Map and Lie Methods for Accelerator Physics: A Tale of Two Symmetries

Alex J. Dragt[†]
University of Maryland

April 17, 2023

Abstract

Charged particle motion in E and B fields is characterized by two symmetries. First is Lorentz invariance, which is well understood. Second is the symplectic symmetry inherent in the Hamiltonian nature of the equations that generate this motion. (The relation between initial and final conditions obtained by integration of Hamilton's equations is what is called a symplectic map. Thus the effect of any beam-line element in an accelerator is a symplectic map.) Its importance was first studied by *Hamilton* and *Jacobi* in the context of generating functions, but all its implications are still not fully appreciated or understood. Since their discovery Lie algebras have been an important branch of Mathematics. Now they are also important for Physics both in Relativity and in Particle Theory. This talk aims to describe how Lie methods can also be used to exploit symplectic symmetry. A major goal of map and Lie methods is to treat linear and nonlinear behavior with equal facility. Lie tools have been developed for representing, multiplying, inverting, computing, and applying symplectic maps. These maps can also be analyzed by normal-form methods (a nonlinear generalization of matrix diagonalization) both to determine and optimize expected accelerator performance and to motivate new designs. Additional applications include fast long-term tracking and the concept and calculation of eigen emittances.

Acknowledgments

Institutions: U.S. Department of Energy, University of Maryland.

Minds: Peter Walstrom, Paul Channell, C. Thomas Mottershead, Filippo Neri, John Finn, David Douglas, Étienne Forest, Liam Healy, Robert Ryne, Govindan Rangarajan, Dan Abell, Marco Venturini, Chad Mitchell.

[†]2023 APS Wilson Prize recipient

Outline

This outline is a 4/17/2023 draft which will polished over the next few weeks.

1. Preliminaries: An Accelerator is a collection of beam-line elements, e.g. magnets, rf cavities, drift spaces, etc. We will employ canonical phase-space variables

$$z = (z_1, z_2 \cdots, z_6) = (q; p) = (x, y, t; p_x, p_y, p_t). \tag{1}$$

Here p_t is a momentum-like variable canonically conjugate to the time t, and time is treated as a coordinate. It turns out that $-p_t$ is the energy. To be more precise, the phase-space variables are *deviation* variables relative to the *design* orbit, and suitably normalized. Subsequently we will use z to denote a *Cartesian* coordinate, and then use the symbol ζ to denote phase-space variables.

2. Let $f(z)$ and $g(z)$ be any two functions of the phase-space variables z. Define their Poisson bracket by the rule

$$[f, g] = \sum_i (\partial f/\partial q_i)(\partial g/\partial p_i) - (\partial f/\partial p_i)(\partial g/\partial q_i). \tag{2}$$

The set of phase-space functions forms a *linear vector space*. (They can be multiplied by scalars and added.) The Poisson bracket may be viewed as a rule for *multiplying* these "vectors". This "multiplication" rule has the properties

$$[f, g] = -[g, f] \text{ (antisymmetry) and } [f, [g, h]] + [g, [h, f]] + [h, [f, g]] = 0 \text{ (Jacobi)}. \tag{3}$$

A multiplication rule with the properties (3) is called a *Lie* product, and we say that these vectors form a *Lie* algebra since they can be both added and "multiplied" to again obtain vectors. Also we have

$$[z_a, z_b] = J_{ab}, \tag{4}$$

where

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}. \tag{5}$$

Finally Let f_n denote any *homogeneous polynomial* of degree n in the variables z. Let deg be a "degree" function that extracts the degree of a homogeneous polynomial so that

$$\deg(f_n) = n. \tag{6}$$

Then, since a Poisson bracket involves a multiplication and two differentiations, there is the result

$$\deg([f_m, g_n]) = m + n - 2. \tag{7}$$

3. The effect of any beam-line element is described by a *transfer map* $\mathcal{M}$ that sends phase space into itself. It sends *initial* points z^{in} into *final* points z^{fin},

$$z^{fin} = \mathcal{M} z^{in}. \tag{8}$$

Let M be the *linear* part of $\mathcal{M}$. It is defined by the matrix

$$M_{ab} = \partial z_a^{fin} / \partial z_b^{in}, \tag{9}$$

and describes how small changes dz^{in} in z^{in} produce small changes dz^{fin} in z^{fin}. Then here is a key result: if $\mathcal{M}$ arises from a Hamiltonian, M must satisfy the relation

$$M^T J M = J \tag{10}$$

where M^T denotes the transpose of M. Matrices M that satisfy (10) are called *symplectic*; and $\mathcal{M}$ is called a *symplectic map*. We may say that the study of Hamiltonian dynamics is equivalent to the study of symplectic maps.

4. Let $f(z)$ be any function of z. Define an associated *differential operator*, denoted by the symbols $: f :$ and called a *Lie operator*, by the rule

$$: f := \sum_i (\partial f / \partial q_i)(\partial / \partial p_i) - (\partial f / \partial p_i)(\partial / \partial q_i). \tag{11}$$

Let $g(z)$ denote any other function of z. Then we have

$$: f : g = [f, g]. \tag{12}$$

Lie operators do not commute. From the Jacobi identity we find

$$: f :: g : - : g :: f := \{: f :, : g :\} =: [f, g] : . \tag{13}$$

5. Define powers of $: f :$ by writing

$$: f :^0 = \mathcal{I} \Leftrightarrow \ : f :^0 g = g, \ : f : g = [f, g], \ : f :^2 g = [f, [f, g]], \cdots . \tag{14}$$

Define the power series $e^{(:f:)} = \exp(: f :)$ by writing

$$\exp(: f :) = \sum_{k=0}^{\infty} : f :^k / k!. \tag{15}$$

Then, for any function g, there is the result

$$\exp(: f :) g = g + [f, g] + (1/2!)[f, [f, g]] + \cdots . \tag{16}$$

The operator $\exp(: f :)$ is called a *Lie transformation*.

6. Suppose a relation/mapping between z^{in} and z^{fin} is specified by writing

$$z^{fin} = \exp(: f :) z^{in} \tag{17}$$

where f is any function $f(z^{in})$. Then we may write (17) in the form

$$z^{fin} = \mathcal{M} z^{in} \tag{18}$$

with

$$\mathcal{M} = \exp(: f :). \tag{19}$$

It is a remarkable fact that any map $\mathcal{M}$ of the form (19) is a *symplectic* map.

7. Conversely, suppose $\mathcal{M}$ is any symplectic map that has a Taylor expansion in the variables z. Then there are *unique homogeneous polynomials* f_n of degree n in the variables z such that $\mathcal{M}$ can be written in the *standard factorized* form

$$\mathcal{M} = \exp(: f_1) \exp(: f_2^c) \exp(: f_2^a) \exp(: f_3) \exp(: f_4) \exp(: f_5) \cdots . \tag{20}$$

(Here f_2^c and f_2^a are quadratic polynomials related to symmetric matrices S^c and S^a with the properties that S^c *commutes* with J and S^a *anticommutes* with J.) Given $\mathcal{M}$, the homogeneous polynomials f_n are *unique*. Also, the product (20) can be truncated at any point and the result remains a symplectic map. We will call the f_n the *generators* of $\mathcal{M}$.

For convenience it is useful to define the map $\mathcal{R}$ by the rule

$$\mathcal{R} = \exp(: f_2^c) \exp(: f_2^a) \tag{21}$$

so that (20) can be rewritten in the form

$$\mathcal{M} = \exp(: f_1 :) \mathcal{R} \exp(: f_3 :) \exp(: f_4) : \exp(: f_5 :) \cdots . \tag{22}$$

It can be shown that $\exp(: f_1 :)$ produces translations in phase space, and therefore describes beam-line element misplacement effects; $\mathcal{R}$ produces linear transformations, typically arising from drifts and quadrupole magnets and rf cavities, and corresponding to symplectic matrices R; $\exp(: f_3 :)$ produces quadratic and higher-order terms corresponding to "sextupole magnet like" effects; $\exp(: f_4 :)$ produces cubic and higher-order terms corresponding to "octupole magnet like" effects; etc. It follows that the effect of any collection of beam-line elements can be described by a set of homogeneous polynomials f_n. How many? Through 4^{th} order $\rightarrow$ 209 coefficients; through 8^{th} order $\rightarrow$ 3002 coefficients. If Taylor series were employed instead, i.e. symplectic symmetry were ignored, $\approx 3\times$ more coefficients would be required.

8. There are Lie algebraic formulas for multiplying/concatenating and inverting symplectic maps. Let $: f :$ and $: g :$ be any two Lie operators and let s and t be any two scalars/parameters. Then, for Lie algebras, there is the *Baker-Campbell-Hausdorff* (BCH) formula which, in the context of Lie operators and Poisson brackets, states that

$$\exp(: sf :) \exp(: tg :) = \exp(: h :) \tag{23}$$

with

$$\begin{aligned} h(s,t) &= sf + tg + (1/2)st[f,g] \\ &+ (1/12)(s^2 t)[f,[f,g]] + (1/12)(st^2)[g,[g,f]] \\ &+ O(s^3 t, s^2 t^2, s t^3). \end{aligned} \tag{24}$$

A remarkable feature of the BCH formula (24), in this context, is that the only multiplicative terms are in the form of single and repeated Poisson brackets.

Introduce the notation

$$\mathcal{M}_f = \exp(:\, f_1)\exp(:\, f_2^c)\exp(:\, f_2^a)\exp(:\, f_3)\exp(:\, f_4)\exp(:\, f_5)\cdots \text{etc.} \tag{25}$$

Then, given polynomials f_n and g_n, it follows from the BCH formula that there are formulas for polynomials h_n such that

$$\mathcal{M}_f\mathcal{M}_g = \mathcal{M}_h. \tag{26}$$

Given, in standard factorized form, the transfer maps for individual beam-line elements, the net map (again in standard factorized form) for a collection of such elements can be found using the results of the form (26).

For map inversion, observe from (23) and (24) that the maps $\exp(:\, f\, :)$ and $\exp(:\, -f\, :)$ are inverses of each other,

$$\exp(:\, f\, :)\exp(:\, -f\, :) = \exp(:\, -f\, :)\exp(:\, f\, :) = \exp(:\, 0\, :) = \mathcal{I}. \tag{27}$$

It follows from (23) and (25) that

$$\mathcal{M}_f^{-1} = \cdots \exp(:\, -f_5\, :)\exp(-\, :\, f_4\, :)\exp(:\, -f_3\, :)\exp(:\, -f_2^a\, :)\exp(:\, -f_2^c\, :)\exp(-\, :\, f_1). \tag{28}$$

Finally, the maps appearing on the right side of (28) can be concatenated to produce a net result for $\mathcal{M}_f^{-1}$ in standard factorized form.

9. Computation of Maps for Beam-Line Elements. According to the discussion so far, the transfer map for any beam-line element, or any sequence of beam-line elements, can be written in the standard factorized form (22). And given any collection/sequence of beam-line elements, with each of their maps presented in standard factorized form, the net map for the collection can be found using concatenation formulas of the form (26). But how does one find the map for any individual beam-line element? How does one find/compute $\mathcal{R}$ and the f_m? That will be the topic of Items 10 through 12 below.

10. Cylindrical Harmonic Expansions. A useful steppingstone would be to have some way of describing the electromagnetic fields associated with a beam-line element. One method is to employ Cylindrical Harmonic Expansions. For example, we will treat the case of static magnetic fields $\boldsymbol{B}$.

    Under the assumption that there are no currents in the interior of an evacuated beam pipe, and that the magnetic field is static, it must satisfy within the pipe the condition

    $$\nabla \times \boldsymbol{B} = 0. \tag{29}$$

    Consequently, there must be a *scalar* potential $\psi(\boldsymbol{r})$ such that, within the pipe,

    $$\boldsymbol{B}(\boldsymbol{r}) = -\nabla\psi(\boldsymbol{r}). \tag{30}$$

    Also, since $\boldsymbol{B}$ must be divergence free, ψ must be harmonic,

    $$\nabla^2\psi = 0. \tag{31}$$

Assume the beam pipe is *straight*. Employ cylindrical coordinates ρ, ϕ, z with the beam pipe centered on the z axis so that

$$x = \rho\cos\phi \text{ and } y = \rho\sin\phi. \tag{32}$$

Then, from the theory of harmonic functions, ψ must have a cylindrical harmonic expansion of the form

$$\begin{aligned}
\psi(x,y,z) &= \psi_0 + \sum_{m=1}^{\infty}\psi_{mc} + \sum_{m=1}^{\infty}\psi_{ms} \\
&= -\sum_{\ell=0}^{\infty}(-1)^{\ell}\frac{1}{2^{2\ell}\ell!\ell!}C_0^{[2\ell]}(z)\rho^{2\ell} \\
&- \sum_{m=1}^{\infty}\cos(m\phi)\sum_{\ell=0}^{\infty}(-1)^{\ell}\frac{m!}{2^{2\ell}\ell!(\ell+m)!}C_{m,c}^{[2\ell]}(z)\rho^{2\ell+m} \\
&- \sum_{m=1}^{\infty}\sin(m\phi)\sum_{\ell=0}^{\infty}(-1)^{\ell}\frac{m!}{2^{2\ell}\ell!(\ell+m)!}C_{m,s}^{[2\ell]}(z)\rho^{2\ell+m}.
\end{aligned} \tag{33}$$

The quantities $C_0^{[0]}(z)$, $C_{m,c}^{[0]}(z)$, and $C_{m,s}^{[0]}(z)$ are called *on-axis gradients*, and there are the recursion relations

$$C_0^{[\ell+1]}(z) = \partial_z C_0^{[\ell]}(z), \tag{34}$$

$$C_{m,c}^{[\ell+1]}(z) = \partial_z C_{m,c}^{[\ell]}(z), \tag{35}$$

$$C_{m,s}^{[\ell+1]}(z) = \partial_z C_{m,s}^{[\ell]}(z). \tag{36}$$

The quantity $C_0^{[0]}(z)$, which has subscript 0, describes a *solenoida*l field. The quantities $C_{m,c}^{[0]}(z)$ and $C_{m,s}^{[0]}(z)$, which have subscript m with $m > 0$, describe *multipole* fields. They have the extra subscripts s and c which denote the cases of normal and skew multipole fields, respectively. Observe that the on-axis gradients serve as *master functions* in that, if they are known, then ψ and hence $\boldsymbol{B}$ are specified by (30) and (33).

Suppose, for example, that only the $C_{2,s}^{[0]}(z)$ related terms are present in (33). In this $m = 2$ case, ψ has an expansion of the form

$$\begin{aligned}
\psi_{2,s}(x,y,z) &= -\rho^2\sin(2\phi)[C_{2,s}^{[0]}(z) - (1/24)(x^2+y^2)C_{2,s}^{[2]}(z) + \cdots] \\
&= -2xy[C_{2,s}^{[0]}(z) - (1/24)(x^2+y^2)C_{2,s}^{[2]}(z) + \cdots],
\end{aligned} \tag{37}$$

and therefore

$$B_x = -\partial_x\psi_{2,s} = 2yC_{2,s}^{[0]}(z) - (1/12)(3x^2y + y^3)C_{2,s}^{[2]}(z) + \cdots, \tag{38}$$

$$B_y = -\partial_y\psi_{2,s} = 2xC_{2,s}^{[0]}(z) - (1/12)(x^3 + 3xy^2)C_{2,s}^{[2]}(z) + \cdots, \tag{39}$$

$$B_z = -\partial_z\psi_{2,s} = 2xy[C_{2,s}^{[1]}(z) - (1/24)(x^2+y^2)C_{2,s}^{[3]}(z) + \cdots]. \tag{40}$$

We see that $\boldsymbol{B}$ is primarily the field of a (normal) quadrupole with a profile given by $Q(z) = 2C^{[0]}_{2,s}(z)$. To lowest order the sizes of its transverse components are specified by $2C^{[0]}_{2,s}(z)$ which, in a long well-made quadrupole, is nearly constant near the center of the quadrupole. As long as $C^{[0]}_{2,s}(z)$ is nearly constant, $C^{[1]}_{2,s}(z)$ and higher derivatives of $C^{[0]}_{2,s}(z)$ will be small, and therefore the other field components will be small away from the ends of the quadrupole. But we also know that $\boldsymbol{B}$ must vanish for large $|z|$, and therefore $C^{[0]}_{2,s}(z)$ must go from a nearly constant value near the center to a vanishing value at the ends of a quadrupole. Consequently $C^{[1]}_{2,s}(z)$ and higher z derivatives of $C^{[0]}_{2,s}(z)$ will be large at the ends of a quadrupole; and therefore B_z will be large at the ends of a quadrupole. Nonzero fringe fields are an unavoidable consequence of the Maxwell equations. Moreover, these fringe fields produce nonlinear terms in the transfer map. We conclude that the transfer map for a real quadrupole must contain nonlinear terms because $C^{[0]}_{2,s}(z)$ must have nonzero derivatives in the fringe-field regions. Similar conclusions hold for all multipoles. One of our goals is to take into account all nonlinear fringe-field and higher-order multipole effects.

11. Computation of On-Axis Gradients. We have seen that on-axis gradients serve as a useful description of beam-line element fields. (However, see also Item 13 to follow.) But how does one compute on-axis gradients? There are several cases to be considered:

    a. Suppose there are no external currents, no iron, and that only rare-earth cobalt (REC) materials are present. Then there are explicit formulas for the $C^{[0]}_{m,c}(z)$ and $C^{[0]}_{m,s}(z)$. We will describe briefly, for example, results for a normal REC quadrupole.

    A 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 for a normal REC quadrupole can be described in terms of a soft-edge bump function which we will call $\mathrm{bump}(z, r_1, r_2, L)$ where the REC material occupies the interval $z \in [0, L]$. That is, the on-axis field gradient $C^{[0]}_{2,s}(z)$ can be written in the form

    $$C^{[0]}_{2,s}(z) = (Q/2)\mathrm{bump}(z, r_1, r_2, L) \tag{41}$$

    where Q is the strength of the REC quadrupole in the infinite length limit. Figure 1 displays, for example, $\mathrm{bump}(z, r_1, r_2, L)$ for a particular case.

    Moreover, the soft-edge bump function can be written in terms of an associated approximating signum function in the form

    $$\mathrm{bump}(z, r_1, r_2, L) = [\mathrm{sgn}(z, r_1, r_2) - \mathrm{sgn}(z - L, r_1, r_2)]/2. \tag{42}$$

    Finally, the approximating signum function $\mathrm{sgn}(z, r_1, r_2)$ is given by the relation

    $$\mathrm{sgn}(z, r_1, r_2) = z[(r_1 + r_2)/(r_1 r_2)][(v_1 v_2)/(v_1 + v_2)][1 + (1/8)v_1 v_2(4 + v_1^2 + v_1 v_2 + v_2^2)] \tag{43}$$

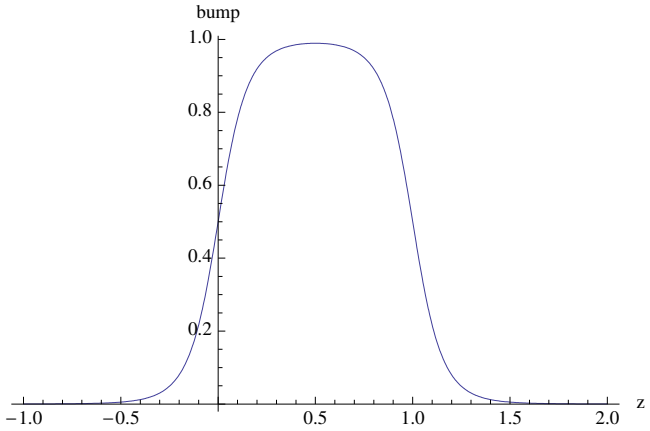

Figure 1: The soft-edge bump function $\text{bump}(z, r_1 r_2, L)$ for a REC quadrupole when $r_1 = .2$, $r_2 = .5$, and $L = 1$.

where v_1 and v_2 are defined by the relations

$$v_1 = 1/\sqrt{1 + (z/r_1)^2}, \tag{44}$$

$$v_2 = 1/\sqrt{1 + (z/r_2)^2}. \tag{45}$$

b. Suppose there are no magnetic materials (REC or iron) present, but only current carrying coils/windings. In this case there are two methods that have been developed to describe possible coil windings and the on-axis gradients they produce. The simplest to visualize/describe is that of Bassetti and Biscari. See Figure 2 which shows a net of n separate single-turn coils placed on a circular cylinder of radius a. We call these windings "rectangular". We note that these coils must be separately powered, which is a possible drawback for the use of rectangular windings. The second method produces more complicated and more general windings that are described by *stream functions*. In this case some windings can be wired in series so that fewer power supplies are needed.

b1. Rectangular Windings

Rectangular windings have the feature that it is possible to compute the $C_{m,s}^{[k]}(z)$ *analytically*. For example, for a quadrupole, if a rectangular winding begins at $z = 0$ and extends to $z = L$, then the on-axis field gradient for such a quadrupole, which we will call an *ideal air-core quadrupole*, can again be described in terms of a soft-edge bump function which we will call $\text{bump}_{\text{iacq}}(z, a, L)$. See Figure 3. That is, the on-axis field gradient $C_{2,s}^{[0]}(z)$ can again be written in the form

$$C_{2,s}^{[0]}(z) = (Q_0/2)\text{bump}_{\text{iacq}}(z, a, L) \tag{46}$$

where Q_0 is again the quadrupole strength in the infinite length limit.

Like the previous example of a soft-edge bump function, the soft-edge bump function for an ideal air-core quadrupole can be written in terms of an associated approximating signum

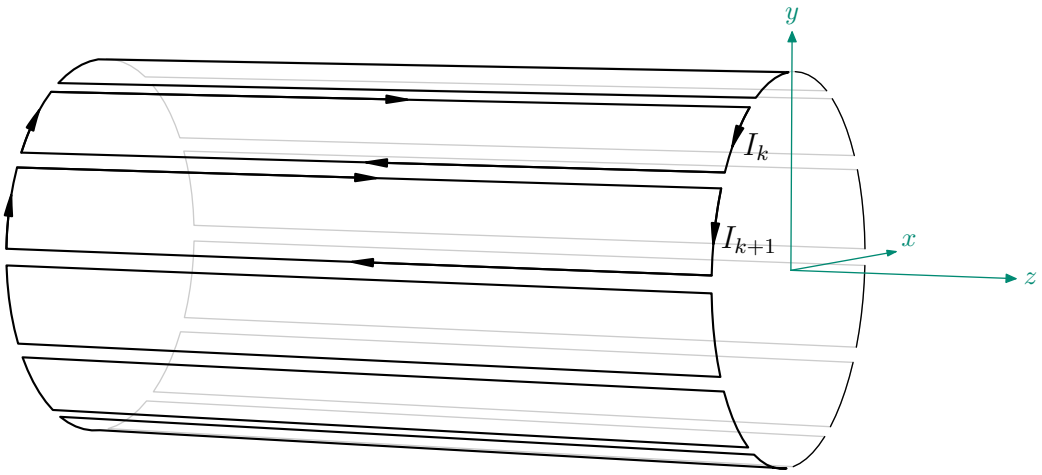

Figure 2: A net of n rectangular coils draped over a cylinder of radius a. The k^{th} coil carries a current I_k.

function sgn_{iacq} in the form

$$\text{bump}_{\text{iacq}}(z, a, L) = [\text{sgn}_{\text{iacq}}(z, a) - \text{sgn}_{\text{iacq}}(z - L, a)]/2. \tag{47}$$

Finally, 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

$$\begin{aligned} \text{sgn}_{\text{iacq}}(z, a) &= [z^5 + (5/2)z^3a^2 + (9/4)za^4]/(z^2 + a^2)^{5/2} \\ &= z[z^4 + (5/2)z^2a^2 + (9/4)a^4]/(z^2 + a^2)^{5/2}. \end{aligned} \tag{48}$$

b2. Stream-Function Windings.

In this section we will describe the use of stream-function windings, which has been extensively discussed by Walstrom. They are more general than Bassetti-Biscari windings, but have the complication that the on-axis gradients and their derivatives they produce, the $C_{m,s}^{[k]}(z)$, are generally available only (but reliably) in numerical form. Moreover, exactly how the windings are to be laid out involves finding stream-function level lines and spacing them appropriately. For simplicity, we will omit describing how this may be done, and will concentrate on the computation of the $C_{m,s}^{[k]}(z)$.

Consider a circular cylinder of radius a. Suppose there is a specified (and time-independent) current $\boldsymbol{J}$ flowing in a winding on the *surface* of the cylinder. Assume also that the electric charge density ρ_e is time independent. Then a static magnetic field $\boldsymbol{B}$ 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 Item 10, 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 $\boldsymbol{J}$.

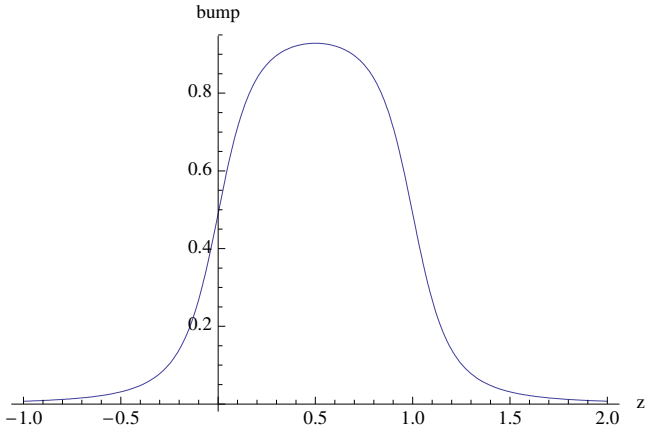

Figure 3: The soft-edge bump function for a Bassetti-Biscari quadrupole when $a = .2$ and $L = 1$.

Since it is assumed that both the current and charge density are time independent, it follows from charge conservation that the current $\boldsymbol{J}$ must be divergence free,

$$\nabla \cdot \boldsymbol{J} = -\partial_t \rho_e = 0. \tag{49}$$

For our purposes we therefore need a way to specify a *divergence free* current $\boldsymbol{J}$ on the surface of a cylinder of radius a.

Consider the vector field $\boldsymbol{F}(\rho, \phi, z)$ defined by

$$\boldsymbol{F}(\rho, \phi, z) = \boldsymbol{e}_\rho \, \delta(\rho - a) V(\phi, z) \tag{50}$$

where V is a function to be determined. Make the Ansatz

$$\boldsymbol{J} = \nabla \times \boldsymbol{F} \tag{51}$$

to find the results

$$J_\rho = 0, \tag{52}$$

$$J_\phi = \delta(\rho - a)\partial_z V, \tag{53}$$

$$J_z = \delta(\rho - a)(-1/\rho)\partial_\phi V = \delta(\rho - a)(-1/a)\partial_\phi V. \tag{54}$$

Evidently $\boldsymbol{J}$ has support only on the cylinder and is tangential to the cylinder. Let us compute $\nabla \cdot \boldsymbol{J}$. We find

$$\nabla \cdot \boldsymbol{J} = (1/\rho)\partial_\phi J_\phi + \partial_z J_z = [(1/\rho)\delta(\rho - a)][\partial_\phi \partial_z - \partial_z \partial_\phi] V = 0, \tag{55}$$

as desired and expected from the Ansatz (51) because the divergence of the curl of any sufficiently differentiable vector field is zero.

Let us write $\boldsymbol{J}$ in the form

$$\boldsymbol{J} = \delta(\rho - a)\boldsymbol{j} \tag{56}$$

where

$$j_\rho = 0, \tag{57}$$

$$j_\phi = \partial_z V, \tag{58}$$

$$j_z = (-1/a)\partial_\phi V. \tag{59}$$

Next introduce quantities Q, P, and H by the rules

$$Q = z/a \Leftrightarrow z = Qa, \tag{60}$$

$$P = \phi, \tag{61}$$

and

$$H(Q,P) = -(1/a)V(P,Qa) = -(1/a)V(\phi,z). \tag{62}$$

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

$$\dot{Q} = \partial H/\partial P \tag{63}$$

and

$$\dot{P} = -\partial H/\partial Q \tag{64}$$

where a dot denotes $(d/d\lambda)$.

Let us see if these Ansätze are sane. From (63), (62), (61), and (59) we find

$$\dot{Q} = \partial H/\partial P = -(1/a)\partial_\phi V = j_z. \tag{65}$$

From (64), (62), (60), and (58) we find

$$\dot{P} = -\partial H/\partial Q = \partial_z V = j_\phi. \tag{66}$$

We see that the flow lines in the Q,P plane generated by H are along the current $\boldsymbol{j}$. Put another way, the tangent vector to the curve $Q(\lambda), P(\lambda)$ is $\boldsymbol{j}$.

Moreover, since H does not depend on λ, it is conserved. It follows that the level lines of H coincide with the flow lines, and hence are also along $\boldsymbol{j}$. We will call $H(Q,P)$ the *canonical stream function*.

Since the current $\boldsymbol{J}$ lies on a cylindrical surface, the function V must be periodic in ϕ with period 2π. It therefore has a Fourier expansion of the form

$$V(\phi,z) = f_0(z) + \sum_{m=1}^{\infty} [f_{ms}(z)\sin(m\phi) + f_{mc}(z)\cos(m\phi)]. \tag{67}$$

We may treat each of the terms in this expansion *separately*. The functions $f_{ms}(z)$ and $f_{mc}(z)$, are called the normal and skew multipole *shape* functions. As the name is meant to imply, they control in part the shape of each coil. We expect (and it can be shown) 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. For simplicity, we will describe only the normal multipole cases with $m > 0$. The case $m = 0$ is that of a solenoid, and can be handled using more elementary methods. The skew cases can easily be related to the normal cases by suitable rotations about the z axis.

With considerable calculation, involving hypergeometric functions or Gegenbauer polynomials, it can be shown that the $C^{[k]}_{m,s}(z)$, the on-axis gradients and their z derivatives, are given in terms of the shape functions $f_{ms}(z)$ by the relations

$$C^{[k]}_{m,s}(z) = -\mu_0(1/m)^2(1/2)^{m+1}[(2m-1)!!/(m-1)!]a^m \int_{-\infty}^{+\infty} dz' f_{ms}(z')\partial_z^k \Lambda(m; z, z'). \tag{68}$$

Here we have introduced/defined what we may call an integration *kernel* function $\Lambda(m; z, z')$ by the rule

$$\Lambda(m; z, z') = [-(m+1)a^2 + m(z-z')^2]/[a^2 + (z-z')^2]^{(m+3/2)}. \tag{69}$$

c. Suppose iron is present (as well as current carrying coils and/or REC material). In that case the magnetic field $\boldsymbol{B}$ can be found numerically at points on a 3-D grid using an electromagnetic code. With a fine enough grid, this data is sufficient to compute individual trajectories. However, to compute the associated transfer maps $\mathcal{M}$, high spatial derivatives of the field are also required. How can these derivatives be obtained to satisfactory accuracy?

It is well known that it is generally difficult to extract reliable information about derivatives from numerical data on a grid. One might imagine fitting the data over some volume to a multivariable polynomial or some other "fitting" function and then differentiating this polynomial/function. However differentiation amplifies whatever truncation and numerical noise that may be present in the data, and should be avoided wherever possible.

Remarkably, this problem can be overcome to some satisfactory aberration order with the use of *surface* data. We will fit field data onto some "imagined" surface, and then use this surface data to compute interior fields. See Figure 4. Specifically, and in summary, 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 see 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.

As an example, 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 5, which illustrates the case of a wiggler.

Suppose the magnetic field $\boldsymbol{B}(x, y, z)$ 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 $\boldsymbol{B}(x, y, z)$ *normal* to the surface. From this known function form the functions $\hat{B}^c_\rho(R, m, k)$ and $\hat{B}^s_\rho(R, m, k)$ by the rules

$$\hat{B}^c_\rho(R, m, k) = [1/(2\pi)]^2 \int_{-\infty}^{\infty} dz \exp(-ikz) \int_0^{2\pi} d\phi \cos(m\phi) B_\rho(R, \phi, z), \tag{70}$$

and

$$\hat{B}^s_\rho(R, m, k) = [1/(2\pi)]^2 \int_{-\infty}^{\infty} dz \exp(-ikz) \int_0^{2\pi} d\phi \sin(m\phi) B_\rho(R, \phi, z), \tag{71}$$

where the indicated integrations may be performed in either order.

With this definition in hand, we are ready to state the desired results: It can be shown (for $m \geq 1$) that there are the results

$$C^{[n]}_{m,\alpha}(z) = i^n (1/2)^m (1/m!) \int_{-\infty}^{\infty} dk [k^{n+m-1}/I'_m(kR)] \hat{B}^\alpha_\rho(R, m, k) \exp(ikz) \tag{72}$$

where $\alpha = s$ or c. [This formula could also be used for $m = 0$ (solenoidal) case, but in that case it is better to employ an analogous formula involving tangential field data because that is then the dominant field component.]

We have found an expression for the on-axis gradients in terms of data for the normal component of the magnetic field on the surface of the cylinder. Equation (72) may be viewed as the convolution of Fourier surface data $\hat{B}^\alpha_\rho(R, m, k)$ with the *inverse Laplacian* kernel $[k^{n+m-1}/I'_m(kR)]$. Moreover, this kernel has a very desirable property. The Bessel functions $I'_m(kR)$ have the asymptotic behavior

$$|I'_m(kR)| \sim \exp(|k|R)/\sqrt{2\pi|k|R} \text{ as } |k| \to \infty. \tag{73}$$

Since $I'_m(kR)$ appears in the denominator of (72), 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 $\hat{B}^\alpha_\rho(R, m, k)$. But

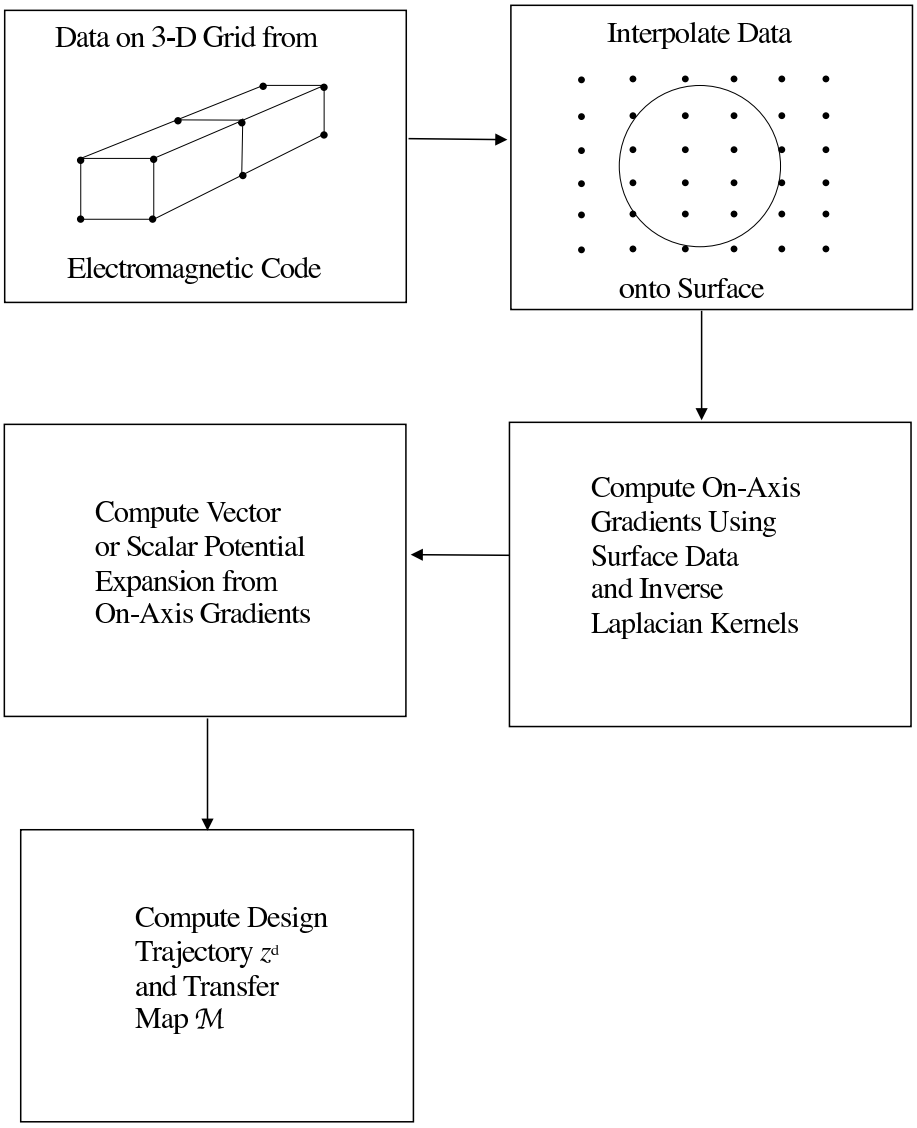

Figure 4: Calculation of realistic design trajectory ζ^d and its associated realistic transfer map $\mathcal{M}$ based on data provided on a 3-dimensional grid for a real beam-line element. Only a few points on the 3-dimensional grid are shown. In this illustration, data from the 3-dimensional grid is 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.

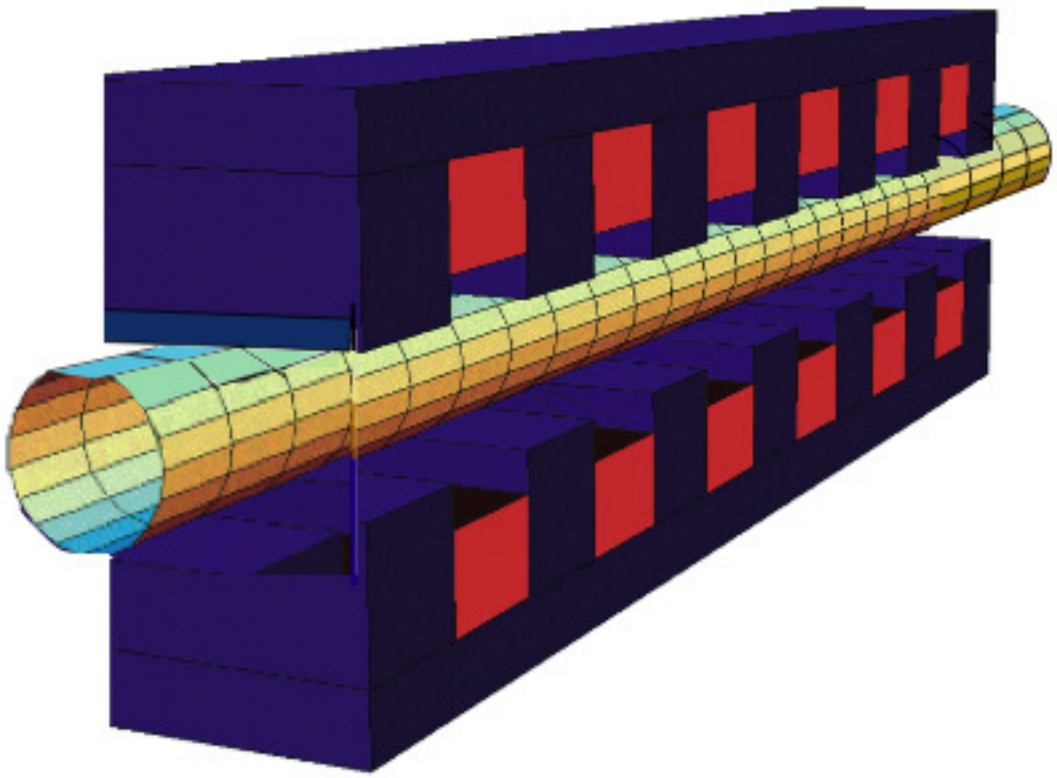

Figure 5: An imagined circular cylinder of radius R, centered on the z-axis, fitting within the bore of a beam-line element, in this case a wiggler, and extending beyond the fringe-field regions at the ends of the beam-line element.

because of the exponential damping arising from $I'_m(kR)$ in the denominator, the effect of this noise is effectively smoothed/filtered out. Moreover, this smoothing action is improved by making R as large as possible.

12. GENMAP Equations for $\mathcal{M}$. Item 11 described how on-axis gradients can be computed for magnetic beam-line elements when the magnetic field is generated by specified REC material and specified coil windings without and with the presence of iron. This item will describe how the *transfer maps* can be computed for these beam-line elements in terms of their associated on-axis gradients. The routine for this purpose is called GENMAP because, given on-axis gradients $C^{[n]}_{m,\alpha}(z)$, they *generate* the associated transfer map $\mathcal{M}$.

    From now on we will use the symbol ζ to denote phase-space variables. Hamilton's equations of motion for a particle moving in phase space read

$$\dot{\zeta} = -[H, \zeta] =: -H : \zeta \tag{74}$$

    where a dot denotes differentiation with respect to the independent variable which we take to be the Cartesian coordinate z. It can be shown that these equations of motion, when combined with the definition (8), yield for $\mathcal{M}$ the operator equation of motion

$$\dot{\mathcal{M}} = \mathcal{M} : -H : . \tag{75}$$

    a. Computation of vector potential. The Hamiltonian description of charged-particle motion in an electromagnetic field requires use of a vector potential $\boldsymbol{A}$. Vector potentials, in turn, can be expressed in terms of the scalar potential ψ, and depend as well on the choice of gauge. Initially, at least, we have found it useful to employ what we call the *symmetric Coulomb* gauge, which is uniquely defined. We will denote this vector potential by the symbol $\hat{\boldsymbol{A}}$. It is given in terms of the on-axis gradients by the relations

$$\hat{\boldsymbol{A}} = \hat{\boldsymbol{A}}^0 + \sum_{m=1}^{\infty} \hat{\boldsymbol{A}}^{m,c} + \sum_{m=1}^{\infty} \hat{\boldsymbol{A}}^{m,s} \tag{76}$$

    where

$$\begin{aligned} \hat{A}^0_x &= -(y/2) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{1}{2^{2\ell} \ell! (\ell+1)!} C_0^{[2\ell+1]}(z)(x^2+y^2)^\ell \\ &= -(y/2)[C_0^{[1]} - (1/8) C_0^{[3]}(x^2+y^2) + \cdots], \end{aligned} \tag{77}$$

$$\begin{aligned} \hat{A}^0_y &= (x/2) \sum_{\ell=0}^{\infty} (-1)^\ell \frac{1}{2^{2\ell} \ell! (\ell+1)!} C_0^{[2\ell+1]}(z)(x^2+y^2)^\ell \\ &= (x/2)[C_0^{[1]} - (1/8) C_0^{[3]}(x^2+y^2) + \cdots], \end{aligned} \tag{78}$$

$$\hat{A}^0_z = 0; \tag{79}$$

$$
\begin{aligned}
\hat{A}_x^{m,c} &= -(1/2)[(\cos\phi)(\sin m\phi) + (\sin\phi)(\cos m\phi)] \times \\
&\quad \sum_{\ell=0}^{\infty}(-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(z)\rho^{2\ell+m+1} \\
&= -(1/2)[\sin(m+1)\phi] \sum_{\ell=0}^{\infty}(-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(z)\rho^{2\ell+m+1}, \\
&= -(1/2)\Im[(x+iy)^{m+1}] \sum_{\ell=0}^{\infty}(-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(z)(x^2+y^2)^\ell, \qquad (80)
\end{aligned}
$$

$$
\begin{aligned}
\hat{A}_y^{m,c} &= (1/2)[-(\sin\phi)(\sin m\phi) + (\cos\phi)(\cos m\phi)] \times \\
&\quad \sum_{\ell=0}^{\infty}(-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(z)\rho^{2\ell+m+1} \\
&= (1/2)[\cos(m+1)\phi] \sum_{\ell=0}^{\infty}(-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(z)\rho^{2\ell+m+1}, \\
&= (1/2)\Re[(x+iy)^{m+1}] \sum_{\ell=0}^{\infty}(-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,c}^{[2\ell+1]}(z)(x^2+y^2)^\ell, \qquad (81)
\end{aligned}
$$

$$
\begin{aligned}
\hat{A}_z^{m,c} &= (\sin m\phi) \sum_{\ell=0}^{\infty}(-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m)!} C_{m,c}^{[2\ell]}(z)\rho^{2\ell+m} \\
&= \Im[(x+iy)^m] \sum_{\ell=0}^{\infty}(-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m)!} C_{m,c}^{[2\ell]}(z)(x^2+y^2)^\ell; \qquad (82)
\end{aligned}
$$

$$
\begin{aligned}
\hat{A}_x^{m,s} &= (1/2)[(\cos\phi)(\cos m\phi) - (\sin\phi)(\sin m\phi)] \times \\
&\quad \sum_{\ell=0}^{\infty}(-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z)\rho^{2\ell+m+1} \\
&= (1/2)[\cos(m+1)\phi] \sum_{\ell=0}^{\infty}(-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z)\rho^{2\ell+m+1} \\
&= (1/2)\Re[(x+iy)^{m+1}] \sum_{\ell=0}^{\infty}(-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z)(x^2+y^2)^\ell, \qquad (83)
\end{aligned}
$$

$$\begin{aligned}
\hat{A}_y^{m,s} &= (1/2)[(\sin\phi)(\cos m\phi) + (\cos\phi)(\sin m\phi)] \times \\
&\quad \sum_{\ell=0}^{\infty}(-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z)\rho^{2\ell+m+1} \\
&= (1/2)[\sin(m+1)\phi] \sum_{\ell=0}^{\infty}(-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z)\rho^{2\ell+m+1} \\
&= (1/2)\Im[(x+iy)^{m+1}] \sum_{\ell=0}^{\infty}(-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m+1)!} C_{m,s}^{[2\ell+1]}(z)(x^2+y^2)^\ell, \qquad (84)
\end{aligned}$$

$$\begin{aligned}
\hat{A}_z^{m,s} &= -(\cos m\phi) \sum_{\ell=0}^{\infty}(-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m)!} C_{m,s}^{[2\ell]}(z)\rho^{2\ell+m} \\
&= -\Re[(x+iy)^m] \sum_{\ell=0}^{\infty}(-1)^\ell \frac{m!}{2^{2\ell}\ell!(\ell+m)!} C_{m,s}^{[2\ell]}(z)(x^2+y^2)^\ell. \qquad (85)
\end{aligned}$$

It can be verified that the vector fields $\hat{\boldsymbol{A}}^0$, $\hat{\boldsymbol{A}}^{m,c}$, and $\hat{\boldsymbol{A}}^{m,s}$ are divergence free and therefore in the Coulomb gauge. Moreover they satisfy the relations

$$\nabla \times \hat{\boldsymbol{A}}^0 = -\nabla\psi_0, \qquad (86)$$

$$\nabla \times \hat{\boldsymbol{A}}^{m,c} = -\nabla\psi_{mc}, \qquad (87)$$

$$\nabla \times \hat{\boldsymbol{A}}^{m,s} = -\nabla\psi_{ms}, \qquad (88)$$

and therefore produce the desired magnetic fields.

In this context, what is the meaning of the adjective"symmetric"? In the case of $\hat{\boldsymbol{A}}^0$ observe that the coefficients of x and $-y$ are the *same*. In the cases of $\hat{\boldsymbol{A}}^{m,\alpha}$ observe that the coefficients of the angular factors in the expressions for the x and y components are the same.

b. Computation of H. When the z coordinate is taken to be the independent variable, charged particle motion in a magnetic field associated with $\hat{\boldsymbol{A}}$ is described by a Hamiltonian, call it K, of the form

$$K = -[(p_t)^2/c^2 - m^2c^2 - (p_x - q\hat{A}_x)^2 - (p_y - q\hat{A}_y)^2]^{1/2} - q\hat{A}_z. \qquad (89)$$

At this point one may introduce suitably scaled deviation variables to obtain H in terms of the variables. Doing so produces a Hamiltonian H of the form

$$H = H_1 + H_2 + H_3 + \cdots \qquad (90)$$

where the H_n are homogeneous polynomials of degree n in the deviation variables. If $\hat{\boldsymbol{A}}$ has no $m = 1$ (dipole) terms, then it can be shown that

$$H_1 = 0. \qquad (91)$$

Corresponding, the design orbit will then lie on the z axis.

For simplicity of presentation, we will limit our further discussion to this case, which embraces solenoids, quadrupoles, sextuples, and all higher-order multipoles.

c. Computation of $\mathcal{M}$. Our goal in this case will be to find $\mathcal{M}$ in terms of H. Begin by making for $\mathcal{M}$ the factorization Ansatz

$$\mathcal{M} = \mathcal{M}_3 \mathcal{M}_4 \cdots \mathcal{R} \tag{92}$$

where, for $n \geq 3$,

$$\mathcal{M}_n = \exp(: g_n :). \tag{93}$$

Then it can be shown that R, the matrix associated with $\mathcal{R}$, satisfies the differential equation

$$\dot{R} = JSR \tag{94}$$

where S is the symmetric matrix associated with the quadratic polynomial H_2,

$$H_2(\zeta; z) = (1/2) \sum_{a,b} S_{ab}(z) \zeta_a \zeta_b. \tag{95}$$

The set of differential equations (94) is to be integrated (generally numerically) from z^{in} to z^{fin} with the initial conditions

$$R(z^{in}) = I. \tag{96}$$

Here z^{in} is the z coordinate where the beam-line element is deemed to begin, and z^{fin} is the z coordinate where the beam-line element is deemed to end, including leading and trailing fringe-field regions.

What remains is to find the polynomials g_n. Define *interaction picture* Hamiltonians by the rule

$$H_n^{int}(\zeta) = H_n(R\zeta). \tag{97}$$

Then it can be shown that the g_n satisfy the differential equations

$$\dot{g}_3 = -H_3^{\rm int}, \tag{98}$$

$$\dot{g}_4 = -H_4^{\rm int} + (: g_3 : /2)(-H_3^{\rm int}), \tag{99}$$

$$\dot{g}_5 = -H_5^{\rm int} - (1/6) : g_3 :^2 (-H_3^{\rm int}) + : g_4 : (-H_3^{\rm int}), \tag{100}$$

$$\begin{aligned} \dot{g}_6 \;=\; & -H_6^{\rm int} + (1/24) : g_3 :^3 (-H_3^{\rm int}) + (1/2) : g_4 : (-H_4^{\rm int}) \\ & -(1/4) : g_4 :: g_3 : (-H_3^{\rm int}) + : g_5 : (-H_3^{\rm int}), \end{aligned} \tag{101}$$

$$\begin{aligned} \dot{g}_7 \;=\; & -H_7^{\rm int} - (1/120) : g_3 :^4 (-H_3^{\rm int}) + (1/6) : g_4 :: g_3 :^2 (-H_3^{\rm int}) \\ & -(1/2) : g_4 :^2 (-H_3^{\rm int}) + : g_5 : (-H_4^{\rm int}) + : g_6 : (-H_3^{\rm int}), \end{aligned} \tag{102}$$

$$\begin{aligned}\dot{g}_8 &= -H_8^{\rm int} + (1/720) : g_3 :^5 (-H_3^{\rm int}) - (1/24) : g_4 :: g_3 :^3 (-H_3^{\rm int}) \\ &\quad -(1/6) : g_4 :^2 (-H_4^{\rm int}) + (1/6) : g_4 :^2 : g_3 : (-H_3^{\rm int}) + (1/2) : g_5 : (-H_5^{\rm int}) \\ &\quad +(1/12) : g_5 :: g_3 :^2 (-H_3^{\rm int}) - (1/2) : g_5 :: g_4 : (-H_3^{\rm int}) + : g_6 : (-H_4^{\rm int}) \\ &\quad + : g_7 : (-H_3^{\rm int}), \end{aligned} \tag{103}$$

$$\dot{g}_m = \text{ expression involving } H_m^{\rm int} \text{ and the } g_\ell \text{ and } H_\ell^{\rm int} \text{ with } \ell < m. \tag{104}$$

These equations are be integrated from z^{in} to z^{fin} with the initial condition

$$g_n(\zeta; z^{in}) = 0. \tag{105}$$

Note that they are *explicit* because of (104). Together the equations (94) through (105) are called the GENMAP equations.

There is one last step required to bring the result for $\mathcal{M}$ to standard factorized product form. Of course, one may simply apply the concatenation procedure to (92) and take note of the result. However, one may also obtain insight by some Lie-algebraic manipulation: Insert factors of $\mathcal{R}\mathcal{R}^{-1} = \mathcal{I}$ into (92) to find the result

$$\begin{aligned}\mathcal{M} &= (\mathcal{R}\mathcal{R}^{-1})\mathcal{M}_3(\mathcal{R}\mathcal{R}^{-1})\mathcal{M}_4 \cdots (\mathcal{R}\mathcal{R}^{-1})\mathcal{R} \\ &= \mathcal{R}(\mathcal{R}^{-1}\mathcal{M}_3\mathcal{R})(\mathcal{R}^{-1}\mathcal{M}_4\mathcal{R}) \cdots . \end{aligned} \tag{106}$$

With the aid of (93) and standard Lie-algebraic manipulations we find

$$\mathcal{R}^{-1}\mathcal{M}_\ell\mathcal{R} = \mathcal{R}^{-1} \exp(: g_\ell :)\mathcal{R} = \exp(: f_\ell :) \tag{107}$$

where

$$f_\ell(\zeta) = g_\ell(R^{-1}\zeta). \tag{108}$$

Consequently, there is the result

$$\mathcal{M} = \mathcal{R} \exp(: f_3 :) \exp(: f_4 :) \cdots \tag{109}$$

with the f_ℓ given in terms of the g_ℓ by (108).

d. Gauge Fixing. We have been using a vector potential in the symmetric Coulomb gauge. Does the choice of gauge matter? It does not if the vector potential is essentially zero at the ends $z = z^{in}$ and $z = z^{fin}$. But in general the choice of gauge does matter. In reality fringe fields, and therefore also vector potentials, never completely vanish at the ends of real beam-line elements. Moreover, without the danger of violating the Maxwell equations, one cannot arbitrarily set $\boldsymbol{B}$ to zero at the ends. But at the expense of introducing fictitious currents at the ends, one can set the vector potential to zero at the ends. (That, in effect, is what happens if one simply integrates the GENMAP equations from z^{in} to z^{fin}.) These currents will be proportional to the actual values of the vector potential at the ends. This observation indicates that in the vicinity of the ends one should employ, consistent with the $\boldsymbol{B}$ fields at the ends, a vector potential that is as *small* as possible.

Remarkably, there is a *unique* choice of gauge that, for a given magnetic field which the vector potential is required to produce, *minimizes* the vector potential. It is the Poincaré-Coulomb gauge. Let $\boldsymbol{A}^{PC}$ be the vector potential in this choice of gauge. In addition to

being divergence free, $\nabla \cdot \boldsymbol{A}^{PC} = 0$, (and producing the desired magnetic field), it satisfies the complementary condition

$$\boldsymbol{r} \cdot \boldsymbol{A}^{PC}(\boldsymbol{r}) = 0 \tag{110}$$

where $\boldsymbol{r}$ is measured from some expansion point, around which $\boldsymbol{A}^{PC}$ is constructed based on the Taylor expansion of $\boldsymbol{B}$ about this point.

Here is the *optimal* strategy for maintaining the fiction of separate function beam-line elements (the case where fringe fields of adjacent elements do not overlap) while minimizing the error associated with this fiction: In the vicinity of the design orbit where $z = z^{in}$ and integration is supposed to begin, compute the vector potential in the Poincaré-Coulomb gauge and also in any other convenient gauge that is to be used for integration over the full orbit. These two vector potentials are related by a gauge transformation. It can be verified that a gauge transformation is a symplectic map, which in this case we will call $\mathcal{M}^{in}$. Similarly, in the vicinity of the design orbit where $z = z^{fin}$ and integration is supposed to end, again compute the vector potential in the Poincaré-Coulomb gauge and also in the gauge that is to be used for integration over the full orbit. These two vector potentials are also related by a gauge transformation for which there is a symplectic map $\mathcal{M}^{fin}$. Then there is an optimal map for the beam-line element, call it $\mathcal{M}^{opt}$, for maintaining the separate-function fiction and minimizing the error associated with truncating the vector potential used for integration over the full orbit. It is given by

$$\mathcal{M}^{opt} = \mathcal{M}^{in}\mathcal{M}\mathcal{M}^{fin}. \tag{111}$$

This is a recipe for "gauge fixing", and yields for $\mathcal{M}^{opt}$ a result that is *independent* of the choice of gauge employed in computing $\mathcal{M}$.

13. An Unsolved Problem. Items 10 through 12 treated the case of straight beam-line elements such as a solenoid, quadrupole, sextupole, octupole, etc., or a wiggler with no net bending. And procedures were described for computing in principle accurate transfer maps that included all fringe-field and higher-order multipole effects to any desired order. All these elements have the common property that the design orbit remains close to the z axis. We expect that a cylindrical harmonic expansion will be applicable/convergent in the vicinity of the z axis. But what happens if, as is the case for dipoles, the design orbit is curved and therefore when followed contains points that are potentially far from the z axis? Straight/unbent dipoles, like all fields in a source free cylindrical region, have cylindrical harmonic expansions. But is such an expansion sufficiently global?

    We observe that a cylindrical harmonic expansion is essentially a power series expansion in the variables x and y with z dependent coefficients. [These coefficients depend on the $C^{[n]}_{m\alpha}(z)$.] And from the theory of several complex variables we may expect that the domain of convergence in the x, y plane will depend on z. And indeed we may expect that the domain of convergence will *enlarge* when z moves evermore into a fringe-field region because such points are farther away from field sources. But is the rate of enlargement sufficient for the domain of convergence to include the design orbit as it moves away from the z axis?

    We know that the domain of convergence of a cylindrical harmonic expansion is independent of the field strength. But the distance the design orbit ultimately moves away from the z axis

depends on the field strength. Therefore if the field strength at its maximum is small (which results in less bending), we might hope that the design orbit (which is essentially straight in the edges of fringe-field regions where the field is sufficiently weak) will remain within the convergence domain all the way to the point where the field strength is negligible. Such is presumably the case for relatively weak steering magnets. But in principle this has to be verified for any proposed dipole with some specific geometry and design net bending angle.

We remark that in the case of the Large Hadron Collider, the dipoles are initially manufactured as being straight and then measured in this condition to ensure that undesired multipole strengths are sufficiently small (at least in the integrated sense and perhaps locally). These dipoles are then *bent* to accommodate the design orbit, at which point the possibility of a cylindrical harmonic expansion is lost, and one still hopes for the best. In practice this seems to work, but it would be good to have a procedure based on the use of field data computed numerically on a grid that would predict success or failure.

To recapitulate, 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 with large bend angles, where no straight cylinder would fit within the aperture.

But it is in principle possible to employ surface methods for geometries that are not cylindrical. For such cases more complicated surfaces are required. See, for example, Figure 6 which 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.

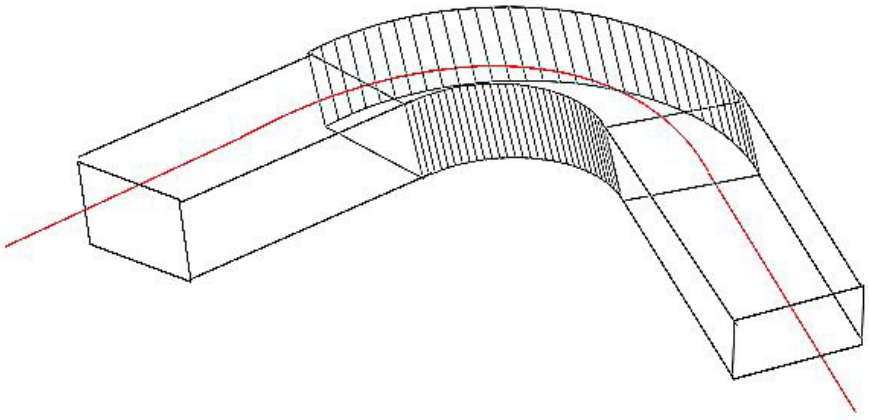

Figure 6: A bent box with straight end legs.

But now there is a complication: The *straight* cylinder methods succeeded because Laplace's equation is separable in cylindrical 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 multipoles 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 grid by some 3-dimensional field solvers, and these data can be interpolated onto the surface.

In the case of a bent box (and several other geometries) an interior vector potential can be computed in terms of the (both normal and tangential) surface field data with the use of Dirac monopoles and their associated strings. However, it is found that this vector potential has the defect that it falls off only very slowly as one leaves the box through its legs, in fact more slowly than the underlying magnetic field falls off. (By contrast, in the case of the symmetric Coulomb gauge or the Poincaré-Coulomb gauge, the vector potential falls off as the magnetic field falls off.) Therefore, for the Dirac based vector potential, which in this case is the only one for which we currently have a computation method, starting the integration at z^{in} and ending at z^{in} can produce serious error. Can this problem be overcome by gauge fixing? Perhaps, but exploration of this possibility suggests that if so the maps $\mathcal{M}^{in}$ and $\mathcal{M}^{fin}$ will involve $\exp(: f_1 :)$ factors with the f_1 large. That would also be potentially problematic due to "feed-down effects". See (7). Of course, there are crude fixes which are commonly employed. One can imagine the $\boldsymbol{B}$ field inside the dipole is constant and then rapidly goes to zero before entry and after exit, thereby producing edge focusing. But this "hard edge" approximation says little about nonlinear effects, and does not attempt to make use of actual numerical field information. Our goal of using surface data for dipoles with large sagitta has not yet been achieved, and awaits further effort.

14. Normal Form Analysis. Recall vector space theory. *Linear* transformations/operators on a vector space can be described by matrices. The matrix for any given operator depends *both* on the operator and the choice of basis for the vector space. Let M be the matrix for some given operator in some basis. Consider all matrices N of the form

$$N = AMA^{-1} \tag{112}$$

where A is any invertible matrix. These are the matrices for the same given operator when the basis vectors are changed by making a linear transformation on them using the linear operator associated with A. Select A to make N as "simple/nice" as possible. Then N is called a *normal form* for M. For example, if the eigenvalues of M are *distinct*, then A can be chosen to make N *diagonal*. If the eigenvalues are not distinct, A can be chosen to make N have *Jordan normal form*. In either case, most of the entries in N are zero. We may regard the nonzero entries in N as being *intrinsic* to the given linear operator, while most to the entries in M are due both to the intrinsic properties of the operator and a nonoptimal choice of basis.

We now seek to extend this approach from the the case of linear transformations to the case of fully nonlinear symplectic maps. Suppose $\mathcal{M}$ is the one-turn map for a ring written in the form (22). For simplicity we assume that $f_1 = 0$ and that the ring is static (no excited rf cavities). But we assure the reader that the general case can also be treated in a similar way.

By analogy to what was done in the linear case, consider all maps of the form

$$\mathcal{N} = \mathcal{A}\mathcal{M}\mathcal{A}^{-1} \tag{113}$$

where $\mathcal{A}$ is symplectic. Assume further that R, the matrix associated with $\mathcal{R}$, has all its eigenvalues on the unit circle and that the tunes associated with these eigenvalues are non-resonant. Then there is a choice of $\mathcal{A}$ such that $\mathcal{N}$ takes the (normal) form

$$\mathcal{N} = \exp(: h :) \tag{114}$$

where

$$\begin{aligned} h &= (w_x + w_x' P_t + w_x'' P_t^2 + \cdots) h_x \\ &+ (w_y + w_y' P_t + w_y'' P_t^2 + \cdots) h_y \\ &+ a h_x^2 + b h_x h_y + c h_y^2 + \cdots \\ &+ \alpha P_t^2 + \beta P_t^3 + \gamma P_t^4 + \cdots . \end{aligned} \tag{115}$$

Here h_x and h_y are the quantities

$$h_x = (1/2)(P_x^2 + X^2), \tag{116}$$

$$h_y = (1/2)(P_y^2 + Y^2). \tag{117}$$

The coefficients in (115) are intrinsic to $\mathcal{M}$ and have the following physical significance:

- w_x and w_y specify the horizontal and vertical tunes (tune=$w/2\pi$),
- w_x' and w_y' specify the first-order horizontal and vertical chromaticities (how the tunes depend on energy),
- w_x'' and w_y'' specify the second-order chromaticities, etc.,
- a, b, c, etc. specify the anharmonicities (how tunes depend on betatron amplitudes),
- α, β, γ, etc. specify linear and higher-order phase slip terms.

Note that these are all *global* properties of a ring. That is, $\mathcal{N}$ describes the global properties of a ring.

By contrast it can be shown that the entries in $\mathcal{A}$, which specify at each point in a ring the choice of basis, describe *local* properties of a ring: horizontal and vertical dispersion functions, lattice functions, and their nonlinear generalizations.

Moreover, it can be shown that the quantities I^x and I_y given by

$$I_x = \mathcal{A}^{-1} h_x \text{ and } I_y = \mathcal{A}^{-1} h_y \tag{118}$$

generalize the Courant-Snyder invariants to the nonlinear case.

Finally, we observe that the maps $\mathcal{M}$, $\mathcal{N}$, and $\mathcal{A}$ are all symplectic, are generally nonlinear, and in principle can be computed to any desired order providing the tunes are non resonant to that order. Thus computation of $\mathcal{N}$ and $\mathcal{A}$ completely extends linear theory to the full nonlinear case.

15. Bending a Beam Without Breaking It. We have seen that normal-form concepts are a nonlinear generalization of linear Courant-Snyder theory. The purpose of this item is to illustrate that the logic behind normal-form theory can also be used to design special purpose compound beam-line elements. In particular we will generalize normal-form concepts to design/construct a Perfect Third-Order Achromat. A *perfect* achromat is a collection of bean-line elements whose net effect is to change the *direction* of a beam but otherwise act as the identity map. It should not produce any linear effects (such as dispersion, focus/defocus effects, time-of-flight effects), nor any nonlinear effects through some order.

    The third-order perfect achromat we will describe uses sextupoles and octupoles and repetitive symmetry to cancel all undesirable terms through third order, and consists of 5 identical cells. Each cell has 32 elements. The content of a cell is as follows:

```
2 normal (not skew) sector bends sbend with a bending angle of 9 degrees

16 drift spaces dr with length .1 meter

3 normal quadrupoles fq, dq, and cq with length .1 meter.  Nominally meant to be
horizontally focusing, horizontally defocussing, and correcting to achieve desired
horizontal and vertical tunes and set the phase-slip factor alpha to 0.
Actually all of them contribute to all three goals.

3 normal sextupoles sext1, sext2, and sext3 with length .1 meter.

 Meant to adjust the contents of f_3.

8 normal octupoles oct1, oct2, oct3, ..., oct8 with length.1 meter.

 Meant to adjust the contents of f_4.

These 32 elements in a cell occur in the following order:

dr     oct1    dr      oct2      dr     fq       dr    sex1
dr     oct3    dr      sex2      dr     oct4     dr    sbend
dr     cq      dr      sbend     dr     sex3     dr    oct5
dr     dq      dr      oct6      dr     oct7     dr    oct8
```

    Before moving on we should say a bit more about the normal form procedure that adjusts the contents of f_3 and f_4 by making a proper choice for $\mathcal{A}$. This is done with a "purifying" algorithm, and examination of this algorithm shows that it cannot remove the entries in (115) for any choice of $\mathcal{A}$. Consequently we must adjust the sextupoles to remove the f_3 entries in (115). Note that there are 3 such terms, which is why there are 3 sextupoles in a cell. Similarly, adjust the octupoles to remove the f_4 entries in (115). Moreover, since the tunes are equal, there is the possibility of difficulty with difference resonance terms. It can be shown

that these are the two difference resonance terms listed below.

The difference resonance terms for f_4 are
$\Re[(X+iP_x)^2(Y-iP_y)^2]$ and $\Im[(X+iP_x)^2(Y-iP_y)^2]$. (119)

We will therefore adjust the octupoles to set to zero the coefficients of the elements in (119) as well. Note that in total there are 8 f_4 terms to be removed: 6 terms in (115) and the 2 difference resonance terms in (119). That is why there are 8 octupoles in a cell.

The stage is now set to proceed. Let $\mathcal{M}_c$ be the map for a cell. The steps for constructing a perfect third-order achromat are as follows:

(a) Compute $\mathcal{M}_c$ using a code that can implement Lie-algebraic concepts and also has fitting capabilities. By varying the strengths of fq, dq, and cq, fit both the horizontal and vertical tunes to 1/5 and the phase slip factor α to 0 (which is equivalent to fitting R_{56} to 0).

(b) With the proper strengths of fq, dq, and cq established and in place, vary the strengths of sext1, sext2, and sext3 to set to 0 the f_3 content of (115).

(c) With the proper strengths of sext1, sext2, and sext3 established and in place, as well, vary the strengths of oct1 through oct8, to set to 0 the f_4 content of (115). Also simultaneously set to 0 the coefficients of the difference resonance terms listed in (119).

(d) With the proper strengths of all the elements of $\mathcal{M}_c$ established and in place, recompute $\mathcal{M}_c$ and with a normal form routine for static maps verify that the normal form associated with $\mathcal{M}_c$, call it $\mathcal{N}_c$, has a 6×6 matrix part R consisting of three 2×2 matrices on the diagonal. The first two should be identical since the horizontal and vertical tunes are identical, and the third should be the 2×2 identity matrix. They should be on the diagonal because the quadrupoles fq, dq, and cq are normal quadrupoles. Finally the f_3 and f_4 generators for $\mathcal{N}_c$ should all vanish. That occurs because the normal form routine removes all the non-resonant terms, and the resonant terms (note that the tunes are the same so that some might have been expected) have been set to zero by properly setting the strengths of the quadrupoles fq, dq, and cq; the sextupoles sext1, sext2, and sext3; and the octupoles oct1 through oct8.

(e) Using the concatenator, compute $\mathcal{M}_{5\ cell}$, the 5-cell map defined by

$$\mathcal{M}_{5\ cell}=\mathcal{M}_c\mathcal{M}_c\mathcal{M}_c\mathcal{M}_c\mathcal{M}_c=\mathcal{M}_c^5. \tag{120}$$

What do we expect to find? From the established form of $\mathcal{N}_c$ we know that

$$\mathcal{N}_c^5=\mathcal{I}+\text{ possible terms with generators } f_n \text{ of order 5 and higher} \tag{121}$$

where $\mathcal{I}$ is the identity map. Here we have used the fact that the tunes associated with $\mathcal{N}_c$ have been fit to 1/5. But from (113) we see that

$$\mathcal{M}_c=\mathcal{A}^{-1}\mathcal{N}_c\mathcal{A} \tag{122}$$

and therefore

$$\begin{aligned}
\mathcal{M}_{5\ cell} &= \mathcal{M}_c^5 = [\mathcal{A}^{-1}\mathcal{N}_c\mathcal{A}][\mathcal{A}^{-1}\mathcal{N}_c\mathcal{A}][\mathcal{A}^{-1}\mathcal{N}_c\mathcal{A}][\mathcal{A}^{-1}\mathcal{N}_c\mathcal{A}][\mathcal{A}^{-1}\mathcal{N}_c\mathcal{A}] \\
&= \mathcal{A}^{-1}\mathcal{N}_c\mathcal{N}_c\mathcal{N}_c\mathcal{N}_c\mathcal{N}_c\mathcal{A} \\
&= \mathcal{A}^{-1}\mathcal{N}_c^5\mathcal{A} \\
&= \mathcal{A}^{-1}[\mathcal{I} + \cdots]\mathcal{A} \\
&= \mathcal{I} + \text{ possible terms with generators } f_n \text{ of order 5 and higher.} \qquad (123)
\end{aligned}$$

We see that the 5-cell combination is indeed a perfect third-order achromat.

And what is its net bending angle? The 5-cell combination contains $10 = 5 \times 2$ sector bend magnets, and each bends by 9 degrees. Therefore the net bending angle is $10 \times 9 = 90$ degrees. Employing the 5-cell combination is a useful way of transporting a beam around a 90 degree corner without (through third order) disturbing it in any way except for a time delay which (again through third order) is the same for all particles.

The general symplectic transfer map through third order, but without translation part, is characterized by $209 - 6 = 203$ numbers. Remarkably, by using $3 + 3 + 8 = 14$ suitably powered correction elements/cell and using the repetition symmetry of 5 identical cells, we have produced a map which has *all* entries zero save for the 6 ones on the diagonal of the matrix part of the map.

16. Fast Long-Term Tracking. Successful accelerator design requires the accurate calculation of particle orbits, often over long periods of time. The most computationally challenging are orbits in *storage rings*, essentially circular machines in which particles continually circulate for very long periods of time. Electron (or positron) storage rings are used to produce intense high-energy X-rays. Electron (or positron) storage rings, as well as proton (or antiproton or ion) storage rings, are used in pairs to produce colliders. In a collider, the beam from one ring collides head-on with the counter-circulating beam in a second ring.

    For example, the Large Hadron Collider (LHC) in CERN (near Geneva, Switzerland), which collides protons on protons, has a circumference of 27 km, and each ring has some 19200 elements (magnets, rf cavities, and intervening drift spaces). Protons moving at essentially the speed of light are stored for about 8 hours, and during this time make approximately 4×10^8 turns around the ring, approximately 8×10^{12} element passages, and approximately 2×10^{10} (betatron) oscillations about the design orbit. Following this number of oscillations is comparable to following the earth's orbit about the sun from the time of the Big Bang. While the number of betatron oscillations in electron (or positron) storage rings is comparable, they need not be followed for as long because those oscillations are damped by the energy loss associated with X-ray emission. Correspondingly, electron/positron storage-ring orbits are less computationally challenging than proton/antiproton/ion storage-ring orbits. This Item is devoted to the most challenging problem of calculating/simulating particle orbits over long periods of time in proton (or antiproton or ion) storage rings.

    With z as the independent variable, the Hamiltonian for charged-particle motion in an electromagnetic field is given by the relation

$$K = -[(p_t + q\psi)^2/c^2 - m^2c^2 - (p_x - qA_x)^2 - (p_y - qA_y)^2]^{1/2} - qA_z. \qquad (124)$$

Here the quantities p_x and p_y denote *canonical* momenta. Note that p_t is usually negative.

$$p_t = -[m^2c^4 + c^2(\boldsymbol{p}^{\rm mech} \cdot \boldsymbol{p}^{\rm mech})]^{1/2} - q\psi = -\gamma mc^2 - q\psi. \tag{125}$$

What we would like to do is to integrate trajectories generated by the Hamiltonian K given by (124) for a very large number of turns in such a way that the symplectic condition is honored.

Can we use symplectic integration? The usual symplectic integration algorithms only work for Hamiltonians of the form $T(p) + V(q)$, and K is not of this form unless $A_x = 0$ and $A_y = 0$ and $\psi = 0$. We may be willing to assume that there is no static electric field so that the electric scalar potential ψ vanishes. But we should not be so willing to assume that $A_x = 0$ and $A_y = 0$. To do so ignores potentially important fringe-field effects. Look at the relations (80) and (83) for A_x and the similar relations for A_y. Note that they involve the quantities $C_{m,\alpha}^{[2\ell+1]}(z)$, and we have seen (because $[2\ell + 1] > 0$) that these quantities are large in the fringe-field regions at magnet ends.

True, there is a symplectic integrator for the 8-dimensional phase-space formulation of charged-particle motion. But this integrator does not preserve particle mass, which would seem to be a good thing to do. It only preserves particle mass to good precision if the step size is sufficiently small, which means that in this case symplectic integration through even a single element may be quite slow.

In the past (and even often now) the Hamiltonian (124) is employed with $A_x = A_y = 0$, in which case fringe field effects are neglected. Also, what we may call the simplest symplectic integrator is often employed. The effects of drifts, dipoles, and quadrupoles are treated in linear matrix approximation. And the effects of sextupoles and higher multipoles (and sometimes as well multipole errors in dipoles and quadrupoles) are treated as single (and sometimes multiple) kicks. This approach is often referred to as *direct*/"*exact*"tracking, although what it actually does is equivalent to implementing a relatively crude, but exactly symplectic, one-turn map. It is relatively slow (for each turn) because of the large number of beam-line elements encountered in going around a large ring.

What we wish to describe in this Item are two methods that employ maps, are faster, and can potentially include fringe-field effects. As will be explained, both are the result considering what may be called the problem of the symplectic completion of symplectic jets. One method employs*generating functions*, and the other employs what are called *Cremona* maps. Since generating function or Cremona tracking begins with *symplectic jets* (to be defined shortly), and these jets can in principle be computed for *realistic* electromagnetic fields, generating function or Cremona tracking can in principle be expected to give more accurate results for realistic machines. However, in order to assure the Accelerator Physics community of their reliability, these methods should also be able to reproduce the results of direct tracking. That is, based on the assumptions made for direct tracking, relatively crude but symplectic maps can be computed for each beam-line element, and these maps can be concatenated to form symplectic maps for lumps (collections of elements) or full one-turn maps. Tracking results using these maps can be compared with those obtained by direct tracking.

Internal consistency tests can also be applied to map methods. The slowest, but presumably most accurate, procedure would be use element maps to track element-by-element. A faster procedure would be to track lump-by-lump using maps for lumps. Its accuracy could be checked by comparing its result to element-by-element tracking results. Even faster and more

daring would be to perform full turn-by-turn tracking using the full one-turn map. Whatever method is selected, it can be applied repeatedly a large number of times to simulate the effect of a large number of turns while being exactly symplectic (to machine precision) and having accuracy through the order of the maps. Since evaluation of the action of maps is fairly fast, with the use of lump or full one turn maps it is possible to track for a relatively large number of turns with relatively modest use of computer time.

Let $\mathcal{M}$ be the one-turn map for a circular machine and suppose we wish to apply $\mathcal{M}$ repeatedly to study long-term tracking behavior. Write $\mathcal{M}$ in the form

$$\mathcal{M} = \exp(: f_1 :)\mathcal{R}\mathcal{N} \tag{126}$$

where here

$$\mathcal{N} = \exp(: f_3 :) \exp(: f_4 :) \cdots . \tag{127}$$

See (22). We will call $\mathcal{N}$ the *nonlinear* part of $\mathcal{M}$.

There is no difficulty in evaluating the quantities $\exp(: f_1 :)\mathcal{R}\zeta_a$ to essentially machine precision since so doing involves simple vector addition and the multiplication of a vector by a matrix. Correspondingly, the symplectic condition is also honored essentially to machine precision. But evaluating the quantities $\mathcal{N}\zeta_a$ in a way that maintains the symplectic condition is much more challenging because doing so to machine precision involves evaluation of *infinite* (very *high* order) series. See (15) and (16).

Suppose we know the f_n in (127) for $n \le maxn$. Define a *truncated* map by the rule

$$\mathcal{N}^{trunc} = \exp(: f_3 :) \cdots \exp(: f_{maxn} :). \tag{128}$$

Consider the polynomials $h^a(\zeta)$ defined by

$$h^a(\zeta) = \mathcal{N}^{trunc}\zeta_a = \zeta_a + h_2^a(\zeta) + \cdots + h_{maxn-1}^a(\zeta) + \text{ higher degree terms.} \tag{129}$$

As the notation in (129) is meant to indicate, there are higher-degree terms in (129). However, these terms would depend on the $f_{n>maxn}$ had we made the truncation at a point farther down the line in the product (127). By contrast, the polynomials displayed in (129) are not dependent on the $f_{n>maxn}$ had we made the truncation at a point farther down the line. The set of polynomials $h_{trunc}^a(\zeta)$ defined by

$$h_{trunc}^a = \zeta_a + h_2^a(\zeta) + \cdots + h_{maxn-1}^a(\zeta) \tag{130}$$

is called the *symplectic jet* associated with $\mathcal{N}^{trunc}$. It is specified by the $f_{\le maxn}$, and contains just the amount of information required to reconstruct them. Calculation shows that they satisfy the Poisson bracket relations

$$[h_{trunc}^a(\zeta), h_{trunc}^b(\zeta)] = J_{ab} + \text{ terms of degree } (maxn - 1) \text{ and higher.} \tag{131}$$

A symplectic jet map is generally not a symplectic map, but can fail to be so only by terms of degree $(maxn - 1)$ and higher.

We are now prepared to discuss symplectification of symplectic jets. For simplicity, we will consider symplectic jets whose expansion is of the form (129). That is, their leading behavior

is the identity map. By symplectification we mean the following: Given a symplectic jet, find a symplectic map whose action is relatively easy to compute and whose Taylor expansion agrees with that of the symplectic jet through terms of degree $(maxn-1)$ and lower. We will now describe two ways of doing so.

The first involves the use of generating functions. It is generally understood (although not always sufficiently stressed in Classical Mechanics texts) that generating functions produce symplectic maps, and acquaintance with the generating functions F_1 through F_4 is part of the common lore. What is not often discussed is that, for the case of a $2n$ dimensional phase space, there is a $2n(4n+1)$ parameter family of generating functions, and F_1 through F_4 are four instances of this family. Thus, for a six dimensional phase space, there is a 78 dimensional family of generating functions.

Given this plethora of generating function types, what type shall we choose? It turns out that there is a unique one with special invariance properties. It is what is called the Poincaré generating function. Here is the setup for use of the Poincaré generating function. As before, define a vector z by the rule

$$z = (z_1, z_2, \cdots, z_{2n}) = (q_1, \cdots, q_n; p_1, \cdots, p_n). \tag{132}$$

Also define a vector Z by the rule

$$Z = (Z_1, Z_2, \cdots, Z_{2n}) = (Q_1, \cdots, Q_n; P_1, \cdots, P_n). \tag{133}$$

Regard the $q's$ and $p's$ that appear on the right side of (132) as being "*old*" variables and the $Q's$ and $P's$ that appear on the right side of (133) as being "*new*" variables. Correspondingly, regard the $z's$ as another name for old variables and the $Z's$ as another name for the new variables. In addition, let u be a set of auxiliary variables labeled by writing

$$u = (u_1, u_2, \cdots, u_{2n}). \tag{134}$$

Next define vectors Σ and Δ by the rules

$$\Sigma = Z + z \tag{135}$$

and

$$\Delta = Z - z. \tag{136}$$

Also, let $G(u)$ be any function of u, which we regard as being a *generating/source* function. Finally, for G to be employed as a Poincaré generating function, make the rule

$$\Delta = J\partial_u G|_{u=\Sigma}. \tag{137}$$

Observe that for any choice of G, (139) specifies Δ as a function of Σ,

$$\Delta = \Delta(\Sigma). \tag{138}$$

These relations, when combined with (135) and (136), provide a set of $2n$ *implicit* equations that can be solved to determine Z in terms of z. It can be shown that, for any choice of G, the resulting relation that specifies Z in terms of z is a symplectic map. For example, if

$$G = 0, \tag{139}$$

then use of (139) gives the result

$$\Delta = 0. \tag{140}$$

Then, we see from (136) that

$$Z = z, \tag{141}$$

which is the identity map.

How do we apply this machinery to the problem at hand? Make the homogeneous polynomial sum Ansatz

$$G(u) = \sum_{n=3}^{maxm} g_n(u) \tag{142}$$

where the $g_n(u)$ to be employed in (142) are to be determined. Let $\hat{\mathcal{N}}$ be the(symplectic) map that results from this choice of G. Write

$$Z = \hat{\mathcal{N}} z, \tag{143}$$

and imagine Taylor expanding in z the relation between Z and z. Require that the terms through order $(maxm - 1)$ *match* those in the jet (130). So doing specifies the $g_n(u)$. Note that the terms in the jet (130) depend in turn on the f_3 through f_{maxn} in (128). Thus, the h_n in (143) are specified by the f_n in (128). In principle one can find formulas that express the g_n in terms of the f_n, and this has been done for the first few g_n. (Analytic results are available for the g_n when $n \leq 8$.) As one might imagine, they become evermore complicated as n increases.

Remarkably such formulas are often not necessary. Suppose the f_n are known in numerical form (sums of monomials with numerical coefficients), as is the case where numerical codes are used to find the f_n. Use of (135), (136), and (142) in (137) yields implicit relations of the form

$$Z_a = z_a + \sum_{n=2}^{*} H_n^a(z + Z). \tag{144}$$

It can be shown that there is a routine for manipulating polynomials in numerical form which computes (in numerical form) the polynomials H_n^a in terms of the f_n. Thus, when the f_n are given in numerical form, one can find the H_n^a in numerical form. In summary, given a symplectic jet in numerical form, one can find [in the implicit form (144)] the symplectic map $\hat{\mathcal{N}}$. We will call this map the (Poincaré generating function) symplectification of the symplectic jet (130).

Of course, given the z as a collection of six numbers, the implicit relations (144) need to be solved numerically to find the six numbers for Z. This is best done using Newton's method starting with the initial guess $Z = z$. Note that if the one-turn map is being employed, it is only necessary to find (and store) the H_n^a once since they can be used for all subsequent turns. If lump maps are being employed, their associated H_n^a need be found and stored only once so they can be reused to compute the phase-space actions of subsequent lumps.

We still need to comment on what makes the Poincaré generating function unique. Suppose $\mathcal{L}$ is any *linear* symplectic map. Use it to transform the f_n appearing in * so that

$$f_n^{tran}(\zeta) = \mathcal{L} f_n(\zeta) = f_n(\mathcal{L}\zeta) = f_n(L\zeta) \tag{145}$$

where L is the symplectic matrix associated with $\mathcal{L}$. Make the definition

$$\mathcal{N}^{tran} = \mathcal{L}\mathcal{N}\mathcal{L}^{-1} \tag{146}$$

Observe that

$$\mathcal{N}^{tran} = \mathcal{L}\mathcal{N}\mathcal{L}^{-1} = \mathcal{L}[\exp(: f_3 :)\exp(: f_4 :)\cdots]\mathcal{L}^{-1} = \exp(f_3^{tran})\exp(f_4^{tran})\cdots . \tag{147}$$

It follows that the symplectic jet associated with $\mathcal{N}_{trunk}^{tran}$, call it $\bar{h}_{trunc}^{a}$, is given by

$$\bar{h}_{trunc}^{a}(\zeta) = \mathcal{L} h_{trunc}^{a}(\zeta) = h_{trunc}^{a}(L\zeta). \tag{148}$$

See (130). Now use the transformed jet $\bar{h}_{trunc}^{a}$ given by (148) and the Poincaré procedure to construct a a symplectic map. Call this map $\hat{\mathcal{N}}^{tran}$. Then it can be shown that

$$\hat{\mathcal{N}}^{tran} = \mathcal{L}\hat{\mathcal{N}}\mathcal{L}^{-1}. \tag{149}$$

Thus one can transform a symplectic jet using a linear map $\mathcal{L}$ and from this transformed jet construct (à la Poincaré) a symplectic map, or one can use the jet as is to construct (à la Poincaré) a symplectic map and then transform this map by sandwiching it between $\mathcal{L}$ and $\mathcal{L}^{-1}$. The results will be the same. We may say that Poincaré symplectification commutes with linear symplectic transformations, and observe that this is a desirable property.

Let us next consider the relation between Poincaré symplectification and map inversion. Suppose $\mathcal{N}$ is some nonlinear symplectic map and from it compute a symplectic jet to some order. From this jet construct (à la Poincaré) the symplectic map $\hat{\mathcal{N}}$. Also compute $\mathcal{N}^{-1}$, and then compute its associated symplectic jet to the same order. From this jet compute (à la Poincaré) a symplectic map and call it $\bar{\mathcal{N}}$. Then it can be shown that

$$\bar{\mathcal{N}} = \hat{\mathcal{N}}^{-1}. \tag{150}$$

We may say that Poincaré symplectification commutes with inversion, and observe that this is also a desirable property.

Finally, it can be shown that symplectification using any other generating function type does not have the properties (149) and (150). In this sense Poincaré symplectification is unique.

We have studied Poincaré symplectification, and have found/stated that, as far as generating function methods are concerned, it is the only method that has the two desirable properties (149) and (150). It also has other pleasant features whose discussion is beyond the scope of these notes.

We now turn to the use of Cremona maps. We define a Cremona map to be a *polynomial* map that is also symplectic. In this context, the goal is as follows: Given a symplectic jet of some order $(maxn - 1)$, try to add to it a *finite* number of terms of higher order than $(maxn - 1)$ in such a way that the resulting polynomial map is exactly symplectic. Thus, if possible, a symplectic jet is converted into a Cremona map, and by construction its jet through order

$(maxn - 1)$ is the given jet. It turns out that this always possible, and in many ways. The challenges are that the added terms should not be too numerous, so that the Cremona map should not be too much harder to numerically evaluate than the given jet map, and that the effect of the added terms is not too large over the phase-space region of interest. It turns out that all but a few of the many ways of converting a symplectic map into a Cremona fail these challenges. We also note that the symplectic maps produced from jet maps with the use of generating functions are not Cremona maps. Their Taylor expansions generally have an infinite number of terms.

A systematic method for constructing general Cremona maps begins with a large supply of simple Cremona maps. Such a supply is provided by what are called *jolt* maps. Let $g(q)$ be any polynomial in the variables q_1, through q_3. Then the map $\exp(:\, g\, :)$, called a *kick* map, is a symplectic map and has the properties

$$\exp(:\, g\, :)q_j = q_j, \tag{151}$$

$$\exp(:\, g\, :)p_j = p_j + \partial g/\partial q_j. \tag{152}$$

Evidently the map $\exp(:\, g\, :)$ is a Cremona map since $g(q)$ is a polynomial. Next, let $\mathcal{L}$ be any *linear* symplectic map. Then it can be verified that the map $\mathcal{J}$ defined by

$$\mathcal{J} = \mathcal{L}\exp(:\, g\, :)\mathcal{L}^{-1}, \tag{153}$$

and called a *jolt* map, is also a Cremona map. To continue, let $\mathcal{C}_1$ and $\mathcal{C}_2$ be any two Cremona maps. Then it is easily verified that their product $\mathcal{C}_1\mathcal{C}_2$ is also a Cremona map. It follows that any product of a finite number of Cremona maps is also a Cremona map, and any finite product of jolt maps is also a Cremona map.

Given a symplectic jet, the main task now is to find a finite collection of jolt maps $\mathcal{J}_1$, $\mathcal{J}_2\cdots$, made using various polynomials $g(q)$ and linear maps $\mathcal{L}$, such that the Cremona map $\mathcal{C}$ given by

$$\mathcal{C} = \mathcal{J}_1\mathcal{J}_2\cdots \tag{154}$$

has, for its Taylor expansion, the given symplectic jet followed by higher-order terms. If this task is accomplished, $\mathcal{C}$ is called a Cremona symplectification of the given symplectic jet. It can be shown that this task can be accomplished by an optimal choice of the various polynomials $g(q)$ and linear maps $\mathcal{L}$. This optimal choice minimizes the number of jolts needed and the size of the higher-order added terms. Construction of this optimal choice involves several technical concepts that lie beyond the scope of this outline. They include a special scalar product on the vector space of polynomials in the phase-space variables z, sensitivity vectors, the Gram matrix they produce, and cubature formulas for the manifolds $SU(2)/SO(2)$ and $SU(3)/SO(3)$.

Prior to the completion of the construction of the LHC, preliminary studies/comparisons of direct and one-turn Cremona map tracking results were carried out for various lattices and nonlinear imperfection models. They show that, even for relatively large betatron amplitudes where nonlinear effects are expected to be important, there is good agreement in dynamic aperture (phase-space region of long-term storage) between direct and one-turn Cremona map tracking results using jets based on polynomials $f_{\leq 8}$ in the standard factored product

Lie representation of the one-turn map. Moreover, in these studies Cremona tracking is approximately 20 times faster than direct tracking. And, if one wishes to simulate only orbits having smaller betatron amplitudes, amplitudes associated with normal LHC operation, then use of generators with $f_{\leq 6}$ appears to be adequate, in which case Cremona tracking is about 60 times faster than direct tracking. Also, the same Cremona map tracking speeds can be achieved to produce accurate results for realistic machines. Finally, we expect that tracking based on use of the Poincaré generating function will have similar good performance. Generating function and Cremona methods are therefore worth further study, development, and implementation.

17. The Symplectic Camel.

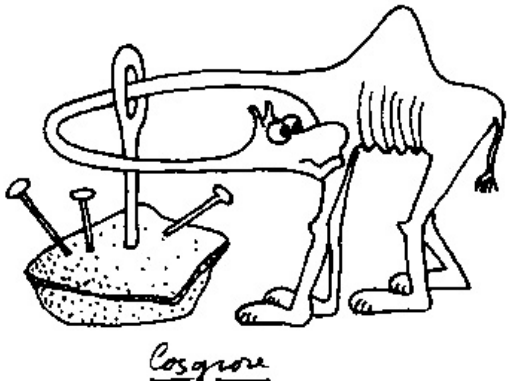

Figure 7: An *ordinary* camel and needle in R^3.

According to Liouville's theorem, if an *initial* region R^i_{2n} of phase space is sent into a *final* region R^f_{2n} of phase space under the action of a symplectic map $\mathcal{M}$, then these two regions must have the same volume. One might wonder about the converse: Given two regions of phase space having the same volume, is there a symplectic map that sends one into the other? The answer is *yes* in the case of two-dimensional phase space ($n = 1$), and, as will be done in the next Item, it is fairly easy to show that the answer is *no* in the case of four or more phase-space dimensions ($n > 1$) if one is restricted to *linear* symplectic maps. But what about the far more complicated case where nonlinear symplectic maps are allowed? The answer to this question was unknown until 1985 when *Gromov* announced his famous *nonsqueezing theorem* and its application to the *symplectic camel*.[1] The proof of his theorem is far beyond the scope of this outline and is part of the deep new field of *symplectic topology*. However, it is easy to state and understand its contents.

In the spirit of the theoretical physicist who instructed the farmer to first consider a spherical cow, the mathematician Gromov considered a spherical region in phase space, the symplectic

[1] "It is easier for a camel to go through the eye of a needle than for a rich man to enter into the kingdom of God", a saying of Jesus as quoted in Matthew 19, Mark 10, and Luke 18.

ball $B^{2n}(r)$ of radius r given by the relation

$$B^{2n}(r) = \{z \in R^{2n} \mid \sum_{j=1}^{n}(p_j^2 + q_j^2) \leq r^2\}. \tag{155}$$

(This ball is called *symplectic* because its definition involves the p_j as well as the q_j.) It is an easy calculation to show that $B^{2n}(r)$ has a finite volume $V(r)$ given by the relation

$$V(r) = r^{2n}\pi^n/[n\Gamma(n)] = r^{2n}\pi^n/n!. \tag{156}$$

Gromov also considered a symplectic cylinder $C_1^{2n}(r')$ of radius r' given by the relation

$$C_1^{2n}(r') = B_1^2(r') \times R^{2n-2}. \tag{157}$$

Here $B_1^2(r')$ is the set

$$(p_1^2 + q_1^2) \leq (r')^2 \tag{158}$$

and, according to (8.9), the remaining variables q_j and p_j for $j > 1$ are allowed to range from minus to plus infinity,

$$q_j \in (-\infty, +\infty),\ p_j \in (-\infty, +\infty) \text{ for } j \in [2, n]. \tag{159}$$

Evidently, because of (8.11), $C_1^{2n}(r')$ has infinite volume. We now ask if there is a symplectic map $\mathcal{M}$, possibly nonlinear, such that when $\mathcal{M}$ is applied to $B^{2n}(r)$ the resulting region lies within (is *embedded* in) $C_1^{2n}(r')$,

$$\mathcal{M}B^{2n}(r) \subset C_1^{2n}(r')? \tag{160}$$

Put another way, if we regard $B^{2n}(r)$ as a "symplectic camel", can this camel be squeezed into the cylinder $C_1^{2n}(r')$ under the action of some symplectic map? Liouville would not object because the volume of the cylinder, being infinite, would certainly exceed the volume of the camel. However, Gromov showed that there was no symplectic $\mathcal{M}$ that would map the camel to a region lying within the cylinder unless the radius of the cylinder equaled or exceeded that of the camel (ball),

$$r' \geq r. \tag{161}$$

A related question, more akin to passing a camel through the eye of a needle, is this: Suppose there is a camel on one side of a wall, and this wall has a hole in it. Is there a continuous family of symplectic maps $\mathcal{M}(\tau)$ such that $\mathcal{M}(0)$ is the identity map $\mathcal{I}$ and the map $\mathcal{M}(1)$ has the property that when it acts on the camel the result is a camel on the other side of the wall? Moreover, is it the case that all points obtained by letting $\mathcal{M}(\tau)$ act on the camel (for $\tau \in [0, 1]$) lie either outside the wall or within the hole in the wall?

Because we are working in dimension four or higher where our intuition may easily fail, let us phrase the question more precisely in mathematical terms. We define the wall W to be the hyperplane $q_1 = 0$,

$$W = \{z \in R^{2n} \mid q_1 = 0\}. \tag{162}$$

We define the hole in the wall, $H(r')$, to be the set

$$H(r') = \{z \in R^{2n} \mid q_1 = 0 \text{ and } \sum_{j=1}^{n}(p_j^2 + q_j^2) \leq (r')^2\}. \tag{163}$$

As for the symplectic camel, we define the two sets $B_+^{2n}(r,a)$ and $B_-^{2n}(r,a)$, with $a > r$, by the rules

$$B_+^{2n}(r,a) = \{z \in R^{2n} \mid (q_1 - a)^2 + p_1^2 + \sum_{j=2}^{n}(p_j^2 + q_j^2) \leq r^2\}, \tag{164}$$

$$B_-^{2n}(r,a) = \{z \in R^{2n} \mid (q_1 + a)^2 + p_1^2 + \sum_{j=2}^{n}(p_j^2 + q_j^2) \leq r^2\}. \tag{165}$$

Evidently $B_+^{2n}(r,a)$ is a camel centered around the point given by $q_1 = a$ with all remaining coordinates being zero, and $B_-^{2n}(r,a)$ is a camel centered around the point given by $q_1 = -a$ with all remaining coordinates being zero. And since we have assumed $r < a$, no part of either camel is in contact with the wall. Therefore $B_+^{2n}(r,a)$ is a camel located on the side of the wall W with $q_1 > 0$, and $B_-^{2n}(r,a)$ is a camel located on the side of the wall W with $q_1 < 0$.

Now suppose the camel is smaller than the hole in the wall, $r < r'$. Then it is easy to see that the camel can be moved through the hole from one side of the wall to the other by a simple translation along the q_1 axis. We can take $\mathcal{M}(\tau)$ to be of the form

$$\mathcal{M}(\tau) = \exp(2\tau a : p_1 :). \tag{166}$$

It easily verified that there are the relations

$$\mathcal{M}(0)B_+^{2n}(r,a) = B_+^{2n}(r,a), \tag{167}$$

$$\mathcal{M}(1)B_+^{2n}(r,a) = B_-^{2n}(r,a). \tag{168}$$

Moreover all the points given by

$$\mathcal{M}(\tau)B_+^{2n}(r,a) = B_+^{2n}(r, a(1-2\tau)) \tag{169}$$

with $q_1 = 0$ satisfy

$$(a(1-2\tau))^2 + p_1^2 + \sum_{j=2}^{n}(p_j^2 + q_j^2) \leq r^2. \tag{170}$$

From (8.22) we conclude that either $q_1 \neq 0$ or

$$q_1 = 0 \text{ and } \sum_{j=1}^{n}(p_j^2 + q_j^2) \leq r^2 - (a(1-2\tau))^2 \leq (r')^2, \tag{171}$$

and therefore all points of the camel are either off the wall ($q_1 \neq 0$) or are within the hole $H(r')$ as the camel passes through the wall under the action of $\mathcal{M}(\tau)$. We have moved the camel from the side with $q_1 > 0$ to the side with $q_1 < 0$.

What happens in the more interesting case where the camel is larger than the hole, $r > r'$? In that case Gromov has shown that there is *no* continuous family of symplectic maps $\mathcal{M}(\tau)$ satisfying (8.19) and (8.20) without some points of $\mathcal{M}(\tau)B_+^{2n}(r,a)$ lying in the wall and outside the hole for some intermediate τ values. Thus for a symplectic camel to pass through the eye of a needle under the action of a continuous family of symplectic maps, the eye of the needle must be larger than the camel.

By contrast, if one is allowed to use maps that are simply volume preserving but not symplectic, it is easy to see that one can pass the camel through the eye of any needle no matter how large the camel is or how small the eye of the needle is. For example, one may first stretch and thin the camel by pulling along her tail in the $+q_1$ direction while holding her nose fixed. (We assume the camel is eyeing the eye with some trepidation, and we plan to pass her through head first.) While increasing her length in the q_1 direction, we appropriately compress her in all other directions so that her volume remains unchanged. Then this thinned camel may be safely passed through the eye of the needle. Finally, the camel can be brought back to her original shape by holding her hind quarters fixed, pushing on her nose thereby compressing her q_1 dimension, and letting her other dimensions expand to their original values.

The discussion so far has been concerned with the 'spherical' camel $B^{2n}(r)$ given by (8.7). It can be extended to the case of a *general elliptic* camel $E^{2n}(r)$. By a general elliptic camel we mean the set defined by the rule

$$E^{2n}(r) = \{z \in R^{2n} \mid (z, Sz) \leq r^2\} \tag{172}$$

where S is a positive-definite matrix.

Evidently the nonsqueezing theorem and the symplectic camel results, which are examples of the general subject of symplectic *capacities*, have important applications to Accelerator Physics. The nonsqueezing theorem has implications for the feasibility of *emittance trading* (the hope that one might be able to concentrate particles in some phase-space plane at the expense of possible dilution in other planes), and the symplectic camel results bear on problems of linear and nonlinear beam transport. Of course one would like to have analogous results for camels and needle eyes with more general shapes than the simple spherical and elliptical and cylindrical shapes assumed in this item. Also, even if all of a camel cannot be squeezed into, say, some cylinder or some other volume, what fraction of the camel can be so squeezed, and how? Some important results have been found in these directions. See the references to Symplectic Geometry and Topology given at the end of this outline.

Finally, we close this item with a related consideration. Suppose $f(z)$ and $g(z)$ are two phase-space distributions. It would be nice to know whether or not two different phase-space distributions could be sent into each other by a symplectic map. For example, given the phase-space distribution coming out of some ion source or electron gun, is there any possible collection of beamline elements that would transform this distribution into some desired distribution at the end of some beamline or accelerator complex? Mathematically stated, one would like to decompose phase-space distributions into equivalence classes. This too is a deep question about which little is known.

18. Eigen Emittances

    This Item requires considerable editing.

Beam Description and Moment Transport

Preliminaries

The previous chapters dealt with *single*-particle orbit theory. In this chapter we will treat *many*-particle distributions in the approximation that the particles are *noninteracting*. Under this noninteracting assumption all results about particle distributions are derivable from properties of the single-particle transfer map $\mathcal{M}$.

Suppose $h(z)$ is some *density* function describing a collection of particles in phase space. That is d^6N, the number of particles in a phase-space volume d^6z, is given by the relation

$$d^6N = h(z)d^6z, \tag{173}$$

and there is the result

$$N = \int d^6z\, h(z) \tag{174}$$

where N is the number of particles under consideration.

More specifically, suppose $h^i(z)$ is a function describing some *initial* distribution of particles in phase space. Next suppose the particle distribution is transported through some system described by a map $\mathcal{M}$. Then, by Liouville's theorem, the *final* distribution $h^f(z)$ at the end of the system is given by the relation

$$h^f(z) = h^i(\mathcal{M}^{-1}z). \tag{175}$$

Also recall that, as sketched in Section 6.8, the problem of determining what distribution can be sent into what under the action of some symplectic map $\mathcal{M}$, which is what (1.4) describes, is deep and only partially understood.

Moments and Moment Transport

Suppose, as before, that $h^i(z)$ is a function describing some *initial* distribution of particles in phase space. Since $h^i(z)$ is a function, generally an infinite number of parameters are required for its specification. One way to *characterize* $h^i(z)$ is in terms of initial moments $Z^i_{abc\cdots}$ defined by the rule

$$Z^i_{abc\cdots} = \langle z_a z_b z_c \cdots \rangle^i = (1/N) \int d^6z\, h^i(z) z_a z_b z_c \cdots . \tag{176}$$

We use the term *characterize* advisedly rather than *specify* because in general the problem of reconstructing (uniquely determining) a function given its moments is ill posed. Nevertheless, moments may provide some useful information about $h^i(z)$.

How are initial and final moments related? To answer this question it is useful to employ a different notation for moments. Let $P_\alpha(z)$, where α is some running index, denote a complete

set of homogeneous polynomials in z through terms of some fixed degree. See Chapter 36. Then one can define initial moments m^i_α by the rule

$$m^i_\alpha = (1/N) \int d^6z \; h^i(z) P_\alpha(z). \tag{177}$$

Correspondingly, the final moments are given by the relation

$$\begin{aligned} m^f_\alpha &= (1/N) \int d^6z \; h^f(z) P_\alpha(z) = (1/N) \int d^6z \; h^i(\mathcal{M}^{-1} z) P_\alpha(z) \\ &= (1/N) \int d^6\bar{z} \; h^i(\bar{z}) P_\alpha(\mathcal{M}\bar{z}). \end{aligned} \tag{178}$$

Here we have used (1.4). And, to obtain the last line, we have changed variables of integration by the rule

$$\bar{z} = \mathcal{M}^{-1} z. \tag{179}$$

Doing so required calculation of the determinant of the Jacobi matrix M associated with $\mathcal{M}$. However, it is a property of symplectic matrices that they all have determinant $+1$. Therefore the determinant of M is $+1$ and need not appear explicitly in (2.3).

Since the P_α are complete, there is an expansion of the form

$$P_\alpha(\mathcal{M}\bar{z}) = \sum_\beta \mathcal{D}_{\alpha\beta}(\mathcal{M}) P_\beta(\bar{z}) \tag{180}$$

where the $\mathcal{D}_{\alpha\beta}(\mathcal{M})$ are coefficients that can be calculated for any transfer map $\mathcal{M}$. Employing (2.5) in (2.3) gives the intermediate result

$$\begin{aligned} m^f_\alpha &= (1/N) \int d^6\bar{z} \; h^i(\bar{z}) P_\alpha(\mathcal{M}\bar{z}) = (1/N) \int d^6\bar{z} \; h^i(\bar{z}) \sum_\beta \mathcal{D}_{\alpha\beta}(\mathcal{M}) P_\beta(\bar{z}) \\ &= \sum_\beta \mathcal{D}_{\alpha\beta}(\mathcal{M}) (1/N) \int d^6\bar{z} \; h^i(\bar{z}) P_\beta(\bar{z}). \end{aligned} \tag{181}$$

It follows that moments transform *linearly* according to the rule

$$m^f_\alpha = \sum_\beta D_{\alpha\beta}(\mathcal{M}) m^i_\beta. \tag{182}$$

Note that by this method one can find the evolution of moments *without* tracking (following orbits of individual particles in) particle distributions.

Kinematic Moment Invariants

Definition

Let m be a vector with components m_α, and let $D(\mathcal{M})$ be a matrix with entries $D_{\alpha\beta}(\mathcal{M})$. Write (2.7) in the more compact form

$$m^f = D(\mathcal{M}) m^i. \tag{183}$$

A function of moments $I[m]$ is said to be a *kinematic moment invariant* if it obeys the relations

$$I[m^f] = I[m^i], \tag{184}$$

or

$$I[D(\mathcal{M})m] = I[m], \tag{185}$$

for *all* symplectic maps $\mathcal{M}$.

Rather little is known about the existence and properties of kinematic moment invariants for the set of all symplectic maps. However, kinematic moment invariants are known to exist and all kinematic moment invariants have been found when the symplectic maps $\mathcal{M}$ are restricted to be those associated with the inhomogeneous symplectic group $ISp(2n)$. Moreover, their existence is a consequence of group theory applied to $ISp(2n)$. We note that, if deviation variables are employed, the full $\mathcal{M}$ is well approximated by translation and linear maps provided excursions about the design orbit are sufficiently small.

Second-Order Moments

Of particular interest are kinematic moment invariants that can be constructed from the second-order moments Z_{ab} with

$$Z_{ab} = \langle z_a z_b \rangle = (1/N) \int d^6 z \; h(z) z_a z_b. \tag{186}$$

For a given particle distribution, let Z be the matrix with entries Z_{ab}. In the case of a 1-degree of freedom system phase space is 2 dimensional, and the matrix Z in this case is 2×2. It is easily verified that in this case a kinematic moment invariant [under the action of $Sp(2)$] is given by the rule

$$I[Z] = \mathrm{tr}[(ZJ_2)^2] = 2[(Z_{12})^2 - Z_{11}Z_{22}] = -2(\langle q^2 \rangle \langle p^2 \rangle - \langle qp \rangle^2). \tag{187}$$

See Exercise 6.1 where this result is verified and shown to be a consequence of group theory applied to $Sp(2)$. Note that in this case I is proportional to the *mean square emittance* ϵ^2 *defined* by the rule

$$\epsilon^2 = \langle q^2 \rangle \langle p^2 \rangle - \langle qp \rangle^2. \tag{188}$$

In the case of a 3-degree of freedom system it can be shown that there are 3 such functionally independent invariants given by the rules

$$I^{(n)}[Z] = \mathrm{tr}[(ZJ)^n], \quad n = 2, 4, 6; \tag{189}$$

and all other invariants constructed from second-order moments are functions of these invariants. See Exercises 6.2 and 6.3.

Some Properties of Second-Order Moments

In this section we will explore various properties of Z.

Positive Definite Property

We begin by showing that the matrix Z, which is obviously real and symmetric, is also positive definite. Since $h(z)$ is a phase-space density, it is positive or zero for all z,

$$h(z) \geq 0 \text{ for all } z; \tag{190}$$

and it follows from (6.1) and continuity that there must be some finite phase-space volume for which $h(z) > 0$. Next let u be any real six-dimensional nonzero vector. Form the function $(u, z)^2$. It has the property

$$(u, z)^2 \geq 0 \text{ for all } z. \tag{191}$$

Moreover, in the volume where $h(z) > 0$, there must be some subvolume where $(u, z)^2 > 0$. It follows that there is the result

$$\begin{aligned} (u, Zu) &= \sum_{ab} u_a Z_{ab} u_b = (1/N) \int d^6z \; h(z) \sum_{ab} u_a z_a u_b z_b \\ &= (1/N) \int d^6z \; h(z)(u, z)^2 > 0. \end{aligned} \tag{192}$$

Transformation Properties

Properties under Linear Symplectic Maps

Let us next find the transformation properties of second-order moments under the action of a linear symplectic map $\mathcal{R}$ described by the symplectic matrix R. From the definition (4.1) and (4.13) we see that the transformed moments $\langle z_a z_b \rangle'$ are given by the relation

$$\langle z_a z_b \rangle' = (1/N) \int d^6z \; h'(z) z_a z_b = (1/N) \int d^6z \; h(R^{-1}z) z_a z_b. \tag{193}$$

Again introduce new variables $\bar{z}$ by the rules (4.15) through (4.17) and supplement (4.17) with the relation

$$z_b = \sum_d R_{bd} \bar{z}_d. \tag{194}$$

Also employ the relation (4.18). With these tools we see that (6.6) can be rewritten in the form

$$\begin{aligned} \langle z_a z_b \rangle' &= (1/N) \int d^6\bar{z} \; \sum_{cd} R_{ac} R_{bd} \bar{z}_c \bar{z}_d h(\bar{z}) \\ &= \sum_{cd} R_{ac} R_{bd} (1/N) \int d^6\bar{z} \; \bar{z}_c \bar{z}_d h(\bar{z}) \\ &= \sum_{cd} R_{ac} R_{bd} \langle z_c z_d \rangle. \end{aligned} \tag{195}$$

In terms of the notation employed in (5.4), the relation (6.8) can be rewritten in the component form

$$Z'_{ab} = \sum_{cd} R_{ac} R_{bd} Z_{cd} = \sum_{cd} R_{ac} Z_{cd} (R^T)_{db}, \tag{196}$$

which has the more compact matrix form

$$Z' = RZR^T. \tag{197}$$

This matrix relation specifies how second-order moments transform under a linear symplectic map. We observe, in particular, that the transformation rule is the same for all distributions having the same second-order moments.

Williamson Normal Form

Even more can be said. Since the matrix Z is real, symmetric, and positive definite, according to a theorem of Williamson there is a symplectic matrix A such that

$$AZA^T = D \tag{198}$$

where D is the *diagonal* matrix

$$D = \mathrm{diag}\{\lambda_1, \lambda_1, \lambda_2, \lambda_2, \lambda_3, \lambda_3\} \tag{199}$$

with all $\lambda_j > 0$. The right side of (6.11) is called the *Williamson normal form* of Z.

Two things should be noted about this remarkable result. Define the matrix Z^{norm} by the rule

$$Z^{\mathrm{norm}} = AZA^T = D. \tag{200}$$

Then, from (6.12 and (6.13), we see that there are the results

$$\langle q_j q_k \rangle^{\mathrm{norm}} = \langle p_j p_k \rangle^{\mathrm{norm}} = 0 \text{ if } j \neq k, \tag{201}$$

$$\langle q_j^2 \rangle^{\mathrm{norm}} = \langle p_j^2 \rangle^{\mathrm{norm}} = \lambda_j, \tag{202}$$

$$\langle q_j p_k \rangle^{\mathrm{norm}} = 0. \tag{203}$$

Also, we observe that (6.11) is of the form (6.10) with $R = A$. Thus, if a beam transport system can be found whose transfer matrix is A, then this transport system will bring the second-order moments to the normal form given by (6.14) through (6.16).

Eigen Emittances

We will next see that two second-order moment matrices Z' and Z have the *same* Williamson normal form if they are connected by a relation of the form (6.10). Indeed, observe that we may write the relation

$$\begin{aligned} AR^{-1}Z'(AR^{-1})^T &= AR^{-1}RZR^T(R^T)^{-1}A^T \\ &= AZA^T = D. \end{aligned} \tag{204}$$

[Here we have used the result $(R^{-1})^T = (R^T)^{-1}$ which holds for any invertible matrix.] But, by the group property of symplectic matrices, the matrix AR^{-1} is symplectic if the matrices A and R are symplectic. We see from (6.17) that the symplectic matrix AR^{-1} brings Z' to Williamson normal form and, according to (6.13), this normal form is the same as that for Z. The quantities λ_j^2 are called mean-square *eigen* emittances, or simply eigen remittances. It follows that while the entries in Z evolve as a particle distribution propagates through various elements, see (6.10), the eigen emittances remain *unchanged* (in the linear approximation). Thus, given an initial particle distribution, one can compute the initial second moments $\langle z_a z_b \rangle^i$, and from them the eigen emitances. And these eigen emittances will remain unchanged (in the linear approximation) as the particle distribution evolves.

It can be shown that the eigen emittances generalize the 1-degree of freedom mean-square emittance given by (5.6) to the fully coupled case. Indeed, it can be shown that in terms of the λ_j the kinematic invariants $I^{(n)}$ given by (5.7) have the values

$$I^{(n)} = 2(-1)^{n/2}(\lambda_1^n + \lambda_2^n + \lambda_3^n). \tag{205}$$

There are symplectic matrix routines that, given Z, find A and the λ_j. If only the λ_j are required, they can be found from the eigenvalues of JZ. Note that JZ is a Hamiltonian matrix.

To see that the λ_j can be found from the eigenvalues of JZ, suppose both sides of (6.11) are multiplied by J to give the result

$$JAZA^T = JD. \tag{206}$$

From the symplectic condition for A it follows that there is the relation

$$JA = (A^T)^{-1}J. \tag{207}$$

Consequently (6.19) can be rewritten in the form

$$(A^T)^{-1}JZA^T = JD, \tag{208}$$

which reveals that the matrices JZ and JD are related by a *similarity* transformation, and therefore have the same eigenvalues. See Exercise 3.7.16.

What remains is to find the eigenvalues of JD which, according to (3.2.10), (3.2.11), and (6.12) can be written in the block form

$$JD = \begin{pmatrix} \lambda_1 J_2 & & \\ & \lambda_2 J_2 & \\ & & \lambda_3 J_2 \end{pmatrix}. \tag{209}$$

Let W_2 be the unitary and (complex) symplectic 2×2 matrix

$$W_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}. \tag{210}$$

[See (3.9.12).] It has the property

$$W_2^{-1} J_2 W_2 = iK_2 \tag{211}$$

where K_2 is the matrix

$$K_2 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{212}$$

From W_2 construct the the 6×6 matrix W given in block form by the rule

$$W = \begin{pmatrix} W_2 & & \\ & W_2 & \\ & & W_2 \end{pmatrix}. \tag{213}$$

It follows from (6.22) and (6.24) that there is the relation

$$W^{-1}JDW = \begin{pmatrix} i\lambda_1 K_2 & & \\ & i\lambda_2 K_2 & \\ & & i\lambda_3 K_2 \end{pmatrix}. \tag{214}$$

We see that the eigenvalues of JD, and hence JZ, are pure imaginary and come in the $\pm$ pairs

$$\sigma_j = \pm i\lambda_j. \tag{215}$$

Conversely, if the eigenvalues $\pm\sigma_j$ of JZ are computed, then the eigen emittances are given by the relation

$$\lambda_j = |\sigma_j]. \tag{216}$$

Suppose we combine the relations (6.21) and (6.27) to find the result

$$W^{-1}(A^T)^{-1}JZA^TW = \begin{pmatrix} i\lambda_1 K_2 & & \\ & i\lambda_2 K_2 & \\ & & i\lambda_3 K_2 \end{pmatrix} = \mathrm{diag}\{-i\lambda_1, i\lambda_1, -i\lambda_2, i\lambda_2, -i\lambda_3, i\lambda_3\}. \tag{217}$$

Let A' be the matrix defined by the rule

$$A' = A^TW, \;\; (A')^{-1} = W^{-1}(A^T)^{-1}. \tag{218}$$

It will be symplectic since A (and hence A^T) and W are symplectic. With the aid of A' the relation (6.30) takes the form

$$(A')^{-1}JZA' = \mathrm{diag}\{-i\lambda_1, i\lambda_1, -i\lambda_2, i\lambda_2, -i\lambda_3, i\lambda_3\}. \tag{219}$$

We observe that since Z is positive definite, JZ is a particular/special kind of Hamiltonian matrix. As a consequence of Williamson's theorem we have seen that it can be diagonalized by a similarity transformation even if its eigenvalues are not distinct; moreover the diagonalizing matrix A' can be chosen to be symplectic. And, as stated earlier, the eigenvalues of JZ are pure imaginary.

Finally, we note that multiplying both sides of (6.10) by J produces the relation

$$JZ' = JRZR^T = (R^T)^{-1}JZR^T. \tag{220}$$

Here we have used the fact that R is symplectic. We see that with the use of J the evolution rule (6.10) for Z becomes the *similarity* transformation rule (6.33) for JZ. Since eigenvalues are preserved by similarity transformations, we have found an alternative explanation of why the eigen emittances remain unchanged as a particle distribution evolves.

Classical Uncertainty Principle

Statement

The results of the previous subsection can be used to derive a *classical* uncertainty principle. What we will show is that there is the inequality

$$\langle q_i^2\rangle\langle p_i^2\rangle \geq \lambda_{\rm min}^2,\ i=1,2,3, \tag{221}$$

where $\lambda_{\rm min}$ is the minimum of the λ_k. No matter what is done to a beam (ignoring nonlinear and nonsymplectic effects), the products of the mean-square deviations in q_i and p_i for any plane must exceed, or at best equal, $\lambda_{\rm min}^2$.

Proof

Begin by rewriting (6.13) in the form

$$Z = A^{-1}Z^{\rm norm}(A^{-1})^T = N^T Z^{\rm norm} N = N^T D N \tag{222}$$

where we have made the definition

$$N = (A^{-1})^T. \tag{223}$$

We note that N will be symplectic if A is symplectic, and conversely.
Let us compute the $\langle q_i^2\rangle$ and $\langle p_i^2\rangle$. To compute the $\langle q_i^2\rangle$ set

$$a = j \text{ with } j = 1,3,5 \text{ when } i = 1,2,3. \tag{224}$$

We then find from (6.35) that

$$\begin{aligned}\langle q_i^2\rangle &= Z_{aa} = (N^T D N)_{aa} = \sum_{cd}(N^T)_{ac}D_{cd}N_{da}\\ &= \sum_c N_{ca}D_{cc}N_{ca} = \sum_c (N_{ca})^2 D_{cc}\\ &\geq \lambda_{\rm min}\sum_c (N_{ca})^2.\end{aligned} \tag{225}$$

Similarly, to compute the $\langle p_i^2\rangle$, upon setting

$$b = j+1, \tag{226}$$

we find that

$$\begin{aligned}\langle p_i^2\rangle &= Z_{bb} = (N^T D N)_{bb} = \sum_{cd}(N^T)_{bc}D_{cd}N_{db}\\ &= \sum_d N_{db}D_{dd}N_{db} = \sum_d (N_{db})^2 D_{dd}\\ &\geq \lambda_{\rm min}\sum_d (N_{db})^2.\end{aligned} \tag{227}$$

It follows that

$$\langle q_i^2 \rangle \langle p_i^2 \rangle \geq \lambda_{\min}^2 \left[\sum_c (N_{ca})^2 \right] \left[\sum_d (N_{db})^2 \right]. \tag{228}$$

To proceed further, let u^a and u^b be vectors with the entries

$$u_c^a = N_{ca}, \tag{229}$$

$$u_d^b = N_{db}. \tag{230}$$

Evidently u^a and u^b are the a^{th} and b^{th} columns of N. With these definitions we may write (6.41) in the more compact form

$$\langle q_i^2 \rangle \langle p_i^2 \rangle \geq \lambda_{\min}^2 ||u^a||^2 \, ||u^b||^2 \tag{231}$$

where $|| * ||$ denotes the Euclidean norm. Since N is a symplectic matrix, it follows from the symplectic condition that there is also the relation

$$(u^a, Ju^b) = 1. \tag{232}$$

See Exercise 3.6.13. It can be shown using the spectral norm for J that (6.45) in turn entails the inequality.

$$||u^a|| \, ||u^b|| \geq 1. \tag{233}$$

See Exercise 3.7.1. Upon combining (6.44) and (6.46) we find the advertised result (6.34).

Minimum Emittance Theorem

Statement

The classical uncertainty principle shows that (in the linear approximation) no matter how a beam is transformed, the product of the spreads in position and the conjugate momentum must satisfy the relation (6.34). There is a related constraint on the mean-square emittances ϵ_i defined by

$$\epsilon_i^2 = \langle q_i^2 \rangle \langle p_i^2 \rangle - \langle q_i p_i \rangle^2. \tag{234}$$

What we will show is that (in the linear approximation) no matter how a beam is transformed (symplectically) there is the constraint

$$\epsilon_i^2 \geq \lambda_{\min}^2, \; i = 1, 2, 3. \tag{235}$$

Together the information provided by the classical uncertainty principle and the minimum emittance theorem is useful when designing a beam line to perform emittance manipulations because it sets lower limits on what one can hope to achieve.

Proof

Suppose, in the 6×6 case under consideration, that we partition Z into nine 2×2 blocks by writing

$$Z = \begin{pmatrix} Z^{11} & Z^{12} & Z^{13} \\ Z^{21} & Z^{22} & Z^{23} \\ Z^{31} & Z^{32} & Z^{33} \end{pmatrix}. \tag{236}$$

Because Z is symmetric, the blocks will satisfy the relations

$$(Z^{ij})^T = Z^{ji}. \tag{237}$$

Let R be a 6×6 matrix having the block form

$$R = \begin{pmatrix} A & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix}. \tag{238}$$

It will be symplectic if A is symplectic. Its use in (6.10) produces a Z' given by

$$Z' = \begin{pmatrix} A & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} Z^{11} & Z^{12} & Z^{13} \\ Z^{21} & Z^{22} & Z^{23} \\ Z^{31} & Z^{32} & Z^{33} \end{pmatrix} \begin{pmatrix} A^T & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix}. \tag{239}$$

Carrying out the indicated multiplication gives the result

$$Z' = \begin{pmatrix} AZ^{11}A^T & AZ^{12} & AZ^{13} \\ Z^{21}A^T & Z^{22} & Z^{23} \\ Z^{31}A^T & Z^{32} & Z^{33} \end{pmatrix}. \tag{240}$$

In particular, we see that

$$(Z')^{11} = AZ^{11}A^T. \tag{241}$$

We will now seek a symplectic A that brings Z^{11} to Williamson normal form. Define the quantity ϵ_1 by the rules

$$\epsilon_1^2 = Z_{11}^{11}Z_{22}^{11} - (Z_{12}^{11})^2 = \langle q_1^2 \rangle \langle p_1^2 \rangle - \langle q_1 p_1 \rangle^2, \tag{242}$$

$$\epsilon_1 = +\sqrt{\epsilon_1^2}. \tag{243}$$

It follows from the Schwarz inequality that there is the relation

$$\langle q_1 p_1 \rangle^2 \leq \langle q_1^2 \rangle \langle p_1^2 \rangle \tag{244}$$

and therefore the right side of (6.55) can never be negative. See Exercise 6.4. Consequently ϵ_1 is well defined by (6.55) and (6.56), and is positive. Next define "beam" betatron functions α, β, γ by the rules

$$\alpha = -Z_{12}^{11}/\epsilon_1 = -\langle q_1 p_1 \rangle / \epsilon_1, \tag{245}$$

$$\beta = Z_{11}^{11}/\epsilon_1 = \langle q_1^2 \rangle / \epsilon_1, \tag{246}$$

$$\gamma = Z_{22}^{11}/\epsilon_1 = \langle p_1^2 \rangle / \epsilon_1. \tag{247}$$

In terms of these definitions, Z^{11} takes the form

$$Z^{11} = \epsilon_1 \begin{pmatrix} \beta & -\alpha \\ -\alpha & \gamma \end{pmatrix}. \tag{248}$$

From (6.59) and (6.60) there are the inequalities $\beta \geq 0$ and $\gamma \geq 0$. And, from (6.55) through (6.60), there is the relation

$$1 = \beta\gamma - \alpha^2. \tag{249}$$

Finally, define the matrix A by the rule

$$A = \begin{pmatrix} 1/\sqrt{\beta} & 0 \\ \alpha/\sqrt{\beta} & \sqrt{\beta} \end{pmatrix}. \tag{250}$$

Since A is 2×2 and evidently has unit determinant, it is symplectic. Correspondingly, the R given by (6.51) is symplectic. And, from the definitions made and executing the matrix multiplications $AZ^{11}A^T$ indicated in (6.54), we find that

$$(Z')^{11} = AZ^{11}A^T = \mathrm{diag}(\epsilon_1, \epsilon_1). \tag{251}$$

We will now exploit these results. From (6.64) we find that

$$\langle q_1^2 \rangle' = (Z')_{11}^{11} = \epsilon_1, \tag{252}$$

$$\langle p_1^2 \rangle' = (Z')_{22}^{11} = \epsilon_1. \tag{253}$$

It follows that

$$\langle q_1^2 \rangle' \, \langle p_1^2 \rangle' = \epsilon_1^2. \tag{254}$$

But we also have the relation (6.34). We conclude that there is the inequality

$$\epsilon_1^2 \geq \lambda_{\min}^2, \tag{255}$$

in accord with (6.48). Analogous results hold for the other planes.

Sharpening

We close this subsection by noting that the minimum emittance theorem (6.48) sharpens the classical uncertainty principle (6.34). Indeed, combining (6.47) and (6.48) produces the result

$$\langle q_i^2 \rangle \langle p_i^2 \rangle \geq \lambda_{\min}^2 + \langle q_i p_i \rangle^2, \; i = 1, 2, 3. \tag{256}$$

We see that to minimize $\langle q_i^2 \rangle \langle p_i^2 \rangle$ we must insure that $\langle q_i p_i \rangle$ vanishes.

Summary of What We Have Learned

The information provided by the classical uncertainty principle and the minimum emittance theorem is useful when designing a beam line to perform emittance manipulations on a beam because it sets lower limits on what one can hope to achieve. It should also be useful in analyzing the results of beam cooling experiments. In this case one can measure all quadratic moments before and after a cooling channel. Next compute the eigen emittances of Z before and Z after. Ideally, one would like to find that all the λ_j^2 have decreased, or at least the minimum of the λ_j^2 has decreased.

We have seen that, in considering what can be achieved under beam transport (in the linear approximation), what counts are the eigen emittances, and these can be viewed as properties of the *initial* particle distribution. Moreover, according to (6.14) through (6.16), the best that can be achieved are the spread relations

$$\langle q_i^2\rangle\langle p_i^2\rangle = \lambda_i^2,\ i = 1, 2, 3 \tag{257}$$

where the λ_i are the eigen emittances in some order. Thus, in the combined context of both source and beam-line design, the challenge is to produce an initial particle distribution having optimal eigen emittances and to then transform the initial particle distribution in such a way that the optimal spread relations are realized in the desired planes. The next sections will describe various methods for producing initial particle distributions having optimal eigen emittances and how to then transform these distributions in such a way that the optimal spread relations are realized in the desired planes.

19. Lie Algebraic Theory of Geometric Light Optics

Lie methods can also be applied to the subject of geometric light optics. Consider the optical system illustrated schematically in Figure 7. A ray originates at the general *initial* point P^i with spatial coordinate $\boldsymbol{r}^i$ and moves in an initial direction specified by the unit vector $\hat{\boldsymbol{s}}^i$. After passing through an optical device it arrives at the *final* point P^f with with spatial coordinate $\boldsymbol{r}^f$ and moves in a final direction specified by the unit vector $\hat{\boldsymbol{s}}^f$. Given the *initial* quantities $(\boldsymbol{r}^i, \hat{\boldsymbol{s}}^i)$, the fundamental problem of geometrical optics is to determine the *final* quantities $(\boldsymbol{r}^f, \hat{\boldsymbol{s}}^f)$ and to design an optical device in such a way that the relation between the initial and final ray quantities has various desired properties.

Suppose the z coordinates of the initial and final points P^i and $P^f$are held fixed. In some instances the planes $z = z^i$ and $z = z^f$ can be viewed as object and image planes, respectively. But in other cases they simply serve as convenient reference planes. Further, suppose the general light ray from P^i to P^f is parameterized using z as an *independent/time-like* variable. That is, the path of a general ray is described by specifying the two functions $x(z)$ and $y(z)$. Then the element of path length ds along a ray is given by the expression

$$ds = [(dz)^2 + (dx)^2 + (dy)^2]^{1/2} = [1 + (x')^2 + (y')^2]^{1/2}dz. \tag{258}$$

Here a prime denotes the differentiation d/dz. Consequently the *optical* path length along a

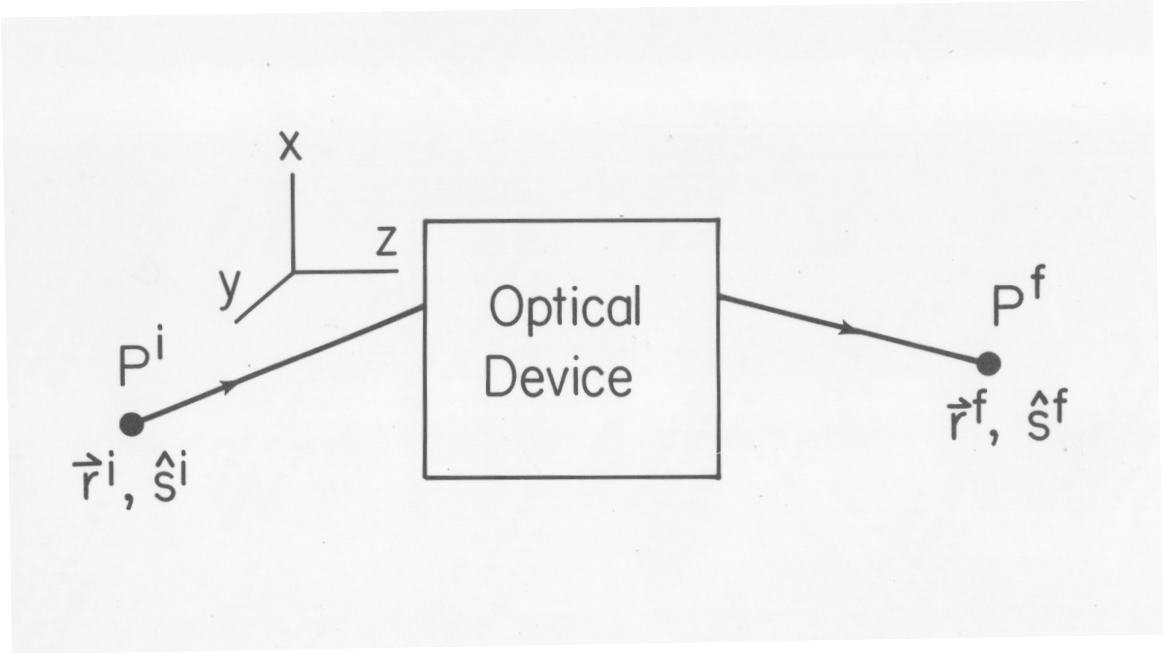

Figure 8: Optical system consisting of an optical device preceded and followed by simple transit. A ray originates at P^i with *initial* location $\boldsymbol{r}^i$ and *initial* direction $\hat{\boldsymbol{s}}^i$, and terminates at P^f with *final* location $\boldsymbol{r}^f$ and *final* direction $\hat{\boldsymbol{s}}^f$

.

ray from P^i to P^f is given by the integral

$$A = \int_{P^i}^{P^f} cdt = \int_{P^i}^{P^f} c(dt/ds)ds = \int_{P^i}^{P^f} c(n/c)ds$$
$$= \int_{P^i}^{P^f} nds = \int_{z^i}^{z^f} n(x,y,z)[1+(x')^2+(y')^2]^{1/2}dz. \tag{259}$$

Here the function $n(x,y,z) = n(\boldsymbol{r})$ specifies the index of refraction at each point before and after the optical device and in the device itself.

Fermat's principle requires that A be an extremum for the path of an actual ray. Therefore the ray path satisfies the Euler-Lagrange equations

$$d/dz(\partial L/\partial x') - \partial L/\partial x = 0, \tag{260}$$

$$d/dz(\partial L/\partial y') - \partial L/\partial y = 0, \tag{261}$$

with a Lagrangian L given by the expression

$$L = n(x,y,z)[1+(x')^2+(y')^2]^{1/2}. \tag{262}$$

To proceed further, it is useful to pass from a Lagrangian formulation to a Hamiltonian formulation. Introduce two momenta p_x and p_y conjugate to the coordinates x and y by the rule

$$p_x = \partial L/\partial x', \tag{263}$$

$$p_y = \partial L/\partial y', \tag{264}$$

with the explicit results that

$$p_x = n(\boldsymbol{r})x'/[1+(x')^2+(y')^2]^{1/2}, \tag{265}$$

$$p_y = n(\boldsymbol{r})y'/[1+(x')^2+(y')^2]^{1/2}. \tag{266}$$

The Hamiltonian H is defined in terms of the Lagrangian L by the Legendre transformation

$$H(x,y,p_x,p_y;z) = p_x x' + p_y y' - L. \tag{267}$$

It follows from (1.5) through (1.10) that in our case H is given by the relation

$$H = -[n^2(\boldsymbol{r}) - p_x^2 - p_y^2]^{1/2}. \tag{268}$$

Let $\boldsymbol{q}$ be a two-component vector with entries $q_x = x$ and $q_y = y$, and let $\boldsymbol{p}$ be a two-component vector with entries p_x and p_y. Evidently, a ray leaving the initial point P^i is characterized by the quantities z^i, $\boldsymbol{q}^i$, and $\boldsymbol{p}^i$. The quantity $\boldsymbol{q}^i$ specifies the initial point of origin on the plane $z = z^i$ and, according to (1.8) and (1.9), $\boldsymbol{p}^i$ describes the initial direction of the ray. Similarly, $\boldsymbol{q}^f$ and $\boldsymbol{p}^f$ characterize the ray as it arrives at the final point P^f in the plane $z = z^f$. Finally, the relation between the initial conditions $\boldsymbol{q}^i$ and $\boldsymbol{p}^i$ and the final conditions $\boldsymbol{q}^f$ and $\boldsymbol{p}^f$ is

given by following from $z = z^i$ to $z = z^f$ a trajectory $\boldsymbol{q}(z)$, $\boldsymbol{p}(z)$ governed by the Hamiltonian H.

At this point it is convenient to introduce a four-component vector $\boldsymbol{w}$ with entries $\boldsymbol{q}$, $\boldsymbol{p}$:

$$\boldsymbol{w} = (w_1, w_2, w_3, w_4) = (q_x, p_x, q_y, p_y). \tag{269}$$

Also, let $\boldsymbol{w}^i$ and $\boldsymbol{w}^f$ denote initial and final values of $\boldsymbol{w}$. The fact that initial conditions *determine* the final conditions can be expressed in terms of a functional relationship or *mapping* $\mathcal{M}$. This relationship can be defined formally by writing the expression

$$\boldsymbol{w}^f = \mathcal{M}\boldsymbol{w}^i. \tag{270}$$

As has been seen, Fermat's principle is equivalent to the statement that the initial conditions $\boldsymbol{w}^i$ and the final conditions $\boldsymbol{w}^f$ are related by following a trajectory governed by a Hamiltonian, namely the Hamiltonian (1.11). This statement is equivalent in turn to the statement that $\mathcal{M}$ is a *symplectic* map.

At this point we assume that the optical device has *axial/rotational symmetry* about the z axis. Introduce the definitions

$$q^2 = (q_x)^2 + (q_y)^2 = \boldsymbol{q} \cdot \boldsymbol{q}, \tag{271}$$

$$p^2 = (p_x)^2 + (p_y)^2 = \boldsymbol{p} \cdot \boldsymbol{p}, \tag{272}$$

$$\boldsymbol{p} \cdot \boldsymbol{q} = p_x q_x + p_y q_y. \tag{273}$$

(Note that this notation can be misleading since, for example, q^2 is not the square of any Cartesian coordinate.) From (1.11) we see that the optical Hamiltonian depends on $\boldsymbol{p}$ only through the quantity p^2. To enforce axial symmetry, we assume that $n(\boldsymbol{r})$ is of the functional form

$$n(\boldsymbol{r}) = \hat{n}(q^2, z) \tag{274}$$

so that the index of refraction also has axial symmetry.

It can be shown that, under the assumption of axial symmetry, the transfer map $\mathcal{M}$ must be of the form

$$\mathcal{M} = \exp(:\, f_2 \,:) \exp(:\, f_4 \,:) \exp(:\, f_6 \,:) \exp(:\, f_8 \,:) \cdots . \tag{275}$$

That is, only the f_m with *even* m can occur in the factored product representation of $\mathcal{M}$. Moreover, the f_m can only depend on the quantities $\boldsymbol{q}$ and $\boldsymbol{p}$ in an axial-rotation invariant way. For example, f_4 must be of the form

$$f_4 = A(p^2)^2 + Bp^2(\boldsymbol{p} \cdot \boldsymbol{q}) + C(\boldsymbol{p} \cdot \boldsymbol{q})^2 + Dp^2q^2 + E(\boldsymbol{p} \cdot \boldsymbol{q})q^2 + F(q^2)^2 \tag{276}$$

where the coefficients A through F are to be determined.

For an imaging system it can be demonstrated that the polynomials in (2.19), shown multiplied by the coefficients A through E, are related to the *Seidel* aberrations called spherical

aberration, coma, astigmatism, curvature of field, and distortion, respectively:[2]

$$A(p^2)^2 \leftrightarrow \text{spherical aberration}, \tag{277}$$

$$Bp^2(\boldsymbol{p}\cdot\boldsymbol{q}) \leftrightarrow \text{coma}, \tag{278}$$

$$C(\boldsymbol{p}\cdot\boldsymbol{q})^2 \leftrightarrow \text{astigmatism}, \tag{279}$$

$$Dp^2q^2 \leftrightarrow \text{curvature of field}, \tag{280}$$

$$E(\boldsymbol{p}\cdot\boldsymbol{q})q^2 \leftrightarrow \text{distortion}, \tag{281}$$

$$F(q^2)^2 \leftrightarrow \text{pocus}. \tag{282}$$

It can be shown that the *net* "pocus" aberration of an imaging system does not affect its ability to form images. (It does not affect the positions where rays arrive on the image plane, but only affects their arrival directions.) Therefore it is less commonly discussed, and indeed its name has been coined only recently. But it can be important for other systems.

Because net pocus does not affect imaging, it might be viewed as being magical: hence another explanation for the whimsical name *pocus* because of its association with the incantation *hocus-pocus*. Moreover, as we will see, pocus is the *only* third-order aberration that can be controllably "injected" into a system by employing lenses with aspherical surfaces. And as can be further shown, pocus then morphs into other aberrations in the course of paraxial transit and paraxial lensing, and these aberrations can be used to cancel other existing aberrations.

We also remark that the use of the name *astigmatism* can be confusing. In the present context it refers to a particular third-order aberration. With regard to vision, the name astigmatism usually refers to a possible lack of axial symmetry of a visual system in terms of its paraxial (linear) properties, and it is this defect (as well as focal length) that eye glasses are principally designed to correct.

We now present some formulas for the generators of maps associated with lens surfaces. In accord with the assumption of axial symmetry, we require that the surfaces to be employed are surfaces of revolution about the z axis, and therefore depend on x and y only through the variable $x^2 + y^2 = q^2$. We also require that they be *analytic* functions of the variable q^2, and pass through the origin. Such surfaces are described by relations of the form

$$z = \beta_2(q^2) + \beta_4(q^2)^2 + \beta_6(q^2)^3 + \beta_8(q^2)^4 + \cdots . \tag{283}$$

If only β_2 is nonvanishing, the surface is a parabola of revolution opening either to the left or the right depending on the sign of β_2.

[2]Philipp Ludwig von Seidel (1821-1896) classified and described the possible third-order geometric aberrations for axially symmetric optical systems. Subsequent extensive work on aberrations was done by many others including definitive work by Karl Schwarzschild (1873-1916). This work took into account the symplectic condition by employing $F_1(q, Q)$ generating functions, called *point* characteristic functions in the optics literature, and $F_4(p, P)$ generating functions, called *angular* characteristic functions in the optics literature. [$F_1(q, Q)$ generating functions are not applicable to imaging systems when object and image planes are employed, but are applicable when object and some aperture planes are employed. $F_4(p, P)$ generating functions are applicable to imaging systems when object and image planes are employed. It can be shown that, despite the plethora of generating function types, there is no one generating function type that works for (is compatible with) all symplectic maps.] After his optics work, in 1915 while serving in the army in World War I, Schwarzschild (in his spare time) did his celebrated work in General Relativity to find the Schwarzschild metric, the first exact solution to the Einstein field equations and the key ingredient for the understanding of black holes.

It is assumed that a light ray crosses a surface from a medium having index of refraction n^- into a medium having index n^+. The map associated with crossing such a surface is factorized in the form

$$\mathcal{M} = \exp(: f_2 :) \exp(: f_4 :) \exp(: f_6 :) \cdots . \tag{284}$$

Formulas for the f_n are available through f_8. Below are the expressions for f_2 and f_4:

$$f_2 = \beta_2(n^- - n^+)q^2 \Rightarrow f_2 = -[1/(2f)]q^2 \tag{285}$$

with the *focal length* f given by

$$1/f = -2\beta_2(n^- - n^+). \tag{286}$$

A surface injects

$$\begin{aligned} f_4 = &-\beta_2^3[(n^- - n^+)/n^-]\{n^-[2 - (\beta_4/\beta_2^3)] - 2n^+\}(q^2)^2 \\ &+ 2\beta_2^2[(n^- - n^+)/n^-]q^2(\boldsymbol{p}\cdot\boldsymbol{q}) \\ &+ \beta_2[(n^- - n^+)/(2n^-n^+)]q^2p^2, \end{aligned} \tag{287}$$

which contains pocus, distortion,and curvature of field, but not spherical, coma, and astigmatism. *Only* pocus is adjustable (depends on β_4) by making surface suitably aspherical.

Drifting injects $f_2 \simeq (p^2)$ and $f_4 \simeq (p^2)^2$.

We will now illustrate how some of our results can be used to design a doublet imaging system that is free of all third-order aberrations. This system is illustrated schematically in Figure 8. It consists of four surfaces separated by drift spaces either in air or two possibly different refractive media. Between the object plane and *Surface* S^1 there is a *left-side* drift space (in air) of on-axis length D_L. Surfaces S^1 and S^2, with an on-axis separation of thickness t_L, constitute a first lens made of a medium with refractive index n_L. Surfaces S^3 and S^4, with an on-axis separation of thickness t_R, constitute a second lens made of a medium with refractive index n_R. In the *middle*, between surfaces S^2 and S^3, there is a drift space (in air) of on-axis length D_{MID}. Finally, between S^4 and the image plane there is a *right-side* drift space (in air) of on-axis length D_R.

We will seek to design the system so that the aberration coefficients A through E vanish. Calculation shows that the quantity (C-2D), called the *Petzval* strength, depends *only* on lengths and refractive indices and the β_2 for the lens surfaces, and not on the β_4, β_6, $\cdots$.[3]

[3]Joseph Maximilian Petzval (1807-1891) was a German-Hungarian mathematician/physicist who, among other things, performed and oversaw early analytic work (directed by himself and involving 8 artillery gunners and 3 corporals, serving as computers, supplied by Archduke Louis of Austria) on geometric aberrations. He also designed the Petzval lens system, which was the first lens system to be designed using analytical methods and was very important in the development of high quality portrait photography. Such lens systems are still for sale today, and are valued for their high speed and the pleasant bokeh/background that they produce in the out-of focus area around and behind the subject due to an exquisite mix of aberrations that come into play at points far from the paraxial focal point. Opera glasses were another of his inventions.

In 1859 his home was broken into and his manuscripts, a result of many years of research, were destroyed. His most refined book on optics, lost with his manuscripts, would never appear in print. Much of what we do know of his work is found in the written descriptions of others who had some familiarity with some of his work. Also, apparently Petzval independently discovered Laplace transforms while working on differential equations. But the name *Laplace* for these transforms, also of course discovered by Laplace, was coined by Poincaré and they have been so called ever since.

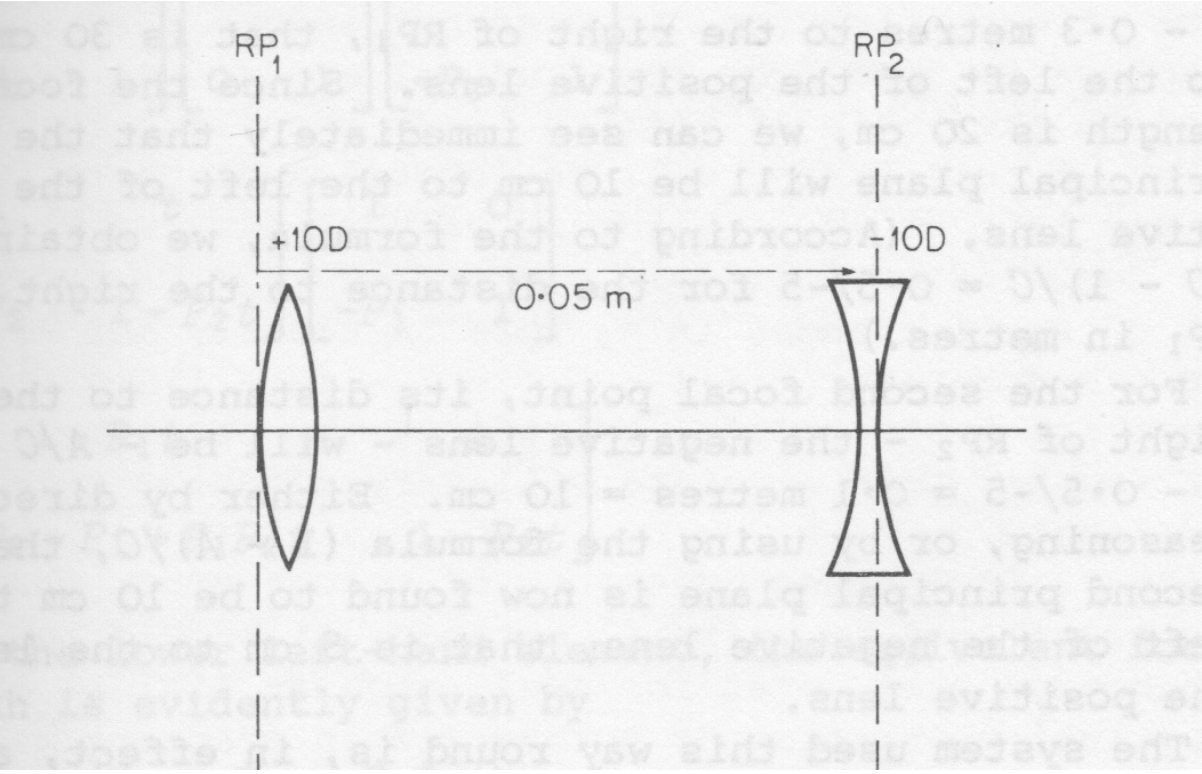

Figure 9: Layout of imaging doublet system that is free of all third-order aberrations. The object plane is on the far left and the image plane is on the far right (so that both are not visible), and only the shapes of the various lens surfaces and the lens thicknesses and spacings are illustrated. The *reference plane* RP_1 is at the beginning of the convex lens, and RP_2 is at the end of the concave lens. In the thin-lens approximation the convex and concave lenses have powers of $1/f = 10$ and $1/f = -10$, respectively.

It turns out that requiring (C-2D) vanish requires that one of the lenses be focusing and the other defocusing; and we will do so while still achieving the net desired paraxial properties of the system.

As stated earlier, ideally we would like to have vanishing values for *all* the coefficients A through E appearing in (2.30) through (2.34). There are five such coefficients, but we have already caused the Petzval combination $(C - 2D)$ to vanish thereby leaving four more goals to be met. We also observe that there are four surfaces S^1 through S^4 for which values β_4^1 through β_4^4 can be assigned. Can they be set in such a way that all the f_4 save for the F terms vanish? The answer is *yes*!

Below are values for the f_4 when the β_4^j are optimally set. Evidently all entries *vanish* save for the coefficients of q_x^4, $q_x^2 q_y^2$, and q_y^4. Examine the entries for $f(84)$, $f(95)$, and $f(175)$. Note that $f(95) = 2f(84) = 2f(175)$, and therefore q_x^4, $q_x^2 q_y^2$, and q_y^4 appear in the *pocus* combination $(q^2)^2$. All third-order aberrations that affect image formation have been caused to vanish.

Exhibit 1: $\mathcal{M}$ after offensive third-order aberration correction.

```
matrix for map is :

-5.00000E-01 -6.39488E-14  0.00000E+00  0.00000E+00
-5.00000E-02 -2.00000E+00  0.00000E+00  0.00000E+00
 0.00000E+00  0.00000E+00 -5.00000E-01 -6.39488E-14
 0.00000E+00  0.00000E+00 -5.00000E-02 -2.00000E+00

nonzero elements in generating polynomial are :

 f( 84)=f( 40 00 00 )=-1.73560018907580E-05
 f( 95)=f( 20 20 00 )=-3.47120037815021E-05
 f(175)=f( 00 40 00 )=-1.73560018907580E-05
```

References

[1] É. Forest, *Beam Dynamics: A New Attitude and Framework*, Harwood Academic Publishers (1998).

[2] É. Forest, *From Tracking Code to Analysis: Generalized Courant-Snyder Theory for Any Accelerator Model*, Springer Japan (2016).

[3] A. Chao, M. Tigner, H. Weise, and F. Zimmermann, Edit., *Handbook of Accelerator Physics and Engineering*, Third Edition, World Scientific (2023).

[4] A. Dragt, *Lie Methods for Nonlinear Dynamics with Applications to Accelerator Physics*, (2726 pp); A. Dragt et al., *MaryLie 3.0 Users' Manual: A Program for Charged Particle Beam Transport Based on Lie Algebraic Methods*, (901 pp). See the Web site http://www.physics.umd.edu/dsat.

[5] A. Dragt and E. Forest, "Computation of nonlinear behavior of Hamiltonian systems using Lie algebraic methods", *J. Math. Phys.* **24**, 2734 (1983).

[6] A. Dragt, F. Neri, G. Rangarajan, D.R. Douglas, L.M. Healy, and R.D. Ryne, "Lie Algebraic Treatment of Linear and Nonlinear Beam Dynamics", *Ann. Rev. Nucl. Part. Sci.* **38**, pp. 455-496 (1988).

[7] A. Dragt, "Numerical third-order transfer map for solenoid", *Nuclear Instruments and Methods in Physics Research* **A298**, p. 441 (1990).

[8] K. Halbach, "Physical and Optical Properties of Rare Earth Cobalt Magnets", *Nuclear Instruments and Methods* **187**, pp. 109-117 (1981).

[9] 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

[10] 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.

[11] P. Walstrom, "Magnetic Field Models for Multipole Magnets in Cylindrical Coordinates", 2 January 2022 preprint.

[12] M. Venturini and A. Dragt, "Accurate Computation of Transfer Maps from Magnetic Field Data", *Nuclear Instruments and Methods* **A427**, p. 387 (1991).

[13] 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).

[14] C. Mitchell, "Calculation of Realistic Charged-Particle Transfer Maps", University of Maryland Physics Department Ph.D. Thesis (2007).

[15] C. Mitchell and A. Dragt, "Accurate Transfer Maps for Realistic Beamline Elements: Part I, Straight Elements", 19 pages, *Phys. Rev. ST Accel. Beams* **13**, 064001 (2010).

[16] A. Dragt and D. Abell, "Symplectic maps and computation of orbits in particle accelerators", *Fields Inst. Commun.* **10**, 59-85 (1995).

[17] D. Abell and A. Dragt, "Structure-preserving techniques in accelerator physics", see the Web site https://arxiv.org/pdf/2211.00252.pdf.

[18] To find Web links to Symplectic Camels, Google "Symplectic Camel" and "Non-squeezing Theorem".

[19] M. Berz, *Modern Map Methods in Particle Beam Physics*, Volume 108 of *Advances in Imaging and Electron Physics*, Academic Press (1999).

[20] A. Dragt, "A Lie Algebraic Theory of Geometrical Optics and Optical Aberrations", *J. Opt. Sci. Am.* **72**, p. 372 (1982).

[21] A. Dragt, E. Forest, and K. Wolf, "Foundations of a Lie Algebraic Theory of Geometrical Optics", *Lie Methods in Optics*, J.S. Mondragon and K.B. Wolf, Edit., Springer-Verlag (1986).