Symplecticity in the Numerical Integration of Linear Beam Optics

Preserving Accuracy and Symplecticity of Simulated Particle Motion

Christopher K. Allen
Spallation Neutron Source
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831

1 Introduction

When simulating particle dynamics it is typically more important to preserve the symplectic nature of the motion over the overall accuracy of the simulation. Preserving the symplectic relationship between velocity and position is necessary to ensure that the motion is physical. A simulation may be numerically accurate, but if the symplectic condition is violated the results are suspect.

The paper centers on two results: a leapfrog technique for including space charge in the dynamics calculations which is shown to be both third-order accurate and symplectic, and methods for adding field imperfections which are also symplectic with varying integration accuracy. The material is presented in the context of Lie groups and algebras in order to demonstrate symplecticity of the methods. Thus, a basic background on Lie methods is included as it applies to beam dynamics. Also covered are some more general facts on matrix theory and differential equations that we need for the development. The idea here is that the material is somewhat self-contained and can be extended at a later date.

We restrict ourselves to linear beam dynamics including space charge effects. Thus, the usual beam optics matrix technique for "integrating" the dynamics equations is valid. However, we present the material in the context the symplectic group $Sp(n,\mathbb{R})$ composed of real $n{\times}n$ symplectic matrices, and its Lie algebra $sp(n,\mathbb{R})$, also represented by real $n{\times}n$ matrices.

2 Background

2.1 The Symplectic Group

For simplicity we start with the set $\mathbb{R}^{2\times2}$ of real, $2x2$ matrices. Larger symplectic groups have similar properties, but are embedded in some $\mathbb{R}^{2n\times2n}$ where n is an integer greater than 1. Define the matrix $\mathbf{J}\in\mathbb{R}^{2\times2}$ as

$$\mathbf{J} \triangleq \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{1}$$

Then a matrix $\mathbf{A} \in \mathbb{R}^{2\times 2}$ is *symplectic* if

$$\mathbf{A}^T \mathbf{J} \mathbf{A} = \mathbf{J}\,. \tag{2}$$

The set of symplectic matrices in $\mathbb{R}^{2\times 2}$ is denoted $Sp(2,\mathbb{R})$. It is straightforward to show that the set of symplectic matrices form a group; that is if $\mathbf{A}$ and $\mathbf{B}$ are both in $Sp(2,\mathbb{R})$, then so is $\mathbf{AB}$. It can also be shown that $\det \mathbf{A} = 1\ \forall \mathbf{A} \in Sp(2,\mathbb{R})$. For the special case of 2×2 matrices the symplectic group is also the special linear group $Sl(2,\mathbb{R})$.

The general n-dimensional symplectic group is denoted $Sp(2n,\mathbb{R})$. For any $\mathbf{A} \in Sp(2n,\mathbb{R})$ the relation (2) still holds, however, the symplectic matrix $\mathbf{J}$ has the form

$$\mathbf{J} \triangleq \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & 0 \end{pmatrix} \tag{3}$$

where $\mathbf{I}$ is the $n\times n$ identity matrix.

2.2 Lie Groups and Lie Algebras

A *Lie group* is a differentiable manifold that is also a mathematical group. The group of symplectic matrices is a Lie group. The group operation (matrix multiplication) respects the differentiable structure of the group as a set; this is an important fact when constructing the *Lie algebra* of a Lie group.

With every Lie group there is an associated Lie algebra, however, multiple Lie groups may project to the same Lie algebra. Let L be a Lie group in $\mathbb{R}^{2n\times 2n}$ and $\mathfrak{L}$ be its Lie algebra. The vector space of $\mathfrak{L}$ is identified as the tangent plane of L at the identity element $\mathbf{I} \in \mathbb{R}^{2n\times 2n}$. Specifically, if $\mathbf{\Phi}(\cdot)$ is any smooth curve on $L \subset \mathbb{R}^{2n\times 2n}$ such that $\mathbf{\Phi}(0) = \mathbf{I}$, then $\mathbf{\Phi}'(0) \triangleq \lim_{s\to 0} \frac{\mathbf{\Phi}(s)-\mathbf{I}}{s} \in \mathbb{R}^{2n\times 2n}$ is in the tangent plane of L at $\mathbf{I}$, and thus, in $\mathfrak{L}$. The algebra of $\mathfrak{L}$ is formed by adding the multiplication operation, given by the matrix commutator $[\cdot,\cdot]$ where $[\mathbf{A}, \boldsymbol{B}] = \mathbf{AB} - \mathbf{BA}$. The properties of $[\cdot,\cdot]$ are what differentiate Lie algebras from ordinary algebras.

As shown above we can identify the set of all smooth curves in the neighborhood of $\mathbf{I}$ in L with the set of all elements in the Lie algebra $\mathfrak{L}$. Conversely, if $\mathbf{A} \in \mathfrak{L} \subset \mathbb{R}^{2n\times 2n}$ is constant then

$$\mathbf{\Phi}_{\mathbf{A}}(s) \triangleq e^{s\mathbf{A}} \tag{4}$$

is a smooth curve in $\mathbb{R}^{2n\times 2n}$ in L. It is then easy to show that $\{\mathbf{\Phi}_{\mathbf{A}}(s) | s \in \mathbb{R}\}$ is a one-parameter subgroup of L where $\mathbf{A}$ is the generator of this subgroup and $\exp(\cdot)$ is the matrix exponential (see below). Moreover, $\mathbf{\Phi}_{\mathbf{A}}(\cdot) \subset L$ is the solution to the linear, first-order differential system

$$\mathbf{\Phi}_{\mathbf{A}}'(s) = \mathbf{A}\mathbf{\Phi}_{\mathbf{A}}(s)\,, \tag{5}$$

Every $\mathbf{A}$ in a Lie algebra $\mathfrak{L}$ generates a solution $\mathbf{\Phi}_{\mathbf{A}}$ to the above differential equation that is contained in the corresponding Lie group L. The image of $\exp \mathfrak{L} \to L$ for small s is called the *lift* of $\mathfrak{L}$ in L. As mentioned before multiple Lie groups can project to the same Lie algebra; likewise, a single Lie algebra may lift to multiple Lie groups.

If a Lie group L is defined by some conservation property, then the property manifests itself in the Lie algebra $\mathfrak{L}$. For example, consider the symplectic group $Sp(2n,\mathbb{R})$. If $\mathbf{\Phi}(\cdot)$ is a curve on $Sp(2n,\mathbb{R})$ passing through the identity at $s=0$ and with $\mathbf{A} \triangleq \mathbf{\Phi}'(0)$ in $sp(2n,\mathbb{R})$ the Lie algebra of $Sp(n,\mathbb{R})$, then the symplectic condition (2) implies

$$\frac{d}{ds}[\mathbf{\Phi}^T(s)\mathbf{J}\mathbf{\Phi}(s)=\mathbf{J}]_{s=0} \Longrightarrow \mathbf{\Phi}^{T'}(0)\mathbf{J}\mathbf{\Phi}(0)+\mathbf{\Phi}^T(0)\mathbf{J}\mathbf{\Phi}'(0)=0,$$

or

$$\mathbf{A}^T\mathbf{J}+\mathbf{J}\mathbf{A}=0. \tag{6}$$

Since the choice of $\mathbf{\Phi}(\cdot)$ was arbitrary, the above must be a necessary condition for every $\mathbf{A}$ in $sp(2n,\mathbb{R})$, the Lie algebra of $Sp(2n,\mathbb{R})$.

2.3 Matrix Exponential Map

The matrix exponent $e^{\mathbf{A}}$ is defined by its Taylor series

$$e^{\mathbf{A}} \triangleq \mathbf{I}+\mathbf{A}+\frac{1}{2}\mathbf{A}^2+\frac{1}{2\cdot 3}\mathbf{A}^3+\cdots. \tag{7}$$

If $\|\cdot\|$ is any matrix norm then it is straightforward to show $\left\|e^{\mathbf{A}}\right\| \le e^{\|\mathbf{A}\|}$ and, consequently, the series converges in $\|\cdot\|$. Also, if λ is an eigenvalue of $\mathbf{A}$ then e^{λ} is an eigenvalue of $e^{\mathbf{A}}$.

The matrix exponential $e^{\mathbf{A}}$ can be computed numerically using formula (7) since it is convergent for all $\mathbf{A}$ such that $\|\mathbf{A}\|<\infty$. However, it is best to condition $\mathbf{A}$ before doing so. For example, first find the smallest integer m such that $\frac{1}{2^m}\|\mathbf{A}\|<1$, or $m=\lceil \log\|\mathbf{A}\| / \log 2\rceil$. Next we compute the exponential of $\frac{1}{2^m}\mathbf{A}$ using formula (7) to achieve $e^{\frac{1}{2^m}\mathbf{A}}$. The formula will converge rapidly and can be further expedited by accumulating the term $\frac{2^{-mn}}{n!}\mathbf{A}^n$ by multiplication with $\frac{2^{-m}}{n}\mathbf{A}$. Matrix $e^{\mathbf{A}}$ is found by squaring $e^{\frac{1}{2^m}\mathbf{A}}$ m times.

Using the matrix exponential definition it is easy to show that the commutator in the Lie algebra can be interpreted as "loop deficiencies" in the Lie groups. Explicitly, for any $\mathbf{A},\mathbf{B}\in\mathfrak{L}$ and small $s\in\mathbb{R}$, consider the path in L defined by $e^{s\mathbf{A}}e^{s\mathbf{B}}e^{-s\mathbf{A}}e^{-s\mathbf{B}}$ starting at the identity. This path moves in the $-\mathbf{B}$ direction a distance s then an equal distance in the $-\mathbf{A}$ direction (note that $e^{-s\mathbf{A}}=[e^{s\mathbf{A}}]^{-1}$). The loop then winds back following $\mathbf{B}$ and $\mathbf{A}$. After expanding, the terminal location of the path is

$$e^{s\mathbf{A}}e^{s\mathbf{B}}e^{-s\mathbf{A}}e^{-s\mathbf{B}}=\left[\mathbf{I}+s\mathbf{A}+\frac{s^2}{2}\mathbf{A}^2+\cdots\right]\left[\mathbf{I}+s\mathbf{B}+\frac{s^2}{2}\mathbf{B}^2+\cdots\right] \tag{8}$$

Thus, the commutator in $\mathfrak{L}$ indicates the "energy" obtained from traversing loops in the group; it is analogous to the outer product of vector mechanics (actually, Euclidean 3-space with outer product is a Lie algebra).

2.4 Some Mathematical Facts

Here we state some mathematical facts that are needed in the sequel. The following theorem on the matrix exponential that has particular relevance to accelerator physics and Lie groups:

Theorem 1 (Campbell-Baker-Hausdorff): Given two matrices $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n\times n}$ sufficiently close to the origin $\mathbf{0}$, there is a well-defined matrix $\mathbf{C} \in \mathbb{C}^{n\times n}$ such that

$$e^{\mathbf{A}}e^{\mathbf{B}} = e^{\mathbf{C}}. \tag{9}$$

The first few terms of the expansion for $\mathbf{C}$ are

$$\mathbf{C} = \mathbf{A} + \mathbf{B} + \frac{1}{2}[\mathbf{A}, \mathbf{B}] + \frac{1}{12}\big([\mathbf{A}, [\mathbf{A}, \mathbf{B}]] - [\mathbf{B}, [\mathbf{B}, \mathbf{A}]]\big) + \cdots \tag{10}$$

Here we see both that the smooth mapping $\exp \mathfrak{L} \to L$ does not respect the group operation of vector space $\mathfrak{L}$. However, the multiplication on $\mathfrak{L}$ provides the first-order degree by which it is violated.

Proof: See [1].

Our next fact concerns the solutions of linear matrix differential equations with variable coefficients.

Theorem 2 (Peano-Baker): Given the linear differential system

$$\begin{aligned}\mathbf{\Phi}'(s) &= \mathbf{A}(s)\mathbf{\Phi}(s),\\ \mathbf{\Phi}(0) &= \mathbf{I}\,,\end{aligned} \tag{11}$$

where $\mathbf{A}(\cdot)$ is an integrable curve on $\mathbb{R}^{n\times n}$, that is, $\mathbf{A}(\cdot) \in L_p(\mathbb{R} \to \mathbb{R}^{n\times n})$, then the solution $\mathbf{\Phi}(\cdot)$ can be represented by the series

$$\begin{aligned}\mathbf{\Phi}(s) = \mathbf{I} + \int_0^s \mathbf{A}(s_1)ds_1 + \int_0^s\int_0^{s_1} \mathbf{A}(s_1)\mathbf{A}(s_2)ds_2ds_1 \\ + \int_0^s\int_0^{s_1}\int_0^{s_2} \mathbf{A}(s_1)\mathbf{A}(s_2)\mathbf{A}(s_3)\, ds_3ds_2ds_1 + \cdots\end{aligned} \tag{12}$$

where the series continues *ad infinitum*.

Proof: Direct differentiation of the above.

It can be readily shown that the above formula reduces to the usual matrix exponential when $\mathbf{A}$ is a constant.

Another standard result concerning transfer matrices is

Theorem 3 (Semi-group property): Let $\mathbf{\Phi}(s, s_0)$ denote the solution to Eqs. (11) starting at some $s_0 \geq 0$ so that $\mathbf{\Phi}(s_0, s_0) = \mathbf{I}$. Then for any $s > s_0 > 0$

$$\mathbf{\Phi}(s, 0) = \mathbf{\Phi}(s, s_0)\mathbf{\Phi}(s_0, 0). \tag{13}$$

Proof: See [2].

Note that the above relation can be applied multiple times. For example, if we divide the interval $[0, s]$ into subintervals at locations $0 < s_0 < s_1 < \cdots < s_N < s$ then $\mathbf{\Phi}(s, 0) = \mathbf{\Phi}(s, s_N)\mathbf{\Phi}(s_N, s_{N-1}) \ldots \mathbf{\Phi}(s_1, s_0)\mathbf{\Phi}(s_0, 0)$. In other words, the mapping $s_0 + s_1 \longmapsto \mathbf{\Phi}(s_1, s_0)\mathbf{\Phi}(s_0, 0)$ respects addition on the real line.

The next lemma follows from the Peano-Baker series and a commutator requirement.

Lemma 4: Let $\mathbf{A}: \mathbb{R}_+ \to \mathbb{R}^{n \times x}$ be an integrable matrix function and denote its (Riemann) integral by $\aleph(s)$, specifically,

$$\aleph(s) \triangleq \int_0^s \mathbf{A}(\sigma) d\sigma \ . \tag{14}$$

If $\mathbf{A}(s)$ and $\aleph(s)$ commute for all s, that is, $[\mathbf{A}(s), \aleph(s)] = 0 \ \forall s \in \mathbb{R}_+$, then the solution to system (11) is

$$\mathbf{\Phi}(s) = e^{\aleph(s)} = e^{\int_0^s \mathbf{A}(\sigma) d\sigma}. \tag{15}$$

Proof (Sketch): The proof is inductive, applied to each successive term in (12). First consider

$$\int_0^s [\aleph^2(s_1)]' ds_1 = \int_0^s [\mathbf{A}(s_1)\aleph(s_1) + \aleph(s_1)\mathbf{A}(s_1)] = 2\int_0^s \int_0^{s_1} \mathbf{A}(s_1)\mathbf{A}(s_2) ds_2 ds_1,$$

where the last equality follows from the condition $[\mathbf{A}(s), \aleph(s)] = 0$ and the definition of $\aleph$**.** From the above we can identify the third term in the Peano-Baker series (12)

$$\int_0^s \int_0^{s_1} \mathbf{A}(s_1)\mathbf{A}(s_2) ds_s ds_1 = \frac{1}{2!}\aleph^2(s) \ . \tag{16}$$

Likewise, for the fourth term consider

$$\int_0^s [\aleph^3(s_1)]' ds_1 = \int_0^s [\mathbf{A}(s_1)\aleph^2(s_1) + \aleph(s_1)\boldsymbol{A}(s_1)\aleph(s_1) + \aleph^2(s_1)\boldsymbol{A}(s_1)] ds_1 ,$$
$$= 3\int_0^s \mathbf{A}(s_1)\aleph^2(s_1) ds_1 \ ,$$
$$= 3 \cdot 2 \int_0^s \int_0^{s_1} \int_0^{s_2} \mathbf{A}(s_1)\mathbf{A}(s_2)\mathbf{A}(s_3) ds_3 ds_2 ds_1 \, .$$

where the second line follows from the commutator relationship and the third line upon substituting the previous result (16). Analogously each term in the Peano-Baker series is generated from the previous. The resulting general formula for the n^{th} repeated integral is

$$\int_0^s \dots \int_0^{s_{n-1}} \mathbf{A}(s_1) \dots \mathbf{A}(s_n) ds_1 \dots ds_n = \frac{1}{n!}\aleph^n(s), \tag{17}$$

which, when substituted into (12), yields

$$\begin{aligned} \mathbf{\Phi}(s) &= \mathbf{I} + \aleph(s) + \frac{1}{2}\aleph^2(s) + \frac{1}{3!}\aleph^3(s) + \cdots \\ &= e^{\aleph(s)} \, , \end{aligned} \tag{18}$$

completing the proof.

Solution (15) is in direct analogue with the scalar case. Note that the condition $[\mathbf{A}(s), \aleph(s)] = 0$ in the above lemma is very restrictive. The most common application is when $\mathbf{A}(s) = k(s)\mathbf{G}$ where $k(\cdot)$ is an integrable function and $\mathbf{G}$ is a constant matrix.

3 Mechanics

Linear beam optics, whether derived from the equations of motion or Hamiltonian formalism, can be represented as a first-order, matrix-vector differential equation. Considering only the horizontal phase plane we have

$$\mathbf{x}'(s) = \boldsymbol{G}(s)\boldsymbol{x}(s) \, , \tag{19}$$

where $\mathbf{x} \triangleq (x, x')^T$ is the particle phase vector, $\mathbf{G}(s) \in \mathbb{R}^{2\times 2}$ describes the dynamics, and $\mathbf{x}_0$ is the initial condition of the particle at s = 0. The matrix $\mathbf{G}$ is termed the *generator matrix* for the dynamics. The solution to (19) is

$$\mathbf{x}(s) = \mathbf{\Phi}_{\mathbf{G}}(s)\mathbf{x}_0 \, , \tag{20}$$

where $\mathbf{\Phi}_{\mathbf{G}}(s) \triangleq e^{\mathbf{G}s}$ when $\mathbf{G}$ is constant and given by the Peano-Baker series (12) when not. The matrix $\mathbf{\Phi}_{\mathbf{G}}(s)$ is referred to as the *transfer matrix* for system (19).

For beam envelope simulation where $\boldsymbol{\sigma}$ is the symmetric matrix of second order moments (given by $\boldsymbol{\sigma} = \langle \mathbf{x}\mathbf{x}^T \rangle$), the dynamics are

$$\boldsymbol{\sigma}(s) = \boldsymbol{\Phi}_{\mathbf{G}}(s)\boldsymbol{\sigma}_0\boldsymbol{\Phi}_{\mathbf{G}}^T(s), \tag{21}$$

where $\boldsymbol{\sigma}_0$ is the initial value. In each case the dynamics are governed by $\mathbf{G}(\cdot)$ and, consequently, $\boldsymbol{\Phi}_{\mathbf{G}}(\cdot)$.

In beam optics $\mathbf{G}$ is typically one of the following constant matrices:

$$\mathbf{G}_0 \triangleq \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{G}_K \triangleq \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{G}_F(k) \triangleq \begin{pmatrix} 0 & 1 \\ -k^2 & 0 \end{pmatrix}, \quad \mathbf{G}_D(k) \triangleq \begin{pmatrix} 0 & 1 \\ k^2 & 0 \end{pmatrix}, \tag{22}$$

where k is the "focusing constant." The subscripts 0, K, F, and D refer to "drift", "kick", "focus", and "defocus", respectively. It can be confirmed that $\mathbf{G}_0, \mathbf{G}_K, \mathbf{G}_F, \mathbf{G}_D$ all satisfy the symplectic algebra condition (6). With the addition of matrix

$$\mathbf{E} \triangleq \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{23}$$

the set $span\{\mathbf{G}_0, \mathbf{G}_K, \mathbf{G}_F, \mathbf{G}_D, \mathbf{E}\}$ is complete under the commutator $[\cdot,\cdot]$ and, thus, generates the symplectic lie algebra $sp(2,\mathbb{R})$. Since $\mathbf{G}_F(k) = \mathbf{G}_0 - k^2\mathbf{G}_K$ and $\mathbf{G}_D(k) = \mathbf{G}_0 + k^2\mathbf{G}_K$ the set $\{\mathbf{G}_0, \mathbf{G}_K, \mathbf{E}\}$ is really all that is needed to generate the algebra. The algebra generator relations are

$$[\mathbf{G}_0, \mathbf{G}_K] = \mathbf{E}\,, \qquad [\mathbf{E}, \mathbf{G}_0] = 2\mathbf{G}_0\,, \qquad [\mathbf{E}, \mathbf{G}_K] = -2\mathbf{G}_K\,, \tag{24}$$

which yields

$$\begin{aligned} &[\mathbf{G}_F(k), \mathbf{G}_0] = +k^2\mathbf{E}, \quad [\mathbf{G}_F(k), \mathbf{G}_K] = \mathbf{E}, \quad [\mathbf{E}, \mathbf{G}_F(k)] = 2\mathbf{G}_F(k), \\ &[\mathbf{G}_D(k), \mathbf{G}_0] = -k^2\mathbf{E}, \quad [\mathbf{G}_D(k), \mathbf{G}_K] = \mathbf{E}, \quad [\mathbf{E}, \mathbf{G}_D(k)] = 2\mathbf{G}_D(k). \end{aligned} \tag{25}$$

Finally,

$$[\mathbf{G}_F(k_F), \mathbf{G}_D(k_D)] = (k_F^2 + k_D^2)\mathbf{E}\,. \tag{26}$$

Note that equations (24) are all that is necessary to define the algebra $sp(2,\mathbb{R})$, the rest are listed for the sake of completeness.

Before closing the section we present the generator matrix and transfer matrix for an RF gap modeled as a thin lens. Since the gap involves a change in beam energy it cannot be symplectic, however, the same basic relation exists between generator and transfer matrix. Let $\eta \triangleq \beta_0\gamma_0/\beta_1\gamma_1$ be the ratio of pre-gap particle momentum to post-gap particle momentum (β being the normalized particle velocity and γ the relativistic factor). Then the generator matrix $\mathbf{G}_{RF}$ for the gap in the transverse directions is

$$\mathbf{G}_{RF} \triangleq \begin{pmatrix} 0 & 0 \\ -\theta\dfrac{\log\eta}{1-\eta} & \eta \end{pmatrix}, \tag{27}$$

where θ is the transverse phase advance through the gap. The transfer matrix $\boldsymbol{\Phi}_{RF}$ for the transverse directions is given by

$$\boldsymbol{\Phi}_{RF} = e^{\mathbf{G}_{RF}s} = \begin{pmatrix} 1 & 0 \\ \theta & \eta \end{pmatrix}. \tag{28}$$

For the longitudinal plane replace η by $\eta\gamma_0^2/\gamma_1^2$ and replace θ by $-2\theta\bar{\gamma}^2/\gamma_1^2$ where $\bar{\gamma}$ is the average of the pre-gap relativistic factor γ_0 and the post-gap relativistic factor γ_1.

4 Numerical Integration

When numerically integrating the dynamics equations often a stepping procedure is used. To solve the dynamics over a distance L we divide the path $[s_0, s_{N-1}]$ into N subsections of length $h_n = s_{n+1} - s_n$ each (lengths h_n need not be the same size). The objective here is that, typically, the generator matrix $\mathbf{G}$ is a function of path length s. By choosing the interval $I_n \triangleq [s_n, s_{n+1}]$ small enough the matrix $\mathbf{G}(s)$ does not change significantly enough to affect the dynamics and we may approximate $\mathbf{G}(s) = \mathbf{G}(s_n) + O(h_n^K)$ for $s \in [s_n, s_{n+1}]$ and some $K > 0$. We select h_n small enough to hold the error term $O(h_n^K)$ in the above approximation below a predetermined error ϵ. The integration then proceeds in steps

$$\mathbf{x}_{n+1} = \mathbf{\Phi}_n \mathbf{x}_n \,, \tag{29}$$

where, since $\mathbf{G}$ is approximately constant over the interval $[s_n, s_{n+1}]$,

$$\mathbf{x}_n \triangleq \mathbf{x}(s_n), \qquad \text{and} \qquad \mathbf{\Phi}_n \triangleq e^{h_n \mathbf{G}(s_n)} \,. \tag{30}$$

Thus we have a set of discrete transfer matrices $\{\mathbf{\Phi}_n\}$ which transport the beam in steps $\{h_n\}$ down the beamline. The accuracy order K is typically determined by the choice of integration technique. (By order of accuracy we mean the numerical error is of order $O(h^K)$.)

Computing the matrix exponent $e^{h_n \mathbf{G}(s_n)}$ for each n can be an expensive procedure. Typically for any simulation the dynamics generator $\mathbf{G}(s)$ varies between one of a handful of known constant matrices $\{\mathbf{G}_1, \mathbf{G}_2, \dots\}$ representing different beamline elements. Thus, the transfer matrices $\{\mathbf{\Phi}_1(h), \mathbf{\Phi}_2(h), \dots\}$ are computed *a priori* and used in the dynamics calculations as necessary.

4.1 Space Charge

Space charge can be modeled as a defocusing force that is dependent upon the size and shape of the beam (computing the magnitude of this force is beyond our scope). Thus, space charge forces can be represented with the generator matrix $\mathbf{G}_D[k_{sc}(s)]$ where $k_{sc}(\cdot)$ is the defocusing "constant" originating from the beam's self forces. However, the beam typically experiences external forces in addition to space charge forces such as focusing and defocusing from quadrupole magnets. Moreover, the effects of space charge are often modeled as kicks with amplitude $k_{sc,n}$ computed at each step h_n.

The beam dynamics generator matrix $\mathbf{G}(s)$ in the presence of space charge can be written in the form

$$\mathbf{G}(s) = \mathbf{G}_{ext}(s) + k_{sc}^2(s)\mathbf{G}_K, \tag{31}$$

where $\mathbf{G}_{ext}$ is the generator matrix for external forces and $k_{sc}(\cdot)$ is the defocusing function arising from space charge. We can assume that $\mathbf{G}_{ext}$ is one of the known generator matrices $\{\boldsymbol{G}_1, \mathbf{G}_2, \dots\}$and has the known transfer matrix function $\mathbf{\Phi}_{ext}(s)$ that is know *a priori*. The transfer matrix $\mathbf{\Phi}_{sc}(s)$ for $k_{sc}^2(s)\mathbf{G}_K$ is trivial since $\mathbf{G}_K^2 = 0$. We have

$$\mathbf{\Phi}_{sc}(s) = \mathbf{I} + k_{sc}^2(s)\mathbf{G}_K \,. \tag{32}$$

If we approximate the transfer matrix $\mathbf{\Phi}_n \triangleq \mathbf{\Phi}(h_n)$ for integration step k as

$$\mathbf{\Phi}_n \approx \mathbf{\Phi}_{sc}(\bar{s}_n)\mathbf{\Phi}_{ext}(h_n), \tag{33}$$

where $\bar{s}_n$ is some s on the interval $[s_n, s_{n+1}]$ so that $\bar{k}_{sc} \triangleq k_{sc}(\bar{s}_n)$ is representative of the average value of k_{sc}, and h_n is the integration step size. Note that $\boldsymbol{G}_{ext}(\bar{s}_n)$ is necessarily representative of $\mathbf{G}_{ext}(s)$ on $[s_n, s_{n+1}]$ generating the transfer matrix $\mathbf{\Phi}_{ext}(h_n)$. Using Campbell-Baker-Hausdorff (10)

$$\mathbf{\Phi}_{sc}(\bar{s}_n)\mathbf{\Phi}_{ext}(h_n) = e^{h_n\left(\bar{k}_{sc}^2\mathbf{G}_K + \mathbf{G}_{ext}(\bar{s})\right) + \frac{h_n^2}{2}\bar{k}_{sc}^2[\mathbf{G}_K, \mathbf{G}_{ext}(\bar{s})] + O\left(h_n^3\right)}. \tag{34}$$

Thus we see that $\mathbf{\Phi}_{sc}(\bar{s}_n)\mathbf{\Phi}_{ext}(h_n)$ is only a first-order accurate approximate for $\mathbf{\Phi}_k$; that is, the error term is of order h_n^2.

Consider now the approximate

$$\mathbf{\Phi}_k \approx \mathbf{\Phi}_{ext}\left(\frac{h_n}{2}\right)\mathbf{\Phi}_{sc}(\bar{s}_n)\mathbf{\Phi}_{ext}\left(\frac{h_n}{2}\right). \tag{35}$$

Multiplying the matrix exponents as before we have

$$\mathbf{\Phi}_{ext}\left(\frac{h_n}{2}\right)\mathbf{\Phi}_{sc}(\bar{s}_n)\mathbf{\Phi}_{ext}\left(\frac{h_n}{2}\right) = e^{h_n\left(\bar{k}_{sc}^2\mathbf{G}_K + \mathbf{G}_{ext}(\bar{s})\right) + O\left(h_n^4\right)}, \tag{36}$$

Thus we have gained two orders of accuracy by adding the extra matrix multiplication.

The above result suggests the use of a leapfrog method when traversing a finite-length beamline element. This technique reduces the number of matrix multiplications by combining the multiplications by $\boldsymbol{\Phi}_{ext}(h_n/2)$ at the end of each integration step resulting in two simultaneous integrations offset by $h_n/2$. Say $h = h_n$ is a constant for each n. Then two steps n and $n+1$ through the element have the transfer matrix $\mathbf{\Phi}(2h)$ given by

$$\mathbf{\Phi}(2h) = \mathbf{\Phi}_{ext}\left(\frac{h}{2}\right)\mathbf{\Phi}_{sc}(\bar{s}_{n+1})\mathbf{\Phi}_{ext}\left(\frac{h}{2}\right)\mathbf{\Phi}_{ext}\left(\frac{h}{2}\right)\mathbf{\Phi}_{sc}(\bar{s}_n)\mathbf{\Phi}_{ext}\left(\frac{h}{2}\right), \tag{37}$$

Alternatively multiplying by $\mathbf{\Phi}_{ext}(h)$ and $\mathbf{\Phi}_{sc}(\bar{s}_n)$ throughout the element gives an integration scheme which is both symplectic and third-order accurate, so long as one uses $\mathbf{\Phi}_{ext}\left(\frac{h}{2}\right)$ at the entrance and exit. This scheme is a similar to the trapezoidal rule for function integration.

4.2 Field Imperfections

Fringe fields are magnetic fields that deviate from the ideal flat top situation. Instead of falling abruptly to zero, there is a finite region of falloff at the edge of a magnet. Typically the fringe field is completely contained within a drift space. However, *leakage* fields occur when the fringing effect is so dramatic that the "fringe fields" of one magnet leak into the region of an adjacent magnet. We refer to both fringe fields and leakage fields as *field imperfections*.

We have a set of transfer matrices $\{\mathbf{\Phi}_m(s)\}$ for the ideal modeling elements $\{m\}$ and we want to include any field imperfections. Denote by $\boldsymbol{G}_\delta(s)$ the generator matrix for these field imperfections, specifically, the deviation of real-world fields from the ideal fields. For example, $\mathbf{G}_\delta(s)$ could represent the fringe fields in a drift space beyond the hard edge of an ideal magnet. The generator matrix $\mathbf{G}(s)$ for the real-world fields is then given by

$$\mathbf{G}(s) = \mathbf{G}_\delta(s) + \mathbf{G}_m(s), \tag{38}$$

where $\mathbf{G}_m(s)$ is the generator matrix for the ideal beamline element (i.e., the model element). The prescription that most naturally fits into this representation is

$$\mathbf{G}_\delta(s) = k_\delta^2(s)\mathbf{G}_K \,, \tag{39}$$

where $k_\delta^2(\cdot)$ is the focusing function for the field deviation. For example, a focusing quadrupole Q might be modeled with the generator $\mathbf{G}_Q(s) = k_\delta^2(s)\mathbf{G}_K + \mathbf{G}_F(k_F)$ inside the flattop region and $\mathbf{G}_Q(s) = k_\delta^2(s)\mathbf{G}_K + \mathbf{G}_0$ outside.

4.2.1 Third-Order Technique

As with space charge, a simple, third order method for including the field imperfections is to apply a leapfrog integration technique using the decomposition of Eq. (38). This is a viable technique due to the simple form of $\mathbf{\Phi}_\delta(\cdot)$, the field imperfection transfer matrix. The approach requires that we re-compute $\mathbf{G}_\delta(\bar{s}_n) = k_\delta^2(\bar{s}_n)$ at each position $\bar{s}_n \in [s_n, s_{n+1}]$ and each step length h_n, which in and of itself is not difficult. The difficulty arises in bookkeeping; assigning a k_δ to an element m that is generated by an adjacent element, say m+1. For example, consider the leakage fields in one quadrupole magnet that originate from another quadrupole magnetic.

4.2.2 First-Order Technique

It would be convenient to treat field imperfections as a separate modeling element $\mathbf{\Phi}_\delta$. Define

$$\mathbf{\Gamma}_\delta(s) \triangleq \int_0^s \mathbf{G}_\delta(s_1)ds_1 = \int_0^s k_\delta^2(s_1)ds_1\,\mathbf{G}_K \,, \tag{40}$$

then $[\mathbf{G}_\delta(s), \mathbf{\Gamma}_\delta(s)] = 0$ for all s, so by Lemma 1

$$\mathbf{\Phi}_\delta(s) = \exp\left[\int_0^s k_\delta^2(s_1)ds_1\,\mathbf{G}_K\right], \tag{41}$$

where the second line follows from the idempotency of $\mathbf{G}_K$. By defining

$$\kappa_\delta(s) \triangleq \int_0^s k_\delta^2(s_1)ds_1, \tag{42}$$

The modeling element has the convenient form

$$\mathbf{\Phi}_\delta(s) = \mathbf{I} + \kappa_\delta(s)\mathbf{G}_K. \tag{43}$$

Let us now find an error estimate when modeling the real-world transfer matrix $\mathbf{\Phi}(s)$ by the composite

$$\mathbf{\Phi}(s) \approx \mathbf{\Phi}_\delta(s)\mathbf{\Phi}_m(s). \tag{44}$$

This is a convenient model since we can simulate the beamline element $\mathbf{\Phi}(s)$ with the ideal element $\boldsymbol{\Phi}_m(s)$ after which we include the field imperfections using $\mathbf{\Phi}_\delta(s)$. Assume that the ideal beamline element generator $\mathbf{G}_m$ is constant (this is not unreasonable since the term "ideal" usually refers to exactly this condition). Define the *corrector* matrix $\boldsymbol{\chi}(s)$ as

$$\boldsymbol{\chi}(s) \triangleq \mathbf{\Phi}(s)\mathbf{\Phi}_m^{-1}(s)\mathbf{\Phi}_\delta^{-1}(s) \tag{45}$$

Applying Eqs. (10), (38), and (43) and expanding everything in site yields

$$\boldsymbol{\chi}(s) = \left(e^{\kappa_\delta(s)\mathbf{G}_K + s\mathbf{G}_m}\right)\left(e^{-s\mathbf{G}_m}\right)\left(e^{-\kappa_\delta(s)\mathbf{G}_K}\right), \tag{46}$$

where the ellipsis refer to terms of third order and higher in s and $\kappa_\delta(s)$. Notice the similarity between Eq. (8) and the above. The difference $\||\boldsymbol{\chi}(s) - \mathrm{I}|\|$ is a measure of the inaccuracy of the approximation $\mathbf{\Phi}_\delta(s)\mathbf{\Phi}_m(s)$. Not only does $\boldsymbol{\chi}(s)$ contain the deviation from the true solution $\mathbf{\Phi}(s)$, it can correct this deflection; indeed, from its definition we have $\mathbf{\Phi}(s) = \boldsymbol{\chi}(s)\mathbf{\Phi}_\delta(s)\mathbf{\Phi}_m(s)$. By verifying that $[\mathbf{G}_m, \mathbf{G}_K]$ satisfies Eq (6) we see that $\boldsymbol{\chi}(s)$ is symplectic and, thus, can be safely used as a corrector matrix for the integration. However, in the case of space charge smooth changes in beam envelope cannot be accurately reproduced when $s\kappa_\delta(s) \gg 0$ because the field imperfections are represented as an impulse.

4.2.3 Example

Let us work out the example of a quadrupole magnet fringe field impinging upon a drift space, or into another quadrupole. Referring to Eqs. (24) and (25) the examples have the same result. The value of $[\boldsymbol{G}_m, \boldsymbol{G}_K]$ is constant and represents the direction of the deviation from ideal (in the Lie algebra), which is $\mathbf{E}$. The factor $\frac{1}{2}s\kappa_\delta(s)$ is the magnitude of this deviation. If we assume a simple linear profile $k_\delta^2(s)$ with field magnitude B_0 and fringe field length l, then

$$\kappa_\delta(s) = \begin{cases} \displaystyle\int_0^s B_0\left(1 - \frac{\sigma}{l}\right)d\sigma & \text{for} \quad s < l \\ \displaystyle\int_0^l B_0\left(1 - \frac{\sigma}{l}\right)d\sigma & \text{for} \quad s > l \end{cases} = \begin{cases} \dfrac{sB_0}{2}\left(2 - \dfrac{s}{l}\right) & \text{for} \quad s < l \\ \dfrac{lB_0}{2} & \text{for} \quad s > l \end{cases} \tag{47}$$

From here we get

$$\boldsymbol{\chi}(s) = \begin{cases} e^{\frac{s^2 B_0}{4}\left(2-\frac{s}{l}\right)\mathbf{E}+\cdots} & \text{for} \quad s < l \\ e^{\frac{slB_0}{4}\mathbf{E}+\cdots} & \text{for} \quad s > l \end{cases} = \begin{cases} \begin{pmatrix} e^{\frac{s^2 B_0}{4}\left(2-\frac{s}{l}\right)} & 0 \\ 0 & e^{-\frac{s^2 B_0}{4}\left(2-\frac{s}{l}\right)} \end{pmatrix} & \text{for} \quad s < l \\ \begin{pmatrix} e^{\frac{slB_0}{4}} & 0 \\ 0 & e^{-\frac{slB_0}{4}} \end{pmatrix} & \text{for} \quad s > l \end{cases} \tag{48}$$

where we have keep only the first-order error term. We see the correction involves increasing the magnitude of particle position and decreasing the magnitude of particle momentum. Thus, the approximation $\boldsymbol{\Phi}_\delta(s)\boldsymbol{\Phi}_m(s)$ for $\boldsymbol{\Phi}(s)$ tends to under-represent the position coordinate and over-represent the momentum coordinate. We can also see that $\boldsymbol{\Phi}_\delta(s)\boldsymbol{\Phi}_m(s)$ is, in general, not an overly accurate model for the true fields. For example, if we add the effects of a fringe field after a drift of length L, then the magnitude of the error in $\boldsymbol{\Phi}_\delta(L)\boldsymbol{\Phi}_0(L)$ is $e^{\frac{lLB_0}{4}} - 1$, which could be quite large depending upon the leakage l and field strength B_0. This is, however, a worst-case scenario since field imperfections are generally hyper linear meaning there is less mass under the integral of $\kappa_\delta(\cdot)$ for the same cutoff L.

4.2.4 Hybrid Technique

It is possible to improve the third order technique using the principles discussed above. Previously we approximated the field imperfection matrix $\boldsymbol{\Phi}_\delta(\bar{s}_n)$ at integration step n by evaluating the generator matrix $\mathbf{G}_\delta(s) = k_\delta^2(s)\mathbf{G}_K$ at some $\bar{s}_n \in [s_n, s_{n+1}]$. Assume the profile $k_\delta(\cdot)$ is known over $[s_n, s_{n+1}]$ we can compute a theoretically more accurate value using Def. (42) and Eq. (43). We have $\boldsymbol{\Phi}_{\delta,n} \triangleq \mathbf{I} + \int_{s_n}^{s_{n+1}} k_\delta^2(\sigma)d\sigma\, \mathbf{G}_K$. Now the value of the field imperfection matrix is first order accurate with error given by $e^{\frac{B_0}{4}h_n^2} - 1 = \frac{B_0}{4}h_n^2 + \frac{B_0^2}{32}h_n^4 + O(h_n^6)$ where B_0 is the magnitude of the leakage field and h_n is the step length at n. Thus, the computation of $\boldsymbol{\Phi}_{\delta,n}$ is first-order accurate and the integration technique is third-order accurate.

5 Conclusion

We have presented a Lie theoretic, matrix-based approach to numerical integration of beam optics equations. The approach is rooted in maintaining the symplectic condition of the simulation while exposing the numerical accuracy of the underlying technique. A leapfrog integration technique for including space charge effects in the beam optics simulation is shown to be third-order accurate and preserves the symplectic requirement of mechanics. An analogous technique can be used to include field imperfections in the simulation. We explored a first-order technique that was easier to implement, however, its applicability is limited due to large errors that can quickly accumulate.

6 References

1. Sattinger, D.H. and O.L. Weaver, *Lie Groups and Algebras with Applications to Physics, Geometry, and Mechanics*1986, New York: Springer-Verlang.
2. Coddington, E.A. and N. Levinson, *Theory of Ordinary Differential Equations*1955, New York: McGraw Hill.