s/2)\sigma^3] \\ &= I \cosh(s/2) + \sigma^3 \sinh(s/2), \end{aligned} \quad (28.0.381)$$

$$\exp(s\hat{N}^1) = I \cosh(s/2) + \sigma^1 \sinh(s/2), \quad (28.0.382)$$

$$\exp(-s\hat{N}^3) = I \cosh(s/2) - \sigma^3 \sinh(s/2), \quad (28.0.383)$$

$$\exp(-s\hat{N}^1) = I \cosh(s/2) - \sigma^1 \sinh(s/2); \quad (28.0.384)$$

$$\begin{aligned} \exp(s\hat{N}^1) \exp(s\hat{N}^3) &= [I \cosh(s/2) + \sigma^1 \sinh(s/2)][I \cosh(s/2) + \sigma^3 \sinh(s/2)] \\ &= I \cosh^2(s/2) + (\sigma^3 + \sigma^1) \cosh(s/2) \sinh(s/2) + \sigma^1 \sigma^3 \sinh^2(s/2) \\ &= I \cosh^2(s/2) + (1/2)(\sigma^3 + \sigma^1) \sinh(s) - i\sigma^2 \sinh^2(s/2), \end{aligned} \quad (28.0.385)$$

$$\begin{aligned} \exp(-s\hat{N}^1) \exp(-s\hat{N}^3) &= \\ &I \cosh^2(s/2) - (1/2)(\sigma^3 + \sigma^1) \sinh(s) - i\sigma^2 \sinh^2(s/2). \end{aligned} \quad (28.0.386)$$

Next show that combining (3.263) and (3.264) yields the result

$$\begin{aligned} \exp(-s\hat{N}^1) \exp(-s\hat{N}^3) \exp(s\hat{N}^1) \exp(s\hat{N}^3) &= \\ &[I \cosh^2(s/2) - (1/2)(\sigma^3 + \sigma^1) \sinh(s) - i\sigma^2 \sinh^2(s/2)] \times \\ &[I \cosh^2(s/2) + (1/2)(\sigma^3 + \sigma^1) \sinh(s) - i\sigma^2 \sinh^2(s/2)] = \\ &\left[ [I \cosh^2(s/2) - i\sigma^2 \sinh^2(s/2)] - [(1/2)(\sigma^3 + \sigma^1) \sinh(s)] \right] \times \\ &\left[ [I \cosh^2(s/2) - i\sigma^2 \sinh^2(s/2)] + [(1/2)(\sigma^3 + \sigma^1) \sinh(s)] \right] = \\ &[I \cosh^2(s/2) - i\sigma^2 \sinh^2(s/2)]^2 - [(1/2)(\sigma^3 + \sigma^1) \sinh(s)]^2 \\ &- (i/2)\{\sigma^2, (\sigma^3 + \sigma^1)\} \sinh^2(s/2) \sinh(s). \end{aligned} \quad (28.0.387)$$

Look at the contents of the last line of (3.265). It has three pieces, each of which can be expanded/simplified. Verify that for the first piece there is the result

$$\begin{aligned} &[I \cosh^2(s/2) - i\sigma^2 \sinh^2(s/2)]^2 = \\ &I \cosh^4(s/2) - I \sinh^4(s/2) - 2i\sigma^2 \cosh^2(s/2) \sinh^2(s/2) = \\ &I[\cosh^2(s/2) - \sinh^2(s/2)][\cosh^2(s/2) + \sinh^2(s/2)] \\ &- 2i\sigma^2[\cosh(s/2) \sinh(s/2)]^2 = I \cosh(s) - (i/2)\sigma^2 \sinh^2(s). \end{aligned} \quad (28.0.388)$$

Verify that for the second and third pieces there are the results

$$-[(1/2)(\sigma^3 + \sigma^1) \sinh(s)]^2 = -I(1/2) \sinh^2(s) \quad (28.0.389)$$

and

$$-(i/2)\{\sigma^2, (\sigma^3 + \sigma^1)\} \sinh^2(s/2) \sinh(s) = -(\sigma^3 - \sigma^1) \sinh^2(s/2) \sinh(s). \quad (28.0.390)$$

Show that adding the results for these three pieces yields the final result

$$\begin{aligned} \hat{h}(s) &= \exp(-s\hat{N}^1) \exp(-s\hat{N}^3) \exp(s\hat{N}^1) \exp(s\hat{N}^3) = \\ &I \cosh(s) - (i/2)\sigma^2 \sinh^2(s) \\ &- I(1/2) \sinh^2(s) \\ &- (\sigma^3 - \sigma^1) \sinh^2(s/2) \sinh(s) = \\ &I[\cosh(s) - (1/2) \sinh^2(s)] - (i/2)\sigma^2 \sinh^2(s) \\ &- (\sigma^3 - \sigma^1) \sinh^2(s/2) \sinh(s) = \\ &I[\cosh(s) - (1/2) \sinh^2(s)] + \hat{L}^2 \sinh^2(s) \\ &- 2(\hat{N}^3 - \hat{N}^1) \sinh^2(s/2) \sinh(s). \end{aligned} \quad (28.0.391)$$

Examination of the far right side of the result (3.269) shows signs of a rotation about the 2 axis and a boost in the  $-\mathbf{e}_3 + \mathbf{e}_1$  direction. But the result as it stands is not particularly illuminating because it is not written in polar form. What is needed to further clarify the situation is a polar decomposition for  $\hat{h}(s)$  of the form

$$\hat{h}(s) = \exp(\lambda_3 \hat{N}^3 + \lambda_1 \hat{N}^1) \exp(\theta \hat{L}^2). \quad (28.0.392)$$

In this decomposition, which we know by Section 4.2 is always possible, the quantities  $(\lambda_3, \lambda_1)$  specify the net boost, and  $\theta$  specifies the net rotation. Only then can definitive statements be made.<sup>11</sup> You will have the privilege of making this decomposition in the next exercise. Also, a numerical calculation will be described that confirms the correctness of our results.

In the meantime, your next subtask is to show that (3.257) and (3.269) agree through terms of order  $s^3$ . Verify, through terms of order  $s^3$ , that (3.269) has the expansion

$$\begin{aligned} \hat{h}(s) &= \exp(-s\hat{N}^1) \exp(-s\hat{N}^3) \exp(s\hat{N}^1) \exp(s\hat{N}^3) = \\ &I[\cosh(s) - (1/2) \sinh^2(s)] + \hat{L}^2 \sinh^2(s) \\ &- 2(\hat{N}^3 - \hat{N}^1) \sinh^2(s/2) \sinh(s) = \\ &I + s^2 \hat{L}^2 - (s^3/2)(\hat{N}^3 - \hat{N}^1) + O(s^4). \end{aligned} \quad (28.0.393)$$

Using the BCH formula show that (3.271) can be rewritten in the form

$$\begin{aligned} \hat{h}(s) &= \exp(-s\hat{N}^1) \exp(-s\hat{N}^3) \exp(s\hat{N}^1) \exp(s\hat{N}^3) = \\ &I + s^2 \hat{L}^2 - (s^3/2)(\hat{N}^3 - \hat{N}^1) + O(s^4) = \\ &[I - (s^3/2)\hat{N}^3 + (s^3/2)\hat{N}^1 + O(s^4)][I + s^2 \hat{L}^2 + O(s^4)] = \\ &\exp[-(s^3/2)\hat{N}^3 + (s^3/2)\hat{N}^1 + O(s^4)] \exp[s^2 \hat{L}^2 + O(s^4)], \end{aligned} \quad (28.0.394)$$

---

<sup>11</sup>Moreover, although the matrices  $(N^3, N^1, L^2)$  and  $(\hat{N}^3, \hat{N}^1, \hat{L}^2)$  obey the same Lie algebra, they do not obey the same “multiplicative” algebra. For example,  $(\hat{N}^1)^2 = I/4$  but  $(N^1)^2 \neq I/4$ . Therefore, we should only expect agreement/correspondence when comparing Lie algebraic quantities in some Lie algebraic form, such as comparing  $h(s)$  and  $\hat{h}(s)$  in polar form.

in agreement with (3.257). Verity, therefore, that

$$\lambda_3 = -s^3/2 + O(s^4), \quad (28.0.395)$$

$$\lambda_1 = s^3/2 + O(s^4), \quad (28.0.396)$$

$$\theta = s^2 + O(s^4). \quad (28.0.397)$$

**28.0.21.** Review Exercise 3.31. As described at the end of that exercise, what would be definitive would be to have a polar decomposition for  $\hat{h}(s)$  of the form (3.270). The purpose of this exercise is to find and check this desired polar decomposition.

In view of (3.232) through (3.237), let us make a few cosmetic changes in (3.270) that will simplify our computations. What we will seek are quantities  $(\bar{\lambda}_3, \bar{\lambda}_1, \bar{\theta})$  such that

$$\hat{h}(s) = \exp(\bar{\lambda}_3 \sigma^3 + \bar{\lambda}_1 \sigma^1) \exp(\bar{\theta} i \sigma^2). \quad (28.0.398)$$

That is, in passing from (3.270) to (3.276) and in view of (3.232) through (3.237), we have made the (temporary) substitutions

$$\bar{\lambda}_3 = (1/2)\lambda_3 \Leftrightarrow \lambda_3 = 2\bar{\lambda}_3, \quad (28.0.399)$$

$$\bar{\lambda}_1 = (1/2)\lambda_1 \Leftrightarrow \lambda_1 = 2\bar{\lambda}_1, \quad (28.0.400)$$

$$\bar{\theta} = -(1/2)\theta \Leftrightarrow \theta = -2\bar{\theta}. \quad (28.0.401)$$

We will also use the notation

$$\lambda = (\lambda_3^2 + \lambda_1^2)^{1/2}, \quad (28.0.402)$$

$$\bar{\lambda} = (\bar{\lambda}_3^2 + \bar{\lambda}_1^2)^{1/2}, \quad (28.0.403)$$

so that

$$\bar{\lambda} = (1/2)\lambda \Leftrightarrow \lambda = 2\bar{\lambda}. \quad (28.0.404)$$

At this point we could invoke the machinery of Exercise 4.2.5 with the hope of working out the desired results. Instead, let us take a different tack. Observe that the penultimate equation in (3.269), which reads

$$\begin{aligned} \hat{h}(s) &= \\ &I[\cosh(s) - (1/2)\sinh^2(s)] \\ &- (i/2)\sigma^2 \sinh^2(s) \\ &- \sigma^3 \sinh^2(s/2) \sinh(s) \\ &+ \sigma^1 \sinh^2(s/2) \sinh(s), \end{aligned} \quad (28.0.405)$$

provides a Pauli matrix expansion for  $\hat{h}(s)$ . We will expand  $\hat{h}(s)$  as given by (3.276) in terms of the Pauli matrices and then use the orthogonality properties (5.7.42) to match coefficients in the anticipated Pauli matrix expansion and that given by (3.283).

In order to expand (3.276) here are things for you to check: First verify that

$$\begin{aligned} \exp(\bar{\theta} i \sigma^2) &= I + (\bar{\theta} i \sigma^2) + (\bar{\theta} i \sigma^2)^2/2! + (\bar{\theta} i \sigma^2)^3/3! + \dots = \\ &I + \bar{\theta}(i \sigma^2) - I \bar{\theta}^2/2! - (i \sigma^2)(\bar{\theta}^3/3!) + \dots = \\ &I(1 - \bar{\theta}^2/2! + \dots) + i \sigma^2(\bar{\theta} - \bar{\theta}^3/3! + \dots) = \\ &I \cos(\bar{\theta}) + i \sigma^2 \sin(\bar{\theta}). \end{aligned} \quad (28.0.406)$$

Next verify that

$$\begin{aligned}
 \exp(\bar{\lambda}_3\sigma^3 + \bar{\lambda}_1\sigma^1) &= I + (\bar{\lambda}_3\sigma^3 + \bar{\lambda}_1\sigma^1) + (\bar{\lambda}_3\sigma^3 + \bar{\lambda}_1\sigma^1)^2/2! + (\bar{\lambda}_3\sigma^3 + \bar{\lambda}_1\sigma^1)^3/3! + \dots = \\
 I + (\bar{\lambda}_3\sigma^3 + \bar{\lambda}_1\sigma^1) + I(\bar{\lambda}_3^2 + \bar{\lambda}_1^2)/2! + (\bar{\lambda}_3\sigma^1 + \bar{\lambda}_1\sigma^1)(\bar{\lambda}_3^2 + \bar{\lambda}_1^2)/3! + \dots = \\
 I \cosh[(\bar{\lambda}_3^2 + \bar{\lambda}_1^2)^{1/2}] + (\bar{\lambda}_3\sigma^3 + \bar{\lambda}_1\sigma^1)(\bar{\lambda}_3^2 + \bar{\lambda}_1^2)^{-1/2} \sinh[(\bar{\lambda}_3^2 + \bar{\lambda}_1^2)^{1/2}] = \\
 I \cosh(\bar{\lambda}) + (\bar{\lambda}_3\sigma^3 + \bar{\lambda}_1\sigma^1)(1/\bar{\lambda}) \sinh(\bar{\lambda}). \tag{28.0.407}
 \end{aligned}$$

Finally, verify that

$$\begin{aligned}
 \hat{h}(s) &= \exp(\bar{\lambda}_3\sigma^3 + \bar{\lambda}_1\sigma^1) \exp(\bar{\theta}i\sigma^2) = \\
 [I \cosh(\bar{\lambda}) + (\bar{\lambda}_3\sigma^3 + \bar{\lambda}_1\sigma^1)(1/\bar{\lambda}) \sinh(\bar{\lambda})][I \cos(\bar{\theta}) + i\sigma^2 \sin(\bar{\theta})] = \\
 I \cosh(\bar{\lambda}) \cos(\bar{\theta}) \\
 + i\sigma^2 \cosh(\bar{\lambda}) \sin(\bar{\theta}) \\
 + (\bar{\lambda}_3\sigma^3 + \bar{\lambda}_1\sigma^1)(1/\bar{\lambda}) \sinh(\bar{\lambda}) \cos(\bar{\theta}) + \\
 (\bar{\lambda}_3\sigma^3 + \bar{\lambda}_1\sigma^1)(1/\bar{\lambda}) \sinh(\bar{\lambda}) i\sigma^2 \sin(\bar{\theta}) = \\
 I \cosh(\bar{\lambda}) \cos(\bar{\theta}) \\
 + i\sigma^2 \cosh(\bar{\lambda}) \sin(\bar{\theta}) \\
 + \sigma^3[\bar{\lambda}_3 \cos(\bar{\theta}) - \bar{\lambda}_1 \sin(\bar{\theta})](1/\bar{\lambda}) \sinh(\bar{\lambda}) \\
 + \sigma^1[\bar{\lambda}_3 \sin(\bar{\theta}) + \bar{\lambda}_1 \cos(\bar{\theta})](1/\bar{\lambda}) \sinh(\bar{\lambda}). \tag{28.0.408}
 \end{aligned}$$

Now equate terms in the Pauli matrix expansion (3.283) with like terms in the Pauli matrix expansion (3.286). Show that so doing yields the relations

$$\cosh(\bar{\lambda}) \cos(\bar{\theta}) = \cosh(s) - (1/2) \sinh^2(s), \tag{28.0.409}$$

$$\cosh(\bar{\lambda}) \sin(\bar{\theta}) = -(1/2) \sinh^2(s), \tag{28.0.410}$$

$$[\bar{\lambda}_3 \cos(\bar{\theta}) - \bar{\lambda}_1 \sin(\bar{\theta})](1/\bar{\lambda}) \sinh(\bar{\lambda}) = -\sinh^2(s/2) \sinh(s), \tag{28.0.411}$$

$$[\bar{\lambda}_3 \sin(\bar{\theta}) + \bar{\lambda}_1 \cos(\bar{\theta})](1/\bar{\lambda}) \sinh(\bar{\lambda}) = \sinh^2(s/2) \sinh(s). \tag{28.0.412}$$

The terms on the left sides of these relations, which are the unknown terms, come from (3.286). And the terms on the right sides, which are known, come from (3.283).

The last step is to solve (3.287) through (3.290) for  $\bar{\lambda}_3$ ,  $\bar{\lambda}_1$ , and  $\bar{\theta}$ . Upon dividing (3.288) by (3.287), show that

$$\tan(\bar{\theta}) = -(1/2) \sinh^2(s)/[\cosh(s) - (1/2) \sinh^2(s)]. \tag{28.0.413}$$

Upon squaring (3.287) and (3.288) and adding the results, show that

$$\cosh^2(\bar{\lambda}) = [\cosh(s) - (1/2) \sinh^2(s)]^2 + (1/4)[\sinh(s)]^4. \tag{28.0.414}$$

The results (3.291) and (3.292) determine  $\bar{\theta}$  and  $\bar{\lambda}$  as functions of  $s$ .

We would also like formulas for  $\bar{\lambda}_3$  and  $\bar{\lambda}_1$  as functions of  $s$ . Upon multiplying (3.289) by  $\cos(\bar{\theta})$  and (3.290) by  $\sin(\bar{\theta})$  and adding the results, show that

$$\bar{\lambda}_3(1/\bar{\lambda}) \sinh(\bar{\lambda}) = [-\cos(\bar{\theta}) + \sin(\bar{\theta})] \sinh^2(s/2) \sinh(s) \tag{28.0.415}$$

so that

$$\bar{\lambda}_3 = [-\cos(\bar{\theta}) + \sin(\bar{\theta})][\sinh^2(s/2) \sinh(s)]/[(1/\bar{\lambda}) \sinh(\bar{\lambda})]. \quad (28.0.416)$$

Upon multiplying (3.289) by  $[-\sin(\bar{\theta})]$  and (3.290) by  $\cos(\bar{\theta})$  and adding the results, show that

$$\bar{\lambda}_1(1/\bar{\lambda}) \sinh(\bar{\lambda}) = [\cos(\bar{\theta}) + \sin(\bar{\theta})] \sinh^2(s/2) \sinh(s) \quad (28.0.417)$$

so that

$$\bar{\lambda}_1 = [\cos(\bar{\theta}) + \sin(\bar{\theta})][\sinh^2(s/2) \sinh(s)]/[(1/\bar{\lambda}) \sinh(\bar{\lambda})]. \quad (28.0.418)$$

Considerable algebra has gone by in passing from (3.270) to (3.296). How do we know there have been no mistakes in deriving the results (3.291), (3.292), (3.294), and (3.296)? Shortly, as a sanity check, we will make low-order expansions in  $s$  for these results to verify that these expansions agree with what we already know. Then we will also describe numerical checks.

But before doing so, and assuming the correctness of (3.294) and (3.296), it is instructive to examine the direction of the net boost associated with the Lorentz group commutator. In view of (3.277), (3.278), and (3.280) through (3.282), define an angle  $\chi$  by the relations

$$\cos(\chi) = \bar{\lambda}_3/\bar{\lambda} = \lambda_3/\lambda, \quad (28.0.419)$$

$$\sin(\chi) = \bar{\lambda}_1/\bar{\lambda} = \lambda_1/\lambda, \quad (28.0.420)$$

$$\tan(\chi) = \bar{\lambda}_1/\bar{\lambda}_3 = \lambda_1/\lambda_3. \quad (28.0.421)$$

Show from (3.294), (3.296), and (3.299) that

$$\begin{aligned} \tan(\chi) &= -[\cos(\bar{\theta}) + \sin(\bar{\theta})]/[\cos(\bar{\theta}) - \sin(\bar{\theta})] = \\ &= -[\cos(\theta/2) - \sin(\theta/2)]/[\cos(\theta/2) + \sin(\theta/2)]. \end{aligned} \quad (28.0.422)$$

We have learned that the direction of the net boost depends simply on  $\bar{\theta}$  or, equivalently, on  $\theta$ .

As promised, we now turn to making low-order expansions in  $s$ . Begin by verifying the expansions

$$\tan^{-1}(\psi) = \psi - \psi^3/3 + \dots, \quad (28.0.423)$$

$$-(1/2) \sinh^2(s)/[\cosh(s) - (1/2) \sinh^2(s)] = -(1/2)s^2 + O(s^4). \quad (28.0.424)$$

Show if follows from (3.291) that

$$\bar{\theta} = -(1/2)s^2 + O(s^4). \quad (28.0.425)$$

Verify that employing (3.279) in (3.299) yields the result

$$\theta = s^2 + O(s^4), \quad (28.0.426)$$

in agreement with (3.275).

Next consider the relation (3.292). Verify, say with the use of *Mathematica* or by hand calculation, that there is the expansion

$$[\cosh(s) - (1/2) \sinh^2(s)]^2 + (1/4)[\sinh(s)]^4 = 1 + s^6/8 + O(s^8). \quad (28.0.427)$$

Note the remarkable fact that the coefficients of  $s^2$  and  $s^4$  vanish! Verify also the expansions

$$\cosh(\bar{\lambda}) = 1 + \bar{\lambda}^2/2! + O(\bar{\lambda}^4), \quad (28.0.428)$$

$$\cosh^2(\bar{\lambda}) = 1 + \bar{\lambda}^2 + O(\bar{\lambda}^4). \quad (28.0.429)$$

Show it follows, using (3.292) and (3.301 through (3.303), that

$$\bar{\lambda}^2 = s^6/8 + O(s^8), \quad (28.0.430)$$

and consequently

$$\bar{\lambda}(s) = |s^3|(1/\sqrt{8}) + O(s^5). \quad (28.0.431)$$

Finally, with the use of (3.282), show that

$$\lambda(s) = |s^3|(1/\sqrt{2}) + O(s^5). \quad (28.0.432)$$

How does the result (3.061) compare with (3.273) and (3.274)? From (3.273), (3.274), and (3.280) show that

$$\lambda(s) = |s^3|(1/\sqrt{2}) + O(s^4). \quad (28.0.433)$$

Evidently (3.306) and (3.307) are consistent.

What remains is to examine the small  $s$  behavior of  $\bar{\lambda}_3$  and  $\bar{\lambda}_1$  as given by (3.294) and (3.296). Begin by verifying the expansions

$$\cos(\bar{\theta}) = 1 - \bar{\theta}^2/2 + \dots = 1 - (1/8)s^4 + O(s^6), \quad (28.0.434)$$

$$\sin(\bar{\theta}) = \bar{\theta} - (\bar{\theta})^3/6 + \dots = -(1/2)s^2 + O(s^4), \quad (28.0.435)$$

$$[\mp \cos(\bar{\theta}) + \sin(\bar{\theta})] = \mp 1 - (1/2)s^2 + O(s^4), \quad (28.0.436)$$

$$\sinh^2(s/2) \sinh(s) = s^3/4 + O(s^5), \quad (28.0.437)$$

$$(1/\bar{\lambda}) \sinh(\bar{\lambda}) = 1 + O(s^6). \quad (28.0.438)$$

Show, therefore, that (3.299) and (3.301) have the expansions

$$\bar{\lambda}_3 = -s^3/4 - s^5/8 + \dots, \quad (28.0.439)$$

$$\bar{\lambda}_1 = s^3/4 - s^5/8 + \dots, \quad (28.0.440)$$

and consequently  $\lambda_3$  and  $\lambda_1$  have the expansions

$$\lambda_3 = -s^3/2 - s^5/4 + \dots, \quad (28.0.441)$$

$$\lambda_1 = s^3/2 - s^5/4 + \dots, \quad (28.0.442)$$

in agreement with (3.278) and (3.279).

The last item of interest with regard to low-order expansions is the behavior of  $\chi(s)$ . From (3.319) and (3.320) we see that

$$\begin{aligned} \tan(\chi) &= \lambda_1/\lambda_3 = (s^3/2 - s^5/4 + \dots)/(-s^3/2 - s^5/4 + \dots) \\ &= -(1 - s^2/2 + \dots)/(1 + s^2/2 + \dots) = -(1 - s^2 + \dots) \\ &= -1 + s^2 + \dots. \end{aligned} \quad (28.0.443)$$

As expected from (3.300), (3.303), and (3.304),  $\chi$  is indeed  $s$  dependent.

To fulfill our last promise, we now describe how the results (3.291), (3.292), (3.294), and (3.296) can also be checked numerically for any  $s$  using the charged particle beam transport code MaryLie. Because it is based on Lie-algebraic methods, MaryLie is capable of performing various operations related to the symplectic group and, as we have seen, what we are seeking are various relations among elements of  $Sp(2, \mathbb{R})$ . In particular for our present purposes, MaryLie can perform the following operations numerically:

1. Given a quadratic polynomial  $f_2$ , it can compute the symplectic matrix  $M$  associated with the linear symplectic map  $\mathcal{M}$  in the relation

$$\mathcal{M} = \exp(: f_2 :). \quad (28.0.444)$$

2. It can multiply symplectic matrices thereby implementing group-element multiplication for the symplectic group.
3. Given a symplectic matrix  $M$ , let  $\mathcal{M}$  be the associated linear symplectic map. With  $M$  as input, MaryLie can compute the quadratic polynomials  $f_2^a$  and  $f_2^c$  in the decomposition

$$\mathcal{M} = \exp(: f_2^c :)\exp(: f_2^a :). \quad (28.0.445)$$

See Section 7.6. That is, MaryLie can carry out (orthogonal) polar decomposition of symplectic matrices.

How can these tools be employed in the present context? Examination of (5.6.6), (5.6.7), and (5.6.11) through (5.6.14) shows that there are the following correspondences between Pauli matrices and quadratic polynomials:

$$f = (1/2)(-q^2 + p^2) \leftrightarrow \sigma^1, \quad (28.0.446)$$

$$b_0 = (1/2)(q^2 + p^2) \leftrightarrow J = i\sigma^2, \quad (28.0.447)$$

$$g = qp \leftrightarrow \sigma^3. \quad (28.0.448)$$

(Recall that these correspondences were set up in Section 5.5.) Consequently, there is also the correspondence

$$\begin{aligned} \hat{h}(s) &= \exp(-s\hat{N}^1)\exp(-s\hat{N}^3)\exp(s\hat{N}^1)\exp(s\hat{N}^3) = \\ &\exp[(s/2)\sigma^1]\exp[(s/2)\sigma^3]\exp[(-s/2)\sigma^1]\exp[(-s/2)\sigma^3] \leftrightarrow \\ &\exp[(s/2) : qp :] \exp[(s/4) : -q^2 + p^2 :] \exp[(-s/2) : qp :] \exp[(-s/4) : -q^2 + p^2 :]. \end{aligned} \quad (28.0.449)$$

[Note that the order of the Lie transformations appearing in the right side of the correspondence (3.327) is opposite to the order of the related matrices on the left side of (3.327). This reversal in order is to be expected. See the discussion in Section 8.3.] What we will use MaryLie to carry out numerically are the operations

$$\begin{aligned} &\exp[(s/2) : qp :] \exp[(s/4) : -q^2 + p^2 :] \exp[(-s/2) : qp :] \exp[(-s/4) : -q^2 + p^2 :] \\ &= \exp(: f_2^c :)\exp(: f_2^a :). \end{aligned} \quad (28.0.450)$$

That is, it will compute and multiply the four maps on the left side of (3.328) and then express the result in the factored product form shown on the right side. Specifically, it will output the quadratic polynomials  $f_2^c$  and  $f_2^a$ . When this is done, in view of (3.276) and the correspondences (3.324) through (3.326), we expect  $f_2^c$  and  $f_2^a$  will be given by the relations

$$f_2^c = \bar{\theta}b_0 = \bar{\theta}(1/2)(q^2 + p^2) \quad (28.0.451)$$

and

$$f_2^a = \bar{\lambda}_3g + \bar{\lambda}_1f = \bar{\lambda}_3qp + \bar{\lambda}_1(1/2)(-q^2 + p^2). \quad (28.0.452)$$

#### Exhibit 7.3.1: Sample MaryLie Run

**28.0.22.** Review Exercises 3.27, 3.31, and 3.32. Exercise 3.27 found, among other things, the concatenation rule for two collinear boosts: rapidities simply add. And Exercises 3.31 and 3.32 found the concatenation rule for four boosts along the sides of a square. The purpose of this exercise is to find the concatenation rule for two non-collinear boosts. Specifically, given two real three-component vectors  $\mu$  and  $\nu$ , we wish to study group element  $k(\mu, \nu)$  defined by the product

$$k(\mu, \nu) = \exp(\nu \cdot N) \exp(\mu \cdot N). \quad (28.0.453)$$

Observe that the vectors  $\mu$  and  $\nu$  determine (or in the collinear case lie in) a plane which for convenience, and without loss of generality, may be taken to be the 3,1 plane. Therefore we may make the decompositions

$$\mu = \mu_3 e_3 + \mu_1 e_1, \quad (28.0.454)$$

$$\nu = \nu_3 e_3 + \nu_1 e_1, \quad (28.0.455)$$

so that

$$\mu \cdot N = \mu_3 N^3 + \mu_1 N^1, \quad (28.0.456)$$

$$\nu \cdot N = \nu_3 N^3 + \nu_1 N^1, \quad (28.0.457)$$

and

$$\mu \times \nu = (\mu_3 \nu_1 - \mu_1 \nu_3) e_2. \quad (28.0.458)$$

What we are interested in is finding the vector

$$\lambda = \lambda_3 e_3 + \lambda_1 e_1 \quad (28.0.459)$$

and the angle  $\theta$  such that

$$\exp(\nu \cdot N) \exp(\mu \cdot N) = \exp(\lambda \cdot N) \exp(\theta L^2). \quad (28.0.460)$$

To get a feel for what to expect, we may use the BCH formula to combine exponents on the left side of (3.338). Verify, through terms quadratic in the components of  $\mu$  and  $\nu$ , that

$$\begin{aligned} \exp(\nu \cdot N) \exp(\mu \cdot N) &= \exp[(\mu + \nu) \cdot N + \{\nu \cdot N, \mu \cdot N\}/2 + \dots] = \\ &\exp[(\mu + \nu) \cdot N + \dots] \exp[\{\nu \cdot N, \mu \cdot N\}/2 + \dots]. \end{aligned} \quad (28.0.461)$$

Next show that

$$\{\boldsymbol{\nu} \cdot \mathbf{N}, \boldsymbol{\mu} \cdot \mathbf{N}\} = (\mu_3\nu_1 - \mu_1\nu_3)L^2 = (\boldsymbol{\mu} \times \boldsymbol{\nu}) \cdot \mathbf{L}. \quad (28.0.462)$$

Conclude, upon comparing the right sides of (3.338) and (3.339), that there are the results

$$\boldsymbol{\lambda} = \boldsymbol{\mu} + \boldsymbol{\nu} + \dots, \quad (28.0.463)$$

$$\theta = (1/2)(\boldsymbol{\mu} \times \boldsymbol{\nu}) \cdot \mathbf{e}_2 + \dots = (1/2)(\mu_3\nu_1 - \mu_1\nu_3) + \dots. \quad (28.0.464)$$

What we are next interested in are the higher-order terms in (3.341) and (3.342).

Again, since what really matters in this situation are the commutation rules which we know are the same for the Lorentz group and  $SL(2, \mathbb{C})$ , we can equally well evaluate the simpler associated  $SL(2, \mathbb{C})$  [and  $Sp(2, \mathbb{R})$ ] function

$$\hat{k}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \exp(\boldsymbol{\nu} \cdot \hat{\mathbf{N}}) \exp(\boldsymbol{\mu} \cdot \hat{\mathbf{N}}). \quad (28.0.465)$$

In this case (3.338) becomes

$$\exp(\boldsymbol{\nu} \cdot \hat{\mathbf{N}}) \exp(\boldsymbol{\mu} \cdot \hat{\mathbf{N}}) = \exp(\boldsymbol{\lambda} \cdot \hat{\mathbf{N}}) \exp(\theta \hat{L}^2) \quad (28.0.466)$$

or, in terms of Pauli matrices,

$$\exp(\boldsymbol{\nu} \cdot \boldsymbol{\sigma}/2) \exp(\boldsymbol{\mu} \cdot \boldsymbol{\sigma}/2) = \exp(\boldsymbol{\lambda} \cdot \boldsymbol{\sigma}/2) \exp(-\theta i \sigma^2/2). \quad (28.0.467)$$

To further simplify calculations, and in analogy to what was done in Exercise 3.32, we will also make the (temporary) substitutions

$$\bar{\boldsymbol{\mu}} = (1/2)\boldsymbol{\mu} \Leftrightarrow \boldsymbol{\mu} = 2\bar{\boldsymbol{\mu}}, \text{ etc.}; \quad (28.0.468)$$

$$\bar{\theta} = -(1/2)\theta \Leftrightarrow \theta = -2\bar{\theta}, \quad (28.0.469)$$

so that

$$\exp(\boldsymbol{\nu} \cdot \boldsymbol{\sigma}/2) = \exp(\bar{\nu}_3\sigma^3 + \bar{\nu}_1\sigma^1), \text{ etc.} \quad (28.0.470)$$

and

$$\exp(-\theta i \sigma^2/2) = \exp(\bar{\theta} i \sigma^2). \quad (28.0.471)$$

Moreover, we will use the notation

$$\mu = (\mu_3^2 + \mu_1^2)^{1/2}, \text{ etc.}, \quad (28.0.472)$$

$$\bar{\mu} = (\bar{\mu}_3^2 + \bar{\mu}_1^2)^{1/2}, \text{ etc.}, \quad (28.0.473)$$

so that

$$\bar{\boldsymbol{\mu}} = (1/2)\boldsymbol{\mu} \Leftrightarrow \boldsymbol{\mu} = 2\bar{\boldsymbol{\mu}}, \text{ etc.} \quad (28.0.474)$$

What we will seek to find are the quantities  $(\bar{\lambda}_3, \bar{\lambda}_1, \bar{\theta})$  such that

$$\exp(\bar{\nu}_3\sigma^3 + \bar{\nu}_1\sigma^1) \exp(\bar{\mu}_3\sigma^3 + \bar{\mu}_1\sigma^1) = \exp(\bar{\lambda}_3\sigma^3 + \bar{\lambda}_1\sigma^1) \exp(\bar{\theta} i \sigma^2). \quad (28.0.475)$$

Consult again Exercise 3.32 to observe that it contains almost all the results necessary to complete the current exercise. For example, in view of (3.285), there is the analogous result

$$\begin{aligned} \exp(\bar{\nu}_3\sigma^3 + \bar{\nu}_1\sigma^1) &= \exp(\bar{\nu} \cdot \boldsymbol{\sigma}) = \\ I \cosh(\bar{\nu}) + (\bar{\nu}_3\sigma^3 + \bar{\nu}_1\sigma^1)(1/\bar{\nu}) \sinh(\bar{\nu}) &= \\ I \cosh(\bar{\nu}) + \bar{\nu} \cdot \boldsymbol{\sigma}(1/\bar{\nu}) \sinh(\bar{\nu}), \end{aligned} \quad (28.0.476)$$

and there is an analogous result involving  $\bar{\mu}$ . Finally, the right side of (3.353) has already been treated in (3.286).

Carry out the next step by showing that

$$\begin{aligned} \exp(\bar{\nu}_3\sigma^3 + \bar{\nu}_1\sigma^1) \exp(\bar{\mu}_3\sigma^3 + \bar{\mu}_1\sigma^1) &= \exp(\bar{\nu} \cdot \boldsymbol{\sigma}) \exp(\bar{\mu} \cdot \boldsymbol{\sigma}) = \\ [I \cosh(\bar{\nu}) + \bar{\nu} \cdot \boldsymbol{\sigma}(1/\bar{\nu}) \sinh(\bar{\nu})][I \cosh(\bar{\mu}) + \bar{\mu} \cdot \boldsymbol{\sigma}(1/\bar{\mu}) \sinh(\bar{\mu})] &= \\ I \cosh(\bar{\nu}) \cosh(\bar{\mu}) & \\ + \bar{\mu} \cdot \boldsymbol{\sigma} \cosh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu}) & \\ + \bar{\nu} \cdot \boldsymbol{\sigma} \cosh(\bar{\mu})(1/\bar{\nu}) \sinh(\bar{\nu}) & \\ + (\bar{\nu} \cdot \boldsymbol{\sigma}) \cdot (\bar{\mu} \cdot \boldsymbol{\sigma})(1/\bar{\nu}) \sinh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu}) &= \\ I \cosh(\bar{\nu}) \cosh(\bar{\mu}) & \\ + [\bar{\mu} \cosh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu}) + \bar{\nu} \cosh(\bar{\mu})(1/\bar{\nu}) \sinh(\bar{\nu})] \cdot \boldsymbol{\sigma} & \\ + [I \bar{\nu} \cdot \bar{\mu} + i(\bar{\nu} \times \bar{\mu}) \cdot \boldsymbol{\sigma}](1/\bar{\nu}) \sinh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu}) &= \\ I[\cosh(\bar{\nu}) \cosh(\bar{\mu}) + \bar{\nu} \cdot \bar{\mu}(1/\bar{\nu}) \sinh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu})] & \\ + i\sigma^2(\bar{\nu}_3\bar{\mu}_1 - \bar{\nu}_1\bar{\mu}_3)(1/\bar{\nu}) \sinh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu}) & \\ + [\bar{\mu}_3 \cosh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu}) + \bar{\nu}_3 \cosh(\bar{\mu})(1/\bar{\nu}) \sinh(\bar{\nu})]\sigma^3 & \\ + [\bar{\mu}_1 \cosh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu}) + \bar{\nu}_1 \cosh(\bar{\mu})(1/\bar{\nu}) \sinh(\bar{\nu})]\sigma^1. \end{aligned} \quad (28.0.477)$$

Now equate terms in the Pauli matrix expansion (3.286) with like terms in the Pauli matrix expansion (3.355). Show that so doing yields the relations

$$\cosh(\bar{\lambda}) \cos(\bar{\theta}) = \cosh(\bar{\nu}) \cosh(\bar{\mu}) + \bar{\nu} \cdot \bar{\mu}(1/\bar{\nu}) \sinh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu}), \quad (28.0.478)$$

$$\cosh(\bar{\lambda}) \sin(\bar{\theta}) = (\bar{\nu}_3\bar{\mu}_1 - \bar{\nu}_1\bar{\mu}_3)(1/\bar{\nu}) \sinh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu}), \quad (28.0.479)$$

$$[\bar{\lambda}_3 \cos(\bar{\theta}) - \bar{\lambda}_1 \sin(\bar{\theta})](1/\bar{\lambda}) \sinh(\bar{\lambda}) = \bar{\mu}_3 \cosh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu}) + \bar{\nu}_3 \cosh(\bar{\mu})(1/\bar{\nu}) \sinh(\bar{\nu}), \quad (28.0.480)$$

$$[\bar{\lambda}_3 \sin(\bar{\theta}) + \bar{\lambda}_1 \cos(\bar{\theta})](1/\bar{\lambda}) \sinh(\bar{\lambda}) = \bar{\mu}_1 \cosh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu}) + \bar{\nu}_1 \cosh(\bar{\mu})(1/\bar{\nu}) \sinh(\bar{\nu}). \quad (28.0.481)$$

The terms on the left sides of these relations, which are the unknown terms, come from (3.286). And the terms on the right sides, which are known, come from (3.355).

The last step is to solve (3.356) through (3.359) for  $\bar{\lambda}_3$ ,  $\bar{\lambda}_1$ , and  $\bar{\theta}$ . Upon dividing (3.357) by (3.356), show that

$$\begin{aligned} \tan(\bar{\theta}) &= (\bar{\nu}_3\bar{\mu}_1 - \bar{\nu}_1\bar{\mu}_3)(1/\bar{\nu}) \sinh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu}) \times \\ &\quad [\cosh(\bar{\nu}) \cosh(\bar{\mu}) + \bar{\nu} \cdot \bar{\mu}(1/\bar{\nu}) \sinh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu})]^{-1}. \end{aligned} \quad (28.0.482)$$

Upon squaring (3.356) and (3.357) and adding the results, show that

$$\begin{aligned}\cosh^2(\bar{\lambda}) &= [\cosh(\bar{\nu}) \cosh(\bar{\mu}) + \bar{\nu} \cdot \bar{\mu}(1/\bar{\nu}) \sinh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu})]^2 \\ &\quad + [(\bar{\nu}_3\bar{\mu}_1 - \bar{\nu}_1\bar{\mu}_3)(1/\bar{\nu}) \sinh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu})]^2.\end{aligned}\quad (28.0.483)$$

The results (3.360) and (3.361) determine  $\bar{\theta}$  and  $\bar{\lambda}$  as functions of  $\bar{\mu}$  and  $\bar{\nu}$ .

We would also like formulas for  $\bar{\lambda}_3$  and  $\bar{\lambda}_1$  as functions of  $\bar{\mu}$  and  $\bar{\nu}$ . Upon multiplying (3.358) by  $\cos(\bar{\theta})$  and (3.359) by  $\sin(\bar{\theta})$  and adding the results, show that

$$\begin{aligned}\bar{\lambda}_3(1/\bar{\lambda}) \sinh(\bar{\lambda}) &= [\bar{\mu}_3 \cos(\bar{\theta}) + \bar{\mu}_1 \sin(\bar{\theta})] \cosh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu}) \\ &\quad + [\bar{\nu}_3 \cos(\bar{\theta}) + \bar{\nu}_1 \sin(\bar{\theta})] \cosh(\bar{\mu})(1/\bar{\nu}) \sinh(\bar{\nu})\end{aligned}\quad (28.0.484)$$

so that

$$\begin{aligned}\bar{\lambda}_3 &= \{[\bar{\mu}_3 \cos(\bar{\theta}) + \bar{\mu}_1 \sin(\bar{\theta})] \cosh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu}) \\ &\quad + [\bar{\nu}_3 \cos(\bar{\theta}) + \bar{\nu}_1 \sin(\bar{\theta})] \cosh(\bar{\mu})(1/\bar{\nu}) \sinh(\bar{\nu})\} \times \\ &\quad [(1/\bar{\lambda}) \sinh(\bar{\lambda})]^{-1}.\end{aligned}\quad (28.0.485)$$

Upon multiplying (3.358) by  $[-\sin(\bar{\theta})]$  and (3.359) by  $\cos(\bar{\theta})$  and adding the results, show that

$$\begin{aligned}\bar{\lambda}_1(1/\bar{\lambda}) \sinh(\bar{\lambda}) &= [-\bar{\mu}_3 \sin(\bar{\theta}) + \bar{\mu}_1 \cos(\bar{\theta})] \cosh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu}) \\ &\quad + [-\bar{\nu}_3 \sin(\bar{\theta}) + \bar{\nu}_1 \cos(\bar{\theta})] \cosh(\bar{\mu})(1/\bar{\nu}) \sinh(\bar{\nu})\end{aligned}\quad (28.0.486)$$

so that

$$\begin{aligned}\bar{\lambda}_1 &= \{[-\bar{\mu}_3 \sin(\bar{\theta}) + \bar{\mu}_1 \cos(\bar{\theta})] \cosh(\bar{\nu})(1/\bar{\mu}) \sinh(\bar{\mu}) \\ &\quad + [-\bar{\nu}_3 \sin(\bar{\theta}) + \bar{\nu}_1 \cos(\bar{\theta})] \cosh(\bar{\mu})(1/\bar{\nu}) \sinh(\bar{\nu})\} \times \\ &\quad [(1/\bar{\lambda}) \sinh(\bar{\lambda})]^{-1}.\end{aligned}\quad (28.0.487)$$

The results (3.360), (3.361), (3.363), and (3.365) can also be checked numerically for any  $\mu$  and  $\nu$  using the charged particle beam transport code MaryLie. How can MaryLie tools be employed in the present context? Based on the correspondences (3.324) through (3.326) there is also the correspondence

$$\begin{aligned}\exp(\nu \cdot \hat{N}) \exp(\mu \cdot \hat{N}) &= \exp(\nu \cdot \sigma/2) \exp(\mu \cdot \sigma/2) = \\ \exp(\bar{\nu}_3 \sigma^3 + \bar{\nu}_1 \sigma^1) \exp(\bar{\mu}_3 \sigma^3 + \bar{\mu}_1 \sigma^1) &\leftrightarrow \\ \exp[: \bar{\mu}_3 qp + \bar{\mu}_1(1/2)(-q^2 + p^2) :] \exp[: \bar{\nu}_3 qp + \bar{\nu}_1(1/2)(-q^2 + p^2) :].\end{aligned}\quad (28.0.488)$$

What we will use MaryLie to carry out numerically are the operations

$$\begin{aligned}\exp[: \bar{\mu}_3 qp + \bar{\mu}_1(1/2)(-q^2 + p^2) :] \exp[: \bar{\nu}_3 qp + \bar{\nu}_1(1/2)(-q^2 + p^2) :] \\ = \exp(: f_2^c : ) \exp(: f_2^a : ).\end{aligned}\quad (28.0.489)$$

That is, it will compute and multiply the two maps on the left side of (3.367) and then express the result in the factored product form shown on the right side. Specifically, it will

output the quadratic polynomials  $f_2^c$  and  $f_2^a$ . When this is done, as in (3.329) and (3.330), we expect  $f_2^c$  and  $f_2^a$  will be given by the relations

$$f_2^c = \bar{\theta}b_0 = \bar{\theta}(1/2)(q^2 + p^2) \quad (28.0.490)$$

and

$$f_2^a = \bar{\lambda}_3g + \bar{\lambda}_1f = \bar{\lambda}_3qp + \bar{\lambda}_1(1/2)(-q^2 + p^2). \quad (28.0.491)$$

#### Exhibit 7.3.2: Sample MaryLie Run

**28.0.23.** Review Exercise 3.7.31. 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  $SO(3, \mathbb{R})$ . The purpose of this exercise is to show that there is a two-to-one homomorphism between  $SU(2)$  and  $SO(3, \mathbb{R})$ . We will also find several formulas, involving  $SU(2)$  and  $SO(3, \mathbb{R})$  and their Lie algebras, that will be useful for later work.

Suppose  $v \in SU(2)$ . Consider matrices  $\bar{K}^\alpha(v)$  defined by the relation

$$\bar{K}^\alpha(v) = v^\dagger K^\alpha v. \quad (28.0.492)$$

Verify that the  $\bar{K}^\alpha(v)$  are anti-Hermitian and traceless. It follows, since the  $K^\beta$  form a basis for the set of  $2 \times 2$  traceless anti-Hermitian matrices, that there must be a relation of the form

$$\bar{K}^\alpha(v) = \sum_\beta M_{\alpha\beta}(v) K^\beta \quad (28.0.493)$$

where  $M(v)$  is a  $3 \times 3$  matrix to be determined. Show, in view of the definitions (3.7.169) through (3.7.171), that (2.32) is equivalent to the relations

$$v^\dagger \sigma^\alpha v = \sum_\beta M(v)_{\alpha\beta} \sigma^\beta, \quad (28.0.494)$$

which may be viewed as defining  $M(v)$ . Indeed, from this result deduce, with the aid of (3.7.168), the relation

$$M_{\alpha\beta}(v) = (1/2)\text{tr}(v^\dagger \sigma^\alpha v \sigma^\beta). \quad (28.0.495)$$

Let us find some of the properties of  $M(v)$ . Verify that

$$M(I) = I, \quad (28.0.496)$$

$$M(-I) = I, \quad (28.0.497)$$

and, more generally,

$$M(-v) = M(v). \quad (28.0.498)$$

Verify also that

$$M_{\beta\alpha}(v) = (1/2)\text{tr}(v^\dagger \sigma^\beta v \sigma^\alpha) = (1/2)\text{tr}(v \sigma^\alpha v^\dagger \sigma^\beta) = M_{\alpha\beta}(v^\dagger) \quad (28.0.499)$$

so that there is the relation

$$M^T(v) = M(v^\dagger). \quad (28.0.500)$$

[Note that  $-v \in SU(2)$  and  $v^\dagger \in SU(2)$  if  $v \in SU(2)$ .]

Evidently (2.34) is a rule that sends any matrix  $v \in SU(2)$  to a corresponding matrix  $M(v)$ . Moreover, for any two  $SU(2)$  elements  $v_1$  and  $v_2$ , this rule has the property

$$M(v_1 v_2) = M(v_1) M(v_2), \quad (28.0.501)$$

and therefore is in fact a group homomorphism. To verify this assertion, show that

$$\begin{aligned} M_{\alpha\beta}(v_1 v_2) &= (1/2)\text{tr}[(v_1 v_2)^\dagger \sigma^\alpha v_1 v_2 \sigma^\beta] = (1/2)\text{tr}[(v_2)^\dagger (v_1)^\dagger \sigma^\alpha v_1 v_2 \sigma^\beta] \\ &= (1/2) \sum_\gamma M(v_1)_{\alpha\gamma} \text{tr}(v_2^\dagger \sigma^\gamma v_2 \sigma^\beta) = \sum_\gamma M(v_1)_{\alpha\gamma} M(v_2)_{\gamma\beta} = [M(v_1) M(v_2)]_{\alpha\beta}. \end{aligned} \quad (28.0.502)$$

Check the chain of deductions

$$I = M(I) = M(v^\dagger v) = M(v^\dagger) M(v) = M^T(v) M(v) \quad (28.0.503)$$

to conclude that  $M$  is orthogonal and that

$$M^{-1}(v) = M(v^{-1}). \quad (28.0.504)$$

Argue, based on (2.35) and the topology (connectedness) of  $SU(2)$ , that, by continuity,  $M$  must have determinant +1, and therefore  $M \in SO(3, \mathbb{R})$ . [We will see below that  $M(v)$  is a real matrix.] Thus, (2.34) provides a map from  $SU(2)$  to  $SO(3, \mathbb{R})$  and, in view of (2.37) and (2.40), this map is a two-to-one homomorphism.

Even more explicit results are possible: Suppose that  $v$  is parameterized as in (3.7.187). Then (2.31) and (2.32) can be rewritten in the form

$$\begin{aligned} \bar{K}^\alpha(\theta, \mathbf{n}) &= v(\theta, \mathbf{n})^\dagger K^\alpha v(\theta, \mathbf{n}) \\ &= \exp(-\theta \mathbf{n} \cdot \mathbf{K}) K^\alpha \exp(\theta \mathbf{n} \cdot \mathbf{K}) \\ &= \sum_\beta M_{\alpha\beta}(\theta, \mathbf{n}) K^\beta, \end{aligned} \quad (28.0.505)$$

and (2.35) takes the form

$$M(0, \mathbf{n}) = I. \quad (28.0.506)$$

Next we will find a differential equation for  $M$ . Show that differentiating the first and third terms in (2.44) yields the result

$$\begin{aligned} \partial_\theta \bar{K}^\alpha(\theta, \mathbf{n}) &= \partial_\theta [\exp(-\theta \mathbf{n} \cdot \mathbf{K}) K^\alpha \exp(\theta \mathbf{n} \cdot \mathbf{K})] \\ &= \exp(-\theta \mathbf{n} \cdot \mathbf{K}) \{K^\alpha, \mathbf{n} \cdot \mathbf{K}\} \exp(\theta \mathbf{n} \cdot \mathbf{K}). \end{aligned} \quad (28.0.507)$$

But, according to (3.7.183) and (3.7.201), there is the relation

$$\begin{aligned} \{K^\alpha, \mathbf{n} \cdot \mathbf{K}\} &= \{\mathbf{e}_\alpha \cdot \mathbf{K}, \mathbf{n} \cdot \mathbf{K}\} = (\mathbf{e}_\alpha \times \mathbf{n}) \cdot \mathbf{K} \\ &= -(\mathbf{n} \times \mathbf{e}_\alpha) \cdot \mathbf{K} = -[(\mathbf{n} \cdot \mathbf{L}) \mathbf{e}_\alpha] \cdot \mathbf{K} \end{aligned} \quad (28.0.508)$$

so that (2.46) can also be written in the form

$$\begin{aligned}
 \partial_\theta \bar{K}^\alpha(\theta, \mathbf{n}) &= -\exp(-\theta \mathbf{n} \cdot \mathbf{K})[(\mathbf{n} \cdot \mathbf{L})\mathbf{e}_\alpha] \cdot \mathbf{K} \exp(\theta \mathbf{n} \cdot \mathbf{K}) \\
 &= -\sum_\beta \exp(-\theta \mathbf{n} \cdot \mathbf{K})[(\mathbf{n} \cdot \mathbf{L})\mathbf{e}_\alpha]_\beta \mathbf{K}^\beta \exp(\theta \mathbf{n} \cdot \mathbf{K}) \\
 &= \sum_\beta (\mathbf{n} \cdot \mathbf{L})_{\alpha\beta} \exp(-\theta \mathbf{n} \cdot \mathbf{K}) \mathbf{K}^\beta \exp(\theta \mathbf{n} \cdot \mathbf{K}) \\
 &= \sum_{\beta\gamma} (\mathbf{n} \cdot \mathbf{L})_{\alpha\beta} M_{\beta\gamma}(\theta, \mathbf{n}) K^\gamma = \sum_\gamma [(\mathbf{n} \cdot \mathbf{L})M(\theta, \mathbf{n})]_{\alpha\gamma} K^\gamma. \tag{28.0.509}
 \end{aligned}$$

Verify, by working with components, that here we have used the antisymmetry of the  $L^\alpha$  to correctly make the calculation

$$\begin{aligned}
 -[(\mathbf{n} \cdot \mathbf{L})\mathbf{e}_\alpha]_\beta &= -(\mathbf{e}_\beta, [\mathbf{n} \cdot \mathbf{L}]\mathbf{e}_\alpha) \\
 &= -[\mathbf{n} \cdot \mathbf{L}]_{\beta\alpha} = [\mathbf{n} \cdot \mathbf{L}]_{\alpha\beta}. \tag{28.0.510}
 \end{aligned}$$

On the other hand, differentiating the first and last terms in (2.44) and changing summation indices yields the result

$$\partial_\theta \bar{K}^\alpha(\theta, \mathbf{n}) = \sum_\gamma [\partial_\theta M(\theta, \mathbf{n})]_{\alpha\gamma} K^\gamma. \tag{28.0.511}$$

By comparing (2.48) and (2.50) conclude that  $M$  satisfies the differential equation

$$\partial_\theta M(\theta, \mathbf{n}) = (\mathbf{n} \cdot \mathbf{L})M(\theta, \mathbf{n}). \tag{28.0.512}$$

Show that (2.51) with the initial condition (2.45) has the unique solution

$$M(\theta, \mathbf{n}) = \exp(\theta \mathbf{n} \cdot \mathbf{L}) = R(\theta, \mathbf{n}). \tag{28.0.513}$$

You have demonstrated that

$$\exp(-\theta \mathbf{n} \cdot \mathbf{K}) K^\alpha \exp(\theta \mathbf{n} \cdot \mathbf{K}) = \sum_\beta R(\theta, \mathbf{n})_{\alpha\beta} K^\beta, \tag{28.0.514}$$

and (2.34) becomes

$$R_{\alpha\beta}(v) = (1/2)\text{tr}(v^\dagger \sigma^\alpha v \sigma^\beta) \Leftrightarrow R[\exp(\theta \mathbf{n} \cdot \mathbf{K})] = \exp(\theta \mathbf{n} \cdot \mathbf{L}). \tag{28.0.515}$$

(Here the symbol  $\Leftrightarrow$  is used to indicate logical implication in both directions.) Should a relation of the form (2.54) be surprising? Not from a group-theoretic perspective. The matrices  $v$  and  $v^\dagger$  on the right side of (2.54) each carry a spin 1/2 representation of  $SU(2)$ . The matrix  $R$  on the left carries a spin 1 representation. We know that two spin 1/2 representations can be combined to produce a spin 1 representation, and evidently the Pauli matrices  $\sigma^\alpha$  and  $\sigma^\beta$  on the right act as Clebsch-Gordan coefficients to pick out this representation.

To find further consequences of (2.53), multiply both sides by  $a_\alpha$  and sum over  $\alpha$  to get the result

$$\begin{aligned} & \exp(-\theta \mathbf{n} \cdot \mathbf{K}) (\sum_{\alpha} a_{\alpha} K^{\alpha}) \exp(\theta \mathbf{n} \cdot \mathbf{K}) = \sum_{\alpha\beta} R(\theta, \mathbf{n})_{\alpha\beta} a_{\alpha} K^{\beta} \\ &= \sum_{\alpha\beta} R^T(\theta, \mathbf{n})_{\beta\alpha} a_{\alpha} K^{\beta} = \sum_{\alpha\beta} R^{-1}(\theta, \mathbf{n})_{\beta\alpha} a_{\alpha} K^{\beta}. \end{aligned} \quad (28.0.516)$$

Verify that (2.55) can be written in the more compact form

$$\begin{aligned} & \exp(-\theta \mathbf{n} \cdot \mathbf{K}) (\mathbf{a} \cdot \mathbf{K}) \exp(\theta \mathbf{n} \cdot \mathbf{K}) = \sum_{\alpha\beta} R^{-1}(\theta, \mathbf{n})_{\beta\alpha} a_{\alpha} K^{\beta} \\ &= \sum_{\beta} [R^{-1}(\theta, \mathbf{n}) \mathbf{a}]_{\beta} K^{\beta} = [R^{-1}(\theta, \mathbf{n}) \mathbf{a}] \cdot \mathbf{K}. \end{aligned} \quad (28.0.517)$$

Show, by making the replacement  $\theta \rightarrow -\theta$ , that there is also the general result

$$\exp(\theta \mathbf{n} \cdot \mathbf{K}) (\mathbf{a} \cdot \mathbf{K}) \exp(-\theta \mathbf{n} \cdot \mathbf{K}) = [R(\theta, \mathbf{n}) \mathbf{a}] \cdot \mathbf{K}. \quad (28.0.518)$$

Note that, according to (3.7.200), the matrix  $R$  is produced/generated by exponentiating elements in the adjoint representation of  $SU(2)$ . Recall from Section 3.7.7 that the adjoint representation is defined completely in terms of the structure constants. It can be shown that the relations (2.56) and (2.57) hold for *any* set of matrices  $K^\alpha$  that satisfy the commutation relations (3.7.173). See Section 8.1. Show, for example, that

$$\exp(-\psi L^j) L^k \exp(\psi L^j) = L^k \cos \psi - \{L^j, L^k\} \sin \psi \text{ for } j \neq k. \quad (28.0.519)$$

Verify the general result

$$R(\mathbf{a} \cdot \mathbf{L}) R^{-1} = (Ra) \cdot \mathbf{L}. \quad (28.0.520)$$

Suppose that  $\mathbf{a}$  and  $\mathbf{b}$  are any three-component vectors. As an application of (2.59), verify that

$$R(\mathbf{a} \times \mathbf{b}) = R(\mathbf{a} \cdot \mathbf{L}) \mathbf{b} = R(\mathbf{a} \cdot \mathbf{L}) R^{-1} R \mathbf{b} = [(Ra) \cdot \mathbf{L}] R \mathbf{b} = (Ra) \times (R \mathbf{b}), \quad (28.0.521)$$

as expected from the geometric definition of the cross product. Here we have also used (3.7.201). Suppose  $O$  is a  $3 \times 3$  orthogonal matrix. Show that

$$\begin{aligned} O(\mathbf{a} \times \mathbf{b}) &= O(\mathbf{a} \cdot \mathbf{L}) \mathbf{b} = O(\mathbf{a} \cdot \mathbf{L}) O^{-1} O \mathbf{b} \\ &= \det(O) [(Oa) \cdot \mathbf{L}] O \mathbf{b} = \det(O) [(Oa) \times (Ob)], \end{aligned} \quad (28.0.522)$$

on account of which  $\mathbf{a} \times \mathbf{b}$  is called a *pseudo* vector if  $\mathbf{a}$  and  $\mathbf{b}$  are vectors.

Observe also that, according to (3.7.189) through (3.7.191), for fixed  $\mathbf{n}$  we must have  $\theta \in [0, 4\pi]$  to achieve a closed path in  $SU(2)$ ; and, by (3.7.200), (3.7.203), and (3.7.204), in doing so an associated closed path in  $SO(3, \mathbb{R})$  gets covered *twice*. Therefore, as already asserted, the homomorphism between  $SU(2)$  and  $SO(3, \mathbb{R})$  is two to one. For a further discussion of the topologies of  $SU(2)$  and  $SO(3, \mathbb{R})$ , see Exercise 8.2.11.

Let us use the relation (2.54) to explore once again the relation between  $su(2)$  and  $so(3, \mathbb{R})$ . This can be done by studying (2.54) for elements  $v$  near the identity. Suppose  $v$  is written in the form

$$v = \exp(\epsilon K) = I + \epsilon K + O(\epsilon^2) \quad (28.0.523)$$

where  $\epsilon$  is small and  $K \in su(2)$  is therefore any  $2 \times 2$  traceless anti-Hermitian matrix. That is, following the terminology of Exercise 3.7.31,  $K$  can be written in the form

$$K = \mathbf{a} \cdot \mathbf{K} = \sum_{\gamma=1}^3 a_\gamma K^\gamma = \sum_{\gamma=1}^3 a_\gamma (-i/2) \sigma^\gamma \quad (28.0.524)$$

where  $\mathbf{a}$  is a real vector. Then we have

$$v^\dagger = \exp(\epsilon K^\dagger) = I + \epsilon K^\dagger + O(\epsilon^2), \quad (28.0.525)$$

and (2.54) yields

$$\begin{aligned} R_{\alpha\beta}(v) &= (1/2)\text{tr}[(I + \epsilon K^\dagger)\sigma^\alpha(I + \epsilon K)\sigma^\beta] + O(\epsilon^2) \\ &= (1/2)\text{tr}(\sigma^\alpha\sigma^\beta) + (\epsilon/2)\text{tr}(\sigma^\alpha K\sigma^\beta + K^\dagger\sigma^\alpha\sigma^\beta) + O(\epsilon^2) \\ &= \delta_{\alpha\beta} + \epsilon L_{\alpha\beta} + O(\epsilon^2) = [\exp(\epsilon L)]_{\alpha\beta} + O(\epsilon^2) \end{aligned} \quad (28.0.526)$$

where

$$L_{\alpha\beta}(K) = (1/2)\text{tr}(\sigma^\alpha K\sigma^\beta + K^\dagger\sigma^\alpha\sigma^\beta). \quad (28.0.527)$$

Show, using the fact that  $K$  is anti-Hermitian and the properties of the trace operation and of the Pauli matrices (see Exercise 5.7.7), that (2.66) can also be written in the form

$$\begin{aligned} L_{\alpha\beta}(\mathbf{a} \cdot \mathbf{K}) &= L_{\alpha\beta}(K) = (1/2)\text{tr}(\sigma^\alpha K\sigma^\beta + K^\dagger\sigma^\alpha\sigma^\beta) = (1/2)\text{tr}(K\sigma^\beta\sigma^\alpha - K\sigma^\alpha\sigma^\beta) \\ &= (-1/2)(-i/2) \sum_{\gamma=1}^3 a_\gamma \text{tr}(\sigma^\gamma \{\sigma^\alpha, \sigma^\beta\}) = (-4i)(-1/2)(-i/2) \sum_{\gamma=1}^3 a_\gamma (L^\gamma)_{\alpha\beta} \\ &= (\mathbf{a} \cdot \mathbf{L})_{\alpha\beta}. \end{aligned} \quad (28.0.528)$$

Thus, the relations (2.66) and (2.67) provide an explicit isomorphism between  $su(2)$  and  $so(3, \mathbb{R})$ . Indeed, we have the relations

$$\{L(K), L(K')\} = L(\{K, K'\}) \text{ and, specifically, } \{L(K^\alpha), L(K^\beta)\} = L(\{K^\alpha, K^\beta\}). \quad (28.0.529)$$

**28.0.24.** The purpose of this exercise is to study the topology of  $SU(2)$  and  $SO(3, \mathbb{R})$ . Consider all points in three-dimensional space of the form  $\theta \mathbf{n}$  where  $\mathbf{n}$  is an arbitrary unit vector and  $0 \leq \theta \leq \theta_{\max}$ . They evidently comprise the interior and surface of a ball in 3-dimensional space with radius  $\theta_{\max}$ . By looking at (3.7.203), show that the topology of  $SO(3, \mathbb{R})$  is the same as that of the interior and surface of a ball of radius  $\pi$  with opposite pairs of points on the surface identified. By looking at (3.7.189), show that the topology of  $SU(2)$  is the same as that of the interior and surface of a ball of radius  $2\pi$  with all points on the surface identified.

The topology of  $SU(2)$  can also be examined without use of the exponential function. Suppose  $v$  is any  $2 \times 2$  matrix with complex entries written in the form

$$v = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}. \quad (28.0.530)$$

Now require that  $v$  be unitary. Show that the conditions

$$v^\dagger v = vv^\dagger = I \quad (28.0.531)$$

yield, among others, the relations

$$\delta = \bar{\alpha} \quad (28.0.532)$$

and

$$\gamma = -\bar{\beta} \quad (28.0.533)$$

so that  $v$  takes the form

$$v = \begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix}. \quad (28.0.534)$$

Next introduce the general parameterizations

$$\alpha = w_0 + iw_3 \quad (28.0.535)$$

and

$$\beta = w_2 + iw_1 \quad (28.0.536)$$

where all the  $w_j$  are real. Show that, in terms of these parameters,  $v$  takes the form

$$v = \begin{pmatrix} w_0 + iw_3 & w_2 + iw_1 \\ -w_2 + iw_1 & w_0 - iw_3 \end{pmatrix}. \quad (28.0.537)$$

Verify that  $v$  can also be written in the form

$$v = w_0 \sigma^0 + i \sum_{j=1}^3 w_j \sigma^j. \quad (28.0.538)$$

Lastly, show that requiring  $v$  to have determinant 1 yields the relation

$$\det(v) = \sum_{j=0}^3 w_j^2 = 1, \quad (28.0.539)$$

and that this relation also guarantees that  $v$  as given by (2.76) or (2.77) is unitary.

Equation (2.78) is that for  $S^3$ , the 3-dimensional surface of a sphere in 4-dimensional space. Thus,  $SU(2)$  has the topology of  $S^3$ .<sup>12</sup> This manifold is known to be *simply connected*, and therefore  $SU(2)$  is simply connected. Simply connected means that any closed curve can be shrunk to a point. By contrast,  $SO(3, \mathbb{R})$  is not simply connected. Consider the ball

---

<sup>12</sup>Put another way, the manifold  $S^3$  can be given a group structure, namely that of  $SU(2)$ . It can be shown that  $S^1$  and  $S^3$  are the only spheres that can be given a group structure.

with radius  $\pi$  that parameterizes  $SO(3, \mathbb{R})$ . Let  $P$  be any path in the ball that stretches between antipodal points on the surface of the ball. Since antipodal points on the surface are to be identified, this path is a closed curve. Show that it cannot be shrunk to a point while remaining closed. A more detailed study shows that  $SO(3, \mathbb{R})$  is doubly connected. Given a multiply connected manifold, there is a standard procedure in topology for constructing an associated singly-connected manifold. This singly-connected manifold is called the *covering* manifold of the original manifold. In the same spirit,  $SU(2)$  is said to be the covering group of  $SO(3, \mathbb{R})$ .

It can be shown that all the  $SO(n, \mathbb{R})$  for  $n \geq 3$  are doubly connected. [We have already learned that  $SO(2, \mathbb{R})$  is infinitely connected. See Section 5.9.1.] Consequently each has a two-fold covering group. These groups are called  $Spin(n, \mathbb{R})$ . For small  $n$  there are the redundancies  $Spin(3, \mathbb{R}) = SU(2)$ ,  $Spin(4, \mathbb{R}) = SU(2) \times SU(2)$ ,  $Spin(5, \mathbb{R}) = USp(4)$ , and  $Spin(6, \mathbb{R}) = SU(4)$ .

Let return to the case of  $SO(3, \mathbb{R})$ . Comparison of (5.10.22) and (2.78), and reference to Exercises 5.10.13 and 5.10.14, show that  $v$  is a unit quaternion matrix. For this reason, the quantities  $w_0 \cdots w_3$  are sometimes called quaternion parameters. The quaternion parameterization of  $SU(2)$  can be extended to a quaternion parameterization of  $SO(3, \mathbb{R})$  with the aid of (2.54) and (2.77).<sup>13</sup> Show that doing so gives the result

$$R_{\alpha\beta}(w) = \delta_{\alpha\beta} (w_0^2 - \sum_{\gamma=1}^3 w_\gamma^2) + 2w_\alpha w_\beta + 2w_0 \sum_{\gamma=1}^3 \epsilon_{\alpha\beta\gamma} w_\gamma. \quad (28.0.540)$$

Students of dynamics or quantum mechanics may be familiar with the use of Euler angles to parameterize elements in both  $SU(2)$  and  $SO(3, \mathbb{R})$ . For example, we may write

$$R(\phi, \theta, \psi) = \exp(\phi L^3) \exp(\theta L^2) \exp(\psi L^3). \quad (28.0.541)$$

See (3.7.195) and (3.7.208). However, when studying rigid-body dynamics, this is not always a good idea because the Euler angles  $\phi$  and  $\psi$  are not uniquely defined when  $\theta = 0$  and  $\theta = \pi$ . [Only the quantity  $(\phi + \psi)$  plays a role when  $\theta = 0$ , and only the quantity  $(\phi - \psi)$  plays a role when  $\theta = \pi$ .] That is, the quantities  $\phi, \theta, \psi$  do not provide good coordinate patches in the neighborhoods  $\theta \simeq 0$  and  $\theta \simeq \pi$ . Correspondingly, the equations of motion for rigid-body motion in terms of Euler angles have singularities at these values of  $\theta$ , and are therefore not well suited for numerical integration.<sup>14</sup> By contrast, the equations of motion are regular everywhere when quaternion parameters are employed. The only penalty to be paid for this advantage is that equations of motion must be integrated for four parameters instead of three. Moreover, the equations of motion preserve the relation (2.78), and the extent to which numerical integration preserves this relation can be used as a check on the accuracy of the procedure. For further discussion, see Section 11.1.

**28.0.25.** In Exercise 3.7.35 you verified that the Lie algebras  $su(4)$  and  $so(6, \mathbb{R})$  have the same dimension, namely dimension 15. The purpose of this exercise is to verify that  $su(4)$

---

<sup>13</sup>The so called *Cayley-Klein* parameters for specifying rotations are closely related to quaternion parameters. Also, sometimes quaternion parameters are called *Euler-Rodrigues* parameters.

<sup>14</sup>The associated problems encountered in numerical integration are sometimes referred to as *gimbal lock*. Google the words *Euler angle evil*.

and  $so(6, \mathbb{R})$  are in fact the same (equivalent) over the real field. Moreover, we will learn that 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 3.7.31, 8.2.10, and 8.2.11.

We begin by exploiting a mathematical fact familiar from the relativistic treatment of electromagnetism. Let  $A$  be a general antisymmetric  $4 \times 4$  matrix. It can be written in the form

$$A = \begin{pmatrix} 0 & -B_z & B_y & E_x/c \\ B_z & 0 & -B_x & E_y/c \\ -B_y & B_x & 0 & E_z/c \\ -E_x/c & -E_y/c & -E_z/c & 0 \end{pmatrix} \quad (28.0.542)$$

where the quantities  $E_\alpha/c$  and  $B_\alpha$  are arbitrary. See (1.6.56). Then we know from its use in relativistic electromagnetic theory that there is the *mathematical* identity

$$\det(A) = [(1/c)\mathbf{E} \cdot \mathbf{B}]^2. \quad (28.0.543)$$

See Exercise 1.6.17. At this point introduce variables  $z_1, z_2, \dots, z_6$  by the rules

$$E_z/c = iz_1 + z_2, \quad (28.0.544)$$

$$B_z = -iz_1 + z_2, \quad (28.0.545)$$

$$E_x/c = iz_3 + z_4, \quad (28.0.546)$$

$$B_x = -iz_3 + z_4, \quad (28.0.547)$$

$$E_y/c = iz_5 + z_6, \quad (28.0.548)$$

$$B_y = -iz_5 + z_6. \quad (28.0.549)$$

(While workable, this ordering may seem a little strange. It will be of use in Exercise 27.5.4.) Verify that in terms of these variables there is the relation

$$(1/c)\mathbf{E} \cdot \mathbf{B} = \sum_{\alpha=1}^6 z_\alpha^2. \quad (28.0.550)$$

Let us write

$$A = A(z) = A(z_1 \cdots z_6) = \sum_{\alpha=1}^6 z_\alpha A^\alpha \quad (28.0.551)$$

where the  $A^\alpha$  are matrices to be determined. Show, using (2.82) and (2.89), that

$$\det(A) = \left[ \sum_{\alpha=1}^6 z_\alpha^2 \right]^2. \quad (28.0.552)$$

Verify that

$$A(z) = \begin{pmatrix} 0 & iz_1 - z_2 & -iz_5 + z_6 & iz_3 + z_4 \\ -iz_1 + z_2 & 0 & iz_3 - z_4 & iz_5 + z_6 \\ iz_5 - z_6 & -iz_3 + z_4 & 0 & iz_1 + z_2 \\ -iz_3 - z_4 & -iz_5 - z_6 & -iz_1 - z_2 & 0 \end{pmatrix} \quad (28.0.553)$$

so that the  $A^\alpha$  are given by the relations

$$A^1 = \begin{pmatrix} 0 & i & 0 & 0 \\ -i & 0 & -0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & -i & 0 \end{pmatrix} = - \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}, \quad (28.0.554)$$

$$A^2 = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} = i \begin{pmatrix} -\sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}, \quad (28.0.555)$$

$$A^3 = \begin{pmatrix} 0 & 0 & 0 & i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ -i & 0 & 0 & 0 \end{pmatrix} = i \begin{pmatrix} 0 & \sigma^1 \\ -\sigma^1 & 0 \end{pmatrix}, \quad (28.0.556)$$

$$A^4 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} = i \begin{pmatrix} 0 & \sigma^2 \\ \sigma^2 & 0 \end{pmatrix}, \quad (28.0.557)$$

$$A^5 = \begin{pmatrix} 0 & 0 & -i & 0 \\ 0 & 0 & 0 & i \\ i & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \end{pmatrix} = i \begin{pmatrix} 0 & -\sigma^3 \\ \sigma^3 & 0 \end{pmatrix}, \quad (28.0.558)$$

$$A^6 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \sigma^0 \\ -\sigma^0 & 0 \end{pmatrix}. \quad (28.0.559)$$

Here use has been made of the Pauli matrices given by (5.7.3). Evidently, the  $A^\alpha$  span the space of  $4 \times 4$  antisymmetric matrices when working over the complex field.

Verify that the  $A^\alpha$  have the properties

$$(A^\alpha)^T = -A^\alpha, \quad (28.0.560)$$

$$(A^\alpha)^\dagger = -(-1)^\alpha A^\alpha, \quad (28.0.561)$$

and that they obey the multiplication rules

$$A^\alpha (A^\alpha)^\dagger = (A^\alpha)^\dagger A^\alpha = I, \quad (28.0.562)$$

$$A^1 A^2 = A^2 A^1 = i \begin{pmatrix} \sigma^0 & 0 \\ 0 & -\sigma^0 \end{pmatrix}, \quad (28.0.563)$$

$$A^1 A^3 = -A^3 A^1 = -i A^5, \quad (28.0.564)$$

$$A^1 A^4 = A^4 A^1 = -i \begin{pmatrix} 0 & \sigma^0 \\ \sigma^0 & 0 \end{pmatrix}, \quad (28.0.565)$$

$$A^1 A^5 = -A^5 A^1 = i A^3, \quad (28.0.566)$$

$$A^1 A^6 = A^6 A^1 = \begin{pmatrix} 0 & -\sigma^2 \\ \sigma^2 & 0 \end{pmatrix}, \quad (28.0.567)$$

$$A^2 A^3 = A^3 A^2 = -i \begin{pmatrix} 0 & \sigma^3 \\ \sigma^3 & 0 \end{pmatrix}, \quad (28.0.568)$$

$$A^2 A^4 = -A^4 A^2 = A^6, \quad (28.0.569)$$

$$A^2 A^5 = A^5 A^2 = -i \begin{pmatrix} 0 & \sigma^1 \\ \sigma^1 & 0 \end{pmatrix}, \quad (28.0.570)$$

$$A^2 A^6 = -A^6 A^2 = -A^4, \quad (28.0.571)$$

$$A^3 A^4 = A^4 A^3 = i \begin{pmatrix} -\sigma^3 & 0 \\ 0 & \sigma^3 \end{pmatrix}, \quad (28.0.572)$$

$$A^3 A^5 = -A^5 A^3 = -i A^1, \quad (28.0.573)$$

$$A^3 A^6 = A^6 A^3 = -i \begin{pmatrix} \sigma^1 & 0 \\ 0 & \sigma^1 \end{pmatrix}, \quad (28.0.574)$$

$$A^4 A^5 = A^5 A^4 = i \begin{pmatrix} -\sigma^1 & 0 \\ 0 & \sigma^1 \end{pmatrix}, \quad (28.0.575)$$

$$A^4 A^6 = -A^6 A^4 = A^2, \quad (28.0.576)$$

$$A^5 A^6 = A^6 A^5 = i \begin{pmatrix} \sigma^3 & 0 \\ 0 & \sigma^3 \end{pmatrix}. \quad (28.0.577)$$

Verify, by looking at (2.102) through (2.116), that there is the rule

$$A^\alpha A^\beta = (-1)(-1)^{\alpha+\beta} A^\beta A^\alpha \text{ for } \alpha \neq \beta. \quad (28.0.578)$$

Show that combining (2.100) and (2.117) gives the relations

$$A^\alpha (A^\beta)^\dagger = -A^\beta (A^\alpha)^\dagger \text{ for } \alpha \neq \beta, \quad (28.0.579)$$

$$(A^\alpha)^\dagger A^\beta = -(A^\beta)^\dagger A^\alpha \text{ for } \alpha \neq \beta. \quad (28.0.580)$$

Show that combining (2.101), (2.118), and (2.119) gives the relations

$$A^\alpha (A^\beta)^\dagger + A^\beta (A^\alpha)^\dagger = (A^\alpha)^\dagger A^\beta + (A^\beta)^\dagger A^\alpha = 2\delta_{\alpha\beta} I. \quad (28.0.581)$$

Finally, verify that the right sides of (2.102) through (2.116) are traceless. Combine this fact with (2.101) to derive the result

$$\text{tr}[A^\alpha (A^\beta)^\dagger] = 4\delta_{\alpha\beta}. \quad (28.0.582)$$

For the moment, let  $v$  be any  $4 \times 4$  matrix. Use it to define quantities  $\bar{A}^\alpha(v)$  by the rule

$$\bar{A}^\alpha(v) = v^T A^\alpha v. \quad (28.0.583)$$

Note the similarity of (2.122) to (2.31) except that  $\dagger$  has been replaced by  $T$ . Verify that the  $\bar{A}^\alpha(v)$  are antisymmetric for any choice of  $v$ . It follows, since the  $A^\alpha$  form a basis for the set of  $4 \times 4$  antisymmetric matrices, that there must be a relation of the form

$$\bar{A}^\alpha(v) = v^T A^\alpha v = \sum_\beta R_{\alpha\beta}(v) A^\beta \quad (28.0.584)$$

where  $R(v)$  is a  $6 \times 6$  matrix to be determined. Verify, in view of (2.121), that there is the explicit formula

$$R_{\alpha\beta}(v) = (1/4)\text{tr}[v^T A^\alpha v (A^\beta)^\dagger]. \quad (28.0.585)$$

Let us find some of the properties of  $R(v)$ . Verify that

$$R(I) = I, \quad (28.0.586)$$

$$R(-I) = I, \quad (28.0.587)$$

and, more generally,

$$R(-v) = R(v). \quad (28.0.588)$$

The rule (2.124) is also a homomorphism,

$$R(v_1 v_2) = R(v_1) R(v_2). \quad (28.0.589)$$

Check this assertion by verifying the computation

$$\begin{aligned} R_{\alpha\beta}(v_1 v_2) &= (1/4)\text{tr}[(v_1 v_2)^T A^\alpha v_1 v_2 (A^\beta)^\dagger] = (1/4)\text{tr}[v_2^T v_1^T A^\alpha v_1 v_2 (A^\beta)^\dagger] \\ &= (1/4) \sum_\gamma R(v_1)_{\alpha\gamma} \text{tr}[v_2^T A^\gamma v_2 (A^\beta)^\dagger] = \sum_\gamma R(v_1)_{\alpha\gamma} R(v_2)_{\gamma\beta} = [R(v_1) R(v_2)]_{\alpha\beta}. \end{aligned} \quad (28.0.590)$$

From (2.123) deduce the relation

$$v^T \left[ \sum_\alpha z_\alpha A^\alpha \right] v = \sum_\alpha z_\alpha v^T A^\alpha v = \sum_\beta \left[ \sum_\alpha z_\alpha R_{\alpha\beta}(v) \right] A^\beta. \quad (28.0.591)$$

Define variables  $\hat{z}_\beta$  by writing

$$\hat{z}_\beta = \sum_\alpha z_\alpha R_{\alpha\beta}(v). \quad (28.0.592)$$

Show that (2.130) can be written more compactly in the form

$$v^T A(z) v = A(\hat{z}). \quad (28.0.593)$$

Take the determinant of both sides of (2.132). Show that doing so yields, in view of (2.91), the result

$$[\det(v)]^2 \left[ \sum_\alpha z_\alpha^2 \right]^2 = \left[ \sum_\beta \hat{z}_\beta^2 \right]^2. \quad (28.0.594)$$

Take the square roots of both sides of (2.133) to find the result

$$[\det(v)] \left[ \sum_{\alpha} z_{\alpha}^2 \right] = \left[ \sum_{\beta} \hat{z}_{\beta}^2 \right]. \quad (28.0.595)$$

In taking this square root there is, of course, a sign ambiguity. This ambiguity is overcome by employing (2.125) and imposing continuity. For a Pfaffian-based approach that avoids this ambiguity, see Exercise 8.2.13.

Next manipulate the ingredients on the right side of (2.134) to show that

$$\begin{aligned} \sum_{\beta} \hat{z}_{\beta}^2 &= \sum_{\beta} \left[ \sum_{\alpha} z_{\alpha} R_{\alpha\beta} \right] \left[ \sum_{\gamma} z_{\gamma} R_{\gamma\beta} \right] \\ &= \sum_{\alpha\beta\gamma} z_{\alpha} R_{\alpha\beta} z_{\gamma} R_{\gamma\beta} = \sum_{\alpha\beta\gamma} z_{\alpha} R_{\alpha\beta} (R^T)_{\beta\gamma} z_{\gamma} = \sum_{\alpha\gamma} z_{\alpha} (RR^T)_{\alpha\gamma} z_{\gamma}. \end{aligned} \quad (28.0.596)$$

It follows that there is the relation

$$\det(v) \sum_{\alpha} z_{\alpha}^2 = \sum_{\alpha\gamma} z_{\alpha} (RR^T)_{\alpha\gamma} z_{\gamma}. \quad (28.0.597)$$

Prove from (2.136) that

$$R(v)R^T(v) = \det(v)I. \quad (28.0.598)$$

Finally, assume that  $v$  has unit determinant. Then we find that

$$R(v)R^T(v) = I, \quad (28.0.599)$$

the matrix  $R$  is *orthogonal*. At this stage we have found that (2.124) provides a homomorphism of  $SL(4, C)$  into  $SO(6, C)$ . Verify, as a sanity check, that both  $SL(4, C)$  and  $SO(6, C)$  have dimension 30.

Next assume that  $v$  is also unitary so that  $v \in SU(4)$ . Then deduce the chain of relations

$$I = R(I) = R(v^{\dagger}v) = R(v^{\dagger})R(v) \quad (28.0.600)$$

to conclude that

$$R^T(v) = R(v^{\dagger}). \quad (28.0.601)$$

Also, we claim that  $R(v)$  is *real* if  $v \in SU(4)$ . Verify that taking the complex conjugate of both sides of (2.124) and employing the invariance of the trace under transposing and cyclic permutation gives (with a \* denoting complex conjugation) the result

$$\begin{aligned} R_{\alpha\beta}^*(v) &= (1/4)\text{tr}[v^{\dagger}(A^{\alpha})^*v^*(A^{\beta})^T] = (1/4)\text{tr}[A^{\beta}v^{\dagger}(A^{\alpha})^{\dagger}v^*] \\ &= (1/4)\text{tr}[v^*A^{\beta}v^{\dagger}(A^{\alpha})^{\dagger}] = (1/4)\text{tr}[(v^{\dagger})^TA^{\beta}v^{\dagger}(A^{\alpha})^{\dagger}] \\ &= R_{\beta\alpha}(v^{\dagger}) = [R(v^{\dagger})]_{\beta\alpha} = [R^T(v)]_{\beta\alpha} = R_{\alpha\beta}(v). \end{aligned} \quad (28.0.602)$$

Thus the mapping (2.124) has the property that  $R(v) \in SO(6, \mathbb{R})$  if  $v \in SU(4)$ . In view of (2.127) and (2.128), this mapping is a two-to-one homomorphism.

Finally, we remark that  $SU(4)$  is known to be simply connected. It follows that  $SO(6, \mathbb{R})$  cannot be simply connected. Indeed,  $SO(6, \mathbb{R})$  is known to be doubly connected, and its covering group is  $SU(4)$ .

At this point one might wonder again about employing the  $T$  operation in (2.123) and (2.124) rather than  $\dagger$  operation as was done in (2.34). The reason for this choice lies in the Clebsch-Gordan series for  $SU(4)$ . The lowest dimensional representations for  $SU(4)$ , those of dimension 4, are *not* self conjugate. Rather, there are two distinct representations which we may call 4 and  $\bar{4}$ . The Clebsch-Gordan series for the direct products of these two representations, and it is something like a direct product that is going on in (2.124), are

$$4 \times 4 = 6 + 10, \quad (28.0.603)$$

$$4 \times \bar{4} = 1 + 15. \quad (28.0.604)$$

Thus, only by avoiding complex conjugation do we have any hope of obtaining something of dimension 6, the dimension that is required for the lowest dimensional representation of  $SO(6, \mathbb{R})$ .

We still have to address the relation between the two Lie algebras  $su(4)$  and  $so(6, \mathbb{R})$ . This can be done by studying (2.124) for elements  $v$  near the identity. Suppose  $v$  is written in the form

$$v = \exp(\epsilon K) = I + \epsilon K + O(\epsilon^2) \quad (28.0.605)$$

where  $\epsilon$  is small and  $K \in su(4)$  is therefore any  $4 \times 4$  traceless anti-Hermitian matrix. Then we have

$$v^T = \exp(\epsilon K^T) = I + \epsilon K^T + O(\epsilon^2), \quad (28.0.606)$$

and (2.124) yields

$$\begin{aligned} R_{\alpha\beta}(v) &= (1/4)\text{tr}[(I + \epsilon K^T)A^\alpha(I + \epsilon K)(A^\beta)^\dagger] + O(\epsilon^2) \\ &= (1/4)\text{tr}[A^\alpha(A^\beta)^\dagger] + (\epsilon/4)\text{tr}[A^\alpha K(A^\beta)^\dagger + K^T A^\alpha (A^\beta)^\dagger] + O(\epsilon^2) \\ &= \delta_{\alpha\beta} + \epsilon L_{\alpha\beta} + O(\epsilon^2) = [\exp(\epsilon L)]_{\alpha\beta} + O(\epsilon^2) \\ \Leftrightarrow R[\exp(\epsilon K)] &= \exp(\epsilon L) + O(\epsilon^2) \end{aligned} \quad (28.0.607)$$

where

$$L_{\alpha\beta}(K) = (1/4)\text{tr}[A^\alpha K(A^\beta)^\dagger + K^T A^\alpha (A^\beta)^\dagger] = (1/4)\text{tr}[K(A^\beta)^\dagger A^\alpha + K^T A^\alpha (A^\beta)^\dagger]. \quad (28.0.608)$$

Here we have made use of the trace property (3.6.130).

We note that, from the homomorphism property (2.128) and the infinitesimal relation (2.146), it follows that there is also the global result

$$R[\exp(K)] = \exp(L). \quad (28.0.609)$$

To verify this claim, show that

$$\begin{aligned} R[\exp(K)] &= R\{\exp(K/\ell)\}^\ell = \{R[\exp(K/\ell)]\}^\ell \\ &= \{\exp(L/\ell) + O[(1/\ell)^2]\}^\ell \\ &= \{\exp(L/\ell)\}^\ell + \ell O[(1/\ell)^2] \\ &= \exp(L) + \ell O[(1/\ell)^2]. \end{aligned} \quad (28.0.610)$$

Now let  $\ell \rightarrow \infty$  in (2.149) to obtain the global result (2.148).

Let us examine the properties of  $L$ . We already know that  $R$  is orthogonal and real, and therefore  $L$  must be antisymmetric and real. It is valuable to check that these results can also be verified directly from (2.147). Show from (2.101) and (2.147) that

$$L_{\alpha\alpha}(K) = (1/4)\text{tr}[KI + K^T I] = (1/2) \text{tr}(K) = 0 \quad (28.0.611)$$

because  $K$  must be traceless to be in  $su(4)$ . Next show using (2.18) and (2.19) that, for  $\alpha \neq \beta$ ,

$$\begin{aligned} L_{\alpha\beta}(K) &= (1/4)\text{tr}[K(A^\beta)^\dagger A^\alpha + K^T A^\alpha (A^\beta)^\dagger] \\ &= -(1/4)\text{tr}[K(A^\alpha)^\dagger A^\beta + K^T A^\beta (A^\alpha)^\dagger] \\ &= -L_{\beta\alpha}(K). \end{aligned} \quad (28.0.612)$$

Taken together, (2.150) and (2.151) show that  $L$  is antisymmetric. Next work on showing that  $L$  is real. Let  $*$  denote the operation of complex conjugation. To show that  $L$  is real, verify the chain of deductions

$$\begin{aligned} [L_{\alpha\beta}(K)]^* &= (1/4)\{\text{tr}[K(A^\beta)^\dagger A^\alpha + K^T A^\alpha (A^\beta)^\dagger]\}^* \\ &= (1/4) \text{tr}[K^*(A^\beta)^T (A^\alpha)^* + K^\dagger (A^\alpha)^* (A^\beta)^T] \\ &= (1/4) \text{tr}\{(K^\dagger)^T (A^\beta)^T (A^\alpha)^* + K^\dagger [(A^\alpha)^T]^\dagger (A^\beta)^T\} \\ &= -(1/4) \text{tr}\{K^T A^\beta (A^\alpha)^\dagger + K (A^\alpha)^\dagger A^\beta\} \\ &= -L_{\beta\alpha}(K) = L_{\alpha\beta}(K). \end{aligned} \quad (28.0.613)$$

Here we have used (2.99) and the fact that  $K$  must be anti-Hermitian to be in  $su(4)$ ,

$$K^\dagger = -K, \quad (28.0.614)$$

and the antisymmetry conditions (2.150) and (2.151).

It remains to be verified that  $L(K)$  is a homomorphism (and potentially an isomorphism). Let  $(K^1, K^2, \dots, K^{15})$  be a set of basis elements for  $su(4)$ . Form *group commutator* elements  $v$  by the rule

$$v = \exp(\epsilon K^\alpha) \exp(\epsilon K^\beta) \exp(-\epsilon K^\alpha) \exp(-\epsilon K^\beta). \quad (28.0.615)$$

Recall Exercise 3.7.41. Show, using the BCH series (3.7.41), that

$$v = \exp(\epsilon^2 \{K^\alpha, K^\beta\}) + O(\epsilon^3), \quad (28.0.616)$$

from which it follows, using (2.148), that

$$R(v) = \exp[\epsilon^2 L(\{K^\alpha, K^\beta\})] + O(\epsilon^3). \quad (28.0.617)$$

Show from (2.128), (2.148), and BCH that

$$\begin{aligned} R(v) &= R[\exp(\epsilon K^\alpha)] R[\exp(\epsilon K^\beta)] R[\exp(-\epsilon K^\alpha)] R[\exp(-\epsilon K^\beta)] \\ &= \exp[\epsilon L(K^\alpha)] \exp[\epsilon L(K^\beta)] \exp[-\epsilon L(K^\alpha)] \exp[-\epsilon L(K^\beta)] \\ &= \exp[\epsilon^2 \{L(K^\alpha), L(K^\beta)\}] + O(\epsilon^3). \end{aligned} \quad (28.0.618)$$

By equating powers of  $\epsilon$  in (2.156) and (2.157), show that

$$\{L(K^\alpha), L(K^\beta)\} = L(\{K^\alpha, K^\beta\}). \quad (28.0.619)$$

Suppose the quantities  $c_{\alpha\beta}^\gamma$  are structure constants for  $su(4)$  so that

$$\{K^\alpha, K^\beta\} = \sum_\gamma c_{\alpha\beta}^\gamma K^\gamma. \quad (28.0.620)$$

Show from (2.158) and (2.159) that there is the relation

$$\{L(K^\alpha), L(K^\beta)\} = \sum_\gamma c_{\alpha\beta}^\gamma L(K^\gamma), \quad (28.0.621)$$

thereby verifying that, for suitable basis choices,  $su(4)$  and  $so(6, \mathbb{R})$  have the same structure constants.

We still want to know what particular  $K \in su(4)$  produces what  $L \in so(6, \mathbb{R})$ , and we want to verify that every  $L \in so(6, \mathbb{R})$  arises from some  $K \in su(4)$  so that (2.150) is, in fact, an isomorphism. We begin this task by looking at specific cases. Listed below are three typical elements in  $su(4)$ :

$$K^1 = \begin{pmatrix} 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = i \begin{pmatrix} \sigma^1 & 0 \\ 0 & 0 \end{pmatrix}, \quad (28.0.622)$$

$$K^2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = i \begin{pmatrix} \sigma^2 & 0 \\ 0 & 0 \end{pmatrix}, \quad (28.0.623)$$

$$K^3 = \begin{pmatrix} i & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = i \begin{pmatrix} \sigma^3 & 0 \\ 0 & 0 \end{pmatrix}. \quad (28.0.624)$$

The matrix  $K^1$  is symmetric, pure imaginary, and has zeroes on the diagonal, which makes it anti-Hermitian and traceless. There are 6 linearly independent matrices of this kind in  $su(4)$ . The matrix  $K^2$  is antisymmetric and real, which makes it anti-Hermitian and traceless. There are also 6 linearly independent matrices of this kind in  $su(4)$ . The element  $K^3$  is diagonal and pure imaginary, which makes it anti-Hermitian, and it is traceless. There are 3 linearly independent matrices of this kind in  $su(4)$  for a total count of  $6 + 6 + 3 = 15$ , the dimension of  $su(4)$ . Show that the associated  $L$  matrices are given by the relations

$$L^1 = L(K^1) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}, \quad (28.0.625)$$

$$L^2 = L(K^2) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \end{pmatrix}, \quad (28.0.626)$$

$$L^3 = L(K^3) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 \end{pmatrix}. \quad (28.0.627)$$

We see that the associated  $L$  matrices are real and antisymmetric, as expected. In particular,  $L^1$  is a linear combination of generators for rotations in the 3,6 and 4,5 planes;  $L^2$  is a linear combination of generators for rotations in the 3,5 and 4,6 planes; and  $L^3$  is a linear combination of generators for rotations in the 3,4 and 5,6 planes. Verify that the two generators in each linear combination commute. Verify that

$$\{K^1, K^2\} = -2K^3, \text{ etc.} \quad (28.0.628)$$

so that the elements  $K^1$  through  $K^3$  form some kind of  $su(2)$  [or  $so(3, R)$ ] within  $su(4)$ . Verify, in accord with (2.158), the relations

$$\{L(K^1), L(K^2)\} = -2L(K^3) = L(\{K^1, K^2\}), \text{ etc.} \quad (28.0.629)$$

Suppose the remaining elements of  $su(4)$  are also considered so that we are working with a complete set of basis elements  $(K^1, K^2, \dots, K^{15})$  for  $su(4)$ . Then presumably their associated matrices  $L^\alpha = L(K^\alpha)$  form a basis for  $so(6, \mathbb{R})$ . In fact we know from general principles that this must be the case. These principles are described in Section 8.9, and applied to the problem at hand in Exercise 8.9.19.

**28.0.26.** Review Section 3.13.3 and Exercise 8.2.12. The purpose of this exercise is to derive various results of Exercise 8.2.12 with the aid of Pfaffians. Let  $A$  be a general  $4 \times 4$  antisymmetric matrix written in the form

$$A = \begin{pmatrix} 0 & a & b & c \\ -a & 0 & d & e \\ -b & -d & 0 & f \\ -c & -e & -f & 0 \end{pmatrix}. \quad (28.0.630)$$

Then the Pfaffian of  $A$  is given by the relation

$$\text{Pf}(A) = af - be + dc. \quad (28.0.631)$$

Suppose  $A$  is given in the form (2.81). Show, for this parameterization, that

$$\text{Pf}(A) = -(1/c)\mathbf{E} \cdot \mathbf{B}. \quad (28.0.632)$$

Use the Pfaffian property (3.13.65) to derive (2.82). Use the Pfaffian property (3.13.66) to derive the relation

$$[\det(v)] \left[ \sum_{\alpha} z_{\alpha}^2 \right] = \left[ \sum_{\beta} \tilde{z}_{\beta}^2 \right], \quad (28.0.633)$$

which is a specific square root of the relation (2.133), and yields (2.134) directly without any ambiguity in sign.

**28.0.27.** Review Exercise 2.10 that studied the relation between  $SU(2)$  and  $SO(3, \mathbb{R})$ . In particular, it derived the formula (2.54) that maps  $SU(2)$  to  $SO(3, \mathbb{R})$  thereby demonstrating that  $SU(2)$  is the covering group for  $SO(3, \mathbb{R})$ . The purpose of this exercise is to find analogous results for the relation between  $SL(2, \mathbb{C})$  and the Lorentz group.

Begin by setting up some notation and definitions for later use. Let  $\sigma^{\alpha}$  for  $\alpha = 1, 2, 3$  be the usual Pauli matrices and let  $\sigma^4$  be the  $2 \times 2$  identity matrix,

$$\sigma^4 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (28.0.634)$$

Verify, in accord with Exercise 5.7.7, that there are the relations

$$\text{tr } \sigma^{\alpha} \sigma^{\beta} = 2\delta^{\alpha\beta} \text{ for } \alpha, \beta = 1, 2, 3, 4. \quad (28.0.635)$$

Let  $x^{\mu}$  and  $y^{\mu}$  be any two four-vectors. Make the definition

$$x \star y = \sum_{\mu=1}^4 x^{\mu} y^{\mu}. \quad (28.0.636)$$

Note that (2.175) is just the ordinary Euclidean scalar product  $(x, y)$ . By extension of notation, make the definition

$$x \star \sigma = \sum_{\mu=1}^4 x^{\mu} \sigma^{\mu}. \quad (28.0.637)$$

Verify that  $x \star \sigma$  is Hermitian if  $x$  is real, and anti-Hermitian if  $x$  is pure imaginary. Verify that

$$\text{tr}(x \star \sigma) = 2x^4. \quad (28.0.638)$$

Verify that

$$\text{tr}[(x \star \sigma)(y \star \sigma)] = 2(x \star y) \quad (28.0.639)$$

and, as a special case,

$$\text{tr}[(x \star \sigma)^2] = 2(x \star x). \quad (28.0.640)$$

Show that  $x \star \sigma$  has the explicit matrix form

$$x \star \sigma = \begin{pmatrix} x^4 + x^3 & x^1 - ix^2 \\ x^1 + ix^2 & x^4 - x^3 \end{pmatrix}. \quad (28.0.641)$$

Given a  $2 \times 2$  matrix  $M$ , is there a four-vector  $x$  such that

$$M = x \star \sigma? \quad (28.0.642)$$

Any  $2 \times 2$  matrix  $M$  can be written in the form

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (28.0.643)$$

where the quantities  $a$  through  $d$  are arbitrary. Verify, upon examination of (2.180) and (2.182), that for (2.181) to hold it must be possible to satisfy the relations

$$x^4 + x^3 = a, \quad (28.0.644)$$

$$x^1 - ix^2 = b, \quad (28.0.645)$$

$$x^1 + ix^2 = c, \quad (28.0.646)$$

$$x^4 - x^3 = d. \quad (28.0.647)$$

Show that the equation set (2.183) through (2.186) has the unique solution

$$x^1 = (1/2)(b + c), \quad (28.0.648)$$

$$x^2 = (i/2)(b - c), \quad (28.0.649)$$

$$x^3 = (1/2)(a - d), \quad (28.0.650)$$

$$x^4 = (1/2)(a + d). \quad (28.0.651)$$

Consequently, for any  $M$ , there exists a unique  $x$  such that (2.181) holds. That is, the  $\sigma^\alpha$  form a basis for the set of all  $2 \times 2$  matrices. Show that  $x$  is given in terms of  $M$  by the relations

$$x^\mu = (1/2) \operatorname{tr}(\sigma^\mu M). \quad (28.0.652)$$

Show that  $x$  is real if  $M$  is Hermitian, and is pure imaginary if  $M$  is anti-Hermitian.

Finally, make the definition

$$x \cdot y = x^4 y^4 - \sum_{\mu=1}^3 x^\mu y^\mu. \quad (28.0.653)$$

Note that (2.192) is the Lorentz inner/scalar product.

We are now ready to make yet another remarkable statement about the Pauli matrices: Verify from (2.180) by explicit calculation that there is the relation

$$\det(x \star \sigma) = x \cdot x. \quad (28.0.654)$$

How could we have guessed that at least something like (2.193) should hold? Review Exercise 3.7.17. According to (3.7.150), for any  $2 \times 2$  matrix  $A$ , there is the relation

$$\det(A) = \{[\operatorname{tr}(A)]^2 - \operatorname{tr}(A^2)\}/2. \quad (28.0.655)$$

Make the substitution

$$A = x \star \sigma. \quad (28.0.656)$$

Verify from (2.177) and (2.179) that

$$[\mathrm{tr}(A)]^2 = 4(x^4)^2 \quad (28.0.657)$$

and

$$\mathrm{tr}(A^2) = 2(x \star x). \quad (28.0.658)$$

Show, therefore, that use of (2.194) yields the result

$$\det(x \star \sigma) = [4(x^4)^2 - 2(x \star x)]/2 = [2(x^4)^2 - 2 \sum_{\mu=1}^3 (x^\mu)^2]/2 = x \cdot x, \quad (28.0.659)$$

in agreement with (2.193).

With the definitions and results just developed now at hand, we are ready to further explore the connection between the Lorentz group and  $SL(2, \mathbb{C})$ . Let  $v$  be any element of  $SL(2, \mathbb{C})$  so that

$$\det v = 1. \quad (28.0.660)$$

Verify that then  $v^\dagger$  is also in  $SL(2, \mathbb{C})$  so that

$$\det v^\dagger = 1. \quad (28.0.661)$$

Next let  $x$  be any real four-vector. Consider the matrix  $v(x \star \sigma)v^\dagger$ . It is evidently  $2 \times 2$ . Show that it is also Hermitian,

$$[v(x \star \sigma)v^\dagger]^\dagger = v(x \star \sigma)v^\dagger. \quad (28.0.662)$$

Since  $v(x \star \sigma)v^\dagger$  is a  $2 \times 2$  Hermitian matrix, there must be a real and unique four-vector  $\hat{x}$  such that

$$\hat{x} \star \sigma = v(x \star \sigma)v^\dagger. \quad (28.0.663)$$

Show, by taking determinants of both sides of (2.202), that there is the relation

$$\hat{x} \cdot \hat{x} = x \cdot x. \quad (28.0.664)$$

It follows that  $\hat{x}$  and  $x$  are related by a Lorentz transformation! Thus, for each element  $v \in SL(2, \mathbb{C})$  there is a Lorentz transformation  $\Lambda(v)$ .

How might one have guessed that this should be the case? We know from Exercise 7.3.30 that the  $v \in SL(2, \mathbb{C})$  carry the representation  $\Gamma(0, 1/2)$  and we might expect, as can be proved, that the  $v^\dagger$  would carry the representation  $\Gamma(1/2, 0)$ . Since the right side of (2.202) involves both  $v$  and  $v^\dagger$  in a “multiplicative” way, we might expect that what we are doing in (2.202) would involve the representation  $\Gamma(0, 1/2) \times \Gamma(1/2, 0)$ . But for the Lorentz group it is easy to see that there is the Clebsch-Gordan result

$$\Gamma(0, 1/2) \times \Gamma(1/2, 0) = \Gamma(1/2, 1/2). \quad (28.0.665)$$

And, according to Exercise 7.3.29,  $\Gamma(1/2, 1/2)$  is the representation carried by Lorentz transformation matrices  $\Lambda$  acting on four-vectors.

Your next task is to find the matrix elements of  $\Lambda$  as functions of  $v$ . Let us examine and manipulate the “contents” of  $v(x \star \sigma)v^\dagger$ , the right side of (2.202). Verify that

$$v(x \star \sigma)v^\dagger = v\left(\sum_\nu x^\nu \sigma^\nu\right)v^\dagger = \sum_\nu (v\sigma^\nu v^\dagger)x^\nu. \quad (28.0.666)$$

Show that the matrices  $v\sigma^\nu v^\dagger$  are Hermitian,

$$(v\sigma^\nu v^\dagger)^\dagger = v\sigma^\nu v^\dagger. \quad (28.0.667)$$

Verify it follows that there are *real* coefficients, call them  $\Lambda^{\xi\nu}(v)$ , such that

$$v\sigma^\nu v^\dagger = \sum_{\xi=1}^4 \Lambda^{\xi\nu}(v)\sigma^\xi. \quad (28.0.668)$$

Show that, correspondingly, there is the relation

$$\Lambda^{\mu\nu}(v) = (1/2) \operatorname{tr}(\sigma^\mu v \sigma^\nu v^\dagger). \quad (28.0.669)$$

We already know that  $\Lambda$  is a real matrix. Still, it would be good to reverify directly that  $\Lambda$  as given by (2.208) is real even though some of the matrices appearing on the right side of (2.208) may be complex. Show, using (3.6.129) and (3.6.130), that

$$\begin{aligned} [\Lambda^{\mu\nu}(v)]^* &= (1/2) \operatorname{tr}[(\sigma^\mu v \sigma^\nu v^\dagger)^\dagger] = (1/2) \operatorname{tr}[v \sigma^\nu v^\dagger \sigma^\mu] \\ &= (1/2) \operatorname{tr}[\sigma^\mu v \sigma^\nu v^\dagger] = \Lambda^{\mu\nu}(v). \end{aligned} \quad (28.0.670)$$

Now what can be said about the vector  $\hat{x}$  and its relation to  $x$ ? Verify, in view of (2.202), that the components of  $\hat{x}$  are given by the relations

$$\hat{x}^\mu = (1/2) \operatorname{tr}[\sigma^\mu v(x \star \sigma)v^\dagger]. \quad (28.0.671)$$

Next use (2.205) to find that

$$(1/2) \operatorname{tr}[\sigma^\mu v(x \star \sigma)v^\dagger] = (1/2) \sum_\nu [\operatorname{tr}(\sigma^\mu v \sigma^\nu v^\dagger)]x^\nu. \quad (28.0.672)$$

Finally, upon combining (2.208), (2.210), and (2.211), show that

$$\hat{x}^\mu = \sum_\nu \Lambda^{\mu\nu}(v)x^\nu \text{ or, in matrix/vector form, } \hat{x} = \Lambda x. \quad (28.0.673)$$

We have learned, as anticipated by our notation, that the quantities  $\Lambda^{\mu\nu}(v)$  defined by (2.208) are the entries of a Lorentz transformation matrix.

What can be said about group properties? Suppose, to begin, that  $v$  is the  $2 \times 2$  identity matrix  $\sigma^4$ . Verify that in this case use of (2.208) gives the result

$$\Lambda^{\mu\nu}(\sigma^4) = (1/2) \operatorname{tr}[\sigma^\mu \sigma^4 \sigma^\nu (\sigma^4)^\dagger] = (1/2) \operatorname{tr}[\sigma^\mu \sigma^\nu] = \delta^{\mu\nu}. \quad (28.0.674)$$

That is, the image of the identity element in  $SL(2, \mathbb{C})$  is the identity matrix in the Lorentz group. Next suppose that  $v$  is of the product form

$$v = uw \quad (28.0.675)$$

where  $u$  and  $w$  are both elements of  $SL(2, \mathbb{C})$ . In this case verify that use of (2.208) gives the result

$$\Lambda^{\mu\nu}(uw) = (1/2) \operatorname{tr}[\sigma^\mu uw \sigma^\nu (uw)^\dagger] = (1/2) \operatorname{tr}[\sigma^\mu u (w \sigma^\nu w^\dagger) u^\dagger]. \quad (28.0.676)$$

But, in analogy to (2.207), verify that

$$w \sigma^\nu w^\dagger = \sum_{\xi=1}^4 \Lambda^{\xi\nu}(w) \sigma^\xi. \quad (28.0.677)$$

Show that combining (2.215) and (2.216) gives the intermediate results

$$\Lambda^{\mu\nu}(uw) = (1/2) \operatorname{tr}[\sigma^\mu u (\sum_{\xi=1}^4 \Lambda^{\xi\nu}(w) \sigma^\xi) u^\dagger] = \sum_{\xi=1}^4 \Lambda^{\xi\nu}(w) (1/2) \operatorname{tr}[\sigma^\mu (u \sigma^\xi u^\dagger)]. \quad (28.0.678)$$

But, again in analogy to (2.207), verify that

$$u \sigma^\xi u^\dagger = \sum_{\rho=1}^4 \Lambda^{\rho\xi}(u) \sigma^\rho. \quad (28.0.679)$$

Show that combining (2.217) and (2.218) gives the final result

$$\begin{aligned} \Lambda^{\mu\nu}(uw) &= \sum_{\xi=1}^4 \Lambda^{\xi\nu}(w) (1/2) \operatorname{tr}[\sigma^\mu (u \sigma^\xi u^\dagger)] \\ &= \sum_{\xi=1}^4 \Lambda^{\xi\nu}(w) (1/2) \operatorname{tr}[\sigma^\mu \sum_{\rho=1}^4 \Lambda^{\rho\xi}(u) \sigma^\rho] \\ &= \sum_{\xi=1}^4 \sum_{\rho=1}^4 \Lambda^{\rho\xi}(u) \Lambda^{\xi\nu}(w) (1/2) \operatorname{tr}[\sigma^\mu \sigma^\rho] \\ &= \sum_{\xi=1}^4 \sum_{\rho=1}^4 \Lambda^{\rho\xi}(u) \Lambda^{\xi\nu}(w) \delta^{\mu\rho} \\ &= \sum_{\xi=1}^4 \Lambda^{\mu\xi}(u) \Lambda^{\xi\nu}(w) \end{aligned} \quad (28.0.680)$$

or, in index-free notation,

$$\Lambda(uw) = \Lambda(u)\Lambda(w). \quad (28.0.681)$$

Complete our group property study with two calculations: First set

$$w = \sigma^4 \quad (28.0.682)$$

in (2.220) and deduce that

$$\Lambda(\sigma^4) = I \quad (28.0.683)$$

in agreement with (2.213). Second, set

$$w = v^{-1} \quad (28.0.684)$$

in (2.220) and deduce that

$$\Lambda(v^{-1}) = [\Lambda(v)]^{-1}. \quad (28.0.685)$$

The relation (2.220) and those that follow from it show that Lorentz transformation matrices  $\Lambda$  provide a representation of  $SL(2, \mathbb{C})$ . That is, the map (2.208) that sends elements of  $SL(2, \mathbb{C})$  into elements of the Lorentz group is a homomorphism. At this point it is important to observe, by inspection, that the map (2.208) has the two-to-one property

$$\Lambda(-v) = \Lambda(v). \quad (28.0.686)$$

Therefore (2.208) is not an isomorphism. As we will see,  $SL(2, \mathbb{C})$  is the covering group of the Lorentz group.

In Exercises 7.7.27 and 7.7.30 it was shown that the Lorentz group and  $SL(2, \mathbb{C})$  have identical Lie algebras and analogous polar decompositions. And in this exercise we have seen that (2.208) provides a two-to-one homomorphic relation between the Lorentz group and  $SL(2, \mathbb{C})$ . The remainder of this exercise explores how these results fit together.

Suppose, employing the polar decomposition (7.3.241), that  $v$  is written in the form

$$v = \exp(\lambda \mathbf{m} \cdot \hat{\mathbf{N}}) \exp(\theta \mathbf{n} \cdot \hat{\mathbf{L}}). \quad (28.0.687)$$

Show that employing this factorization in (2.220) yields the result

$$\Lambda(v) = \Lambda[\exp(\lambda \mathbf{m} \cdot \hat{\mathbf{N}})] \Lambda[\exp(\theta \mathbf{n} \cdot \hat{\mathbf{L}})]. \quad (28.0.688)$$

Your next tasks will be to work out results for each factor on the right side of (2.227).

## Begin Evaluation of Second Factor

Begin with the second factor. For it you will need to work out

$$\Lambda^{\mu\nu}(w) = (1/2) \text{tr}(\sigma^\mu w \sigma^\nu w^\dagger) \quad (28.0.689)$$

with

$$w = \exp(\theta \mathbf{n} \cdot \hat{\mathbf{L}}) = \exp(\theta \mathbf{n} \cdot \mathbf{K}) = \exp[\theta(-i/2) \mathbf{n} \cdot \boldsymbol{\sigma}]. \quad (28.0.690)$$

See (7.3.232) through (7.3.234). Verify that in this case  $w$  is unitary.

The simplest matrix element to work out is  $\Lambda^{44}$ . Verify using (2.208) that

$$\Lambda^{44}(w) = (1/2) \text{tr}(\sigma^4 w \sigma^4 w^\dagger) = (1/2) \text{tr}(ww^\dagger) = (1/2) \text{tr}(\sigma^4) = 1. \quad (28.0.691)$$

The next simplest cases are the  $\Lambda^{\alpha 4}$  and  $\Lambda^{4\alpha}$  with  $\alpha = 1, 2, 3$ . Verify that for these  $\alpha$

$$\Lambda^{\alpha 4}(w) = (1/2) \text{tr}(\sigma^\alpha w \sigma^4 w^\dagger) = (1/2) \text{tr}(\sigma^\alpha) = 0 \quad (28.0.692)$$

and

$$\Lambda^{4\alpha}(w) = (1/2) \operatorname{tr}(\sigma^4 w \sigma^\alpha w^\dagger) = (1/2) \operatorname{tr}(w^\dagger w \sigma^\alpha) = (1/2) \operatorname{tr}(\sigma^\alpha) = 0. \quad (28.0.693)$$

[Here, in writing (2.232), use has been made of the trace relation (3.6.130).] So it has now been established that  $\Lambda(w)$  is of the form

$$\Lambda(w) = \begin{pmatrix} * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (28.0.694)$$

To fill in the missing entries in (2.233) the simplest way at this point is to use what has already been accomplished in Exercises 3.7.31 and 2.10, which you should review. Note, again using (3.6.130), that (2.208) can be rewritten in the form

$$\Lambda^{\mu\nu}(v) = (1/2) \operatorname{tr}(v^\dagger \sigma^\mu v \sigma^\nu), \quad (28.0.695)$$

which has the same form as (2.54). Consequently, the missing entries are the elements of the  $3 \times 3$  matrix given by (2.54). It follows, as expected, that for the  $w$  given by (2.229)  $\Lambda(w)$  is given by the relation

$$\Lambda(w) = \exp(\theta \mathbf{n} \cdot \mathbf{L}) \quad (28.0.696)$$

or

$$\Lambda[\exp(\theta \mathbf{n} \cdot \hat{\mathbf{L}})] = \exp(\theta \mathbf{n} \cdot \mathbf{L}) \quad (28.0.697)$$

where the  $\mathbf{L}$  are the matrices given by (7.3.177) through (7.3.179) and whose upper left  $3 \times 3$  submatrices are the matrices given by (3.7.178) through (3.7.180).

As a sanity check of (2.236), consider the simple case where

$$\mathbf{n} = \mathbf{e}_3 \quad (28.0.698)$$

so that

$$w = \exp(\theta \mathbf{n} \cdot \hat{\mathbf{L}}) = \exp(\theta \hat{L}^3) = \exp[\theta(-i/2)\sigma^3]. \quad (28.0.699)$$

Verify that

$$w = \exp[\theta(-i/2)\sigma^3] = \exp(\theta K^3) = \begin{pmatrix} \exp(-i\theta/2) & 0 \\ 0 & \exp(i\theta/2) \end{pmatrix}. \quad (28.0.700)$$

See (3.7.171) and (3.7.194). Also, compare (2.236) with the analogous result (2.54). Show that for the  $w$  given by (2.235) and (2.236) there is the result

$$\Lambda(w) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 & 0 \\ \sin(\theta) & \cos(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \exp(\theta L^3) \quad (28.0.701)$$

with  $L^3$  given by (7.3.179). See (3.7.207). Note that, according to (2.240),  $\Lambda(w)$  is periodic in  $\theta$  with period  $2\pi$ ,

$$\Lambda(w)|_{\theta+2\pi} = \Lambda(w)|_\theta. \quad (28.0.702)$$

But, according to (2.239),  $w$  is not; it has period  $4\pi$ ,

$$w|_{\theta+4\pi} = w|_\theta, \quad (28.0.703)$$

and

$$w|_{\theta+2\pi} = -w|_\theta. \quad (28.0.704)$$

Show that this result is consistent with (2.225). Show also that results completely analogous to (2.251) through (2.243) hold for any choice of the unit vector  $\mathbf{n}$  and are consistent with (2.225).

The work of Exercise 2.10, which is what we have been using, involved the formation and solution of a differential equation. Another way to proceed, which is the same in essence, is to evaluate (2.229) and then (2.228) for small  $\theta$ , and then build up to large values of  $\theta$  by repeatedly using the group property (2.220). Let us explore this approach, which will turn out to be simpler.

Suppose, in view of (2.229), we work with a  $w$  and a  $w^\dagger$  of the forms

$$w = \exp(\epsilon \mathbf{n} \cdot \hat{\mathbf{L}}) = \exp[\epsilon(-i/2)\mathbf{n} \cdot \boldsymbol{\sigma}] = I + \epsilon(-i/2)\mathbf{n} \cdot \boldsymbol{\sigma} + O(\epsilon^2), \quad (28.0.705)$$

$$w^\dagger = I + \epsilon(+i/2)\mathbf{n} \cdot \boldsymbol{\sigma} + O(\epsilon^2). \quad (28.0.706)$$

Show that employing this  $w$  and  $w^\dagger$  pair in (2.228) gives the result

$$\begin{aligned} \Lambda^{\mu\nu}(w) &= (1/2) \operatorname{tr}(\sigma^\mu w \sigma^\nu w^\dagger) \\ &= (1/2) \operatorname{tr}[\sigma^\mu(I + \epsilon(-i/2)\mathbf{n} \cdot \boldsymbol{\sigma})\sigma^\nu(I + \epsilon(+i/2)\mathbf{n} \cdot \boldsymbol{\sigma})] + O(\epsilon^2) \\ &= (1/2) \operatorname{tr}\{\sigma^\mu \sigma^\nu + \epsilon(-i/2)\sigma^\mu[(\mathbf{n} \cdot \boldsymbol{\sigma})\sigma^\nu - \sigma^\nu(\mathbf{n} \cdot \boldsymbol{\sigma})]\} + O(\epsilon^2) \\ &= \delta^{\mu\nu} + \epsilon(-i/4) \operatorname{tr}\{\sigma^\mu[(\mathbf{n} \cdot \boldsymbol{\sigma})\sigma^\nu - \sigma^\nu(\mathbf{n} \cdot \boldsymbol{\sigma})]\} + O(\epsilon^2). \end{aligned} \quad (28.0.707)$$

Verify that

$$(\mathbf{n} \cdot \boldsymbol{\sigma})\sigma^\nu - \sigma^\nu(\mathbf{n} \cdot \boldsymbol{\sigma}) = \sum_{\xi=1}^3 n_\xi \{\sigma^\xi, \sigma^\nu\}. \quad (28.0.708)$$

Make the definition

$$n_4 = 0 \quad (28.0.709)$$

so that (2.247) can also be written in the form

$$(\mathbf{n} \cdot \boldsymbol{\sigma})\sigma^\nu - \sigma^\nu(\mathbf{n} \cdot \boldsymbol{\sigma}) = \sum_{\xi=1}^4 n_\xi \{\sigma^\xi, \sigma^\nu\}. \quad (28.0.710)$$

## Pause to Develop Needed Mathematical Results

At this point we pause in our present calculations to define and develop the properties of a remarkable tensor that will be of subsequent use. Let  $U^{\alpha\beta\gamma}$  be the tensor defined in terms of the Pauli matrices  $\sigma^1$  through  $\sigma^4$  by the rule

$$U^{\alpha\beta\gamma} = \operatorname{tr}[\sigma^\alpha(\sigma^\beta\sigma^\gamma - \sigma^\gamma\sigma^\beta)]. \quad (28.0.711)$$

Evidently  $U$  has the symmetry property

$$U^{\alpha\gamma\beta} = -U^{\alpha\beta\gamma}. \quad (28.0.712)$$

That is,  $U$  is antisymmetric under the interchange of its last two indices. Next verify, because of the trace relation (3.6.130), that

$$\begin{aligned} U^{\alpha\beta\gamma} &= \text{tr}[\sigma^\alpha(\sigma^\beta\sigma^\gamma - \sigma^\gamma\sigma^\beta)] = \text{tr}[\sigma^\alpha\sigma^\beta\sigma^\gamma - \sigma^\alpha\sigma^\gamma\sigma^\beta] \\ &= \text{tr}[\sigma^\beta\sigma^\gamma\sigma^\alpha - \sigma^\gamma\sigma^\beta\sigma^\alpha] = \text{tr}[\sigma^\beta\sigma^\gamma\sigma^\alpha - \sigma^\beta\sigma^\alpha\sigma^\gamma] \\ &= \text{tr}[\sigma^\beta(\sigma^\gamma\sigma^\alpha - \sigma^\alpha\sigma^\gamma)] = -U^{\beta\alpha\gamma}. \end{aligned} \quad (28.0.713)$$

Thus,  $U$  is also antisymmetric under the interchange of its first two indices. Show that it follows from (2.251) and (2.252) that  $U$  is *completely* antisymmetric: That is,  $U$  changes sign under the interchange of any pair of indices.

To continue, define an antisymmetric tensor  $A$  with Pauli matrix entries by the rule

$$A^{\mu\nu} = \{\sigma^\mu, \sigma^\nu\}. \quad (28.0.714)$$

Show that

$$A = 2i \begin{pmatrix} 0 & \sigma^3 & -\sigma^2 & 0 \\ -\sigma^3 & 0 & \sigma^1 & 0 \\ \sigma^2 & -\sigma^1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (28.0.715)$$

Look at the matrices  $L^\alpha$  for  $\alpha = 1, 2, 3$  defined by (3.180) through (3.182). Also define a matrix  $L^4$  by the rule

$$L^4 = \mathbf{0} \quad (28.0.716)$$

where  $\mathbf{0}$  is the  $4 \times 4$  matrix with all entries having value zero. Show that

$$A = -2i \sum_{\alpha=1}^4 L^\alpha \sigma^\alpha. \quad (28.0.717)$$

Consequently there is the associated component relation

$$\{\sigma^\mu, \sigma^\nu\} = A^{\mu\nu} = -2i \sum_{\alpha=1}^4 (L^\alpha)^{\mu\nu} \sigma^\alpha \quad (28.0.718)$$

from which it follows that

$$\text{tr}[\sigma^\beta(\sigma^\mu\sigma^\nu - \sigma^\nu\sigma^\mu)] = \text{tr}[\sigma^\beta\{\sigma^\mu, \sigma^\nu\}] = -2i \sum_{\alpha=1}^4 (L^\alpha)^{\mu\nu} \text{tr}(\sigma^\beta\sigma^\alpha) = -4i(L^\beta)^{\mu\nu}. \quad (28.0.719)$$

Finally, show that comparison of (2.249) and (2.257) gives the result

$$A^{\beta\mu\nu} = -4i(L^\beta)^{\mu\nu}. \quad (28.0.720)$$

With these definitions before us, show that

$$(\mathbf{n} \cdot \boldsymbol{\sigma})\sigma^\nu - \sigma^\nu(\mathbf{n} \cdot \boldsymbol{\sigma}) = \sum_{\xi=1}^4 n_\xi \{\sigma^\xi, \sigma^\nu\} = \sum_{\xi=1}^4 n_\xi A^{\xi\nu} = -2i \sum_{\xi=1}^4 \sum_{\alpha=1}^4 n_\xi (L^\alpha)^{\xi\nu} \sigma^\alpha. \quad (28.0.721)$$

Continue on to show that

$$\text{tr}\{\sigma^\mu[(\mathbf{n} \cdot \boldsymbol{\sigma})\sigma^\nu - \sigma^\nu(\mathbf{n} \cdot \boldsymbol{\sigma})]\} = -2i \sum_{\xi=1}^4 \sum_{\alpha=1}^4 n_\xi (L^\alpha)^{\xi\nu} \text{tr}(\sigma^\mu \sigma^\alpha) = -4i \sum_{\xi=1}^4 n_\xi (L^\mu)^{\xi\nu}. \quad (28.0.722)$$

But, from the relation (2.259) and the symmetry properties of  $A$ , it follows that

$$(L^\mu)^{\xi\nu} = -(L^\xi)^{\mu\nu}. \quad (28.0.723)$$

Show, therefore, that (2.261) can be rewritten in the form

$$\text{tr}\{\sigma^\mu[(\mathbf{n} \cdot \boldsymbol{\sigma})\sigma^\nu - \sigma^\nu(\mathbf{n} \cdot \boldsymbol{\sigma})]\} = -4i \sum_{\xi=1}^4 n_\xi (L^\mu)^{\xi\nu} = 4i \sum_{\xi=1}^4 n_\xi (L^\xi)^{\mu\nu} = 4i(\mathbf{n} \cdot \mathbf{L})^{\mu\nu}. \quad (28.0.724)$$

## Resume and Complete Evaluation of Second Factor

Verify, as a result of these intervening calculations, that combining (2.246) and (2.263) yields the pleasingly simple result

$$\Lambda^{\mu\nu}(w) = \delta^{\mu\nu} + \epsilon(\mathbf{n} \cdot \mathbf{L})^{\mu\nu} + O(\epsilon^2) \quad (28.0.725)$$

or, in matrix form,

$$\Lambda(w) = I + \epsilon \mathbf{n} \cdot \mathbf{L} + O(\epsilon^2). \quad (28.0.726)$$

Show, using (2.244) and (2.265), that there is the relation

$$\Lambda[\exp(\epsilon \mathbf{n} \cdot \hat{\mathbf{L}})] = \exp(\epsilon \mathbf{n} \cdot \mathbf{L}) + O(\epsilon^2). \quad (28.0.727)$$

As promised, we now intend to use (2.266) and the group property (2.219) to work out results for finite values of  $\theta$ . Let  $\ell$  be a positive integer and set

$$\epsilon = \theta/\ell. \quad (28.0.728)$$

Show that

$$\begin{aligned} \Lambda[\exp(\theta \mathbf{n} \cdot \hat{\mathbf{L}})] &= \Lambda[\exp(\ell \epsilon \mathbf{n} \cdot \hat{\mathbf{L}})] = \{\Lambda[\exp(\epsilon \mathbf{n} \cdot \hat{\mathbf{L}})]\}^\ell \\ &= [\exp(\epsilon \mathbf{n} \cdot \mathbf{L}) + O(\epsilon^2)]^\ell = \{\exp(\epsilon \mathbf{n} \cdot \mathbf{L}) + O[(\theta/\ell)^2]\}^\ell \\ &= \exp(\ell \epsilon \mathbf{n} \cdot \mathbf{L}) + \ell O[(\theta/\ell)^2] = \exp(\theta \mathbf{n} \cdot \mathbf{L}) + O(\theta^2/\ell). \end{aligned} \quad (28.0.729)$$

Verify that taking the limit  $\ell \rightarrow \infty$  in (2.268) yields the expected result

$$\Lambda[\exp(\theta \mathbf{n} \cdot \hat{\mathbf{L}})] = \exp(\theta \mathbf{n} \cdot \mathbf{L}). \quad (28.0.730)$$

## Begin Evaluation of First Factor

What remains is to work out results for the first factor on the right side of (2.226). For it you will need to work out  $\Lambda(u)$  with

$$u = \exp(\lambda \mathbf{m} \cdot \hat{\mathbf{N}}) = \exp[\lambda(1/2)\mathbf{m} \cdot \boldsymbol{\sigma}]. \quad (28.0.731)$$

See (7.3.235) through (7.3.237). Verify that  $u$  is the exponential of a Hermitian matrix and is therefore Hermitian and positive definite.<sup>15</sup> See Exercise 3.7.44. According to (2.208) what we now need to compute are the entries

$$\Lambda^{\mu\nu}(u) = (1/2) \operatorname{tr}(\sigma^\mu u \sigma^\nu u^\dagger) = (1/2) \operatorname{tr}(\sigma^\mu u \sigma^\nu u). \quad (28.0.732)$$

To proceed, we will adopt a strategy analogous to what we employed for the second factor. We will evaluate (2.266) for small  $\lambda$  and then build up to large values of  $\lambda$  by repeatedly using the group property (2.219).

Suppose, in view of (2.242), we work with a  $u$  of the form

$$u = \exp(\epsilon \mathbf{m} \cdot \hat{\mathbf{N}}) = \exp[\epsilon(1/2)\mathbf{m} \cdot \boldsymbol{\sigma}] = I + \epsilon(1/2)\mathbf{m} \cdot \boldsymbol{\sigma} + O(\epsilon^2). \quad (28.0.733)$$

Show that employing this  $u$  in (2.243) gives the result

$$\begin{aligned} \Lambda^{\mu\nu}(u) &= (1/2) \operatorname{tr}(\sigma^\mu u \sigma^\nu u) \\ &= (1/2) \operatorname{tr}[\sigma^\mu(I + \epsilon(1/2)\mathbf{m} \cdot \boldsymbol{\sigma})\sigma^\nu(I + \epsilon(1/2)\mathbf{m} \cdot \boldsymbol{\sigma})] + O(\epsilon^2) \\ &= (1/2) \operatorname{tr}\{\sigma^\mu \sigma^\nu + \epsilon(1/2)\sigma^\mu[(\mathbf{m} \cdot \boldsymbol{\sigma})\sigma^\nu + \sigma^\nu(\mathbf{m} \cdot \boldsymbol{\sigma})]\} + O(\epsilon^2) \\ &= \delta^{\mu\nu} + \epsilon(1/4) \operatorname{tr}\{\sigma^\mu[(\mathbf{m} \cdot \boldsymbol{\sigma})\sigma^\nu + \sigma^\nu(\mathbf{m} \cdot \boldsymbol{\sigma})]\} + O(\epsilon^2). \end{aligned} \quad (28.0.734)$$

Verify that

$$(\mathbf{m} \cdot \boldsymbol{\sigma})\sigma^\nu + \sigma^\nu(\mathbf{m} \cdot \boldsymbol{\sigma}) = \sum_{\xi=1}^3 m_\xi \{\sigma^\xi, \sigma^\nu\}_+. \quad (28.0.735)$$

Make the definition

$$m_4 = 0 \quad (28.0.736)$$

so that (2.246) can also be written in the form

$$(\mathbf{m} \cdot \boldsymbol{\sigma})\sigma^\nu + \sigma^\nu(\mathbf{m} \cdot \boldsymbol{\sigma}) = \sum_{\xi=1}^4 m_\xi \{\sigma^\xi, \sigma^\nu\}_+. \quad (28.0.737)$$

---

<sup>15</sup>Note that, unlike in the context of the second factor where we encounter both  $w$  and  $-w$ , in the context of the first factor we will not encounter both  $u$  and  $-u$  because if  $u$  is positive definite, then  $-u$  is not, and in the context of the first factor we only encounter the positive definite case.

## Again Pause to Develop Needed Mathematical Results

At this point we pause in our present calculations to define and develop the properties of another remarkable tensor that will be of subsequent use. Let  $V^{\alpha\beta\gamma}$  be the tensor defined in terms of the Pauli matrices  $\sigma^1$  through  $\sigma^4$  by the rule

$$V^{\alpha\beta\gamma} = \text{tr}[\sigma^\alpha(\sigma^\beta\sigma^\gamma + \sigma^\gamma\sigma^\beta)]. \quad (28.0.738)$$

Evidently  $V$  has the symmetry property

$$V^{\alpha\gamma\beta} = V^{\alpha\beta\gamma}. \quad (28.0.739)$$

That is,  $V$  is symmetric under the interchange of its last two indices. Next verify, because of the trace relation (3.6.130), that

$$\begin{aligned} V^{\alpha\beta\gamma} &= \text{tr}[\sigma^\alpha(\sigma^\beta\sigma^\gamma + \sigma^\gamma\sigma^\beta)] = \text{tr}[\sigma^\alpha\sigma^\beta\sigma^\gamma + \sigma^\alpha\sigma^\gamma\sigma^\beta] \\ &= \text{tr}[\sigma^\beta\sigma^\gamma\sigma^\alpha + \sigma^\gamma\sigma^\beta\sigma^\alpha] = \text{tr}[\sigma^\beta\sigma^\gamma\sigma^\alpha + \sigma^\beta\sigma^\alpha\sigma^\gamma] \\ &= \text{tr}[\sigma^\beta(\sigma^\gamma\sigma^\alpha + \sigma^\alpha\sigma^\gamma)] = V^{\beta\alpha\gamma}. \end{aligned} \quad (28.0.740)$$

Thus,  $V$  is also symmetric under the interchange of its first two indices. Show that it follows from (2.250) and (2.251) that  $V$  is *completely* symmetric: That is,  $V$  is unchanged under any permutation of its indices.

To continue, define a symmetric tensor  $S$  with Pauli matrix entries by the rule

$$S^{\mu\nu} = \{\sigma^\mu, \sigma^\nu\}_+. \quad (28.0.741)$$

Show that

$$S = 2 \begin{pmatrix} \sigma^4 & 0 & 0 & \sigma^1 \\ 0 & \sigma^4 & 0 & \sigma^2 \\ 0 & 0 & \sigma^4 & \sigma^3 \\ \sigma^1 & \sigma^2 & \sigma^3 & \sigma^4 \end{pmatrix}. \quad (28.0.742)$$

Look at the matrices  $N^\alpha$  for  $\alpha = 1, 2, 3$  defined by (3.180) through (3.182). Also define a matrix  $N^4$  by the rule

$$N^4 = I \quad (28.0.743)$$

where  $I$  is the  $4 \times 4$  identity matrix. Show that

$$S = 2 \sum_{\alpha=1}^4 N^\alpha \sigma^\alpha. \quad (28.0.744)$$

Consequently there is the associated component relation

$$\{\sigma^\mu, \sigma^\nu\}_+ = S^{\mu\nu} = 2 \sum_{\alpha=1}^4 (N^\alpha)^{\mu\nu} \sigma^\alpha \quad (28.0.745)$$

from which it follows that

$$\text{tr}[\sigma^\beta(\sigma^\mu\sigma^\nu + \sigma^\nu\sigma^\mu)] = \text{tr}[\sigma^\beta\{\sigma^\mu, \sigma^\nu\}_+] = 2 \sum_{\alpha=1}^4 (N^\alpha)^{\mu\nu} \text{tr}(\sigma^\beta\sigma^\alpha) = 4(N^\beta)^{\mu\nu}. \quad (28.0.746)$$

Finally, show that comparison of (2.249) and (2.257) gives the result

$$V^{\beta\mu\nu} = 4(N^\beta)^{\mu\nu}. \quad (28.0.747)$$

With these definitions before us, show that

$$(\mathbf{m} \cdot \boldsymbol{\sigma})\sigma^\nu + \sigma^\nu(\mathbf{m} \cdot \boldsymbol{\sigma}) = \sum_{\xi=1}^4 m_\xi \{\sigma^\xi, \sigma^\nu\}_+ = \sum_{\xi=1}^4 m_\xi S^{\xi\nu} = 2 \sum_{\xi=1}^4 \sum_{\alpha=1}^4 m_\xi (N^\alpha)^{\xi\nu} \sigma^\alpha. \quad (28.0.748)$$

Continue on to show that

$$\text{tr}\{\sigma^\mu[(\mathbf{m} \cdot \boldsymbol{\sigma})\sigma^\nu + \sigma^\nu(\mathbf{m} \cdot \boldsymbol{\sigma})]\} = 2 \sum_{\xi=1}^4 \sum_{\alpha=1}^4 m_\xi (N^\alpha)^{\xi\nu} \text{tr}[\sigma^\mu \sigma^\alpha] = 4 \sum_{\xi=1}^4 m_\xi (N^\mu)^{\xi\nu}. \quad (28.0.749)$$

But, from the relation (2.258) and the symmetry properties of  $V$ , it follows that

$$(N^\mu)^{\xi\nu} = (N^\xi)^{\mu\nu}. \quad (28.0.750)$$

Show, therefore, that (2.260) can be rewritten in the form

$$\text{tr}\{\sigma^\mu[(\mathbf{m} \cdot \boldsymbol{\sigma})\sigma^\nu + \sigma^\nu(\mathbf{m} \cdot \boldsymbol{\sigma})]\} = 4 \sum_{\xi=1}^4 m_\xi (N^\mu)^{\xi\nu} = 4 \sum_{\xi=1}^4 m_\xi (N^\xi)^{\mu\nu} = 4(\mathbf{m} \cdot \mathbf{N})^{\mu\nu}. \quad (28.0.751)$$

## Resume and Complete Evaluation of First Factor

Show, as a result of these intervening calculations, that combining (2.269) and (2.286) yields the pleasingly simple result

$$\Lambda^{\mu\nu}(u) = \delta^{\mu\nu} + \epsilon(\mathbf{m} \cdot \mathbf{N})^{\mu\nu} + O(\epsilon^2) \quad (28.0.752)$$

or, in matrix form,

$$\Lambda(u) = I + \epsilon \mathbf{m} \cdot \mathbf{N} + O(\epsilon^2). \quad (28.0.753)$$

[We remark that had we replaced the matrices in (7.3.235) through (7.3.327) by their negatives, which would not have affected the commutation rules (7.328) through (7.3.240), the + signs in (2.291) and (2.292) would have been replaced by - signs. This change would lead to unpleasant consequences in what follows.]

As promised, we now intend to use (2.292) and the group property (2.219) to work out results for finite values of  $\lambda$ . Review the steps (2.266) through (2.269) that led from (2.265) to (2.269). Demonstrate that analogous steps lead from (2.292) to the expected and desired relation

$$\Lambda[\exp(\lambda \mathbf{m} \cdot \hat{\mathbf{N}})] = \exp(\lambda \mathbf{m} \cdot \mathbf{N}). \quad (28.0.754)$$

What would have been the resulting relation had the + sign in (2.292) been a - sign?

## Summary of Results

We began this exercise with the knowledge that the  $SL(2, \mathbb{C})$  and Lorentz groups have the same Lie algebras and analogous polar decompositions. We therefore expect they are closely related. In this exercise we found that (2.208) provides a map that sends elements of  $SL(2, \mathbb{C})$  into elements of the Lorentz group, and (2.219) showed that this map is a homomorphism. Subsequently application of this homomorphism to elements of  $SL(2, \mathbb{C})$  written in polar form produced the factorization (2.226). The results for each factor were then found to be given by (2.269) and (2.293). Verify that combining (2.226), (2.269), and (2.293) produces the relation

$$\begin{aligned}\Lambda[\exp(\lambda \mathbf{m} \cdot \hat{\mathbf{N}}) \exp(\theta \mathbf{n} \cdot \hat{\mathbf{L}})] &= \Lambda[\exp(\lambda \mathbf{m} \cdot \hat{\mathbf{N}})] \Lambda[\exp(\theta \mathbf{n} \cdot \hat{\mathbf{L}})] \\ &= \exp(\lambda \mathbf{m} \cdot \mathbf{N}) \exp(\theta \mathbf{n} \cdot \mathbf{L}).\end{aligned}\quad (28.0.755)$$

This relation shows that the map (2.208) has the property that every element in the Lorentz group has a preimage in  $SL(2, \mathbb{C})$ . From (2.224) it follows that every element of the Lorentz group has (at least) two distinct preimages in  $SL(2, \mathbb{C})$ . Therefore the map (2.208) is not an isomorphism. To be more precise, examination of (2.293) shows that the mapping between the first factors of  $SL(2, \mathbb{C})$  and of the Lorentz group is one to one. And examination of the discussion associated with (2.232) shows that the relation between the second factors of  $SL(2, \mathbb{C})$  and of the Lorentz group is two to one in complete analogy to the relation between  $SU(2)$  and  $SO(3, \mathbb{R})$ . We conclude that the map between  $SL(2, \mathbb{C})$  and the Lorentz group provided by (2.208) and (2.294) is two to one.

What can be said about the topologies of the two groups? Since the topology of  $SL(2, \mathbb{C})$  is  $E^3 \times S^3$ , and both factors are simply connected,  $SL(2, \mathbb{C})$  can be shown to be simply connected. From (7.3.186) we see that the topology of the Lorentz group is  $E^3 \times SO(3, \mathbb{R})$ . Since  $E^3$  is simply connected and  $SO(3, \mathbb{R})$  is doubly connected, the Lorentz group can be shown to be doubly connected. All these topological statements are consistent with the nature of the map between  $SL(2, \mathbb{C})$  and the Lorentz group provided by (2.208) and (2.294). It follows that  $SL(2, \mathbb{C})$  is the covering group of the Lorentz group.

**28.0.28.** Let  $\phi$  and  $\mathbf{k}$  be two real three-component vectors. Use them to parameterize the  $s\ell(2, \mathbb{C})$  element  $\hat{S}$  defined by

$$\hat{S} = \phi \cdot \hat{\mathbf{L}} + \mathbf{k} \cdot \hat{\mathbf{N}}. \quad (28.0.756)$$

In turn, let  $v$  be the  $SL(2, \mathbb{C})$  element defined by

$$v = \exp(\hat{S}). \quad (28.0.757)$$

Prove that for the map (2.208) there is the relation

$$\Lambda(v) = \Lambda[\exp(\hat{S})] = \exp(\phi \cdot \mathbf{L} + \mathbf{k} \cdot \mathbf{N}). \quad (28.0.758)$$

Suggestion: First show that for small  $\epsilon$  there is the result

$$\Lambda[\exp(\epsilon \hat{S})] = \exp[\epsilon(\phi \cdot \mathbf{L} + \mathbf{k} \cdot \mathbf{N})] + O(\epsilon^2). \quad (28.0.759)$$

This small  $\epsilon$  result can be obtained by direct evaluation of (2.208) or, more easily, by use of (2.269), (2.293), and the BCH formula. Then use (2.298) and the homomorphism relation (2.219) to work out results for any finite value of  $\epsilon$  including  $\epsilon = 1$ .

**28.0.29.** When can and when cannot an element in  $SL(2, \mathbb{C})$  be written in single exponential form? And what can be said about the related question for the Lorentz group? Your task for this exercise is to answer these questions.

Begin with the case of  $SL(2, \mathbb{C})$ . Recall that any  $2 \times 2$  matrix  $v$  can be diagonalized if its eigenvalues are distinct. That is, if the eigenvalues of  $v$  are distinct, it can be written in the form

$$v = ada^{-1} \quad (28.0.760)$$

where  $a$  is a nonsingular matrix,  $d$  is the diagonal matrix

$$d = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \quad (28.0.761)$$

and  $\lambda_1, \lambda_2$  are its eigenvalues. Verify that in this case

$$\det(v) = \det(d) = \lambda_1 \lambda_2. \quad (28.0.762)$$

Moreover,

$$\lambda_1 \lambda_2 = 1 \quad (28.0.763)$$

if  $v$  is an element of  $SL(2, \mathbb{C})$ . Verify therefore that, in the  $SL(2, \mathbb{C})$  and distinct eigenvalue case,  $d$  can be written in the form

$$d = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} = \exp(\alpha \sigma^3) \quad (28.0.764)$$

where

$$\alpha = \log(\lambda). \quad (28.0.765)$$

Consequently, verify that any element of  $SL(2, \mathbb{C})$  with distinct eigenvalues can be written in the forms

$$v = ada^{-1} = a[\exp(\alpha \sigma^3)]a^{-1} = \exp(\alpha a \sigma^3 a^{-1}). \quad (28.0.766)$$

Verify that the matrix  $a \sigma^3 a^{-1}$  is traceless, and therefore there is a three-component (possibly complex) vector  $\beta$  such that

$$\alpha a \sigma^3 a^{-1} = \beta \cdot \sigma. \quad (28.0.767)$$

Show that the net result of these deliberations is the relation

$$v = \exp(\beta \cdot \sigma), \quad (28.0.768)$$

which demonstrates that any element of  $SL(2, \mathbb{C})$  with distinct eigenvalues can be written in single exponential form.

There remain the cases in which the eigenvalues of  $v$  are not distinct, in which cases because of (2.302) there are the eigenvalue possibilities  $1, 1$  and  $-1, -1$ . Suppose that in these non distinct cases that  $v$  can nevertheless be diagonalized. Show that then there are the two possibilities

$$v = \sigma^4 = \exp(\mathbf{0}) \quad (28.0.769)$$

and

$$v = -\sigma^4 = \exp(i\pi\sigma^3). \quad (28.0.770)$$

Here, as before,  $\mathbf{0}$  is the matrix with all zero entries. Evidently in both these cases  $v$  has again been written in single exponential form.

The last possibility for these non distinct cases is that  $v$  cannot be diagonalized. Verify that for this possibility there are only the cases

$$v = aj_{\pm}a^{-1} \quad (28.0.771)$$

where  $j_{\pm}$  are the two *Jordan* normal form elements

$$j_+ = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad (28.0.772)$$

and

$$j_- = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}. \quad (28.0.773)$$

Consider the case with eigenvalues 1, 1 and the Jordan normal form  $j_+$  given by (2.311). Show that in this case

$$j_+ = \exp[(1/2)(\sigma^1 + i\sigma^2)] \quad (28.0.774)$$

from which it follows that

$$v = \exp[(1/2)a(\sigma^1 + i\sigma^2)a^{-1}]. \quad (28.0.775)$$

Verify that the exponent appearing in (2.314) is in  $s\ell(2, \mathbb{C})$ . You have demonstrated that in the 1, 1 eigenvalue case every such  $v$  in  $SL(2, \mathbb{C})$  can be written in single exponential form with the exponent being an element in  $s\ell(2, \mathbb{C})$ .

The final remaining case is that with eigenvalues  $-1, -1$  and the Jordan normal form  $j_-$  given by (2.312). From the work of Exercise 3.7.12, which you should review, we know that  $j_-$  cannot be written in single exponential form. Show it follows that all  $SL(2, \mathbb{C})$  elements that have  $j_-$  as their Jordan normal form cannot be written in single exponential form.

The case of  $SL(2, \mathbb{C})$  has been dispatched: All elements in  $SL(2, \mathbb{C})$  can be written in single exponential form except those having  $j_-$  as their Jordan normal form.

We now turn to the case of the Lorentz group. According to Exercise 2.15 above, all the Lorentz group elements associated with the  $SL(2, \mathbb{C})$  elements that can be written in single exponential form can also be written in single exponential form. What about the Lorentz group elements associated with the  $SL(2, \mathbb{C})$  elements whose Jordan normal form is  $j_-$ ? Let  $v_-$  be any such  $SL(2, \mathbb{C})$  element. What we wish to determine is the nature of  $\Lambda(v_-)$ . Show that if a  $v$  in  $SL(2, \mathbb{C})$  cannot be written in single exponential form, then  $-v$  can. In particular, show that  $-j_-$  has Jordan normal form  $j_+$ . (Again see the work of Exercise 3.7.12.) Verify it follows that  $-v_-$  can be written in single exponential form, and therefore  $\Lambda(-v_-)$  can also be written in single exponential form. But from (2.225) we know that

$$\Lambda(v_-) = \Lambda(-v_-). \quad (28.0.776)$$

Therefore  $\Lambda(v_-)$ , which is the same as  $\Lambda(-v_-)$ , can be written in single exponential form. Conclude that *all* Lorentz group elements can be written in single exponential form!

**28.0.30.** Review Exercise 7.3.30. It showed that the use of  $sl(2, \mathbb{C})$  provides the  $\Gamma(1/2, 0)$  and  $\Gamma(0, 1/2)$  representations of the Lorentz group Lie algebra. Review Exercise 7.3.34 that constructed an isomorphism between  $n \times n$  possibly complex matrices and  $2n \times 2n$  real matrices. The purpose of this exercise and the next is to describe how the results of Exercises 7.3.30 and 7.3.34 may be used to characterize/determine the effect of Lorentz transformations on what we will call Dirac 4-spinors.

Recall the  $SL(2, \mathbb{C})$  group elements  $\hat{\Lambda}$  given by (7.3.246). They are  $2 \times 2$  possibly complex matrices. According to Exercise 7.3.30 they carry the representation  $\Gamma(0, 1/2)$  of the Lorentz group. Make the Ansatz

$$k = \hat{\Lambda} \quad (28.0.777)$$

and use (7.3.375) to define associated *real*  $4 \times 4$  matrices. For example, suppose

$$\hat{\Lambda} = \exp(\theta \hat{L}^3). \quad (28.0.778)$$

Verify that in this case

$$\hat{\Lambda} = \exp(\theta \hat{L}^3) = \begin{pmatrix} \exp(-i\theta/2) & 0 \\ 0 & \exp(i\theta/2) \end{pmatrix} = \quad (28.0.779)$$

$$\begin{pmatrix} \cos(\theta/2) & 0 \\ 0 & \cos(\theta/2) \end{pmatrix} + i \begin{pmatrix} -\sin(\theta/2) & 0 \\ 0 & \sin(\theta/2) \end{pmatrix}, \quad (28.0.780)$$

and

$$K(k) = K(\hat{\Lambda}) = K[\exp(\theta \hat{L}^3)] = \begin{pmatrix} \cos(\theta/2) & 0 & \sin(\theta/2) & 0 \\ 0 & \cos(\theta/2) & 0 & -\sin(\theta/2) \\ -\sin(\theta/2) & 0 & \cos(\theta/2) & 0 \\ 0 & \sin(\theta/2) & 0 & \cos(\theta/2) \end{pmatrix}. \quad (28.0.781)$$

Verify that when  $\theta = 2\pi$  there are the results

$$\hat{\Lambda} = \exp(2\pi \hat{L}^3) = -I^{[2]} \quad (28.0.782)$$

and

$$K(\hat{\Lambda}) = K[\exp(2\pi \hat{L}^3)] = -I^{[4]}. \quad (28.0.783)$$

Since the  $\hat{\Lambda}$  are  $2 \times 2$  matrices, they act on two-dimensional objects/arrays. Since the  $K(\hat{\Lambda})$  are  $4 \times 4$  matrices, they act on four-dimensional objects/arrays. But the  $2 \times 2$  and  $4 \times 4$  matrix/group elements given by (2.321) and (2.322) and corresponding to the Lorentz transformation consisting of a  $\theta = 2\pi$  rotation about the  $z$  axis are *not* identity matrices. It follows that the objects/arrays on which they act are *not* vectors.<sup>16</sup> Rather, they are 2-spinors and 4-spinors, respectively.<sup>17</sup> Also note that, by the constructions (7.3.246) and (7.3.375), all elements  $K(\hat{\Lambda})$  are continuously connected to the identity matrix  $I^{[4]}$ . As

<sup>16</sup>By their transformation properties ye shall know them.

<sup>17</sup>Sometimes what we have called 4-spinors are called *bispinors*.

further evidence that the four-component arrays that we have called 4-spinors are not vectors, observe that, according to (7.3.167),  $-I^{[4]}$  is not a Lorentz transformation in the identity component of the Lorentz group.

Let  $b$  be any of the possibly complex  $2 \times 2$   $sl(2, \mathbb{C})$  basis elements (7.3.236) through (7.3.241) and consider the associated *real*  $4 \times 4$  matrices  $K(b)$ . What representation of the Lorentz group Lie algebra is provided by the matrices  $K(\hat{L}^j)$  and  $K(\hat{N}^j)$ ? Verify, using (7.3.383), that

$$WK(b)W^{-1} = \begin{pmatrix} b & 0 \\ 0 & \bar{b} \end{pmatrix}. \quad (28.0.784)$$

We know from the work of Exercise 7.3.30 that the matrices  $b$  provide the  $\Gamma(0, 1/2)$  representation of the Lorentz group Lie algebra. Momentarily we will demonstrate that the matrices  $\bar{b}$  provide the  $\Gamma(1/2, 0)$  representation of the Lorentz group Lie algebra. That is, for the case of  $sl(2, \mathbb{C})$ , complex conjugation is equivalent to the grave operation. It follows from (2.323) that the matrices  $K(b)$  provide the direct sum

$$\Gamma(0, 1/2) \oplus \Gamma(1/2, 0) = \Gamma(0, 1/2) \oplus \bar{\Gamma}(0, 1/2) = \bar{\Gamma}(1/2, 0) \oplus \Gamma(1/2, 0) \quad (28.0.785)$$

representation of the Lorentz group Lie algebra.

Now work on the advertised demonstration: We already know that complex conjugation is the result of the breve operation and that this operation is a conjugacy operation. See (3.7.223). For present purposes what we need to show is that, in the case of  $sl(2, \mathbb{C})$ , the breve/bar operation and the grave operation are equivalent in the sense of (3.7.218). Verify that for the bar and grave operations there are the relations

$$(\hat{\bar{L}}^1, \hat{\bar{L}}^2, \hat{\bar{L}}^3, \hat{\bar{N}}^1, \hat{\bar{N}}^2, \hat{\bar{N}}^3) = (-\hat{L}^1, \hat{L}^2, -\hat{L}^3, \hat{N}^1, -\hat{N}^2, \hat{N}^3), \quad (28.0.786)$$

$$(\hat{\grave{L}}^1, \hat{\grave{L}}^2, \hat{\grave{L}}^3, \hat{\grave{N}}^1, \hat{\grave{N}}^2, \hat{\grave{N}}^3) = (\hat{L}^1, \hat{L}^2, \hat{L}^3, -\hat{N}^1, -\hat{N}^2, -\hat{N}^3). \quad (28.0.787)$$

See (7.3.248) and (7.3.249). We will next show is that the contents of the right sides of (2.325) and (2.326) are related by a similarity transformation. Recall the relation (3.7.234) which involved the matrix

$$J_2 = i\sigma^2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad (28.0.788)$$

Using the anticommutation relations (5.7.41) verify that

$$J_2\sigma^1 J_2^{-1} = -\sigma^1, \quad (28.0.789)$$

$$J_2\sigma^2 J_2^{-1} = \sigma^2, \quad (28.0.790)$$

$$J_2\sigma^3 J_2^{-1} = -\sigma^3. \quad (28.0.791)$$

Show it follows from the definitions (7.3.236) through (7.3.241) and the results (2.328) through (3.230) that there are the similarity relations

$$J_2(-\hat{L}^1, \hat{L}^2, -\hat{L}^3, \hat{N}^1, -\hat{N}^2, \hat{N}^3)J_2^{-1} = (\hat{L}^1, \hat{L}^2, \hat{L}^3, -\hat{N}^1, -\hat{N}^2, -\hat{N}^3). \quad (28.0.792)$$

Thus the right sides of (2.325) and (2.326) are related by a similarity transformation, and therefore the left sides of (2.325) and (2.326) are related by a similarity transformation.

There is a second way of verifying that the matrices  $K(\hat{L}^j)$  and  $K(\hat{N}^j)$  provide a representation of  $sl(2, \mathbb{C})$  that is equivalent to the direct sum (2.324). It essentially amounts to the previous argument, but appears to be more direct. Verify, using (7.3.236) through (7.3.241), that there are the *six* explicit results

$$K(\hat{L}^1) = (1/2)K(-i\sigma^1) = (1/2) \begin{pmatrix} \mathbf{0} & \sigma^1 \\ -\sigma^1 & \mathbf{0} \end{pmatrix}, \quad (28.0.793)$$

$$K(\hat{L}^2) = (1/2)K(-i\sigma^2) = (1/2) \begin{pmatrix} -i\sigma^2 & \mathbf{0} \\ \mathbf{0} & -i\sigma^2 \end{pmatrix}, \quad (28.0.794)$$

$$K(\hat{L}^3) = (1/2)K(-i\sigma^3) = (1/2) \begin{pmatrix} \mathbf{0} & \sigma^3 \\ -\sigma^3 & \mathbf{0} \end{pmatrix}, \quad (28.0.795)$$

$$K(\hat{N}^1) = (1/2)K(\sigma^1) = (1/2) \begin{pmatrix} \sigma^1 & \mathbf{0} \\ \mathbf{0} & \sigma^1 \end{pmatrix}, \quad (28.0.796)$$

$$K(\hat{N}^2) = (1/2)K(\sigma^2) = (1/2) \begin{pmatrix} \mathbf{0} & i\sigma^2 \\ -i\sigma^2 & \mathbf{0} \end{pmatrix}, \quad (28.0.797)$$

$$K(\hat{N}^3) = (1/2)K(\sigma^3) = (1/2) \begin{pmatrix} \sigma^3 & \mathbf{0} \\ \mathbf{0} & \sigma^3 \end{pmatrix}. \quad (28.0.798)$$

Next let  $V$  be the matrix defined by the rule

$$V = (1/\sqrt{2}) \begin{pmatrix} I^{[2]} & iI^{[2]} \\ i\sigma^2 & \sigma^2 \end{pmatrix}. \quad (28.0.799)$$

Note that  $V$  involves  $i\sigma^2$ , which is  $J_2$  in disguise, just as (2.331) involves  $J_2$ . Show that  $V$  is unitary so that

$$V^{-1} = V^\dagger = (1/\sqrt{2}) \begin{pmatrix} I^{[2]} & -i\sigma^2 \\ -iI^{[2]} & \sigma^2 \end{pmatrix}. \quad (28.0.800)$$

Finally, by executing the indicated matrix multiplications, verify the six similarity relation results

$$VK(\hat{L}^1)V^{-1} = (-i/2) \begin{pmatrix} \sigma^1 & \mathbf{0} \\ \mathbf{0} & \sigma^1 \end{pmatrix}, \quad (28.0.801)$$

$$VK(\hat{L}^2)V^{-1} = (-i/2) \begin{pmatrix} \sigma^2 & \mathbf{0} \\ \mathbf{0} & \sigma^2 \end{pmatrix}, \quad (28.0.802)$$

$$VK(\hat{L}^3)V^{-1} = (-i/2) \begin{pmatrix} \sigma^3 & \mathbf{0} \\ \mathbf{0} & \sigma^3 \end{pmatrix}, \quad (28.0.803)$$

$$VK(\hat{N}^1)V^{-1} = (1/2) \begin{pmatrix} \sigma^1 & \mathbf{0} \\ \mathbf{0} & -\sigma^1 \end{pmatrix}, \quad (28.0.804)$$

$$VK(\hat{N}^2)V^{-1} = (1/2) \begin{pmatrix} \sigma^2 & \mathbf{0} \\ \mathbf{0} & -\sigma^2 \end{pmatrix}, \quad (28.0.805)$$

$$VK(\hat{N}^3)V^{-1} = (1/2) \begin{pmatrix} \sigma^3 & \mathbf{0} \\ \mathbf{0} & -\sigma^3 \end{pmatrix}. \quad (28.0.806)$$

In summary, there are the results

$$VK(\hat{L}^j)V^{-1} = \begin{pmatrix} \hat{L}^j & \mathbf{0} \\ \mathbf{0} & \hat{\tilde{L}}^j \end{pmatrix} = \begin{pmatrix} \hat{L}^j & \mathbf{0} \\ \mathbf{0} & \hat{L}^j \end{pmatrix}, \quad (28.0.807)$$

$$VK(\hat{N}^j)V^{-1} = \begin{pmatrix} \hat{N}^j & \mathbf{0} \\ \mathbf{0} & \hat{\tilde{N}}^j \end{pmatrix} = \begin{pmatrix} \hat{N}^j & \mathbf{0} \\ \mathbf{0} & -\hat{N}^j \end{pmatrix}. \quad (28.0.808)$$

Evidently the matrices on the right sides of (2.340) through (2.347) are block diagonal. Moreover, the upper blocks carry the representation  $\Gamma(0, 1/2)$  of  $sl(2, \mathbb{C})$  and the lower blocks carry the representation  $\Gamma(1/2, 0)$ . See Exercise 7.3.30. Thus the full matrices carry the representation

$$\Gamma(0, 1/2) \oplus \Gamma(1/2, 0). \quad (28.0.809)$$

**28.0.31.** This exercise is a continuation of the previous exercise and is devoted to sketching the relation of the  $sl(2, \mathbb{C})$  representation provided by the real  $4 \times 4$  matrices  $K(\hat{L}^j)$  and  $K(\hat{N}^j)$ , see (7.3.385), to some of the mathematical machinery associated with some forms of the *Dirac* (1902-1984) equation.

## Background

Discussion of the Dirac equation often begins with the introduction of four  $4 \times 4$  *gamma* matrices, denoted as  $\gamma^\mu$  with  $\mu = 1 \cdots 4$ , which are required to satisfy the *anticommutation* rules

$$\{\gamma^\mu, \gamma^\nu\}_+ = \gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2g^{\mu\nu}I^{[4]}. \quad (28.0.810)$$

The relations (2.349) are sometimes called the *Dirac algebra*.<sup>18</sup> For our purposes it is convenient to take the  $\gamma^\mu$  to be the matrices

$$\gamma^1 = \begin{pmatrix} \mathbf{0} & i\sigma^3 \\ i\sigma^3 & \mathbf{0} \end{pmatrix} = i \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & -0 \end{pmatrix}, \quad (28.0.811)$$

$$\gamma^2 = \begin{pmatrix} iI^{[2]} & \mathbf{0} \\ \mathbf{0} & -iI^{[2]} \end{pmatrix} = i \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad (28.0.812)$$

$$\gamma^3 = \begin{pmatrix} \mathbf{0} & -i\sigma^1 \\ -i\sigma^1 & \mathbf{0} \end{pmatrix} = i \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}, \quad (28.0.813)$$

---

<sup>18</sup>Dirac algebra is a particular case of *Clifford* (1845-1879) algebra.

$$\gamma^4 = \begin{pmatrix} \mathbf{0} & -\sigma^2 \\ -\sigma^2 & \mathbf{0} \end{pmatrix} = i \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}. \quad (28.0.814)$$

Verify that the  $\gamma^\mu$  do indeed satisfy the Dirac algebra (2.349). Note that they are all *purely imaginary*. Representations of the Dirac algebra that are purely imaginary are called *Majorana* (1906-1938) representations. Thus the Ansätze (2.350) through (2.353) provide a Majorana representation.<sup>19</sup>

Verify that for this Majorana representation there are the results

$$(\gamma^\mu)^\dagger = -\gamma^\mu \text{ for } \mu = 1, 2, 3 \quad (28.0.815)$$

and

$$(\gamma^4)^\dagger = \gamma^4. \quad (28.0.816)$$

Verify that, in view of (2.349), the results (2.354) and (2.355) are equivalent to the relations

$$(\gamma^\mu)^\dagger = \gamma^4 \gamma^\mu \gamma^4 \text{ for } \mu = 1 \dots 4. \quad (28.0.817)$$

Finally, verify that  $\gamma^4$  can be written in the form

$$\gamma^4 = i \begin{pmatrix} \mathbf{0} & J_2 \\ J_2 & \mathbf{0} \end{pmatrix}. \quad (28.0.818)$$

### Definition and some Properties of $\gamma^5$

It is useful to define a fifth  $4 \times 4$  gamma matrix  $\gamma^5$  in terms of the  $\gamma^\mu$  by the rule

$$\gamma^5 = i\gamma^1\gamma^2\gamma^3\gamma^4. \quad (28.0.819)$$

Verify that it too is purely imaginary in a Majorana representation.<sup>20</sup> Show that, for the above Majorana representation given by (2.350) through (2.353), there is the result

$$\gamma^5 = i \begin{pmatrix} \mathbf{0} & -I^{[2]} \\ I^{[2]} & \mathbf{0} \end{pmatrix} = -iJ \quad (28.0.820)$$

where here  $J$  denotes the  $4 \times 4$  version of (3.1.1),

$$J = \begin{pmatrix} \mathbf{0} & I^{[2]} \\ -I^{[2]} & \mathbf{0} \end{pmatrix}. \quad (28.0.821)$$

It follows that  $\gamma^5$  is antisymmetric in our Majorana representation,

$$(\gamma^5)^T = -\gamma^5. \quad (28.0.822)$$

---

<sup>19</sup>There are several possible Majorana representations. As trivial examples of how one Majorana representation may be converted into another, the  $\gamma^\mu$  or any subset of the  $\gamma^\mu$  may be multiplied by  $-1$ . Also, the  $\gamma^\mu$  for  $\mu = 1 \dots 3$  may be permuted among each other. See, in addition, Exercise 2.21.

<sup>20</sup>The notation  $\gamma^5$  can be misleading. Whereas we will see that in some contexts the contravariant index  $\nu$  in  $\gamma^\nu$  for  $\nu = 1 \dots 4$  can be lowered using  $g_{\mu\nu}$ , there is in the Dirac machinery no analogous operation for  $\gamma^5$ . Thus, in the Dirac machinery, the 5 in  $\gamma^5$  is simply a label and not a contravariant index.

The matrix  $\gamma^5$  defined by (2.358) has some other properties that follow from the Dirac algebra relations (2.349) and therefore hold in any representation of the  $\gamma^\mu$ . Verify from (2.349) and (2.358) that

$$\{\gamma^\mu, \gamma^5\}_+ = 0, \quad (28.0.823)$$

$$\{\gamma^\mu \gamma^\nu, \gamma^5\} = 0, \quad (28.0.824)$$

and

$$(\gamma^5)^2 = I^{[4]}. \quad (28.0.825)$$

### Definition of the $\hat{\sigma}^{\mu\nu}$ and their Relation to the $K(\hat{L}^j)$ and $K(\hat{N}^j)$

Also, for  $\mu, \nu = 1 \cdots 4$ , define matrices  $\hat{\sigma}^{\mu\nu}$  by the *commutators*

$$\hat{\sigma}^{\mu\nu} = (1/2)\{\gamma^\mu, \gamma^\nu\} \Rightarrow \hat{\sigma}^{\mu\nu} = \gamma^\mu \gamma^\nu \text{ for } \mu \neq \nu. \quad (28.0.826)$$

[Here, in writing the second relation in (2.365), we have used the Dirac algebra.] Note that, because the  $\gamma^\mu$  are purely imaginary in a Majorana representation, the  $\hat{\sigma}^{\mu\nu}$  are purely real.<sup>21</sup> We will soon see that the  $\hat{\sigma}^{\mu\nu}$  are related to the Lorentz group Lie generator matrices  $K(\hat{L}^j)$  and  $K(\hat{N}^j)$  and, since these  $K$  matrices are real, we would like the  $\hat{\sigma}^{\mu\nu}$  to be real. That is the reason for our use of a Majorana representation for the  $\gamma^\mu$ ; and indeed it will transpire that we have chosen a particular Majorana representation to make things work out neatly.

Verify that there are the *six* results

$$\hat{\sigma}^{12} = \gamma^1 \gamma^2 = \begin{pmatrix} \mathbf{0} & \sigma^3 \\ -\sigma^3 & \mathbf{0} \end{pmatrix}, \quad (28.0.827)$$

$$\hat{\sigma}^{23} = \gamma^2 \gamma^3 = \begin{pmatrix} \mathbf{0} & \sigma^1 \\ -\sigma^1 & \mathbf{0} \end{pmatrix}, \quad (28.0.828)$$

$$\hat{\sigma}^{31} = \gamma^3 \gamma^1 = \begin{pmatrix} -i\sigma^2 & \mathbf{0} \\ \mathbf{0} & -i\sigma^2 \end{pmatrix}, \quad (28.0.829)$$

$$\hat{\sigma}^{41} = \gamma^4 \gamma^1 = \begin{pmatrix} \sigma^1 & \mathbf{0} \\ \mathbf{0} & \sigma^1 \end{pmatrix}, \quad (28.0.830)$$

$$\hat{\sigma}^{42} = \gamma^4 \gamma^2 = \begin{pmatrix} \mathbf{0} & i\sigma^2 \\ -i\sigma^2 & \mathbf{0} \end{pmatrix}, \quad (28.0.831)$$

$$\hat{\sigma}^{43} = \gamma^4 \gamma^3 = \begin{pmatrix} \sigma^3 & \mathbf{0} \\ \mathbf{0} & \sigma^3 \end{pmatrix}, \quad (28.0.832)$$

and that all other  $\hat{\sigma}^{\mu\nu}$  results can be obtained from the antisymmetry relation

$$\hat{\sigma}^{\mu\nu} = -\hat{\sigma}^{\nu\mu}. \quad (28.0.833)$$

---

<sup>21</sup>Observe also that we depart from the usual Dirac literature definition/notation  $\sigma^{\mu\nu} = (i/2)\{\gamma^\mu, \gamma^\nu\}$  by omitting a factor of  $i$  and placing a hat  $\hat{\cdot}$  on  $\sigma$  to indicate that a change of notation has occurred. As described in the end of Exercise 3.7.43, we do this to avoid introducing mathematically unnecessary factors of  $i$ .

For comparison, recall the *six* results (2.332) through (2.337). Verify from (7.3.236) through (7.3.241), (2.332) through (2.337), and (2.366) through (2.371) that there are the relations

$$(1/2)\hat{\sigma}^{12} = (1/2)K(-i\sigma^3) = K(\hat{L}^3), \quad (28.0.834)$$

$$(1/2)\hat{\sigma}^{23} = (1/2)K(-i\sigma^1) = K(\hat{L}^1), \quad (28.0.835)$$

$$(1/2)\hat{\sigma}^{31} = (1/2)K(-i\sigma^2) = K(\hat{L}^2), \quad (28.0.836)$$

$$(1/2)\hat{\sigma}^{41} = (1/2)K(\sigma^1) = K(\hat{N}^1), \quad (28.0.837)$$

$$(1/2)\hat{\sigma}^{42} = (1/2)K(\sigma^2) = K(\hat{N}^2), \quad (28.0.838)$$

$$(1/2)\hat{\sigma}^{43} = (1/2)K(\sigma^3) = K(\hat{N}^3). \quad (28.0.839)$$

We see that, when the Majorana representation (2.350) through (2.353) is employed for the  $\gamma^\mu$ , the matrices  $(1/2)\hat{\sigma}^{\mu\nu}$  are simply a relabeling of the matrices  $K(\hat{L}^j)$  and  $K(\hat{N}^j)$ .

It follows that, at least in this case, the matrices  $(1/2)\hat{\sigma}^{\mu\nu}$  satisfy the Lorentz group Lie algebra commutation rules. For example, verify using (2.373) through (2.375), (7.3.385), and (7.3.243) that there is the commutation rule

$$\{(1/2)\hat{\sigma}^{12}, (1/2)\hat{\sigma}^{23}\} = \{K(\hat{L}^3), K(\hat{L}^1)\} = K(\{\hat{L}^3, \hat{L}^1\}) = K(\hat{L}^2) = (1/2)\hat{\sigma}^{31}. \quad (28.0.840)$$

Actually, the same commutation rules for the matrices  $(1/2)\hat{\sigma}^{\mu\nu}$  hold for any representation choice for the  $\gamma^\mu$ , and are in fact a consequence of the Dirac algebra (2.349). Consider again, for example, the commutator  $\{(1/2)\hat{\sigma}^{12}, (1/2)\hat{\sigma}^{23}\}$ . Verify, using (2.349) and (2.365), that

$$\begin{aligned} \{(1/2)\hat{\sigma}^{12}, (1/2)\hat{\sigma}^{23}\} &= (1/4)\{\gamma^1\gamma^2, \gamma^2\gamma^3\} = (1/4)(\gamma^1\gamma^2\gamma^2\gamma^3 - \gamma^2\gamma^3\gamma^1\gamma^2) = \\ &(1/4)(-\gamma^1\gamma^3 - \gamma^3\gamma^1\gamma^2\gamma^2) = (1/4)(\gamma^3\gamma^1 + \gamma^3\gamma^1) = (1/2)\gamma^3\gamma^1 = (1/2)\hat{\sigma}^{31}, \end{aligned} \quad (28.0.841)$$

as expected. The matrices  $(1/2)\hat{\sigma}^{\mu\nu}$  depend on the choice of representation for the  $\gamma^\mu$ , but their commutation rules do not.

## Additional Notation

At this point it is useful to introduce additional notation. Since the  $\gamma^\mu$  (for  $\mu = 1 \dots 5$ ) are purely imaginary in a Majorana representation, we may write them in the form

$$\gamma^\mu = i\gamma_r^\mu \quad (28.0.842)$$

where the  $\gamma_r^\mu$  are purely *real*. For example, from (2.353) we see that

$$\gamma_r^4 = \begin{pmatrix} \mathbf{0} & J_2 \\ J_2 & \mathbf{0} \end{pmatrix}, \quad (28.0.843)$$

and from (2.359) we see that

$$\gamma_r^5 = -J. \quad (28.0.844)$$

Verify that both  $\gamma_r^4$  and  $\gamma_r^5$  are antisymmetric in our Majorana representation,

$$(\gamma_r^4)^T = -\gamma_r^4 \text{ and } (\gamma_r^5)^T = -\gamma_r^5. \quad (28.0.845)$$

### Lorentz Invariance of $\gamma^5$

We are now prepared to further explore some properties of Dirac matrices and 4-spinors. The first step has to do with  $\gamma^5$  and consists of verifying that it is invariant under the action of the Lorentz group. That is, there is the result

$$\{K(k), \gamma^5\} = 0 \Leftrightarrow K^{-1}\gamma^5 K = \gamma^5 \quad (28.0.846)$$

where  $k$  is any element in  $SL(2, \mathbb{C})$ . We will arrive at this result by several steps.

Begin by verifying from (2.363) and (2.365) that

$$\{\gamma^5, \hat{\sigma}^{\mu\nu}\} = 0. \quad (28.0.847)$$

At this point review Exercise 7.3.30. Use (2.316) and (7.3.246) to write

$$k(\lambda, \mathbf{m}; \theta, \mathbf{n}) = \exp(\lambda \mathbf{m} \cdot \hat{\mathbf{N}}) \exp(\theta \mathbf{n} \cdot \hat{\mathbf{L}}). \quad (28.0.848)$$

Verify it follows from (7.3.378) and (7.3.386) that

$$\begin{aligned} K(k) &= K[\exp(\lambda \mathbf{m} \cdot \hat{\mathbf{N}}) \exp(\theta \mathbf{n} \cdot \hat{\mathbf{L}})] \\ &= K[\exp(\lambda \mathbf{m} \cdot \hat{\mathbf{N}})] K[\exp(\theta \mathbf{n} \cdot \hat{\mathbf{L}})] \\ &= \exp[K(\lambda \mathbf{m} \cdot \hat{\mathbf{N}})] \exp[K(\theta \mathbf{n} \cdot \hat{\mathbf{L}})]. \end{aligned} \quad (28.0.849)$$

As a further step, verify it follows from (2.373) through (2.378) and (2.386) that

$$\{K(\hat{L}^j), \gamma^5\} = 0 \quad (28.0.850)$$

and

$$\{K(\hat{N}^j), \gamma^5\} = 0. \quad (28.0.851)$$

Use these results to show that

$$\{\exp[K(\theta \mathbf{n} \cdot \hat{\mathbf{L}})], \gamma^5\} = 0 \quad (28.0.852)$$

and

$$\{\exp[K(\lambda \mathbf{m} \cdot \hat{\mathbf{N}})], \gamma^5\} = 0; \quad (28.0.853)$$

and therefore it follows from (2.388) that (2.385) holds. Note that, in view of (2.381), there is the equivalent *real* matrix relation

$$\{K(k), \gamma_r^5\} = 0. \quad (28.0.854)$$

### Properties of $\gamma_r^4$

Next, in preparation for future use, let us study various properties of  $\gamma_r^4$ . You have already verified that it is antisymmetric. Recall (2.384). Verify that it also satisfies the relation

$$(\gamma_r^4)^2 = -I^{[4]}. \quad (28.0.855)$$

Verify from (2.381) that

$$(\gamma^\mu)^\dagger = -i(\gamma_r^\mu)^T. \quad (28.0.856)$$

Insert this result into (2.356) to obtain the result

$$-i(\gamma_r^\mu)^T = (i)^3 \gamma_r^4 \gamma_r^\mu \gamma_r^4 \Leftrightarrow (\gamma_r^\mu)^T = \gamma_r^4 \gamma_r^\mu \gamma_r^4. \quad (28.0.857)$$

Finally employ (2.394) in (2.396) to obtain (for  $\mu = 1 \cdots 4$ ) the result

$$(\gamma_r^\mu)^T = -(\gamma_r^4)^{-1} \gamma_r^\mu \gamma_r^4 = -\gamma_r^4 \gamma_r^\mu (\gamma_r^4)^{-1}. \quad (28.0.858)$$

The matrix  $\gamma_r^4$  has one further, and remarkable, property with regard to  $SL(2, \mathbb{C})$ /Lorentz transformations. Namely, suppose  $k$  is any element in  $SL(2, \mathbb{C})$ . Then there is the result

$$K^T(k) \gamma_r^4 K(k) = \gamma_r^4. \quad (28.0.859)$$

We/you will also prove this result in steps.

Begin by verifying the following relations:

$$\hat{\sigma}^{\mu\nu} = \gamma^\mu \gamma^\nu = (i)^2 \gamma_r^\mu \gamma_r^\nu = -\gamma_r^\mu \gamma_r^\nu, \quad (28.0.860)$$

$$\begin{aligned} \gamma_r^4 \hat{\sigma}^{\mu\nu} (\gamma_r^4)^{-1} &= -\gamma_r^4 \gamma_r^\mu \gamma_r^\nu (\gamma_r^4)^{-1} = -\gamma_r^4 \gamma_r^\mu (\gamma_r^4)^{-1} \gamma_r^4 \gamma_r^\nu (\gamma_r^4)^{-1} = \\ &- (\gamma_r^\mu)^T (\gamma_r^\nu)^T = -(\gamma_r^\nu \gamma_r^\mu)^T = (\gamma_r^\mu \gamma_r^\nu)^T = -(\hat{\sigma}^{\mu\nu})^T, \end{aligned} \quad (28.0.861)$$

$$(\hat{\sigma}^{\mu\nu})^T = -\gamma_r^4 \hat{\sigma}^{\mu\nu} (\gamma_r^4)^{-1}. \quad (28.0.862)$$

Recall that the  $\hat{\sigma}^{\mu\nu}$  are proportional to the  $K(\hat{N}^j)$  and the  $K(\hat{L}^j)$ . See (2.373) through (2.378). Show from (2.401) that

$$[K(\hat{N}^j)]^T = -\gamma_r^4 K(\hat{N}^j) (\gamma_r^4)^{-1} \quad (28.0.863)$$

and

$$[K(\hat{L}^j)]^T = -\gamma_r^4 K(\hat{L}^j) (\gamma_r^4)^{-1}. \quad (28.0.864)$$

Let us pause for a moment. Note that (2.402) and (2.403) can be rewritten in the form

$$- [K(\hat{N}^j)]^T = \gamma_r^4 K(\hat{N}^j) (\gamma_r^4)^{-1} \quad (28.0.865)$$

and

$$- [K(\hat{L}^j)]^T = \gamma_r^4 K(\hat{L}^j) (\gamma_r^4)^{-1}. \quad (28.0.866)$$

Observe that the left side of (2.404) is the result of applying the tilde conjugacy operation to  $K(\hat{N}^j)$ , and the left side of (2.405) is the result of applying the tilde conjugacy operation to  $K(\hat{L}^j)$ . Recall (3.7.129). Consequently, the relations (2.404) and (2.405) show that the representation of  $sl(2, \mathbb{C})$  provided by the  $K(\hat{N}^j)$  and the  $K(\hat{L}^j)$  is *self conjugate* (for the tilde conjugacy relation) under the similarity transformation provided by  $\gamma_r^4$ .

Recall (7.3.377). Continue on to verify it follows from (2.402) and (2.403) that

$$[K(\lambda \mathbf{m} \cdot \hat{\mathbf{N}})]^T = -\gamma_r^4 K(\lambda \mathbf{m} \cdot \hat{\mathbf{N}}) (\gamma_r^4)^{-1} \quad (28.0.867)$$

and

$$[K(\theta \mathbf{n} \cdot \hat{\mathbf{L}})]^T = -\gamma_r^4 K(\theta \mathbf{n} \cdot \hat{\mathbf{L}}) (\gamma_r^4)^{-1}. \quad (28.0.868)$$

Using (2.387), (2.406), and (2.407) verify that

$$\begin{aligned}
K^T(k) &= \{\exp[K(\lambda \mathbf{m} \cdot \hat{\mathbf{N}})] \exp[K(\theta \mathbf{n} \cdot \hat{\mathbf{L}})]\}^T = \\
&\{\exp[K(\theta \mathbf{n} \cdot \hat{\mathbf{L}})]\}^T \{\exp[K(\lambda \mathbf{m} \cdot \hat{\mathbf{N}})]\}^T = \\
&\exp\{[K(\theta \mathbf{n} \cdot \hat{\mathbf{L}})]^T\} \exp\{[K(\lambda \mathbf{m} \cdot \hat{\mathbf{N}})]^T\} = \\
&\exp[-\gamma_r^4 K(\theta \mathbf{n} \cdot \hat{\mathbf{L}})(\gamma_r^4)^{-1}] \exp[-\gamma_r^4 K(\lambda \mathbf{m} \cdot \hat{\mathbf{N}})(\gamma_r^4)^{-1}] = \\
&\gamma_r^4 \exp[-K(\theta \mathbf{n} \cdot \hat{\mathbf{L}})](\gamma_r^4)^{-1} \gamma_r^4 \exp[-K(\lambda \mathbf{m} \cdot \hat{\mathbf{N}})](\gamma_r^4)^{-1} = \\
&\gamma_r^4 \{\exp[K(\theta \mathbf{n} \cdot \hat{\mathbf{L}})]\}^{-1} \{\exp[K(\lambda \mathbf{m} \cdot \hat{\mathbf{N}})]\}^{-1} (\gamma_r^4)^{-1} = \\
&\gamma_r^4 \{\exp[K(\lambda \mathbf{m} \cdot \hat{\mathbf{N}})] \exp[K(\theta \mathbf{n} \cdot \hat{\mathbf{L}})]\}^{-1} (\gamma_r^4)^{-1} = \\
&\gamma_r^4 [K(k)]^{-1} (\gamma_r^4)^{-1}, \tag{28.0.869}
\end{aligned}$$

from which it follows that

$$K^T(k) \gamma_r^4 = \gamma_r^4 K^{-1}(k) \tag{28.0.870}$$

and therefore

$$K^T(k) \gamma_r^4 K(k) = \gamma_r^4. \tag{28.0.871}$$

Show, using (2.384), that there is also the relation

$$K^T(k) (\gamma_r^4)^T K(k) = (\gamma_r^4)^T. \tag{28.0.872}$$

Note that (2.408), upon comparing its beginning and end, can be written as

$$K^T(k) = \gamma_r^4 [K(k)]^{-1} (\gamma_r^4)^{-1}. \tag{28.0.873}$$

This relation shows that, when the Majorana matrices (2.350) through (2.353) are employed, the matrices  $K^T(k)$  and  $K^{-1}(k)$  are related by the similarity transformation provided by  $\gamma_r^4$ . Persuade yourself, upon reflection, that this group element relation is a consequence of the Lie algebraic self conjugacy relations (2.404) and (2.405).

The relation (2.411) has an important consequence. Let  $M$  be *any*  $4 \times 4$  matrix. Verify that

$$[\gamma_r^4 K]^T M K = K^T (\gamma_r^4)^T M K = K^T (\gamma_r^4)^T K K^{-1} M K = (\gamma_r^4)^T K^{-1} M K. \tag{28.0.874}$$

This will prove to be a key result.

### Conjugate 4-Spinors and Bilinear Forms

With  $\gamma_r^4$  in mind, we are ready for further definitions/constructions. Suppose  $u$  is some real 4-spinor. Define a related *conjugate* 4-spinor  $\bar{u}$  by the rule

$$\bar{u} = \gamma_r^4 u. \tag{28.0.875}$$

Note, in view of (2.394), that this barring operation is what we might call an *anti-involution*,

$$\bar{\bar{u}} = (\gamma_r^4)^2 u = -I^{[4]} u = -u. \tag{28.0.876}$$

We are going to be working with quantities of the form  $(\bar{u}, v)$  where  $v$  is any other real 4-spinor and the usual real (no complex conjugate) scalar product is employed. Upon employing (2.414) and (2.384), verify that we may write

$$(\bar{u}, v) = (\gamma_r^4 u, v) = (u, [\gamma_r^4]^T v) = -(u, \gamma_r^4 v) = -(u, v)_{\gamma_r^4}. \quad (28.0.877)$$

Here we have introduced, for any matrix  $G$ , the definition

$$(u, v)_G = (u, Gv). \quad (28.0.878)$$

Quantities of the form  $(u, Gv)$  are called *bilinear forms*. Evidently, according to (2.416), the introduction of conjugate 4-spinors is equivalent (apart from a minus sign) to the use of the bilinear form associated with  $\gamma_r^4$ . Finally, suppose  $G$  is a *symmetric* matrix  $S$  or an *antisymmetric* matrix  $A$ . Verify that in these cases there are the relations

$$(u, v)_S = (v, u)_S \text{ and } (u, v)_A = -(v, u)_A. \quad (28.0.879)$$

Because of these symmetries under the interchange of the arguments  $u$  and  $v$ , the bilinear forms  $(u, v)_S$  and  $(u, v)_A$  are said to be *symmetric* and *antisymmetric*, respectively. Note that, by (2.384) and this definition, that the bilinear form  $(u, v)_{\gamma_r^4}$  is antisymmetric.

### Concept of Transformation Properties

With these definitions behind us, suppose  $K(k)$  with  $k \in SL(2, \mathbb{C})$  acts on  $u$  to produce a transformed spinor that we denote as  $\check{u}$ ,

$$\check{u} = K(k)u. \quad (28.0.880)$$

We will also use the notation  $\bar{\check{u}}$  to denote the conjugate of  $\check{u}$ . Therefore we may write

$$\bar{\check{u}} = \gamma_r^4 K(k)u. \quad (28.0.881)$$

Finally, let  $v$  be any other real 4-spinor, and act on it to obtain the *transformed* 4-spinor

$$\check{v} = K(k)v. \quad (28.0.882)$$

Here we again beg the reader's forgiveness for awkward notation. In the present context a bar  $\bar{\phantom{x}}$  simply denotes multiplication by  $\gamma_r^4$  as in (2.414); and a breve  $\check{\phantom{x}}$  is simply a distinguishing mark as in (2.419). Also, we observe that the term *conjugate* has many meanings/applications. Here we speak of conjugate spinors. In Exercise 3.7.36 we spoke of conjugate matrices and representations.

What we/you will soon explore are the *transformation properties* of  $(\bar{\check{u}}, M\check{v})$  under the action of  $K(k)$  where, as in (2.413),  $M$  is any  $4 \times 4$  matrix. Verify, using the notation just introduced, the properties of  $K$  and  $\gamma_r^4$ , and (2.413) that

$$\begin{aligned} (\bar{\check{u}}, M\check{v}) &= (\gamma_r^4 K u, MKv) = (u, [\gamma_r^4 K]^T MKv) = (u, (\gamma_r^4)^T K^{-1} MKv) = \\ &= (\gamma_r^4 u, K^{-1} MKv) = (\bar{u}, K^{-1} MKv). \end{aligned} \quad (28.0.883)$$

By looking at the beginning and end of (2.419) we see that it can be rewritten in the form

$$(\bar{\check{u}}, M\check{v}) = (\bar{u}, K^{-1} MKv). \quad (28.0.884)$$

This, like (2.413), is also a key result.

### Transformation Properties of $I^{[4]}$ and $\gamma^5$

Let us consider various possibilities for the matrix  $M$ . The simplest possibility is  $M = I^{[4]}$ . In this case (2.423) becomes

$$(\bar{u}, I^{[4]}\check{v}) = (\bar{u}, K^{-1}I^{[4]}Kv) = (\bar{u}, I^{[4]}v) \Leftrightarrow (\bar{u}, I^{[4]}\check{v}) = (\bar{u}, I^{[4]}v) \text{ or } (\bar{u}, \check{v}) = (\bar{u}, v). \quad (28.0.885)$$

We may say that  $I^{[4]}$  is *invariant* (behaves like a *scalar*) under the action of  $K$ . Suppose we also describe the result (2.424) in terms of the associated bilinear form  $(u, v)_{\gamma_r^4}$ . Verify that there is the relation

$$\begin{aligned} (\check{u}, I^{[4]}\check{v})_{\gamma_r^4} &= (Ku, Kv)_{\gamma_r^4} = (Ku, \gamma_r^4 Kv) = (u, K^T \gamma_r^4 Kv) = (u, \gamma_r^4 v) = (u, v)_{\gamma_r^4} \\ \Leftrightarrow (\bar{u}, I^{[4]}\check{v}) &= (\bar{u}, I^{[4]}v) \text{ or } (\bar{u}, \check{v}) = (\bar{u}, v) \text{ or } (Ku, Kv)_{\gamma_r^4} = (u, v)_{\gamma_r^4}. \end{aligned} \quad (28.0.886)$$

Verify that

$$(\bar{u}, v) = (\gamma_r^4 u, v) = (u, [\gamma_r^4]^\dagger, v) = -(u, \gamma_r^4 v) = -(u, v)_{\gamma_r^4}. \quad (28.0.887)$$

Out of curiosity, verify that

$$(u, v)_{\gamma_r^4} = -u_4 v_1 + u_3 v_2 - u_2 v_3 + u_1 v_4. \quad (28.0.888)$$

Note that the associated bilinear form is manifestly antisymmetric, as expected from (2.384).

The next more complicated possibility is  $M = \gamma_r^5$ . In this case use of (2.423) and (2.385) gives the result

$$(\bar{u}, \gamma_r^5 \check{v}) = (\bar{u}, K^{-1} \gamma_r^5 Kv) = (\bar{u}, \gamma_r^5 v) \quad (28.0.889)$$

so we may say that  $\gamma_r^5$  is also a scalar/invariant under the action of  $K$ . Verify that, in terms of bilinear forms, there is the associated result

$$\begin{aligned} (\check{u}, \check{v})_{\gamma_r^4 \gamma_r^5} &= (\check{u}, \gamma_r^4 \gamma_r^5 \check{v}) = (Ku, \gamma_r^4 \gamma_r^5 Kv) = (Ku, \gamma_r^4 K \gamma_r^5 v) = \\ (u, K^T \gamma_r^4 K \gamma_r^5 v) &= (u, \gamma_r^4 \gamma_r^5 v) = (u, v)_{\gamma_r^4 \gamma_r^5}. \end{aligned} \quad (28.0.890)$$

Verify that

$$(\bar{u}, \gamma_r^5 v) = (\gamma_r^4 u, \gamma_r^5 v) = -(u, \gamma_r^4 \gamma_r^5 v) = -(u, v)_{\gamma_r^4 \gamma_r^5}. \quad (28.0.891)$$

Verify using (2.362) and (2.384) that the associated bilinear form is antisymmetric. Out of curiosity verify that

$$(u, v)_{\gamma_r^4 \gamma_r^5} = +u_1 v_2 - u_2 v_1 - u_3 v_4 + u_4 v_3, \quad (28.0.892)$$

which is manifestly antisymmetric. We have learned that there are two invariant bilinear forms, namely (2.427) and (2.431), and both are antisymmetric.

### Transformation Properties of the $\gamma^\mu$

With these instructive but relatively simple observations behind us, let us consider other possibilities for the matrix  $M$ . Suppose we select the possibilities  $M = \gamma^\mu$  for  $\mu = 1 \dots 4$ . Your task in this case is to show that

$$(\bar{u}, \gamma^\mu \check{v}) = \sum_\nu \Lambda^{\mu\nu}(k)(\bar{u}, \gamma^\nu v). \quad (28.0.893)$$

Note that, because of (2.381), we may also write the real matrix relations

$$(\bar{u}, \gamma_r^\mu \check{v}) = \sum_\nu \Lambda^{\mu\nu}(k) (\bar{u}, \gamma_r^\nu v). \quad (28.0.894)$$

In view of (2.432), it is commonly (although somewhat imprecisely) said that the  $\gamma^\mu$  (equivalently, the  $\gamma_r^\mu$ ) behave like a four-vector under the action of  $K(k)$  for  $k \in SL(2, \mathbb{C})$  in the same spirit that  $I^{[4]}$  and  $\gamma_r^5$  are said to behave like scalars.<sup>22</sup> According to (2.423) there is the relation

$$(\bar{u}, \gamma^\mu \check{v}) = (\bar{u}, K^{-1} \gamma^\mu K v). \quad (28.0.895)$$

Show that (2.432) is established if there is the relation

$$K^{-1}(k) \gamma^\mu K(k) = \sum_\nu \Lambda^{\mu\nu}(k) \gamma^\nu. \quad (28.0.896)$$

Let us work on verifying (2.435). It helps to separate the problem into two parts. Make the decomposition

$$k = k_b k_r \quad (28.0.897)$$

where

$$k_b = \exp(\lambda \mathbf{m} \cdot \hat{\mathbf{N}}) \quad (28.0.898)$$

is the  $SL(2, \mathbb{C})$  element for a *boost* and

$$k_r = \exp(\theta \mathbf{n} \cdot \hat{\mathbf{L}}) \quad (28.0.899)$$

is the  $SL(2, \mathbb{C})$  element for a *rotation*. Then, by (7.3.378), verify that

$$K(k) = K(k_b k_r) = K(k_b) K(k_r). \quad (28.0.900)$$

Correspondingly, verify that

$$K^{-1}(k) \gamma^\alpha K(k) = K^{-1}(k_r) K^{-1}(k_b) \gamma^\alpha K(k_b) K(k_r). \quad (28.0.901)$$

Conjecture that (2.435) holds for *pure* boosts and for *pure* rotations,

$$K^{-1}(k_b) \gamma^\alpha K(k_b) \stackrel{?}{=} \sum_\beta \Lambda^{\alpha\beta}(k_b) \gamma^\beta \quad \text{and} \quad K^{-1}(k_r) \gamma^\beta K(k_r) \stackrel{?}{=} \sum_\delta \Lambda^{\beta\delta}(k_r) \gamma^\delta. \quad (28.0.902)$$

Verify the desired result (2.435) then follows from (2.440), (2.441), and (2.219):

$$\begin{aligned} K^{-1}(k) \gamma^\alpha K(k) &= K^{-1}(k_r) K^{-1}(k_b) \gamma^\alpha K(k_b) K(k_r) = \\ K^{-1}(k_r) \sum_\beta \Lambda^{\alpha\beta}(k_b) \gamma^\beta K(k_r) &= \sum_\beta \Lambda^{\alpha\beta}(k_b) K^{-1}(k_r) \gamma^\beta K(k_r) = \\ \sum_{\beta\delta} \Lambda^{\alpha\beta}(k_b) \Lambda^{\beta\delta}(k_r) \gamma^\delta &= \sum_\delta \Lambda^{\alpha\delta}(k_b k_r) \gamma^\delta = \sum_\delta \Lambda^{\alpha\delta}(k) \gamma^\delta. \end{aligned} \quad (28.0.903)$$

---

<sup>22</sup>Since the  $\gamma^\mu$  for  $\mu = 1 \dots 4$  behave like four-vectors, it makes sense to raise and lower their indices (if desired) using the metric tensor  $g$ . See, for example, (2.478).

What remains is to verify the conjectures (2.441).

For simplicity, first work on verifying the second conjecture in (2.441). Verify that

$$K(k_r) = K[\exp(\theta \mathbf{n} \cdot \hat{\mathbf{L}})] = \exp[\theta K(\mathbf{n} \cdot \hat{\mathbf{L}})]. \quad (28.0.904)$$

See (7.3.386). Next verify that combining (2.443) and the second conjecture in (2.441) yields the conjecture

$$\exp[-\theta K(\mathbf{n} \cdot \hat{\mathbf{L}})]\gamma^\beta \exp[\theta K(\mathbf{n} \cdot \hat{\mathbf{L}})] \stackrel{?}{=} \sum_{\delta} \Lambda^{\beta\delta}(k_r)\gamma^\delta. \quad (28.0.905)$$

At this point review Sections 8.1 and 8.2 that described the concept of an adjoint Lie operator, using the symbol  $\#$ , in the context of Lie operators :  $f$  :. The same concept can be applied to matrices in terms of commutators. Let  $A$  and  $B$  be any two  $n \times n$  matrices. In this context, define a matrix adjoint operator  $\#A\#$  that acts on matrices  $B$  by the rule

$$\#A\#B = \{A, B\}. \quad (28.0.906)$$

Then, in complete analogy to the discussion in Sections 8.1 and 8.2, there is the relation

$$\exp(-A)B \exp(A) = \exp(-\#A\#)B, \quad (28.0.907)$$

which may be thought of as the matrix version of Hadamard's lemma. See also (27.11.20) through (27.11.24) in Section 27.11 where this result is again used.

Verify that applying the matrix adjoint operator machinery just described produces, in the present context, the relation

$$\exp[-\theta K(\mathbf{n} \cdot \hat{\mathbf{L}})]\gamma^\beta \exp[\theta K(\mathbf{n} \cdot \hat{\mathbf{L}})] = \exp[-\theta \#K(\mathbf{n} \cdot \hat{\mathbf{L}})\#]\gamma^\beta. \quad (28.0.908)$$

Consequently, employing (2.447) in (2.444) produces the conjecture

$$\exp[-\theta \#K(\mathbf{n} \cdot \hat{\mathbf{L}})\#]\gamma^\beta \stackrel{?}{=} \sum_{\delta} \Lambda^{\beta\delta}[\exp(\theta \mathbf{n} \cdot \hat{\mathbf{L}})]\gamma^\delta. \quad (28.0.909)$$

Verify that employing (2.269) in (2.448) produces the logically equivalent conjecture

$$\exp[-\theta \#K(\mathbf{n} \cdot \hat{\mathbf{L}})\#]\gamma^\beta \stackrel{?}{=} \sum_{\delta} [\exp(\theta \mathbf{n} \cdot \hat{\mathbf{L}})]^{\beta\delta}\gamma^\delta, \quad (28.0.910)$$

which, in turn, produces the logically equivalent conjectures

$$-\theta\{K(\mathbf{n} \cdot \hat{\mathbf{L}}), \gamma^\beta\} \stackrel{?}{=} \theta \sum_{\delta} [\mathbf{n} \cdot \hat{\mathbf{L}}]^{\beta\delta}\gamma^\delta, \quad (28.0.911)$$

$$-\{K(\mathbf{n} \cdot \hat{\mathbf{L}}), \gamma^\beta\} \stackrel{?}{=} \sum_{\delta} [\mathbf{n} \cdot \hat{\mathbf{L}}]^{\beta\delta}\gamma^\delta. \quad (28.0.912)$$

Let us seek to verify (2.451) for the three specific cases  $\mathbf{n} = \mathbf{e}_1$ ,  $\mathbf{n} = \mathbf{e}_2$ , and  $\mathbf{n} = \mathbf{e}_3$ . Verify that if we can do so, then (2.451) follows by linearity. Consider, for example, the case  $\mathbf{n} = \mathbf{e}_3$  so that the conjecture (2.451) becomes the conjecture

$$-\{K(\hat{L}^3), \gamma^\beta\} \stackrel{?}{=} \sum_{\delta} [L^3]^{\beta\delta}\gamma^\delta. \quad (28.0.913)$$

Verify that using (2.373) and (2.366) in (2.452) produces the logically equivalent conjectures

$$-(1/2)\{\hat{\sigma}^{12}\}, \gamma^\beta\} \stackrel{?}{=} \sum_\delta [L^3]^{\beta\delta} \gamma^\delta, \quad (28.0.914)$$

$$-(1/2)\{\gamma^1 \gamma^2, \gamma^\beta\} \stackrel{?}{=} \sum_\delta [L^3]^{\beta\delta} \gamma^\delta. \quad (28.0.915)$$

Using the Dirac algebra, verify the following commutation relations:

$$-(1/2)\{\gamma^1 \gamma^2, \gamma^1\} = -\gamma^2, \quad (28.0.916)$$

$$-(1/2)\{\gamma^1 \gamma^2, \gamma^2\} = \gamma^1, \quad (28.0.917)$$

$$-(1/2)\{\gamma^1 \gamma^2, \gamma^3\} = 0, \quad (28.0.918)$$

$$-(1/2)\{\gamma^1 \gamma^2, \gamma^4\} = 0. \quad (28.0.919)$$

Recall from (7.3.182) that there is the result

$$L^3 = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (28.0.920)$$

Verify, upon comparing (2.455) through (2.459), that the conjecture (2.454) is correct, and therefore the conjecture (2.452) is correct. Similarly, verify that (2.451) also holds for  $\mathbf{n} = \mathbf{e}_1$  and  $\mathbf{n} = \mathbf{e}_2$ , and therefore holds for all  $\mathbf{n}$ . Finally, proceed backwards through the chain of logically equivalent conjectures to verify that the initial conjecture, the second conjecture in (2.441), is indeed correct.

Now work on verifying the first conjecture in (2.441). The procedure for doing so is analogous to that just used to verify the second conjecture. Carry out the verification using the relations and conjectures below:

$$K(k_b) = K[\exp(\lambda \mathbf{m} \cdot \hat{\mathbf{N}})] = \exp[\lambda K(\mathbf{m} \cdot \hat{\mathbf{N}})], \quad (28.0.921)$$

$$\exp[-\lambda K(\mathbf{m} \cdot \hat{\mathbf{N}})] \gamma^\beta \exp[\lambda K(\mathbf{m} \cdot \hat{\mathbf{N}})] \stackrel{?}{=} \sum_\delta \Lambda^{\beta\delta}(k_b) \gamma^\delta, \quad (28.0.922)$$

$$\exp[-\lambda K(\mathbf{m} \cdot \hat{\mathbf{N}})] \gamma^\beta \exp[\lambda K(\mathbf{m} \cdot \hat{\mathbf{N}})] = \exp[-\lambda \# K(\mathbf{m} \cdot \hat{\mathbf{N}}) \#] \gamma^\beta, \quad (28.0.923)$$

$$\exp[-\lambda \# K(\mathbf{m} \cdot \hat{\mathbf{N}}) \#] \gamma^\beta \stackrel{?}{=} \sum_\delta \Lambda^{\beta\delta} [\exp(\lambda \mathbf{m} \cdot \hat{\mathbf{N}})] \gamma^\delta, \quad (28.0.924)$$

$$\exp[-\lambda \# K(\mathbf{m} \cdot \hat{\mathbf{N}}) \#] \gamma^\beta \stackrel{?}{=} \sum_\delta [\exp(\lambda \mathbf{m} \cdot \mathbf{N})]^{\beta\delta} \gamma^\delta, \quad (28.0.925)$$

$$-\lambda \{K(\mathbf{m} \cdot \hat{\mathbf{N}}), \gamma^\beta\} \stackrel{?}{=} \lambda \sum_\delta [\mathbf{m} \cdot \mathbf{N}]^{\beta\delta} \gamma^\delta, \quad (28.0.926)$$

$$-\{K(\mathbf{m} \cdot \hat{\mathbf{N}}), \gamma^\beta\} \stackrel{?}{=} \sum_\delta [\mathbf{m} \cdot \mathbf{N}]^{\beta\delta} \gamma^\delta, \quad (28.0.927)$$

$$-\{K(\hat{N}^3), \gamma^\beta\} \stackrel{?}{=} \sum_\delta [N^3]^{\beta\delta} \gamma^\delta, \quad (28.0.928)$$

$$-(1/2)\{\hat{\sigma}^{43}\}, \gamma^\beta\} \stackrel{?}{=} \sum_\delta [N^3]^{\beta\delta} \gamma^\delta, \quad (28.0.929)$$

$$-(1/2)\{\gamma^4\gamma^3, \gamma^\beta\} \stackrel{?}{=} \sum_\delta [N^3]^{\beta\delta} \gamma^\delta, \quad (28.0.930)$$

$$-(1/2)\{\gamma^4\gamma^3, \gamma^1\} = 0, \quad (28.0.931)$$

$$-(1/2)\{\gamma^4\gamma^3, \gamma^2\} = 0, \quad (28.0.932)$$

$$-(1/2)\{\gamma^4\gamma^3, \gamma^3\} = \gamma^4, \quad (28.0.933)$$

$$-(1/2)\{\gamma^4\gamma^3, \gamma^4\} = \gamma^3, \quad (28.0.934)$$

$$N^3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (28.0.935)$$

Let us pause for a moment to reflect on what has been discovered/accomplished here. We were able to contemplate/determine the action of the  $K(k)$  on the  $\gamma^\mu$  by the fact that both  $K(k)$  and the  $\gamma^\mu$  were known  $4 \times 4$  matrices so that the quantities  $K^{-1}\gamma^\mu K$  could be evaluated. Equivalently, because the  $\gamma^\mu$  obeyed the Dirac algebra, we were able to use this algebra to determine the action of the  $K(\hat{L}^j)$  and the  $K(\hat{N}^j)$  on the  $\gamma^\beta$  as in (2.452) through (2.458) and (2.467) through (2.473).

### Transformation Properties of Products of gamma matrices

Consider the case of a product of two *distinct* gamma matrices,  $M = \gamma^\mu\gamma^\nu$  with  $\mu \neq \nu$ . Verify that

$$K^{-1}\gamma^\mu\gamma^\nu K = K^{-1}\gamma^\mu K K^{-1}\gamma^\nu K. \quad (28.0.936)$$

Now use (2.435) in (2.475) to obtain the result

$$K^{-1}\gamma^\mu\gamma^\nu K = \sum_{\alpha\beta} \Lambda^{\mu\alpha}(k) \Lambda^{\nu\beta}(k) \gamma^\alpha \gamma^\beta. \quad (28.0.937)$$

Evidently  $\gamma^\mu\gamma^\nu$  transforms like a second rank tensor. Note, according to (2.365), that  $\hat{\sigma}^{\mu\nu}$  involves the commutator  $\{\gamma^\mu, \gamma^\nu\}$ . Show it follows that

$$K^{-1}\hat{\sigma}^{\mu\nu} K = \sum_{\alpha\beta} \Lambda^{\mu\alpha}(k) \Lambda^{\nu\beta}(k) \hat{\sigma}^{\alpha\beta}, \quad (28.0.938)$$

the  $\hat{\sigma}^{\mu\nu}$  transform like a second rank antisymmetric tensor.

You have determined the transformation properties of the commutator  $\{\gamma^\mu, \gamma^\nu\}$ . What can be said about the transformation properties of the anticommutator  $\{\gamma^\mu, \gamma^\nu\}_+$ ? Evaluate  $K^{-1}\{\gamma^\mu, \gamma^\nu\}_+ K$  using (2.349). Also evaluate it using (2.476), and then simplify your result using (2.349). Verify, with the aid of (6.2.51), that (2.349) and (2.476) are compatible.

Using the Dirac algebra, verify that

$$\sum_{\nu} \gamma_{\nu} \gamma^{\nu} = \sum_{\mu\nu} g_{\mu\nu} \gamma^{\mu} \gamma^{\nu} = 4I^{[4]}, \quad (28.0.939)$$

and observe from the far right side of (2.478) that  $\sum_{\nu} \gamma_{\nu} \gamma^{\nu}$  is invariant, as the notation suggests.

Consider the case of three gamma matrices. Verify, using the Dirac algebra, that the product of any three distinct gamma matrices can be written in the form  $\gamma^5 \gamma^{\mu}$ . Verify that

$$K^{-1} \gamma^5 \gamma^{\mu} K = \gamma^5 K^{-1} \gamma^{\mu} K = \sum_{\nu} \Lambda^{\mu\nu}(k) \gamma^5 \gamma^{\nu}. \quad (28.0.940)$$

Evidently the quantities  $\gamma^5 \gamma^{\mu}$ , like the quantities  $\gamma^{\mu}$ , also behave as four-vectors.

Consider the case of four gamma matrices. Verify that the product of any four distinct gamma matrices must be proportional to  $\gamma^5$ , whose transformation properties are given by (2.385). Even a bit more can be said. From (2.435), and an obvious extension of (2.476) to the case of four gamma matrices, verify that the gamma matrix products  $\gamma^{\mu} \gamma^{\nu} \gamma^{\sigma} \gamma^{\tau}$  transform like a rank four tensor. Verify, moreover, that there is the relation

$$\gamma^5 = (i/4!) \sum_{\mu\nu\sigma\tau} \epsilon_{\mu\nu\sigma\tau} \gamma^{\mu} \gamma^{\nu} \gamma^{\sigma} \gamma^{\tau}. \quad (28.0.941)$$

Observe that the invariance of  $\gamma^5$ , already established, is consistent with the invariant appearance of this relation.

Consider the case of five or more gamma matrices. Verify that any product of five or more gamma matrices may be reduced to the product of four or less gamma matrices using the Dirac algebra, and the transformation properties of these products have already been determined.

### Summary of Results and Relation to the Clebsch-Gordan Series

In summary, we have found the following: two “scalars”,  $I^{[4]}$  and  $\gamma^5$ ; two “four-vectors”,  $\gamma^{\mu}$  and  $\gamma^5 \gamma^{\mu}$ ; and one “antisymmetric tensor”,  $\hat{\sigma}^{\mu\nu}$ . Why should this be? Look again at the bilinear form  $(\bar{u}, M\bar{v})$  that appears on the left side of (2.423). It contains the two 4-spinors  $\bar{u}$  and  $\bar{v}$ , each of which carries the representation  $\Gamma(0, 1/2) \oplus \Gamma(1/2, 0)$ , and a fixed matrix  $M$ . Evidently the quantities  $(\bar{u}, M\bar{v})$  are essentially tensor products and may be expected to contain whatever representations of  $sl(2, \mathbb{C})$  occur in the tensor product of  $\Gamma(0, 1/2) \oplus \Gamma(1/2, 0)$  with itself. Let us see what representations can be found in this tensor product. Verify the tensor product result

$$\begin{aligned} & [\Gamma(0, 1/2) \oplus \Gamma(1/2, 0)] \times [\Gamma(0, 1/2) \oplus \Gamma(1/2, 0)] = \\ & [\Gamma(0, 1/2) \times \Gamma(0, 1/2)] \oplus [\Gamma(0, 1/2) \times \Gamma(1/2, 0)] \oplus \\ & [\Gamma(1/2, 0) \times \Gamma(0, 1/2)] \oplus [\Gamma(1/2, 0) \times \Gamma(1/2, 0)]. \end{aligned} \quad (28.0.942)$$

Next verify the Clebsch-Gordan results

$$\Gamma(0, 1/2) \times \Gamma(0, 1/2) = \Gamma(0, 1) \oplus \Gamma(0, 0), \quad (28.0.943)$$

$$\Gamma(0, 1/2) \times \Gamma(1/2, 0) = \Gamma(1/2, 1/2), \quad (28.0.944)$$

$$\Gamma(1/2, 0) \times \Gamma(0, 1/2) = \Gamma(1/2, 1/2), \quad (28.0.945)$$

$$\Gamma(1/2, 0) \times \Gamma(1/2, 0) = \Gamma(1, 0) \oplus \Gamma(0, 0). \quad (28.0.946)$$

Show it follows that there is the grand Clebsch-Gordan result

$$\begin{aligned} & [\Gamma(0, 1/2) \oplus \Gamma(1/2, 0)] \times [\Gamma(0, 1/2) \oplus \Gamma(1/2, 0)] = \\ & 2\Gamma(0, 0) \oplus 2\Gamma(1/2, 1/2) \oplus [\Gamma(0, 1) \oplus \Gamma(1, 0)], \end{aligned} \quad (28.0.947)$$

which corresponds to the two scalars, two four-vectors, and one antisymmetric tensor, as described above.

### Group-Theoretical Nature of Dirac Gamma Matrices

You have verified that the four matrices  $\gamma^\mu$  given by (2.350) through (2.353) satisfy the Dirac algebra (2.349). Moreover, in view of (2.365) and (2.373) through (2.378), the  $\gamma^\mu$  are related to  $sl(2, \mathbb{C})$  in that they have the further remarkable property that they *factorize* the  $K(\hat{L}^j)$  and  $K(\hat{N}^j)$ . That is, each  $K(\hat{L}^j)$  and  $K(\hat{N}^j)$  can be written as the product of two gamma matrices.

But do the gamma matrices have additional group-theoretical significance? Let  $p$  be a four-vector with covariant components  $p_\beta$  and suppose a Lorentz group element associated with the  $SL(2, \mathbb{C})$  element  $k$  acts on  $p$  to produce the four-vector  $\check{p}$ . Verify that, according to (1.6.286), (1.6.287), and (1.6.289), the four-vector  $\check{p}$  will have covariant components  $\check{p}_\alpha$  given by the relation

$$\check{p}_\alpha = \sum_\beta \{[\Lambda^T(k)]^{-1}\}^{\alpha\beta} p_\beta = \sum_\beta \{[\Lambda^{-1}(k)]^T\}^{\alpha\beta} p_\beta = \sum_\beta [\Lambda^{-1}(k)]^{\beta\alpha} p_\beta. \quad (28.0.948)$$

[Here we must again apologize for confusing notation: The matrix  $K$  defined by (1.6.287) is *not* the matrix  $K(k)$  that appears in (7.3.375).] Now consider the matrix, call it  $C(k)$ , given by the rule

$$C(k) = \sum_\alpha \check{p}_\alpha K^{-1}(k) \gamma^\alpha K(k) = K^{-1}(k) \left[ \sum_\alpha \check{p}_\alpha \gamma^\alpha \right] K(k). \quad (28.0.949)$$

According to (2.435) there is the relation

$$K^{-1}(k) \gamma^\alpha K(k) = \sum_\delta \Lambda^{\alpha\delta}(k) \gamma^\delta. \quad (28.0.950)$$

Now employ (2.487) and then (2.489) in (2.488) to show that

$$\begin{aligned} C(k) &= K^{-1}(k) \left[ \sum_\alpha \check{p}_\alpha \gamma^\alpha \right] K(k) = K^{-1} \left\{ \sum_{\alpha\beta} p_\beta \gamma^\alpha [\Lambda^{-1}(k)]^{\beta\alpha} \right\} K(k) \\ &= \sum_{\alpha\beta\delta} p_\beta \gamma^\delta [\Lambda^{-1}(k)]^{\beta\alpha} \Lambda(k)^{\alpha\delta} = \sum_{\beta\delta} p_\beta \gamma^\delta [\Lambda^{-1}(k) \Lambda(k)]^{\beta\delta} \\ &= \sum_{\beta\delta} p_\beta \gamma^\delta \{I^{[4]}\}^{\beta\delta} = \sum_\delta p_\delta \gamma^\delta. \end{aligned} \quad (28.0.951)$$

You have shown that the matrix  $C(k)$  is, in fact, *independent* of  $k$ .

Now suppose that  $p$  is the momentum four-vector for a particle that has mass  $m$  and is at rest,

$$p_\beta = (0, 0, 0, mc). \quad (28.0.952)$$

In this case there is the relation

$$\sum_\delta p_\delta \gamma^\delta = mc\gamma^4 \quad (28.0.953)$$

so that

$$C(k) = mc\gamma^4. \quad (28.0.954)$$

Next verify from (2.355) that  $\gamma^4$  must be diagonalizable and have real eigenvalues. And verify from the Dirac algebra condition  $(\gamma^4)^2 = I^{[4]}$  that its eigenvalues, call them  $\tau$ , must satisfy  $\tau^2 = 1$ . Let  $v$  be a 4-spinor which is an eigenvector of  $\gamma^4$  with eigenvalue  $\tau$ ,

$$\gamma^4 v = \tau v. \quad (28.0.955)$$

We will determine such eigenvectors shortly. They will turn out to be complex, but that need not concern us. Meanwhile define, as before,  $\check{v}$  by the rule

$$\check{v} = K(k)v. \quad (28.0.956)$$

With these definitions in mind, verify the relations

$$mc\gamma^4 v = mc\tau v \quad (28.0.957)$$

and

$$C(k)v = K^{-1}(k)[\sum_\alpha \check{p}_\alpha \gamma^\alpha] \check{v} = mc\gamma^4 v = \tau mcv. \quad (28.0.958)$$

Show it follows from (2.495) and (2.497) that

$$[\sum_\alpha \check{p}_\alpha \gamma^\alpha] \check{v} = \tau mK(k)v = \tau mcv. \quad (28.0.959)$$

Let us rewrite (2.498) in the form

$$[\sum_\alpha \gamma^\alpha \check{p}_\alpha] \check{v} = \tau mcv. \quad (28.0.960)$$

Observe that on the left side of (2.499) there are the  $k$  dependent quantities  $\check{p}_\alpha$  and  $\check{v}$  that carry the  $\Gamma(1/2, 1/2)$  and  $[\Gamma(0, 1/2) \oplus \Gamma(1/2, 0)]$  representations of  $sl(2, \mathbb{C})$ , respectively. And on the right side of (2.499) we find just  $\check{v}$  which carries the  $[\Gamma(0, 1/2) \oplus \Gamma(1/2, 0)]$  representation of  $sl(2, \mathbb{C})$ . Also observe that matrices  $\gamma^\alpha$  are *three*-index quantities because there is the superscript  $\alpha$  and the two matrix indices for each gamma matrix. Thus, from a group-theoretical perspective, the three-index quantities  $\gamma^\alpha$  are *Clebsch-Gordan coefficients* that couple the representations  $\Gamma(1/2, 1/2)$  and  $[\Gamma(0, 1/2) \oplus \Gamma(1/2, 0)]$  to produce the representation  $[\Gamma(0, 1/2) \oplus \Gamma(1/2, 0)]$ .<sup>23</sup>

<sup>23</sup>Also, from the same group-theoretical perspective, the matrices  $M = I^{[4]}$ ,  $M = \gamma^5$ ,  $M = \gamma^\mu$ ,  $M = \gamma^5 \gamma^\mu$ , and  $M = \hat{\sigma}^{\mu\nu}$  are Clebsch-Gordan coefficients that project out from the tensor product appearing on the left side of (2.481) the representations  $\Gamma(0, 0)$ ,  $\Gamma(1/2, 1/2)$ , and  $[\Gamma(0, 1) \oplus \Gamma(1, 0)]$ .

Let us check that this possibility/conclusion makes group-theoretical sense. Verify the tensor product result

$$\begin{aligned}\Gamma(1/2, 1/2) \times [\Gamma(0, 1/2) \oplus \Gamma(1/2, 0)] &= \\ \Gamma(1/2, 1/2) \times \Gamma(0, 1/2) \oplus \Gamma(1/2, 1/2) \times \Gamma(1/2, 0). &\end{aligned}\quad (28.0.961)$$

Verify the Clebsch-Gordan results

$$\Gamma(1/2, 1/2) \times \Gamma(0, 1/2) = \Gamma(1/2, 0) \oplus \Gamma(1/2, 1), \quad (28.0.962)$$

$$\Gamma(1/2, 1/2) \times \Gamma(1/2, 0) = \Gamma(0, 1/2) \oplus \Gamma(1, 1/2). \quad (28.0.963)$$

Verify, therefore, that there is the grand Clebsch-Gordan result

$$\begin{aligned}\Gamma(1/2, 1/2) \times [\Gamma(0, 1/2) \oplus \Gamma(1/2, 0)] &= \\ \Gamma(1/2, 0) \oplus \Gamma(1/2, 1) \oplus \Gamma(0, 1/2) \oplus \Gamma(1, 1/2) &= \\ [\Gamma(0, 1/2) \oplus \Gamma(1/2, 0)] \oplus \Gamma(1/2, 1) \oplus \Gamma(1, 1/2). &\end{aligned}\quad (28.0.964)$$

Evidently the role of the  $\gamma^\alpha$  is to project out, from the tensor product with which (2.503) begins, the representation that appears in square brackets in the last line of (2.503).

### Application to the Dirac Equation

The Dirac equation for a *free* particle of mass  $m$  reads

$$i\hbar \sum_\mu \gamma^\mu \partial_\mu \psi(x) = mc\psi(x) \quad (28.0.965)$$

where  $\psi(x)$  is a 4-spinor *field* that depends on the space-time coordinates  $x$ . Dirac's equation is often written in natural units (for which  $\hbar = c = 1$ ) and the summation convention is employed so that it takes the elegant form

$$i\gamma^\mu \partial_\mu \psi = m\psi. \quad (28.0.966)$$

One may also employ the *Feynman* (1918-1988) *slash* notation

$$\not{\partial} = \sum_\mu \gamma^\mu \partial_\mu \quad (28.0.967)$$

to achieve the even more elegant form

$$i\not{\partial}\psi = m\psi. \quad (28.0.968)$$

Dirac's memorial stone, near Newton's monument in Westminster Abbey, displays his equation in the form

$$i\gamma \cdot \partial\psi = m\psi. \quad (28.0.969)$$

To see this stone, Google the two words Dirac Westminster.

For pedagogical reasons we will use Dirac's equation in the form (2.504) in order to keep track of dimensions. Note that here again it is evident that the gamma matrices play the role of Clebsch-Gordan coefficients. On the left side of (2.504) we have  $\partial_\mu$  and  $\psi$  which carry the representations  $\Gamma(1/2, 1/2)$  and  $\Gamma(0, 1/2) \oplus \Gamma(1/2, 0)$ , respectively. The  $\gamma^\mu$  couple them to produce the representation  $\Gamma(0, 1/2) \oplus \Gamma(1/2, 0)$ , which is also the representation carried by the  $\psi$  appearing on the right side of (2.504). Thus, Dirac's equation is group-theoretically consistent.

Let us seek a plane-wave solution to (2.504) of the form

$$\psi(x) = w \exp[-i(\sum_\mu p_\mu x^\mu)/\hbar]. \quad (28.0.970)$$

As done before in (2.491), we will take  $p_\mu$  to be specified by the relation

$$p_\mu = (0, 0, 0, mc), \quad (28.0.971)$$

and we will assume  $w$  is a 4-spinor that is independent of  $x$ . This  $\psi$  has no *spatial* dependence ( $\psi$  is translationally invariant) which implies that this proposed Ansatz is intended to describe a free-particle at rest.<sup>24</sup> With regard to temporal dependence, show that

$$\partial_4 \psi = (\partial/\partial x^4)\psi = -i(mc/\hbar)w \exp[-i(\sum_\mu p_\mu x^\mu)/\hbar]. \quad (28.0.972)$$

Therefore, in this case, verify that the left side of (2.504) becomes

$$i\hbar \sum_\mu \gamma^\mu \partial_\mu \psi(x) = mc\gamma^4 \psi(x) \quad (28.0.973)$$

so that (2.504) becomes

$$mc\gamma^4 \psi(x) = mc\psi(x). \quad (28.0.974)$$

Verify that canceling out common factors from both sides of (2.513) yields for  $w$  the relation

$$\gamma^4 w = w. \quad (28.0.975)$$

That is,  $w$  must be an eigenvector of  $\gamma^4$  with eigenvalue +1. We will soon see that  $\gamma^4$  has eigenvectors with eigenvalues  $\pm 1$  so that (2.494) has solutions for  $\tau = \pm 1$ . Note that (2.514) is consistent with (2.494).

Let us pause/digress briefly to discuss some commonly employed terminology for solutions of the Dirac equation. Verify, using (2.510) and (1.6.42), that the the Ansatz (2.509) can be rewritten in the form

$$\psi = w \exp[-i\mathcal{E}_0 t/\hbar] \quad (28.0.976)$$

where  $\mathcal{E}_0$  is the (manifestly positive) rest energy,

$$\mathcal{E}_0 = mc^2. \quad (28.0.977)$$

---

<sup>24</sup>We remark that once “at rest” solutions have been found, all other free-particle solutions can be found by acting on the at rest solutions with suitable Lorentz transformations.

Despite the appearance of a *minus* sign in the exponent appearing in (2.515), and also in (2.509), this proposed solution is called a *positive energy* solution. The reason for this terminology has to do with the analogous case of the nonrelativistic Schrödinger (1887-1961) equation. The Schrödinger equation for the Schrödinger wave function, call it  $\chi(\mathbf{r}, t)$ , reads

$$i\hbar(\partial/\partial t)\chi(\mathbf{r}, t) = H(q, p)\chi(\mathbf{r}, t) \quad (28.0.978)$$

where  $H$  is the (assumed time independent) Hamiltonian. If we make the separation of variables Ansatz

$$\chi(\mathbf{r}, t) = u(\mathbf{r})f(t) \quad (28.0.979)$$

and specify that  $u$  is an eigenfunction of  $H$  with eigenvalue  $E$  so that

$$Hu = Eu, \quad (28.0.980)$$

then (2.517) has the solution

$$\chi(\mathbf{r}, t) = u(\mathbf{r}) \exp[-iEt/\hbar]. \quad (28.0.981)$$

Evidently the argument of the exponential function appearing in (2.520) is *negative* imaginary as the time  $t$  becomes evermore positive providing the energy  $E$  is *positive*, and vice versa. Observe that the argument of the exponential function appearing in (2.509) eventually becomes negative imaginary as  $t$  becomes evermore positive. Correspondingly, (2.509) is called a positive energy solution of the Dirac equation.

Now return to the main discussion. As promised, let us find the eigenvectors of  $\gamma^4$ . Begin by writing  $w$  in the form

$$w = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \quad (28.0.982)$$

where the quantities  $a$  through  $d$  are to be determined from (2.514). Verify using (2.353) that

$$\gamma^4 w = \begin{pmatrix} id \\ -ic \\ ib \\ -ia \end{pmatrix}, \quad (28.0.983)$$

and consequently the condition (2.514) yields the relations

$$d = -ia \quad (28.0.984)$$

$$c = ib. \quad (28.0.985)$$

Therefore any eigenvector of  $\gamma^4$  having eigenvalue +1 must be of the form

$$w = \begin{pmatrix} a \\ b \\ ib \\ -ia \end{pmatrix}. \quad (28.0.986)$$

Let us pause at this point to seek eigenvectors of  $\gamma^4$  with eigenvalue  $-1$ . Take the complex conjugate of both sides of (2.514) to find the relation

$$(\gamma^4)^* w^* = (\gamma^4 w)^* = w^*. \quad (28.0.987)$$

Since  $\gamma^4$  is purely imaginary (in a Majorana representation), there is the relation

$$(\gamma^4)^* = -\gamma^4, \quad (28.0.988)$$

from which it follows that

$$\gamma^4 w^* = -w^*. \quad (28.0.989)$$

Therefore the eigenvectors of  $\gamma^4$  with eigenvalue  $-1$  are of the form

$$w^* = \begin{pmatrix} a^* \\ b^* \\ -ib^* \\ ia^* \end{pmatrix}. \quad (28.0.990)$$

Let us continue on. Evidently there is a two-fold degeneracy for both the eigenvalue  $+1$  and  $-1$  eigenvectors. To break this degeneracy it is convenient to employ the matrix  $M^3$  defined by the rule

$$M^3 = i\hat{\sigma}^{12}. \quad (28.0.991)$$

Verify, using the Dirac algebra and (2.349), that there are the relations

$$(M^3)^\dagger = M^3 \quad (28.0.992)$$

and

$$(M^3)^2 = I^{[4]}. \quad (28.0.993)$$

It follows that  $M^3$  can be diagonalized, has real eigenvalues, and the square of these eigenvalues is  $+1$ . Verify, again using the Dirac algebra, that

$$\{M^3, \gamma^4\} = 0, \quad (28.0.994)$$

which shows that  $\gamma^4$  and  $M^3$  can be diagonalized simultaneously.

Let us find the simultaneous eigenvectors of  $\gamma^4$  and  $M^3$ . For  $w$  of the form (2.521) show, using (2.366), that

$$M^3 w = \begin{pmatrix} ic \\ -id \\ -ia \\ ib \end{pmatrix}. \quad (28.0.995)$$

Therefore, if we demand that

$$M^3 w = w, \quad (28.0.996)$$

verify that there are the relations

$$c = -ia \quad (28.0.997)$$

and

$$d = ib. \quad (28.0.998)$$

Suppose we now seek to satisfy all the demands (2.523), (2.524), (2.536), and (2.537). Verify that combining (2.523) and (2.537) yields the relation

$$b = -a. \quad (28.0.999)$$

Show that inserting (2.538) into (2.525) yields the vector, which we will call  $w^{+\uparrow}$ , given by the relation

$$w^{+\uparrow} = \begin{pmatrix} a \\ -a \\ -ia \\ -ia \end{pmatrix}. \quad (28.0.1000)$$

(Here, in anticipation of subsequent results, we introduce the symbols  $\uparrow$  and  $\downarrow$  to indicate what will soon be identified as having spin up and spin down.) Verify that  $w^{+\uparrow}$  is a simultaneous eigenvector of both  $\gamma^4$  and  $M^3$  with eigenvalues +1 for each.

Alternatively, if we demand that

$$M^3 w = -w, \quad (28.0.1001)$$

verify that there are the relations

$$c = ia \quad (28.0.1002)$$

and

$$d = -ib. \quad (28.0.1003)$$

Verify that combining (2.523) and (2.542) in this case now yields the relation

$$b = a. \quad (28.0.1004)$$

Show that inserting (2.543) into (2.525) yields the vector, which we will call  $w^{+\downarrow}$ , given by the relation

$$w^{+\downarrow} = \begin{pmatrix} a \\ a \\ ia \\ -ia \end{pmatrix}. \quad (28.0.1005)$$

Verify that  $w^{+\downarrow}$  is a simultaneous eigenvector of both  $\gamma^4$  and  $M^3$  with eigenvalue +1 for  $\gamma^4$  and eigenvalue -1 for  $M^3$ . Note that the quantity  $a$  appearing in  $w^{+\uparrow}$  and  $w^{+\downarrow}$  is arbitrary and may, for example, be set to 1.

Finally, let us make the definition

$$S^3 = iK(\hat{L}^3). \quad (28.0.1006)$$

Verify using (2.373) and (2.530) that

$$S^3 = (1/2)M^3. \quad (28.0.1007)$$

Consequently,  $w^{+\uparrow}$  is a simultaneous eigenvector of both  $\gamma^4$  and  $S^3$  with eigenvalue +1 for  $\gamma^4$  and eigenvalue +1/2 for  $S^3$ . And  $w^{+\downarrow}$  is a simultaneous eigenvector of both  $\gamma^4$  and  $S^3$  with eigenvalue +1 for  $\gamma^4$  and eigenvalue -1/2 for  $S^3$ . We may say that  $w^{+\uparrow}$  is the spinor part of the  $\psi$  for a spin 1/2 particle (at rest) having spin up, and  $w^{+\downarrow}$  is the spinor part of the  $\psi$  for a spin 1/2 particle (at rest) having spin down.

So far we have been discussing positive energy solutions of the Dirac equation. We close this exercise with a brief discussion of negative energy solutions, which also exist. Suppose, instead of (2.509), we make the Ansatz

$$\psi(x) = w \exp[+i(\sum_{\mu} p_{\mu} x^{\mu})/\hbar]. \quad (28.0.1008)$$

Observe that the argument of the exponential function appearing in (2.547) eventually becomes *positive* imaginary as  $t$  becomes evermore positive. Correspondingly, (2.547) is called a *negative* energy solution.<sup>25</sup> Show that the Ansatz (2.547) satisfies the Dirac equation (2.504) provided  $w$  satisfies the relation

$$\gamma^4 w = -w. \quad (28.0.1009)$$

That is,  $w$  must be an eigenvector of  $\gamma^4$  with eigenvalue -1.

Consider the vectors (4-spinors)  $w^{-\uparrow}$  and  $w^{-\downarrow}$  defined by

$$w^{-\uparrow} = (w^{+\downarrow})^* \quad (28.0.1010)$$

and

$$w^{-\downarrow} = (w^{+\uparrow})^*. \quad (28.0.1011)$$

Verify that

$$\gamma^4 w^{-\uparrow} = [(\gamma^4)^* w^{+\downarrow}]^* = [-\gamma^4 w^{+\downarrow}]^* = -(w^{+\downarrow})^* = -w^{-\uparrow}. \quad (28.0.1012)$$

Similarly, verify that

$$\gamma^4 w^{-\downarrow} = -w^{-\downarrow}. \quad (28.0.1013)$$

Next, observe from (2.545) that

$$(S^3)^* = -S^3. \quad (28.0.1014)$$

Consequently, verify that

$$S^3 w^{-\uparrow} = [(S^3)^* w^{+\downarrow}]^* = [-S^3 w^{+\downarrow}]^* = (1/2)[w^{+\downarrow}]^* = (1/2)w^{-\uparrow}. \quad (28.0.1015)$$

Similarly, verify that

$$S^3 w^{-\downarrow} = -(1/2)w^{-\downarrow}. \quad (28.0.1016)$$

Thus,  $w^{-\uparrow}$  and  $w^{-\downarrow}$  behave under the action of  $\gamma^4$  and  $S^3$  as their notation suggests. Finally, verify from (2.539), (2.544), (2.549), and (2.550) that there are the explicit results

$$w^{-\uparrow} = \begin{pmatrix} a^* \\ a^* \\ -ia^* \\ ia^* \end{pmatrix}, \quad (28.0.1017)$$

---

<sup>25</sup>Motivated by the relation  $E = h\nu$ , positive energy solutions are also sometimes called *positive frequency* solutions, and negative energy solutions are sometimes called *negative frequency* solutions.

$$w^{-\downarrow} = \begin{pmatrix} a^* \\ -a^* \\ ia^* \\ ia^* \end{pmatrix}. \quad (28.0.1018)$$

Here again the quantity  $a$  is arbitrary and may, for example, be set to 1.

In summary, when both positive and negative energy solutions (for a particle at rest) are considered, we have seen that there are four possibilities for  $w$ , namely  $w^{+\uparrow}$ ,  $w^{+\downarrow}$ ,  $w^{-\uparrow}$ , and  $w^{-\downarrow}$ .

What is the use of the negative energy solutions? Very much further discussion would bring us too far afield. But remarkably it can be shown that, taken together, the positive and negative energy solutions can be used to construct a four-component *quantum* field involving creation and destruction operators for particles and their antiparticles. In this construction destruction operators are associated with positive energy solutions and creation operators are associated with negative energy solutions. The result of this construction is a theory that describes simultaneously both spin 1/2 particles and their antimatter counterparts (for example electrons and positrons), and these particles obey Fermi-Dirac statistics. (Thus the four-fold nature of the Dirac equation associated with the four possibilities for  $w$  is revealed to be related to the four possibilities of spin up and spin down and matter and antimatter.) Moreover, in the full quantum field version of Dirac theory, there is a ground state called the *vacuum* which corresponds to a (unique) state for which there are no particles. And in the quantum field theory version both particle and antiparticle states have positive energies, and there is complete symmetry between matter and antimatter.

**28.0.32.** The two preceding Exercises 2.17 and 2.18 in this chapter have, among other things, explored the relations between the  $2 \times 2$  complex matrices  $\hat{L}^j$  and  $\hat{N}^j$ , that carry the  $\Gamma(0, 1/2)$  of the Lorentz group Lie algebra, and the  $4 \times 4$  real matrices  $K(\hat{L}^j)$  and  $K(\hat{N}^j)$ . According to Exercise 7.3.30, the  $2 \times 2$  complex matrices  $\hat{L}^j$  and  $\hat{N}^j$  carry the  $\Gamma(1/2, 0)$  of the Lorentz group Lie algebra. The purpose of this exercise is to consider the complementary question: what are the relations between the  $2 \times 2$  complex matrices  $\hat{L}^j$  and  $\hat{N}^j$  and the  $4 \times 4$  real matrices  $K(\hat{L}^j)$  and  $K(\hat{N}^j)$ ? We know from the work of Exercise 2.17 that the matrix set  $K(\hat{L}^j), K(\hat{N}^j)$  carries the Lorentz group Lie algebra representation  $\Gamma(0, 1/2) \oplus \Gamma(1/2, 0)$ . Could it be that the matrix set  $K(\hat{L}^j), K(\hat{N}^j)$  also carries this representation? If so, the two sets must be related by a similarity transformation (and vice versa).

Begin our/your investigation by showing from (7.3.248), (7.3.249), and (7.3.376) that there are the relations

$$K(\hat{L}^j) = K(\hat{L}^j) \quad (28.0.1019)$$

and

$$K(\hat{N}^j) = -K(\hat{N}^j). \quad (28.0.1020)$$

Next verify, using (2.366) through (2.368), (2.373) through (2.375), and the Dirac algebra, that

$$(\gamma^4)^{-1} K(\hat{L}^j) \gamma^4 = K(\hat{L}^j) \quad (28.0.1021)$$

from which, by (2.558), it follows that

$$(\gamma^4)^{-1} K(\hat{L}^j) \gamma^4 = K(\hat{L}^j). \quad (28.0.1022)$$

Also verify, using (2.369) through (2.371), (2.376) through (2.378), and the Dirac algebra, that

$$(\gamma^4)^{-1} K(\hat{N}^j) \gamma^4 = -K(\hat{N}^j) \quad (28.0.1023)$$

from which, by (2.559), it follows that

$$(\gamma^4)^{-1} K(\hat{N}^j) \gamma^4 = K(\check{\hat{N}}^j). \quad (28.0.1024)$$

Taken together, (2.561) and (2.563) show that the two sets  $K(\hat{L}^j), K(\hat{N}^j)$  and  $K(\check{\hat{L}}^j), K(\check{\hat{N}}^j)$  are indeed related by a similarity transformation, the similarity transformation provided by  $\gamma^4$ .<sup>26</sup> Finally, verify that the two sets are also related by the similarity transformation provided by  $\gamma^5 \gamma^4$ .

Let us summarize our findings. From Exercise 7.3.30 we learned that the  $\Gamma(0, 1/2)$  representation provided by the matrices  $\hat{L}^j, \hat{N}^j$  is different from (not equivalent to) the  $\Gamma(1/2, 0)$  representation provided by the matrices  $\check{\hat{L}}^j, \check{\hat{N}}^j$ . Now we have learned that the representations provided by the two sets of matrices  $K(\hat{L}^j), K(\hat{N}^j)$  and  $K(\check{\hat{L}}^j), K(\check{\hat{N}}^j)$  are the same/equivalent, namely the representation  $\Gamma(0, 1/2) \oplus \Gamma(1/2, 0)$ .

**28.0.33.** (Under construction) Review Exercise 2.19. The purpose of this exercise is to do something analogous for the representations  $\Gamma(0, 1)$ ,  $\Gamma(1, 0)$ , and  $\Gamma(0, 1) \oplus \Gamma(1, 0)$  using  $K$ .

**28.0.34.** (Under construction) Review Exercise 7.3.34. The purpose of this exercise is to explore, at least to some extent, what special properties matrices  $K(k)$  might have if the matrices  $k$  have special properties. In particular, we will study what can be said about the case

$$k \in SL(2, \mathbb{C}) = Sp(2, \mathbb{C}) \Leftrightarrow k^T J_2 k = J_2 \quad (28.0.1025)$$

without immediately invoking the Dirac machinery.

Apply (7.3.378) to the symplectic condition in (2.557) to show that

$$K(k^T) K(J_2) K(k) = K(J_2). \quad (28.0.1026)$$

Using the definition (7.3.375), verify the following relations:

$$K(J_2) = \begin{pmatrix} J_2 & \mathbf{0} \\ \mathbf{0} & J_2 \end{pmatrix} = J', \quad (28.0.1027)$$

$$K(k^T) = \begin{pmatrix} \Re k^T & -\Im k^T \\ \Im k^T & \Re k^T \end{pmatrix} = \begin{pmatrix} (\Re k)^T & -(\Im k)^T \\ (\Im k)^T & (\Re k)^T \end{pmatrix}, \quad (28.0.1028)$$

and

$$[K(k)]^T = \begin{pmatrix} (\Re k)^T & (\Im k)^T \\ -(\Im k)^T & (\Re k)^T \end{pmatrix}. \quad (28.0.1029)$$

---

<sup>26</sup>Note that the relations (2.561) and (2.563) can also be written in the form  $(\gamma^4)^{-1} K(\check{\hat{L}}^j) \gamma^4 = K(\hat{L}^j)$ , etc. Thus, there is complete symmetry between the use of the  $\Gamma(0, 1/2)$  and the  $\Gamma(1/2, 0)$  representations.

Let  $J$  be the matrix

$$J = \begin{pmatrix} \mathbf{0} & I^{[2]} \\ -I^{[2]} & \mathbf{0} \end{pmatrix}. \quad (28.0.1030)$$

Verify that

$$J^{-1}[K(k)]^T J = K(k^T). \quad (28.0.1031)$$

Show, using (2.559) and (2.563), that (2.558) can be rewritten in the form

$$J^{-1}[K(k)]^T J J' K(k) = J', \quad (28.0.1032)$$

from which it follows that

$$[K(k)]^T [J J'] K(k) = [J J']. \quad (28.0.1033)$$

Define a matrix  $S$  by the rule

$$S = J J' \quad (28.0.1034)$$

so that (2.565) can be written as

$$[K(k)]^T S K(k) = S. \quad (28.0.1035)$$

[Note that, despite our notation, this matrix  $S$  has nothing to do with the matrix  $S^3$  defined by (2.538) and (2.539).] Verify that

$$S = \begin{pmatrix} \mathbf{0} & J_2 \\ -J_2 & \mathbf{0} \end{pmatrix}. \quad (28.0.1036)$$

Verify that  $J$  and  $J'$  commute,

$$\{J, J'\} = 0. \quad (28.0.1037)$$

Show that

$$S^T = S, \quad (28.0.1038)$$

$$\det(S) = 1, \quad (28.0.1039)$$

and

$$S^2 = I^{[4]}. \quad (28.0.1040)$$

Evidently  $S$  is *real*. Verify from (2.570) and (2.572) that  $S$  is also orthogonal,

$$S^T S = I^{[4]}. \quad (28.0.1041)$$

A brute force verification using (2.568), while tedious, is possible. But note that (2.570) through (2.572) follow directly from (2.566) and (2.569) and the relations

$$J^T = -J, \quad (J')^T = -J', \quad \det(J) = \det(J') = 1. \quad (28.0.1042)$$

Upon comparing (2.567) with the symplectic condition in (2.557) we see that they have a similar form but  $J_2$  is antisymmetric while  $S$  is symmetric. Since  $S$  is symmetric, it can be used to define an inner product, and the relation (2.567) can be viewed as a *preservation*

condition for this inner product. Suppose  $u$  and  $v$  are any two four-component real arrays and we define their inner product  $(u, v)_S$  by the rule

$$(u, v)_S = (u, Sv) \quad (28.0.1043)$$

where the inner product on the right side of (2.575) is the usual real inner product. Note that, since  $S$  is real, the quantity  $(u, v)_S$  is real if  $u$  and  $v$  are real. Then, using (2.567), verify that

$$(Ku, Kv)_S = (Ku, SKv) = (u, K^T SKv) = (u, Sv) = (u, v)_S. \quad (28.0.1044)$$

That is, the inner product  $(u, v)_S$  is invariant under the action of  $K(k)$ . Moreover, make the definition

$$(u, v)_{S\gamma_r^5} = (u, S\gamma_r^5 v). \quad (28.0.1045)$$

Note that  $(u, v)_{S\gamma_r^5}$  is also *real*. Then, using (2.567) and (2.384), verify that

$$\begin{aligned} (Ku, Kv)_{S\gamma_r^5} &= (Ku, S\gamma_r^5 Kv) = (Ku, SK\gamma_r^5 v) = \\ (u, K^T SK\gamma_r^5 v) &= (u, S\gamma_r^5 v) = (u, v)_{S\gamma_r^5}. \end{aligned} \quad (28.0.1046)$$

Therefore the inner product  $(u, v)_{S\gamma_r^5}$  is also invariant under the action of  $K(k)$ .

But now we are confronted with an embarrassment of riches!

Out of curiosity, verify that

$$(u, v)_S = (u, Sv) = u_1 v_4 - u_2 v_3 - u_3 v_2 + u_4 v_1. \quad (28.0.1047)$$

**28.0.35.** (Under Construction) Exercise on use of  $S$  rather than  $\gamma_r^4$  and the existence of more bilinear forms.

**28.0.36.** (Under Construction) Other Majorana representations.

What is the nature of the inner product  $(*, *)_S$ ? Evidently  $S$  is real. Verify from (2.329) and (2.331) that  $S$  is also orthogonal,

$$S^T S = I^{[4]}. \quad (28.0.1048)$$

Therefore there must be a real orthogonal matrix  $O$  such that

$$O^T S O = D \Leftrightarrow S = O D O^T \quad (28.0.1049)$$

where  $D$  is diagonal. Verify using (2.335) and (2.331) that there is the relation

$$D^2 = O^T S O O^T S O = O^T S^2 O = O^T O = I^{[4]}, \quad (28.0.1050)$$

and therefore the diagonal entries of  $D$  must have absolute value 1.

Let us try to find  $O$  and  $D$ . To continue, verify that

$$S = K(-iJ_2) = K(\sigma^2). \quad (28.0.1051)$$

Next, using the relation between  $\boldsymbol{\sigma}$  and  $\mathbf{K}$  given by (3.7.169) through (3.7.171), verify that (8.2.57) can also be written in the form

$$\exp[(-i/2)\theta \mathbf{n} \cdot \boldsymbol{\sigma}] (\mathbf{a} \cdot \boldsymbol{\sigma}) \exp[(i/2)\theta \mathbf{n} \cdot \boldsymbol{\sigma}] = [R(\theta, \mathbf{n})\mathbf{a}] \cdot \boldsymbol{\sigma}. \quad (28.0.1052)$$

(Sorry yet again about the possibly confusing notation! Although they may look related, there is *no* connection between the symbols  $\mathbf{K}$  and  $K$ . Sometimes there are not enough symbols to go around.) Evaluate (2.338) for the case  $\mathbf{n} = \mathbf{e}_1$  and  $\mathbf{a} = \mathbf{e}_2$  so that it becomes

$$\exp[(-i/2)\theta \sigma^1](\sigma^2) \exp[(i/2)\theta \sigma^1] = [R(\theta, \mathbf{e}_1)\mathbf{e}_2] \cdot \boldsymbol{\sigma}. \quad (28.0.1053)$$

Verify that

$$[R(\pi/2, \mathbf{e}_1)\mathbf{e}_2] = \mathbf{e}_3. \quad (28.0.1054)$$

Put another way, rotating  $\mathbf{e}_2$  by  $\theta = \pi/2$  about the  $\mathbf{e}_1$  axis yields  $\mathbf{e}_3$ . See (3.7.205). Consequently, for  $\theta = \pi/2$ , (2.339) becomes

$$\exp[-i(\pi/4)\sigma^1](\sigma^2) \exp[i(\pi/4)\sigma^1] = \sigma^3. \quad (28.0.1055)$$

As a sanity check, verify directly that (2.341) holds by evaluating the indicated exponential functions and carrying out the indicated multiplications. In particular, you should find for the exponential functions the results

$$\exp[\pm i(\pi/4)\sigma^1] = \cos(\pi/4)\sigma^0 \pm i \sin(\pi/4)\sigma^1 = (1/\sqrt{2})\sigma^0 \pm i(1/\sqrt{2})\sigma^1. \quad (28.0.1056)$$

See (3.7.192).

Now verify that applying (7.3.378) to both sides of (2.341) yields the result

$$K\{\exp[-i(\pi/4)\sigma^1]\}K(\sigma^2)K\{\exp[i(\pi/4)\sigma^1]\} = K(\sigma^3). \quad (28.0.1057)$$

Make the assignment

$$O = K\{\exp[i(\pi/4)\sigma^1]\} \quad (28.0.1058)$$

and show that

$$O = (1/\sqrt{2}) \begin{pmatrix} \sigma^0 & -\sigma^1 \\ \sigma^1 & \sigma^0 \end{pmatrix}. \quad (28.0.1059)$$

Note that, as expected,  $O$  is real. Also, using (7.3.381), show that

$$O^{-1} = K\{\exp[-i(\pi/4)\sigma^1]\} \quad (28.0.1060)$$

from which it follows that

$$O^{-1} = (1/\sqrt{2}) \begin{pmatrix} \sigma^0 & \sigma^1 \\ -\sigma^1 & \sigma^0 \end{pmatrix}. \quad (28.0.1061)$$

Verify, by comparing (2.345) and (2.347), that

$$O^{-1} = O^T. \quad (28.0.1062)$$

Verify it follows from (2.343) that (2.335) has been achieved with

$$D = K(\sigma^3) = \begin{pmatrix} \sigma_3 & \mathbf{0} \\ \mathbf{0} & \sigma_3 \end{pmatrix} = \text{diag}(1, -1, 1, -1). \quad (28.0.1063)$$

As a final sanity check on our/your work, verify directly that (2.335) has been achieved using (2.327) for  $S$ , (2.345) for  $O$ , (2.347) for  $O^{-1} = O^T$ , and (2.349) for  $D$ .

Define matrices  $M(k)$  by the rule

$$M(k) = O^T K(k) O = O^{-1} K(k) O. \quad (28.0.1064)$$

Verify that, for  $k \in SL(2, \mathbb{C}) = Sp(2, \mathbb{C})$ , they satisfy the relation chain

$$\begin{aligned} M^T(k) D M(k) &= [O^T K(k) O]^T D [O^T K(k) O] = \\ &[O^T K^T(k)] [O D O^T] [K(k) O] = O^T [K^T(k) S K(k)] O = \\ &O^T S O = D. \end{aligned} \quad (28.0.1065)$$

That is, upon comparing the beginning and end of (2.351), we see that there is the relation

$$M^T(k) D M(k) = D. \quad (28.0.1066)$$

Verify also, using (7.3.380), that

$$M(I^{[2]}) = I^{[4]}. \quad (28.0.1067)$$

Because  $D$  has two positive and two negative diagonal entries we conclude that, for  $k \in SL(2, \mathbb{C}) = Sp(2, \mathbb{C})$ , the matrices  $M(k)$  provide a representation of  $SO(2, 2, \mathbb{R})$ . See Exercises 3.7.38 and 3.7.40.

Let  $P$  be the permutation operator that interchanges the two and three axes and leaves the other axes in peace. That is,  $P$  is a linear operator with the actions

$$P\mathbf{e}_1 = \mathbf{e}_1, \quad P\mathbf{e}_2 = \mathbf{e}_3, \quad P\mathbf{e}_3 = \mathbf{e}_2, \quad P\mathbf{e}_4 = \mathbf{e}_4. \quad (28.0.1068)$$

Correspondingly,  $P$  has the matrix representation

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (28.0.1069)$$

Verify that  $P$  is symmetric,  $P^T = P$ ; is an involution,  $P^2 = I^{[4]}$ ; and is therefore orthogonal,  $P^T = P^{-1}$ . Define a linear operator  $\hat{D}$  by the rule

$$\hat{D} = P^T D P. \quad (28.0.1070)$$

Verify that  $\hat{D}$  has the matrix representation

$$\hat{D} = \begin{pmatrix} I^{[2]} & \mathbf{0} \\ \mathbf{0} & -I^{[2]} \end{pmatrix} = \text{diag}(1, 1, -1, -1). \quad (28.0.1071)$$

Define matrices  $\hat{M}(k)$  by the rule

$$\hat{M}(k) = P^T M(k) P. \quad (28.0.1072)$$

Verify that, for  $k \in SL(2, \mathbb{C}) = Sp(2, \mathbb{C})$ , they satisfy the relation chain

$$\begin{aligned} M^T(k)DM(k) &= [O^T K(k)O]^T D [O^T K(k)O] = \\ &[O^T K^T(k)][ODO^T][K(k)O] = O^T[K^T(k)SK(k)]O = \\ &O^T SO = D. \end{aligned} \quad (28.0.1073)$$

That is, upon comparing the beginning and end of (2.351), we see that there is the relation

$$M^T(k)DM(k) = D. \quad (28.0.1074)$$

Verify also, using (7.3.380), that

$$M(I^{[2]}) = I^{[4]}. \quad (28.0.1075)$$

Because  $D$  has two positive and two negative diagonal entries we conclude that, for  $k \in SL(2, \mathbb{C}) = Sp(2, \mathbb{C})$ , the matrices  $M(k)$  provide a representation of  $SO(2, 2, \mathbb{R})$ . See Exercises 3.7.38 and 3.7.40.

**28.0.37.** (Under Construction) Review Exercise 3.7.37 that relates the Lorentz group Lie algebra  $so(3, 1, \mathbb{R})$  to  $so(4, \mathbb{R})$  when working over the complex field, and hence also to  $su(2) \oplus su(2)$ . See also Exercise 6.2.6.

In the case of the Lorentz group there are the  $4 \times 4$  *Dirac gamma* matrices  $\gamma^\mu$ , with  $\mu = 1 \cdots 4$ , that satisfy the anti-commutation rules

$$\{\gamma^\mu, \gamma^\nu\}_+ = \gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2g^{\mu\nu}, \quad (28.0.1076)$$

and transform under the action of the Lorentz group according to the rules

$$* = *. \quad (28.0.1077)$$

For the case of  $SO(4, \mathbb{R})$  there are analogous  $4 \times 4$  matrices; call them  $\Gamma^\mu$ . Define them as follows:

$$\Gamma^1 =, \quad (28.0.1078)$$

$$\Gamma^2 =, \quad (28.0.1079)$$

$$\Gamma^3 =, \quad (28.0.1080)$$

$$\Gamma^4 =. \quad (28.0.1081)$$

Note they are all real and involve  $2 \times 2$  blocks featuring the Pauli matrices  $*$ . Verify that they satisfy the anti-commutation rules

$$\{\Gamma^\mu, \Gamma^\nu\}_+ = \delta^{\mu\nu}, \quad (28.0.1082)$$

and transform under the action of the rotation group  $SO(4, \mathbb{R})$  according to the rules

$$* = *. \quad (28.0.1083)$$

Show that there are six linearly independent matrices of the form

$$\Sigma^{\mu\nu} = \{\Gamma^\mu, \Gamma^\nu\} = \Gamma^\mu \Gamma^\nu - \Gamma^\nu \Gamma^\mu. \quad (28.0.1084)$$

Show that these matrices form a basis for  $so(4, \mathbb{R})$ .

Something about  $SO(n, \mathbb{R})$  and Clifford algebras.

**28.0.38.** (Under Construction) Exercise on the Möbius and Lorentz groups.

**28.0.39.** The goal of this exercise is to relate the Lie algebras  $sp(4)$  and  $so(5)$ , and the Lie groups  $Sp(4)$  and  $SO(5)$ . You already know from Exercise 3.7.31 that they have the same dimension. You also know from Section 5.10.1 that  $sp(2n, \mathbb{R})$  and  $usp(2n)$  are equivalent over the complex field, but not over the real field. In this exercise you will show that  $usp(4)$  and  $so(5, \mathbb{R})$  are isomorphic. Therefore, in so doing, you will show that that  $sp(4, \mathbb{R})$  and  $so(5, \mathbb{R})$  are equivalent over the complex field, but not over the real field.

Review Exercise 8.2.12. There it is shown that if  $K \in su(4)$ , then  $L$  given by

$$L_{\alpha\beta}(K) = -(1/2)\text{tr}[K(A^\alpha)^\dagger A^\beta] \quad (28.0.1085)$$

will have the property  $L \in so(6, \mathbb{R})$ . Correspondingly, from the relation

$$R_{\alpha\beta}(v) = (1/4)\text{tr}[v^T A^\alpha v (A^\beta)^\dagger], \quad (28.0.1086)$$

there will be an  $R \in SO(6, \mathbb{R})$  for every  $v \in SU(4)$ . Now suppose that  $K \in sp(4, \mathbb{C})$  as well so that  $K \in usp(4)$ . Then we will also have  $v \in Sp(4, \mathbb{C})$  so that  $v \in USp(4)$ .

Suppose we set  $\alpha = 6$  in (5.33). From (8.2.98) we see that

$$A^6 = J, \quad (28.0.1087)$$

and therefore, if  $v$  is symplectic, there is the relation

$$v^T A^6 v = v^T J v = J = A^6. \quad (28.0.1088)$$

Consequently, show that there is the result

$$\begin{aligned} R_{6\beta}(v) &= (1/4)\text{tr}[v^T A^6 v (A^\beta)^\dagger] = (1/4)\text{tr}[v^T J v (A^\beta)^\dagger] \\ &= (1/4)\text{tr}[J (A^\beta)^\dagger] = (1/4)\text{tr}[A^6 (A^\beta)^\dagger] = \delta_{6\beta}. \end{aligned} \quad (28.0.1089)$$

Next set  $\beta = 6$  in (5.33). Show that

$$\begin{aligned} R_{\alpha 6}(v) &= (1/4)\text{tr}[v^T A^\alpha v (A^6)^\dagger] = (1/4)\text{tr}[v^T A^\alpha v (J)^\dagger] \\ &= (1/4)\text{tr}[A^\alpha v (J)^\dagger v^T] = (1/4)\text{tr}[A^\alpha J^\dagger] \\ &= (1/4)\text{tr}[A^\alpha (A^6)^\dagger] = \delta_{\alpha 6}. \end{aligned} \quad (28.0.1090)$$

Consequently show that, when  $v \in USp(4)$ ,  $R$  has the block form

$$R(v) = \begin{pmatrix} R_{11} & R_{12} & R_{13} & R_{14} & R_{15} & 0 \\ R_{21} & R_{22} & R_{23} & R_{24} & R_{25} & 0 \\ R_{31} & R_{32} & R_{33} & R_{34} & R_{35} & 0 \\ R_{41} & R_{42} & R_{43} & R_{44} & R_{45} & 0 \\ R_{51} & R_{52} & R_{53} & R_{54} & R_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (28.0.1091)$$

Let  $\hat{R}$  be the  $5 \times 5$  matrix

$$\hat{R} = \begin{pmatrix} R_{11} & R_{12} & R_{13} & R_{14} & R_{15} \\ R_{21} & R_{22} & R_{23} & R_{24} & R_{25} \\ R_{31} & R_{32} & R_{33} & R_{34} & R_{35} \\ R_{41} & R_{42} & R_{43} & R_{44} & R_{45} \\ R_{51} & R_{52} & R_{53} & R_{54} & R_{55} \end{pmatrix}. \quad (28.0.1092)$$

We know that  $R \in SO(6, \mathbb{R})$  when  $v \in USp(4)$ , and we have seen that  $R$  must then also have the form (5.38). Consequently, there is the result that  $\hat{R} \in SO(5, \mathbb{R})$  when  $v \in USp(4)$ . Therefore (5.33) provides a map of  $USp(4)$  into  $SO(5, \mathbb{R})$  when  $\alpha, \beta$  are restricted to range from 1 to 5. Verify that this map is a homomorphism,

$$\hat{R}(v_1 v_2) = \hat{R}(v_1) \hat{R}(v_2), \quad (28.0.1093)$$

and that

$$\hat{R}(-v) = \hat{R}(v) \quad (28.0.1094)$$

so that the homomorphism is two to one.

Finally, we should study the relation between  $usp(4)$  and  $so(5, \mathbb{R})$ . Show from (5.32) that

$$L_{6\beta}(K) = L_{\alpha 6}(K) = 0 \text{ when } K \in usp(4). \quad (28.0.1095)$$

We already know that  $L$  is real and antisymmetric when  $K \in su(4)$  and hence it will be real and antisymmetric when  $K \in usp(4)$  since  $usp(4)$  is a subalgebra of  $su(4)$ . It follows that (5.32) provides a map of  $usp(4)$  into  $so(5, \mathbb{R})$  when  $\alpha, \beta$  are restricted to range from 1 to 5. Your last task is to show that this map is an isomorphism.

As a first step, verify that  $J \in usp(4)$ . Consider computing  $L(J)$ . From (5.32) we know that

$$L_{\alpha\beta}(J) = -(1/2)\text{tr}[J(A^\alpha)^\dagger A^\beta]. \quad (28.0.1096)$$

Examine the products  $(A^\alpha)^\dagger A^\beta$ . See (8.2.112) through (8.2.116). Observe that the products are either symmetric matrices  $S$  or antisymmetric matrices  $A$ . Recall that matrices of the form  $JS$  are traceless. Therefore we only need be concerned with those products that are antisymmetric. But in this case we only need consider those products whose results are proportional to  $J = A^6$  because of the orthogonality condition (8.2.121). From (8.2.108) we see that the only products that contribute are of the form  $A^2 A^4$ . Verify that

$$\begin{aligned} L_{24}(J) &= -(1/2)\text{tr}[J(A^2)^\dagger A^4] = (1/2)\text{tr}[JA^2 A^4] \\ &= (1/2)\text{tr}[JA^6] = -(1/2)\text{tr}[A^6 (A^6)^\dagger] = -2. \end{aligned} \quad (28.0.1097)$$

Thus, show that

$$L(J) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (28.0.1098)$$

Next verify that (5.32) provides a homomorphism of  $usp(4)$  into  $so(5, \mathbb{R})$  when  $\alpha, \beta$  are restricted to range from 1 to 5. Finally, make an argument analogous to that of Exercise 8.9.19 to show that (5.32) provides an isomorphism of  $usp(4)$  into  $so(5, \mathbb{R})$  when  $\alpha, \beta$  are restricted to range from 1 to 5.