Chapter 7

Repetitive Systems

For the past three decades, repetitive beam optical systems such as synchrotrons and storage rings have represented one of the major tools for high-energy physics studies. Recently, synchrotron radiation from electron storage rings has been increasingly used in fields such as condensed matter physics, chemistry, and biological sciences. All these repetitive devices require that the particles remain confined in the ring, which means that the motion is stable in that their coordinates do not exceed a certain bound $d$. Therefore, the study of the dynamics of repetitive systems has long been a key concern of beam physics.

The description of these repetitive systems in terms of maps relating initial phase space coordinates to final phase coordinates is much the same as for the case of single-pass systems. For the repetitive systems, it is most convenient to compute the map for one complete revolution of the device, the one-turn map. Instead of studying the long-term behavior of the differential equations governing the motion, we study the repeated application of a map. This idea has been used since the beginning of the study of dynamical systems, which originated with the extensive work of Henri Poincare (1993), and in acknowledging this is usually referred to as the Poincare Map.

The study of the Poincare map is indeed sufficient for the understanding of the stability of the entire device. Should the repetitive application of the Poincare map result in unstable motion, then surely the underlying differential equations are unstable because the repeated application of the Poincare map merely corresponds to the evaluation of the solutions of the differential equations at certain fixed times $n \cdot \Delta t$. On the other hand, should the repetitive application of the Poincare map show stability for a certain region $G$, then one merely has to verify that this region $G$ transports through the system once without exceeding $d$ in between, which can usually be achieved without high requirements for accuracy.

The questions associated with the stability of repetitive systems are usually of a quite different nature as in the case of single-pass systems. In case a system is transversed only once, the correction or control of aberrations is the dominating question, and the knowledge of aberrations to sufficiently high order allows a complete understanding of the system under consideration. In the case of repetitive systems the situation is more complicated and subtle since for sufficiently large numbers of iterations, even slight approximations that are made in representing the system may build up over many turns.
To understand this fundamentally important effect qualitatively, we consider a repetitive system that is well behaved in the sense that the particles are stable. Consider the motion of one particle through the Poincare map, and record its phase space position over many turns. Now assume the system is perturbed in such a way that after each turn, the particle’s coordinates are moved by a small amount depending on the position of the particle. In some instances, the various displacements may average out, and the qualitative behavior of the particle is the same as before. On the other hand, it may also happen that at the locations the particle visits, the displacements have a tendency to move in the same direction and thus repeatedly enhance each other, which over longer terms may fundamentally change the character of the motion and even lead to its instability. While qualitatively it is very difficult to characterize this effect in more detail, it is to be expected that consequences are more likely if the particle’s motion is restricted to only a few locations in phase space. In this case, the average of the effect has to be taken over a much smaller region, and hence the likelihood of it averaging out decreases. This effect is particularly pronounced if the particle in fact exactly repeats its motion after $k$ turns, which we call a $k$th-order resonance. Many of these effects can be described very well in a quantitative manner with the so-called normal form methods discussed in detail later.

### 7.1 Linear Theory

While the understanding of the detailed effects of repetitive motion is usually very complicated, as long as the motion is linear, it is completely straightforward. Since the work of Courant and Snyder (1958), the linear theory of repetitive systems has been fully developed and has become a basic tool in the design and understanding of circular machines. In this section, we discuss the stability of the linear Hamiltonian motion and the invariant ellipses and the behavior of the elliptical beam phase space. An overview of various types of lattices used in circular machines can be found in (Berz, Makino, and Wan 1999).

#### 7.1.1 The Stability of the Linear Motion

As discussed previously, the stability of the motion of charged particles is the most important problem for a repetitive system because particles usually have to stay in the system for millions of turns. The stability of the linear motion is determined by the eigenvalues of the transfer matrix; if any of them has an absolute value $>1$, the motion is unstable. This can be seen as follows. Suppose the initial state of a particle is an eigenvector $\vec{v}_j$ with eigenvalue $\lambda_j$. After $n$ turns, the final state is

$$\vec{r}_f = \lambda_j^n \cdot \vec{v}_j; \quad (7.1)$$
therefore, by one of the properties of a linear vector-space norm,

$$||\vec{r}_f|| = |\lambda_j|^n \cdot ||\vec{V}_j||.$$  

Hence, if $|\lambda_j| > 1$, the “length” of $\vec{r}_f$ goes to infinity as $n$ becomes large. In the more general case that $\vec{r}_f$ is a linear combination of eigenvectors, $\vec{r}_f = \sum_k a_k \vec{V}_k$, its component in direction of $\vec{V}_j$ will still increase with a factor of $|\lambda_j|^n$; therefore, the motion appears to be “drawn” toward this dominating eigenvector and ultimately will behave very similarly to Eq. (7.1). Hence, if $|\lambda_{\text{max}}| > 1$, some particles will be lost, and the device is **linearly unstable**.

As shown in Chapter 1, Section 1.4.4, the system of a particle moving in electromagnetic field can be described by a Hamiltonian, which according to Liouville’s theorem (Eq. 1.184) entails that the phase space volume is conserved, and according to (Eq. 1.239) the transfer map is symplectic, which means that its Jacobian $\hat{M}$ satisfies

$$\hat{M} \hat{J} \hat{M}^t = \hat{J},$$  \hspace{1cm} (7.2)

where $\hat{J}$ is the matrix from Eq. (1.86).

A direct result of symplecticity is that the eigenvalues always appear in **reciprocal pairs**. This can be shown from the facts that $\det(\hat{M}) = \det(\hat{J}) = 1$, $\hat{J}^2 = -\hat{I}$ and $\det(\hat{M}^t) = \det(\hat{M})$. From the previous symplectic condition, we obtain $\hat{M} = -\hat{J}(\hat{M}^t)^{-1} \hat{J}$. Hence,

$$\det(\hat{M} - \lambda \hat{I}) = \det(-\hat{J}(\hat{M}^t)^{-1} \hat{J} - \lambda \hat{I}) = \det(-(\hat{M}^t)^{-1} + \lambda \hat{I})$$

$$= \det(-\hat{I} + \lambda \hat{M}^t) = \lambda^{2v} \det \left( \hat{M} - \frac{1}{\lambda} \hat{I} \right)$$

where $2v$ is the dimensionality of phase space. Moreover, the fact that $\hat{M}$ is real requires that any complex eigenvalues also appear in conjugate pairs. Therefore, when $|\lambda| \neq 1$ and $\lambda \neq 0$, they are always in **quadruples** arranged symmetrically with respect to the real axis and the unit circle (Fig. 7.1). The discussion at the beginning of this section shows that the system is linearly stable when all eigenvalues of $\hat{M}$ satisfy $|\lambda| \leq 1$. This implies that the system is **stable when all eigenvalues lie on the unit circle**. In this case, reciprocity and conjugacy are the same, and thus the eigenvalues always occur in pairs.

Next, we study the stability of a system with respect to a small **perturbation**, and under what conditions its stability is sensitive to the perturbation. We first discuss the case in which $\hat{M}$ is nondegenerate, i.e., all $2v$ eigenvalues are distinct. Then it is possible to draw circles centered at each eigenvalue such that none of the circles intersect (Fig. 7.2). If the perturbation is small enough, all eigenvalues of the perturbed system still lie in their respective circles. Suppose that one eigenvalue $\lambda$ moves off the unit circle, which means that $|\lambda| \neq 1$. Symplecticity
would then require that another eigenvalue $\lambda' = 1/\lambda$ exists in the same circle in which $\lambda$ lies. As a result, the total number of eigenvalues is $> 2v$, which is not possible. This shows that nondegenerate stable system remains **stable under small perturbations**.

The previous arguments prove that perturbing a nondegenerate stable system causes the eigenvalues to move along the unit circle. What happens when two **eigenvalues collide**? First, when two eigenvalues collide, their complex conjugates also collide. These four eigenvalues can form a quadruple, can leave the unit circle, and can make the system unstable (Fig. 7.3). Collisions also occur when two complex–conjugate eigenvalues meet on the real axis. These collisions may lead to instabilities because $\text{Im} (\lambda)$ vanishes at the collision point (Fig. 7.4).

Now, we discuss a method to **decouple** a general, not necessarily symplectic, linear matrix into $2 \times 2$ blocks, which will make it much easier to study the evolution of the system. In the following, we assume that the linear part of the phase space map has **distinct eigenvalues**, which is the case for most important systems. Then the linear part of the map can be diagonalized. We now group the eigenvalues such that complex–conjugate eigenvalues form pairs; we group any remaining real eigenvalues into pairs by demanding that the elements of a pair have the same sign. This is always possible since the determinant is positive and thus there is an even number of negative eigenvalues.
FIGURE 7.2. Behavior of nondegenerate eigenvalues under small perturbation.

FIGURE 7.3. Possible behavior of degenerate eigenvalues under small perturbation.
Each pair we write as \( r_je^{\pm i\mu_j} \). In the case of a complex pair, this is readily accomplished by choosing \( r_j \) and \( \mu_j \) as the modulus and phase. In the case of a real pair \( R_{1j} \) and \( R_{2j} \), we choose \( r_j = \pm \sqrt{R_{1j}R_{2j}} \), where the sign is determined to be the same as that of \( R_{1j} \) and \( R_{2j} \). We then choose \( \mu_j = i \cdot \log(\sqrt{R_{1j}/R_{2j}}) \). Since the determinant is nonzero and \( R_{1j} \) and \( R_{2j} \) are of the same sign, \( r_j \) and \( \mu_j \) are always well-defined.

Denoting the eigenvectors corresponding to \( rje^{\pm i\mu_j} \) by \( s_j^\pm \), we obtain that in the eigenvectors basis, the linear part of the map has the form

\[
\begin{pmatrix}
    r_1e^{+i\mu_1} & r_1e^{-i\mu_1} & 0 \\
    0 & \ddots & 0 \\
    0 & 0 & r_ve^{+i\mu_v} & r_ve^{-i\mu_v}
\end{pmatrix}
\]  

(7.3)

We note that if the \( j \)th eigenvalue pair consists of complex conjugate eigenvalues, then so are the associated eigenvectors, and if the \( j \)th eigenvalue pair is real, so are the eigenvectors.

We now perform another change of basis after which the matrix is real. For each complex conjugate pair of eigenvalues, we choose the real and imaginary parts of an eigenvector as two basis vectors. For each pair of real eigenvalues, we choose the two real eigenvectors themselves.
The result of this change of basis is a matrix having $2 \times 2$ subblocks along the diagonal. A subblock originating from a complex eigenvalue pair will have four nonzero entries, and a subblock originating from a real eigenvalue pair will be diagonal. Therefore, the matrix has the form

$$
\begin{bmatrix}
    a_1 & b_1 & 0 & 0 \\
    c_1 & d_1 & 0 & 0 \\
    0 & 0 & a_v & b_v \\
    0 & 0 & c_v & d_v
\end{bmatrix}
$$

(7.4)

We note that if the underlying matrix is symplectic, it is possible to scale the transformation matrix such that it is also symplectic via a Gram-Schmidt like procedure with the antisymmetric scalar product $\langle , \rangle$. Since products of symplectic matrices are symplectic, so is the transformed matrix.

This decoupling provides a convenient stepping stone to the computation of the quantities relevant for a study of stability. With the linear matrix decoupled by the method described previously, or by the symmetry of the system (e.g., the midplane symmetry in Section 3.1.2), it is sufficient to study only one of the $N \times N$ blocks to understand the linear behavior of the system. The characteristic polynomial of a $2 \times 2$ matrix is

$$
\lambda^2 - T \cdot \lambda + D = 0,
$$

(7.5)

where $T$ and $D$ are its trace and determinant, respectively. An immediate result is that $\lambda_1 \lambda_2 = D$. When $D > 1$, the system is always unstable because there would be at least one eigenvalue whose absolute value is $>1$. However, in practice this case will never be chosen deliberately in single-particle systems and therefore will not be discussed further. When $D < 1$, the system is stable not only when the eigenvalues form a complex–conjugate pair but also when they are real numbers and both of them have absolute values smaller than 1; this case occurs when damping is present. When $D = 1$, the system is symplectic; it is stable only when the eigenvalues form a complex–conjugate pair.

The eigenvalues of Eq. (7.5) are given by

$$
\lambda_{1,2} = \frac{T}{2} \pm \sqrt{\left(\frac{T}{2}\right)^2 - D}.
$$

(7.6)

The nature of the eigenvalues depends on the relation between $T$ and $D$.

As the first case, we consider $|T| > 2\sqrt{D}$. Then $\lambda_{1,2}$ are real numbers. If $D < 1$, the system can be stable or unstable depending on the value of $T$. If $D = 1$, the system is unstable and a phase space point from it moves along a hyperbola, which can be seen as follows (Fig. 7.5). Considering motion in eigenvector coordinates
\[ \vec{V}_1 \text{ and } \vec{V}_2, \text{ we have} \]
\[ \vec{x}_0 = a_0 \vec{V}_1 + b_0 \vec{V}_2, \]
\[ \vec{x}_1 = \hat{M} \vec{x}_0 \]
\[ = \lambda_1 a_0 \vec{V}_1 + \lambda_2 b_0 \vec{V}_2 \]
\[ = a_1 \vec{V}_1 + b_1 \vec{V}_2. \]

Thus, the relations between the components are
\[ a_1 = \lambda_1 a_0, \]  
\[ b_1 = \lambda_2 b_0. \]  \hfill (7.7)
\hfill (7.8)

Since \( \lambda_1 \cdot \lambda_2 = D = 1 \), we obtain that
\[ a_1 b_1 = a_0 b_0 = \text{const}. \]  \hfill (7.9)

This curve shows that the particle moves further out and the system is **unstable**.

As the **second case**, we consider \( |T| = 2 \sqrt{D} \). The eigenvalues are degenerate, where
\[ \lambda_{1,2} = \pm \sqrt{D}. \]  \hfill (7.10)

Therefore, the motion is stable when \( D < 1 \). If \( D = 1 \), it is **marginally stable**; that is, it can become unstable under perturbation as discussed previously.

As the **third case**, we study \( |T| < 2 \sqrt{D} \). The eigenvalues are
\[ \lambda_{1,2} = \sqrt{D} \left( \frac{T}{2 \sqrt{D}} \pm i \sqrt{1 - \left( \frac{T}{2 \sqrt{D}} \right)^2} \right). \]  \hfill (7.11)
We introduce

\[ r = \sqrt{D} \]
\[ \mu = \text{sign}(b) \cdot \cos \left( \frac{T}{2\sqrt{D}} \right), \tag{7.12} \]

where \( \text{sign}(b) = +1 \) if \( b \geq 0 \), and \( \text{sign}(b) = -1 \) if \( b < 0 \). The solutions can be written as

\[ \lambda_{1,2} = r \cdot e^{\pm i\mu}. \tag{7.13} \]

It is clear that \(|\lambda_{1,2}| = r\), and hence the motion is stable when \( D \leq 1 \).

It is useful and customary (Courant and Snyder 1958) to introduce a set of new quantities, the so-called Twiss parameters

\[ \alpha = \frac{a - d}{2r \sin \mu} \]
\[ \beta = \frac{b}{r \sin \mu} \]
\[ \gamma = \frac{c}{r \sin \mu}. \tag{7.14} \]

The Twiss parameters satisfy \( \beta \gamma - \alpha^2 = 1 \), so they are not independent. Two of them, together with \( r \) and \( \mu \), determine the matrix \( \hat{M} \). We note that \( \beta \) is always positive.

Using the Twiss parameters, the matrix can be written as

\[ \hat{M} = r \cdot \begin{pmatrix} \cos \mu + \alpha \sin \mu & \beta \sin \mu \\ -\gamma \sin \mu & \cos \mu - \alpha \sin \mu \end{pmatrix}. \tag{7.15} \]

For the sake of completeness we note that in the important case of \( \beta \neq 0 \), the eigenvectors assume the compact form

\[ v_{1,2} = (i\beta, -i\alpha \mp 1). \tag{7.16} \]

The Twiss parameters define the similarity transformation in which the map is diagonal, and we obtain

\[ \hat{A} \cdot \hat{M} \cdot \hat{A}^{-1} = \begin{pmatrix} r \cdot e^{+i\mu} & 0 \\ 0 & r \cdot e^{-i\mu} \end{pmatrix}, \tag{7.17} \]

where

\[ \hat{A}^{-1} = \begin{pmatrix} i\beta & i\beta \\ -1 - i\alpha & 1 - i\alpha \end{pmatrix} \quad \text{and} \quad \hat{A} = \begin{pmatrix} (1 - i\alpha)/2i\beta & -1/2 \\ (1 + i\alpha)/2i\beta & +1/2 \end{pmatrix}. \tag{7.18} \]
Since the case \( r > 1 \), and thus \( D > 1 \), leads to unstable motion, it is not of interest to us. When \( r < 1 \), and thus \( D < 1 \), the phase space area shrinks with each turn. This is the case when damping is present. If the damping is strong and \( r \) is significantly smaller than unity, the nonlinear motion is simple, whereas if the damping is weak and thus \( r \) is very close to unity, the motion is similar to the one with \( r = 1 \). Thus, in the rest of this section only stable symplectic motion is studied.

First, the transfer matrix can be written as

\[
\begin{pmatrix}
\cos \mu + \alpha \sin \mu & \beta \sin \mu \\
-\gamma \sin \mu & \cos \mu - \alpha \sin \mu
\end{pmatrix},
\]  

(7.19)

which has the eigenvalues \( \lambda_{1,2} = e^{\pm i\mu} \). Here, the quantity \( \mu \) is a very important characteristic parameter of a repetitive system, and is called the tune. The eigenvectors satisfy the same relation as the eigenvalues, \( \vec{V}_1 = \vec{V}_2 \). We now introduce a new basis that is defined by the unit vectors \( \vec{V}_+ = \text{Re}(\vec{V}_1) \) and \( \vec{V}_- = \text{Im}(\vec{V}_1) \). Expressed in these unit vectors, the motion is said to be in linear normal form. In this coordinate system, we have

\[
\begin{pmatrix}
\cos \mu & \sin \mu \\
-\sin \mu & \cos \mu
\end{pmatrix},
\]  

(7.20)

which means that any particle moves in a circle; therefore, the motion is stable under iteration (Figure 7.6). The angular advance in the normal coordinates is constant for each iteration and equals the tune \( \mu \).

Note that in the original coordinates, the angular advance does not have to be constant from turn to turn; but it can quickly be determined that over many turns, the average angular advance equals the tune. For this purpose we observe that the angle advance in the normal coordinates can be described by the angle advance due to the transformation to linear normal form coordinates plus the angle advance in linear normal form coordinates plus the angle advance of the reverse transformation. Since for many iterations, the transformation advances contribute only an insignificant amount, the average approaches the angle in the normal form coordinates.

It is very important in practice to ensure the stability of the accelerator when it is under small perturbation. As shown previously, this requires that the traces of any submatrix lie between \(-2\) and \(2\) and also that the traces be different from one another. Furthermore, the choice of the traces, and thus the tunes, is crucial to the nonlinear behavior of a repetitive system and often a delicate one in the sense of trying to avoid the so-called resonances.

To conclude this discussion, we address another aspect of repetitive motion, namely, the behavior of off-energy particles. In case there are no accelerating structures, such as in a plain FODO lattice (focusing element, defocusing element, and drifts denoted by O), the energies of the individual particles stay constant in time but affect the motion in that particles with higher energy are bent less
in dipoles and in quadrupoles. Therefore, the existence of dipoles, which are always present in repetitive systems, causes the off-energy particles that move along the center to drift away from it. Likewise, the focusing power of quadrupoles is changed for off-energy particles. Consequently, closed orbit as well as the tunes of the off-energy particles are likely to differ from those of on-energy particles. The tune shifts caused by the energy deviation are customarily called chromaticities.

The presence of different off-energy closed orbits and tunes leads to potentially different stability properties of off-energy particles. Therefore, it is a standard practice in accelerator design to eliminate at least first-order chromaticities, i.e., the linear dependence of the tunes on energy, in order to increase the range of energy acceptance.

We now address the linear behavior of the off-energy particles, which can be described within a matrix theory. The full nonlinear behavior of these particles will be discussed in detail in Section 7.2.3. For a lattice satisfying midplane symmetry, $(y|d)$ and $(b|d)$ vanish; hence, a $3 \times 3$ matrix can describe the motion:

$$\hat{M}_x = \begin{pmatrix} x|x & x|a & x|d \\ a|x & a|a & a|d \\ 0 & 0 & 1 \end{pmatrix}. \quad (7.21)$$

When studying the eigenvalue spectrum of this matrix, different from the $2 \times 2$ case, there is now an additional eigenvalue, which can be read to be $1$. The corresponding eigenvector is a fixed point of the matrix and is denoted as $(\eta, \eta', 1)$. It is easy to show that

$$\eta = \frac{1 - (a|a)(x|d) + (x|a)(a|d)}{2(1 - \cos \mu)}, \quad (7.22)$$

$$\eta' = \frac{1 - (x|x)(a|d) + (a|x)(x|d)}{2(1 - \cos \mu)}. \quad (7.23)$$

The values of $\eta$ and $\eta'$ describe position and momentum of the fixed point of an off-energy particle with $d = 1$. Similarly, the fixed point of an off-energy particle with another value of $d$ is given by $(\eta d, \eta'd)$ because $(\eta d, \eta'd, d)$ is also a fixed point of $\hat{M}_x$. Apparently, the value of $\eta$ for different positions around the accelerator is an important quantity for its description, and it is commonly referred to as the eta-function.

The knowledge of $\eta$ and $\eta'$ allows the introduction of new relative coordinates around the fixed point. In these coordinates, the matrix of the linear motion assumes the form

$$\hat{M}_x^* = \begin{pmatrix} (x|x)^* & (x|a)^* & 0 \\ (a|x)^* & (a|a)^* & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (7.24)$$
and in particular has the same $2 \times 2$ block form as before. This allows the computation of tunes of off-energy particles as $\arccos(((x|x) + (q|q)/2)$ and hence their dependence on different values of $d$.

### 7.1.2 The Invariant Ellipse of Stable Symplectic Motion

In this section we study the dynamics in a linear stable symplectic system. As shown in Eq. (7.19), the matrix has the form

$$
\hat{M} = \begin{pmatrix} 
\cos \mu + \alpha \sin \mu & \beta \sin \mu \\
-\gamma \sin \mu & \cos \mu - \alpha \sin \mu 
\end{pmatrix},
$$

and we have $\det(\hat{M}) = 1$, which implies that $\beta \gamma - \alpha^2 = 1$.

First, we prove that

$$
\hat{M}^t \begin{pmatrix} \gamma & \alpha \\
\alpha & \beta \end{pmatrix} \hat{M} = \begin{pmatrix} \gamma & \alpha \\
\alpha & \beta \end{pmatrix}.
$$

Let $\hat{T} = \begin{pmatrix} \gamma & \alpha \\
\alpha & \beta \end{pmatrix}$. We introduce the matrix $\hat{K} = \begin{pmatrix} \alpha & \beta \\
-\gamma & -\alpha \end{pmatrix} = \hat{J} \hat{T}$.

Then $\hat{M}$ can be expressed in terms of $\hat{K}$ and the $2 \times 2$ unity matrix $\hat{I}$:

$$
\hat{M} = \hat{I} \cos \mu + \hat{K} \sin \mu.
$$
We have

\[
\hat{T}\hat{K} = \left( \begin{array}{cc} \gamma_i & \alpha_i \\ \alpha_i & \beta_i \end{array} \right) \left( \begin{array}{cc} \alpha_i & \beta_i \\ -\gamma_i & -\alpha_i \end{array} \right) = \hat{J}
\]

\[
\hat{K}^t\hat{T} = \hat{K}^t\hat{T}^t = \hat{J}^t = -\hat{J}
\]

\[
\hat{K}^t\hat{T}\hat{K} = -\hat{J}\hat{K} = \hat{T}
\]

and so

\[
\hat{M}^t\hat{T}\hat{M} = (\hat{I} \cos \mu + \hat{K}^t \sin \mu)\hat{T}(\hat{I} \cos \mu + \hat{K} \sin \mu)
\]

\[= \hat{T}.\]

It follows that the quadratic form,

\[
\hat{T} \hat{z} \hat{T} \hat{z} = \varepsilon,
\]

remains invariant under the transformation \(\hat{M}\). This relation entails that if the envelope of a beam at the beginning fills the ellipse described by the quadratic form, it will remain the same after one turn and hence after any number of turns. This ellipse is an invariant of motion, which is called the invariant ellipse or the Courant–Snyder invariant. A beam that fills an invariant ellipse is said to be matched to the repetitive system. If the beam is not matched, the beam size at a certain point will vary from turn to turn, which is referred to as beating (Fig. 7.7).

The invariant ellipse of a stable repetitive system is a characteristic of the system and is as important as the tune. Also, it has universal importance, independent of the particular location of the Poincaré section. Consider two Poincaré sections at different locations around a repetitive system, and let \(\hat{M}_1\) and \(\hat{M}_2\) be the one-turn maps describing the motion from the respective Poincaré sections. Let \(\hat{M}_{12}\) and \(\hat{M}_{21}\) be the maps from section 1 to 2 and from section 2 to 1, respectively. Then \(\hat{M}_1\) and \(\hat{M}_2\) can be expressed in terms of \(\hat{M}_{12}\) and \(\hat{M}_{21}\) via the equations

\[
\hat{M}_1 = \hat{M}_{21} \cdot \hat{M}_{12}
\]

(7.29)

\[
\hat{M}_2 = \hat{M}_{12} \cdot \hat{M}_{21}
\]

(7.30)

By eliminating \(\hat{M}_{21}\), we obtain the following relationship between \(\hat{M}_1\) and \(\hat{M}_2\):

\[
\hat{M}_2 = \hat{M}_{12} \cdot \hat{M}_1 \cdot \hat{M}_{12}^{-1}
\]

(7.31)

This entails that \(\hat{M}_1\) and \(\hat{M}_2\) have the same trace and hence the same tune. Furthermore, if the system is matched at one Poincaré section, it is matched at the other.
7.1.3 Transformations of Elliptical Phase Space

Here, we discuss how a beam traverses phase space. In reality, the shape of the phase space of a beam can be irregular, and there exist various ways to model it. For example, parallelograms and other polygons are sometimes used to represent beams in single-pass systems. The study of the transformation properties of such a polygon is straightforward; since the transformation preserves straight lines, it is sufficient to study the transformation of the corner points of the polygon.

Another important approach is to describe the beam by an ellipse. Similar to polygons, the transformation of ellipses occurs through linear transformations, and it is merely necessary to study the transformation properties of the parameters describing the ellipse. This is particularly advantageous for repetitive systems that are operated most efficiently if the beam fills the invariant ellipse of the system.

In the following we study the propagation of such beam ellipses through linear symplectic transformations. To begin, we study some basic geometric properties of ellipses. An arbitrary ellipse centered at the origin is described by

\[\begin{align*}
(x, a) \begin{pmatrix} \gamma & \alpha \\ \alpha & \beta \end{pmatrix} \begin{pmatrix} x \\ a \end{pmatrix} &= \varepsilon. 
\end{align*}\]  

The quantity \(\varepsilon\) serves as a scaling parameter, and without loss of generality we thus require the determinant of the ellipse to be unity, i.e., \(\beta \gamma - \alpha^2 = 1\). After simple arithmetic, we have

\[\gamma x^2 + 2\alpha x a + \beta a^2 = \varepsilon,\]
from which we determine that the **axis intersections** of the ellipse are

\[ x_0 = \pm \frac{\varepsilon}{\gamma} \quad \text{and} \quad a_0 = \pm \frac{\varepsilon}{\beta}. \] (7.34)

To determine the **maxima** in the \( x \) and \( a \) directions, let

\[ f(x, a) = \gamma x^2 + 2\alpha x a + \beta a^2; \]

then we have

\[ \nabla f = (2\gamma x + 2\alpha a, 2\alpha x + 2\beta a). \] (7.35)

At \( x_m (\nabla f)_a = 0 \); thus, at this point, \( a = -\alpha / \beta \cdot x \). We define the slope \( S \) of the ellipse as

\[ S = \frac{\alpha}{\beta}. \] (7.36)

Inserting \( a \) back into Eq. (7.33), we obtain \( x_m \); in a similar way, we can obtain \( a_m \). Altogether, we have

\[ x_m = \pm \sqrt{\beta \varepsilon} \quad \text{and} \quad a_m = \pm \sqrt{\gamma \varepsilon}. \] (7.37)

As a function of the position \( s \) around the ring, the quantity \( \beta \) is a measure for the beam width and is thus important. The function \( \beta \) is often called the **betatron**
or beta function and \( x_{\text{max}} \) the beam envelope. It is worth noting that the area of the beam ellipse described in Eq. (7.33) is \( \pi \varepsilon \), which is called the emittance (Fig. 7.8).

Now we return to the question of beam ellipses transformations. Suppose at one point a beam fills the ellipse

\[
\begin{pmatrix}
  x_1 \\
  a_1
\end{pmatrix}
\begin{pmatrix}
  \gamma_1 \\
  \alpha_1 \\
  \beta_1
\end{pmatrix}
\begin{pmatrix}
  x_1 \\
  a_1
\end{pmatrix} = \varepsilon. \quad (7.38)
\]

Under a map \( \hat{M} \), the point \((x_1, a_1)\) moves to

\[
\begin{pmatrix}
  x_2 \\
  a_2
\end{pmatrix} = \hat{M} \cdot \begin{pmatrix}
  x_1 \\
  a_1
\end{pmatrix}. \quad (7.39)
\]

Conversely, we have

\[
\begin{pmatrix}
  x_1 \\
  a_1
\end{pmatrix} = \hat{M}^{-1} \cdot \begin{pmatrix}
  x_2 \\
  a_2
\end{pmatrix}. \quad (7.40)
\]

Inserting this relation back into Eq. (7.38), we obtain the new beam ellipse as

\[
(x_2, a_2) \left[ (\hat{M}^{-1})^t \begin{pmatrix}
  \gamma_1 \\
  \alpha_1 \\
  \beta_1
\end{pmatrix} \hat{M}^{-1} \right] \begin{pmatrix}
  x_2 \\
  a_2
\end{pmatrix} = \varepsilon. \quad (7.41)
\]

Since

\[
\left[ (\hat{M}^{-1})^t \begin{pmatrix}
  \gamma_1 \\
  \alpha_1 \\
  \beta_1
\end{pmatrix} \hat{M}^{-1} \right]^t = \left[ (\hat{M}^{-1})^t \begin{pmatrix}
  \gamma_1 \\
  \alpha_1 \\
  \beta_1
\end{pmatrix} \hat{M}^{-1} \right],
\]

the beam still fills an ellipse at point \((x_2, a_2)\). The relations between the initial and final beam parameters are explicitly given as

\[
\begin{pmatrix}
  \beta_2 \\
  a_2 \\
  \gamma_2
\end{pmatrix} = \begin{pmatrix}
  \frac{x|x|^2}{(a|x)^2} & -2(x|x)(a|x) & \frac{(x|a)^2}{(a|a)^2} \\
  -(x|x)(a|x) & (x|x)(a|a) + (x|a)(a|x) & -(x|a)(a|a) \\\n  (a|x)^2 & -2(a|x)(a|a) & (a|a)^2
\end{pmatrix} \begin{pmatrix}
  \beta_1 \\
  \alpha_1 \\
  \gamma_1
\end{pmatrix}. \quad (7.42)
\]

For the special case that the transformation describes one turn around a ring of a beam that is matched to the machine, we would have that \((\beta_2, a_2, \gamma_2) = (\beta_1, a_1, \gamma_1)\). Viewed in terms of Eq. (7.42), this represents an eigenvalue problem and offers another way to compute the invariant ellipse of a given system.
As an example for the transformation properties of a beam ellipse as a function of $s$, we consider the transform through a drift, which has the transfer matrix

$$
\mathbf{M} = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}.
$$

Plugging the matrix elements into Eq. (7.42), we have

$$
\beta_2 = \beta_1 - 2s\alpha_1 + s^2\gamma_1.
$$

Therefore, the beta function of the beam varies quadratically in a drift. The minimum of $\beta_2$ occurs at $s = \alpha_1/\gamma_1$, and this point is called the waist (Fig. 7.9). The betatron function around the waist is often written as

$$
\beta(s) = \beta_* + \frac{s^2}{\beta_*}.
$$

Note that the location of the waist does not have to coincide with the location of an image. Furthermore, we observe that a waist is characterized by $\alpha_2 = 0$.

### 7.2 Parameter-Dependent Linear Theory

So far we have discussed only the linear properties of ideal repetitive systems and their impacts on design considerations. When studying the nonlinear effects of the motion, there are two separate aspects that can be considered. The first is the influence of parameters including energy spread and also magnet misalignments and magnetic field errors on the linear motion, which will be discussed here. The second consideration addresses the full nonlinear dynamics of the particles, which requires more sophisticated tools including normal form methods and is addressed in Section 7.3.
Deviations of various kinds, such as energy spread, magnet misalignments, and magnet field errors, change the linear dynamics in possibly complicated ways. It is very important to study these changes in order to decide on error tolerances for a specific design. Among all the linear quantities discussed previously, changes in tunes usually affect the motion most severely. As discussed previously, it is important to “sample” phase space uniformly in order to not be susceptible to resonances. For tunes, this means that the tune should not be a rational number, and the ratio of tunes should not be rational. Therefore, any shifts in tune due to changes in parameters may potentially destroy the stability of a system, and so it is important to control them. Among all changes in tune, tune shifts caused by energy spread are the most important because the energy spread is usually large compared to other factors.

Differential algebra (DA) methods provide a powerful tool to study the influences of perturbations to linear quantities such as tunes and lattice parameters to very high accuracy. Here, we first study the method of finding parameter-dependent fixed points of a map, which is the foundation for the algorithm of the parameter-dependent tune shifts and normal form theory (see Section 7.3). The algorithm for computing tune shifts is outlined later, as is correction of chromaticities, which is a direct application.

### 7.2.1 The Closed Orbit

As discussed in Section 7.1, the energy dependence of fixed points identifying the closed orbit is very important to accelerator design. Similarly, if large errors are present, it is important to study their influence as well. In general, knowledge of the parameter dependency of fixed points is needed to compute parameter-dependent tune shifts to high orders. Let us assume we are given a set of parameters \( \delta \), which may include the energy of the particle but also important system parameters. As discussed in Chapter 5, the map \( M(\bar{z}_F, \delta) \) of the system depending on the variables \( \bar{z} \) as well as the system parameters \( \delta \) can be obtained to any order using DA methods. Our goal is to establish a nonlinear relationship between the fixed point \( \bar{z}_F \) and the values of the parameters \( \delta \). As always within the perturbative approach, we are interested only in the \( n \)th order expansion of this relationship.

In order to determine how the fixed point \( \bar{z}_F \) depends on the parameters \( \delta \), we observe that the fixed point satisfies

\[
(\bar{z}_F, \delta) = M(\bar{z}_F, \delta).
\]  
(7.45)

This fixed-point equation can be rewritten as

\[
(M - I_H)(\bar{z}_F, \delta) = (\bar{0}, \delta)
\]  
(7.46)

where the map \( I_H \) contains a unity map in the upper block corresponding to the variables \( \bar{z} \) and zeros everywhere else. This form of the fixed-point equation
clearly shows how the parameter-dependent fixed point $\tilde{x}_F$ can be obtained: It is necessary to invert the map $\mathcal{M} - I_H$. Since we are interested only in the properties of the inverse up to order $n$, we work with the equivalence class $[\mathcal{M} - I]_n$ of the map and apply the technique given in Section 2.3.1 to obtain the equivalence class $[\tilde{x}_F]_n$.

As previously determined, we can compute the $n$th order inverse if and only if the linear part of the map is invertible. While this is always the case for transfer maps, here the situation is more subtle; the map is invertible if and only if the phase space part of $\mathcal{M}$ does not have unit eigenvalue. However, since such an eigenvalue can lead to an unstable map under small errors as discussed previously, it is avoided in the design of repetitive systems. Altogether, up to order $n$, the fixed point is given as the upper part of

$$\begin{pmatrix} \tilde{x}_F, \tilde{\delta} \end{pmatrix}_n = [\mathcal{M} - I_H]_n^{-1} (\tilde{\theta}, \tilde{\delta}).$$

(7.47)

7.2.2 Parameter Tune Shifts

As the first step in the computation of tune shifts, we change coordinates by performing a transformation to the parameter-dependent fixed point. In these coordinates, the map is origin preserving, i.e., $\mathcal{M}(\tilde{\theta}, \tilde{\delta}) = \tilde{\theta}$. Furthermore, the linear motion around this new fixed point follows ellipses, as discussed previously. This also implies that all partial derivatives of the final coordinates with respect to parameters alone vanish.

The key consequence of this transformation is that we can now view the map in such a way that the partial derivatives of the final phase space variables with respect to the initial phase space variables and hence the aberrations depend on the system parameters, but the system parameters do not influence the map otherwise. Therefore, in this view, our map now relates initial phase space coordinates to final phase space coordinates, and the expansion coefficients depend on the parameters.

We now compute the linear matrix of motion in which all matrix elements depend on system parameters to order $n$. If the motion is in the form of $2 \times 2$ subblocks, we obtain a matrix $\hat{\mathcal{M}}$ consisting of matrix elements that are now equivalence classes of order $\hat{\mathcal{M}} = n - 1$ depending on the parameters

$$\hat{\mathcal{M}} = \begin{pmatrix} [a]_m & [b]_m \\ [c]_m & [d]_m \end{pmatrix}.$$  

(7.48)

Note that one order is lost in the process: Since $a$ was a first derivative, its $(m)$th derivatives with respect to the parameters are certain $n$th derivatives of the original map $\mathcal{M}$.

The computation of the parameter dependence of the tunes is now quite straightforward. Following common DA practice, we replace all real operations in the computation of the tunes and dampings by the corresponding DA operations.
After the motion is decoupled, i.e., the map is already in $2 \times 2$ block form, this merely involves the computation of the class of $\mu$ from the determinant and trace. In particular, we obtain
\[
[\mu]_m = \text{sign}([b]_m) \cos \left( \frac{[a]_m + [b]_m}{2([a]_m \cdot [d]_m - [b]_m \cdot [c]_m)} \right), \tag{7.49}
\]
where the arcsec and sign operations are acting in DA.

In the case of coupled motion, conceptually the strategy is the same. In addition, one simply replaces all operations in the whole eigenvalue package necessary for bringing the motion to block form by the corresponding ones in DA. Since for low dimensions, good eigenvalue and eigenvector algorithms are very efficient, the additional effort is minor. Altogether, a very compact algorithm results, and compared to the more general DA normal form algorithm discussed later which obtains the parameter-dependent tunes as a by-product, the tune shifts are obtained in an efficient and straightforward way.

### 7.2.3 Chromaticity Correction

As discussed in Section 7.1, the correction of chromaticities plays an important role in accelerator design. Here, we show an immediate and useful application of the algorithm outlined in the previous sections, namely, the correction of chromaticities using system parameters. To this end, we utilize the method of the previous section to write the $\nu$ tunes in terms of the system parameters
\[
\vec{\mu} = T(\vec{\delta}). \tag{7.50}
\]

The map $T$ contains a constant part (the linear tunes) as well as nonlinear parts, and the algorithm of the previous section allowed us to compute the class $[T]_{m-1}$ of $T$.

We now split the parameters into the energy deviation $d_k$ and the true system parameters. For the subsequent discussion, we are interested only in the case of $\nu$ true system parameters, i.e., one for each phase space pair. Furthermore, we choose the parameters such that they do not produce tune shifts by themselves but only in connection with energy deviations. For example, this can be achieved by using the strengths of $\nu$ suitably chosen sextupoles as parameters. (Quadrupoles strengths are not useful because they produce tune shifts even without $d_k$, since they obviously affect the linear tune.) In this case, the tune equations reduce to
\[
\vec{\mu} = \vec{\mu}_0 + d_k \cdot \vec{e} + d_k \cdot S(\vec{\delta}), \tag{7.51}
\]
where $S$ is a nonlinear map. To correct the chromaticities, i.e., make the tune independent of $d_k$, requires that the following be satisfied:
\[
\vec{e} + S(\vec{\delta}) = 0, \tag{7.52}
\]
which can be obtained by choosing
\[ \bar{\delta} = S^{-1}(\bar{\sigma}) \]
if the inverse of \( S \) exists. From \( S^{-1} \) we now pass to its equivalence classes and use the inversion algorithm (Eq. 2.83). This yields the classes \([\bar{\delta}]_m\) and hence the Taylor expansion of the strengths of \( v \) suitably chosen elements to correct the chromaticities. Using these Taylor expansions, an approximate value for \( \delta \) can be computed. Since the missing terms scale with the \( n \)th power of \( \delta \), iterating the procedure yields very fast convergence, requiring only very few steps in practice.

To illustrate the method, we use a simple storage ring designed for 1 GeV protons, with superperiodicity 3 and each cell being a double bend achromat (see (Berz, Makino, and Wan 1999)). The quadrupoles outside the dispersion region help to obtain the designed tunes (Fig. 7.10). From Table I, it is obvious that the chromaticities are large after the first-order design. Note that the so-called first-order chromaticities are in fact second-order quantities since the tunes are first-order quantities. The sextupoles inside the dispersive region are used to correct the first-order chromaticities. Using the methods discussed previously, the strengths of the two sextupoles are adjusted to make the dependencies vanish. After five steps the sequence converges; the results are shown in Table II.
7.3.1 The DA Normal Form Algorithm

The goal of the normal form algorithm is to provide a nonlinear change of variables such that the map in the new variables has a significantly simpler structure...

\[ \begin{array}{c|c|c}
\text{TABLE I} \\
\text{CHROMATICITIES BEFORE CORRECTION} \\
\hline
\text{Energy tune shifts} & \text{Order} & \text{Exponents} \\
\hline
0.65000000000012144 & 0 & 0 0 0 0 \\
-2.487379243931609 & 1 & 0 0 0 1 \\
\hline
0.3399999999961454 & 0 & 0 0 0 0 \\
-1.491350645132962 & 1 & 0 0 0 1 \\
\hline
\end{array} \]

\[ \begin{array}{c|c|c}
\text{TABLE II} \\
\text{CHROMATICITIES AFTER CORRECTION} \\
\hline
\text{Energy tune shifts} & \text{Order} & \text{Exponents} \\
\hline
0.65000000000012144 & 0 & 0 0 0 0 \\
-0.940585791126889E-09 & 1 & 0 0 0 1 \\
\hline
0.3399999999961454 & 0 & 0 0 0 0 \\
-0.2365239133389891E-08 & 1 & 0 0 0 1 \\
\hline
\end{array} \]
than before. In particular, it will attempt to find a transformation such that in the new variables, up to a certain order, the motion is circular with an amplitude-dependent frequency.

Therefore, we assume we are given the transfer map of a particle optical system

\[ \tilde{z}_f = \mathcal{M}(\tilde{z}, \delta), \quad (7.53) \]

where \( \tilde{z} \) are the \( 2\nu \) phase space coordinates and \( \delta \) are system parameters. Using the DA methods, we are able to compute the partial derivatives \( [\mathcal{M}]_n \) of the map to any order \( n \).

The normal form algorithm consists of a sequence of coordinate transformations \( \mathcal{A} \) of the map

\[ \mathcal{A} \circ \mathcal{M} \circ \mathcal{A}^{-1} \quad (7.54) \]

The first such coordinate transformation is the move to the parameter-dependent fixed point \( \tilde{z}_F \), which satisfies

\[ \tilde{z}_F = \mathcal{M}(\tilde{z}_F, \delta), \quad (7.55) \]

This transformation can be performed to arbitrary order using DA methods as shown in Eq. (7.47). After the fixed-point transformation, the map is origin preserving; this means that for any \( \delta \),

\[ \mathcal{M}(\tilde{0}, \delta) = \tilde{0} \quad (7.56) \]

As discussed previously, we note that the fixed-point transformation is possible if and only if 1 is not an eigenvalue of the linear map.

In the next step we perform a coordinate transformation that provides a linear diagonalization of the map. For this process, we have to assume that there are \( 2\nu \) distinct eigenvalues. This together with the fact that no eigenvalue should be unity and that their product is positive are the only requirements we have to demand for the map; under normal conditions, repetitive systems are always designed such that these conditions are met.

After diagonalization, the linear map assumes the form

\[ M = \begin{pmatrix} r_1 e^{+i\mu_1} & 0 & \cdots & 0 \\ r_1 e^{-i\mu_1} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & r_\nu e^{+i\mu_\nu} \end{pmatrix}, \quad (7.57) \]

Here, the linear tunes \( \mu_j \) are either purely real or purely imaginary. For stable systems, none of the \( r_\nu e^{\pm i\mu_\nu} \) must exceed unity in modulus.
For **symplectic** systems, the determinant is unity, which entails that the product of the $r_j$ must be unity. This implies that for symplectic systems, for any $r_j < 1$ there is another with $r_j > 1$. Thus, stable symplectic systems have $r_j = 1$ for all $j$ because otherwise there would be one $j$ for which $r_j$ exceeds unity, and thus at least one $r_j e^{\pm i \mu_j}$ would have modulus larger than unity. This would also happen if a $\mu_j$ were imaginary. Therefore, all $\mu_j$ are real, and they are even nonzero because we demanded distinct eigenvalues.

To the eigenvector pair $s_j^\pm$ belonging to the eigenvalue $r_j e^{\pm i \mu_j}$, we associate another pair $t_j^\pm$ of variables as follows:

$$
\begin{align*}
t_j^+ &= (s_j^+ + s_j^-)/2 \\
t_j^- &= (s_j^+ - s_j^-)/2i.
\end{align*}
\tag{7.58}
$$

In the case of complex $s_j^\pm$, which corresponds to the stable case, the $t_j^\pm$ are the real and imaginary parts and thus are real. In the unstable case, $t_j^+$ is real and $t_j^-$ is imaginary. Obviously, the $s_j^\pm$ can be expressed in terms of the $t_j^\pm$ as

$$
\begin{align*}
s_j^+ &= t_j^+ + i \ t_j^- \\
s_j^- &= t_j^+ - i \ t_j^{-}
\end{align*}
\tag{7.59}
$$

Later, we will perform the manipulations in the $s_j^\pm$, whereas the results are most easily interpreted in the $t_j^\pm$.

We now proceed to a **sequence of order-by-order transformations** that will simplify the nonlinear terms. After the fixed-point transformation and the linear diagonalization, all further steps are purely nonlinear and do not affect the linear part anymore, but the $m$th step transforms only the $m$th order of the map and leaves the lower orders unaffected.

We begin the $m$th step by splitting the momentary map $\mathcal{M}$ into its linear and nonlinear parts $\mathcal{R}$ and $\mathcal{S}_m$, i.e., $\mathcal{M} = \mathcal{R} + \mathcal{S}_m$. The linear part $\mathcal{R}$ has the form of Eq. (7.17). Then we perform a transformation using a map that to $m$th order has the form

$$
\mathcal{A}_m = \mathcal{I} + \mathcal{T}_m,
\tag{7.60}
$$

where $\mathcal{T}_m$ vanishes to order $m - 1$. Because the linear part of $\mathcal{A}_m$ is the unity map, $\mathcal{A}_m$ is invertible. Moreover, inspection of the algorithm to invert transfer maps (Eq. 2.83) reveals that up to order $m$,

$$
\mathcal{A}_m^{-1} = m \mathcal{I} - \mathcal{T}_m.
\tag{7.61}
$$

Of course, the full inversion of $\mathcal{A}_m$ contains higher order terms, which is one of the reasons why iteration is needed.
It is also worth noting that in principle the higher order parts of $\mathcal{T}_m$ can be chosen freely. In most cases it is particularly useful to choose these terms in such a way that they represent the flow of a dynamical system by interpreting $\mathcal{T}$ as the first term in the flow operator (Eq. 2.120) and utilizing the arguments about flow factorizations from Section 4.3.1. This has the advantages that the computation of the inverse is trivial and also that the transformation map will automatically be symplectic as soon as the original map is symplectic.

To study the effect of the transformation, we infer up to order $\mathcal{U}$:

$$\mathcal{A} \circ \mathcal{M} \circ \mathcal{A}^{-1} = (\mathcal{I} + \mathcal{T}_m) \circ (\mathcal{R} + \mathcal{S}_m) \circ (\mathcal{I} - \mathcal{T}_m)$$

$$= (\mathcal{I} + \mathcal{T}_m) \circ (\mathcal{R} + \mathcal{S}_m - \mathcal{R} \circ \mathcal{T}_m)$$

$$= \mathcal{R} + \mathcal{S}_m + (\mathcal{T}_m \circ \mathcal{R} - \mathcal{R} \circ \mathcal{T}_m) \quad (7.62)$$

For the first step, we used $\mathcal{S}_m \circ (\mathcal{I} - \mathcal{T}_m) = \mathcal{S}_m$, which holds because $\mathcal{S}_m$ is nonlinear and $\mathcal{T}_m$ is of order $m$. In the second step we used $\mathcal{T}_m \circ (\mathcal{R} + \mathcal{S}_m - \mathcal{R} \circ \mathcal{T}_m) = \mathcal{T}_m \circ \mathcal{R}$, which holds because $\mathcal{T}_m$ is of exact order $m$ and all variables in the second term are nonlinear except $\mathcal{R}$.

A closer inspection of the last line reveals that $\mathcal{S}_m$ can be simplified by choosing the commutator $\mathcal{C}_m = \{ \mathcal{T}_m, \mathcal{R} \} = (\mathcal{T}_m \circ \mathcal{R} - \mathcal{R} \circ \mathcal{T}_m)$ appropriately. Indeed, if the range of $\mathcal{C}_m$ is the full space, then $\mathcal{S}_m$ can be removed entirely. However, most of the time this is not the case.

Let $(\mathcal{T}_{m,j}^\pm | k^+_1, k^-_1, \ldots, k^+_n, k^-_n)$ be the Taylor expansion coefficient of $\mathcal{T}_{m,j}$ with respect to $(s^+_1)^{k^+_1} (s^-_1)^{k^-_1} \cdots (s^+_n)^{k^+_n} (s^-_n)^{k^-_n}$ in the $j$th component pair of $\mathcal{T}_m$. Therefore, $\mathcal{T}_{m,j}^\pm$ is written as

$$\mathcal{T}_{m,j}^\pm = \sum (\mathcal{T}_{m,j}^\pm | k^+_1, k^-_1, \ldots, k^+_n, k^-_n) \cdot (s^+_1)^{k^+_1} (s^-_1)^{k^-_1} \cdots (s^+_n)^{k^+_n} (s^-_n)^{k^-_n} \quad (7.63)$$

Similarly, we identify the coefficients of $\mathcal{C}_m$ by $(\mathcal{C}_{m,j}^\pm | k^+_1, k^-_1, \ldots, k^+_n, k^-_n)$. Because $\mathcal{R}$ is diagonal, it is possible to express the coefficients of $\mathcal{C}_m$ in terms of the ones of $\mathcal{T}$. One obtains

$$(\mathcal{C}_{m,j}^\pm | k^+_1, k^-_1, \ldots, k^+_n, k^-_n)$$

$$= \left( \prod_{l=1}^{n} \left( k^+_l + k^-_l \right) \right) \cdot \frac{e^{\sum_{j} (\hat{k}^+_j - \hat{k}^-_j) - r_j \cdot e^{\pm i\mu_j}}}{(\mathcal{T}_{m,j}^\pm | k^+_1, k^-_1, \ldots, k^+_n, k^-_n)}$$

$$= C_{m,j}^\pm (\hat{k}^+, \hat{k}^-) \cdot (\mathcal{T}_{m,j}^\pm | k^+_1, k^-_1, \ldots, k^+_n, k^-_n). \quad (7.64)$$
Now it is apparent that a term in $S_{m_j}^\pm$ can be removed if and only if the factor $C_{m_j}^\pm(k_+^+,k_-^-)$ is nonzero; if it is nonzero, then the required term in $T_{m_j}^\pm$ is the negative of the respective term in $S_{m_j}^\pm$ divided by $C_{m_j}^\pm(k_+^+,k_-^-)$.

Therefore, the outcome of the whole normal form transformation depends on the conditions under which the term $C_{m_j}^\pm(k_+^+,k_-^-)$ vanishes. This is obviously the case if and only if the moduli and the arguments of $r_j \cdot e^{\pm i \mu_j}$ and $(\prod_{l=1}^{n} r_l^{(k_l^+ + k_l^-)}) \cdot e^{i \nu_l (k_+^+ - k_-^-)}$ are identical. Next, we discuss the conditions of this to happen for various special cases and draw conclusions.

### 7.3.2 Symplectic Systems

As discussed previously, in the stable symplectic case all the $r_j$ are equal to 1, and the $\mu_j$ are purely real. Therefore, the moduli of the first and second terms in $C_{m_j}^\pm(k_+^+,k_-^-)$ are equal if and only if their phases agree modulo $2\pi$. This is obviously the case if

$$\bar{\mu} \cdot (k_+^+ - k_-^-) = \pm \mu_j \text{(mod } 2\pi\text{)},$$

where the different signs apply for $C_{m_j}^+(k_+^+,k_-^-)$ and $C_{m_j}^-(k_+^+,k_-^-)$, respectively. This can occur in two possible ways:

1. $k_l^+ = k_l^- \forall l \neq j$ and $k_j^+ = k_j^- \pm 1$.
2. $\bar{\mu} \cdot \bar{\nu} = 0 \text{ (mod } 2\pi\text{)}$ has nontrivial solutions.

The first case is of mathematical nature and lies at the heart of the normal form algorithm. It yields terms that are responsible for **amplitude-dependent tune shifts**. We will discuss its consequences later. The second case is equivalent to the system lying on a higher order resonance and is of a more physical nature. In case the second condition is satisfied, there will be resonance-driven terms that cannot be removed and that prevent a direct computation of amplitude tune shifts.

Before proceeding, we note that the second condition entails complications even if it is almost, but not exactly, satisfied. In this case, the removal of the respective term produces a small denominator that generates terms that become increasingly larger, depending on the proximity to the resonance. For the removal process, in Eq. (7.64) this resonance proximity factor is multiplied by the respective expansion coefficient, and therefore this product is obviously an excellent **characteristic of resonance strength**.

With increasingly higher orders, i.e., larger $k_+^+$ and $k_-^-$, the number of relevant resonances increases. Since the resonances lie dense in tune space, eventually the growth of terms is almost inevitable and hence produces a map that is much more nonlinear than the underlying one. As we shall discuss in the next section, this problem is alleviated by damping.
We now discuss the form of the map if no resonances occur. In this case, the transformed map will have the form

\[
\mathcal{M}_j^+ = s_j^+ \cdot f_j(s_1^+, \ldots, s_6^+)
\]
\[
\mathcal{M}_j^- = s_j^- \cdot \tilde{f}_j(s_1^-, \ldots, s_6^-).
\]  

The variables \( s_j^\pm \) are not particularly well suited for the discussion of the result, and we express the map in terms of the adjoined variables \( t_j^\pm \) introduced in Eq. (7.58). Simple arithmetic shows that

\[
s_j^+ \cdot s_j^- = (t_j^+)^2 + (t_j^-)^2.
\]

It is now advantageous to write \( f_j \) in terms of amplitude and phase as \( f_j = a_j \cdot e^{i\phi_j} \). Performing the transformation to the coordinates \( t_j^\pm \), we thus obtain

\[
\mathcal{M}_j^\pm = \begin{pmatrix} 1/2 & 1/2 \\ 1/2i & -1/2i \end{pmatrix} \begin{pmatrix} (t_j^+ + i t_j^-) \cdot f_j[(t_1^+)^2 + (t_1^-)^2, \ldots, (t_6^+)^2 + (t_6^-)^2] \\ (t_j^+ - i t_j^-) \cdot \tilde{f}_j[(t_1^+)^2 + (t_1^-)^2, \ldots, (t_6^+)^2 + (t_6^-)^2] \end{pmatrix} = a_j \cdot \begin{pmatrix} \cos(\phi_j) & -\sin(\phi_j) \\ \sin(\phi_j) & \cos(\phi_j) \end{pmatrix} \begin{pmatrix} t_j^+ \\ t_j^- \end{pmatrix}.
\]  

Here, \( \phi_j = \phi_j[(t_1^+)^2 + (t_1^-)^2, \ldots, (t_6^+)^2 + (t_6^-)^2] \) depends on a rotationally invariant quantity.

Therefore, in these coordinates, the motion is now given by a rotation, the frequency of which depends only on the amplitudes \( (t_j^+)^2 + (t_j^-)^2 \) and some system parameters and thus does not vary from turn to turn. Indeed, these frequencies are precisely the tunes of the nonlinear motion.

For any repetitive system, the tune of one particle is the total polar angle advancement divided by the number of turns in the limit of turn number going to infinity, if this limit exists. If we now express the motion in the new coordinates, we pick up an initial polar angle for the transformation to the new coordinates; then, every turn produces an equal polar angle \( \phi_j \) which depends on the amplitude and parameters of the particle. We produce a final polar angle for the transformation back to the old coordinates.

As the number of turns increases, the contribution of the initial and final polar angles due to the transformation becomes increasingly insignificant, and in the limit the tune results in only \( \phi_j \). Therefore, we showed that the limit exists and that it can be computed analytically as a by-product of the normal form transformation.
To demonstrate the effects of nonlinear normal form transformations in the symplectic case, Fig. 7.11 shows a typical tracking picture for a stable symplectic system. The nonlinear effects disturb the elliptic structure expected from linear theory, and the cross-coupling between the two planes induces a substantial broadening.

On the other hand, Fig. 7.12 shows the same motion displayed in normal form coordinates. The motion is circular up to printer resolution, removing the cross-coupling as well as other nonlinear effects.
7.3.3 Nonsymplectic Systems

In the case of stable, nonsymplectic maps, all $r_j$ must satisfy $r_j \leq 1$ because otherwise at least one of the $r_j e^{\pm i\mu_j}$ is larger than unity in modulus. Since in the normal form transformation, terms can be removed if and only if the phases or amplitudes for the two contributions in $C^\pm_{m_j}(k^+, k^-)$ are different and the amplitudes contribute, more terms can be removed.

Of particular practical interest is the totally damped case in which $r_j < 1$ for all $j$ and all $\mu_j$ are real, which describes damped electron rings. In this case an inspection of Eq. (7.64) reveals that now every nonlinear term can be removed. Then a similar argument as presented in the previous section shows that the motion assumes the form

$$M_j^\pm = r_j \cdot \begin{pmatrix} \cos(\phi_j) & -\sin(\phi_j) \\ \sin(\phi_j) & \cos(\phi_j) \end{pmatrix} \cdot \begin{pmatrix} t_j^+ \\ \tilde{t}_j^- \end{pmatrix},$$

where the angle $\phi_j$ does not depend on the phase space variables anymore but only on the parameters. This means that the normal form transformation of a totally damped system leads to exponential spirals with constant frequency $\phi_j$. In particular, this entails that a totally damped system has no amplitude-dependent tune shifts, and the motion eventually collapses into the origin. Since in practice the damping is of course usually very small, these effects are usually covered by the short-term sensitivity to resonances.

It is quite illuminating to consider the small denominator problem in the case of totally damped systems. Clearly, the denominator can never fall below $1 - \max(r_j)$ in magnitude. This puts a limit on the influence of any low-order resonance on the dynamics; in fact, even being exactly on a low-order resonance does not have any serious consequences if the damping is strong enough. In general, the influence of a resonance now depends on two quantities: the distance in tune space and the contraction strength $r_j$. High-order resonances are suppressed particularly strongly because of the contribution of additional powers of $r_j$.

Because all systems exhibit a residual amount of damping, the arguments presented here are generally relevant. It is especially noteworthy that residual damping suppresses high order resonances by the previous mechanism even for proton machines, which entails that from a theoretical view, ultimately high-order resonances become insignificant.

Clearly, the normal form algorithm also works for unstable maps. The number of terms that can be removed will be at least the same as that in the symplectic case, and sometimes it is possible to remove all terms. Among the many possible combinations of $r_j$ and $\mu_j$, the most common case in which the $\mu_j$ are real is worth studying in more detail. In this case, all terms can be removed unless the
logarithms of the $r_j$ and the tunes satisfy the same resonance condition, i.e.,

$$\vec{n} \cdot (\log(r_1), \ldots, \log(r_v)) = 0$$
$$\vec{n} \cdot \vec{\mu} = 0 \pmod{2\pi}$$  \hspace{1cm} (7.70)

have simultaneous nontrivial solutions. This situation characterizes a new type of resonance, the coupled phase-amplitude resonance.

Phase-amplitude resonances can never occur if all $r_j$ are greater than unity in magnitude. This case corresponds to a totally unbound motion, and the motion in normal form coordinates moves along growing exponential spirals.

Symplectic systems, on the other hand, satisfy

$$\prod_{l=1}^{n} r_l = 1.$$ 

Therefore, if there are $r_j$ with both signs of the logarithm, the possibility for amplitude resonances exists. In fact, any symplectic system lies on the fundamental amplitude resonance characterized by $\vec{n} = (1, 1, \ldots, 1)$. In this light, the stable symplectic case is a degeneracy in which all logarithms vanish and so the system lies on every amplitude resonances. Thus, it is susceptible to any phase resonance, and it suffices to study only these.

To study the nonlinear normal form transformations in the case of damped systems, Fig. 7.13 shows a typical tracking picture of a very weakly damped system. A difference to the symplectic case is not visible, and there is familiar cross-coupling as well as nonlinear distortions.
On the other hand, Fig. 7.14 shows the same motion displayed in normal form coordinates. The cross-coupling and nonlinear distortions are removed, and the motion follows elliptical spirals.

### 7.3.4 Amplitude Tune Shifts and Resonances

It was shown in Eq. (7.68) that in the normal form coordinates, the motion of the particles has the form

\[
\mathcal{M}_j^{\pm} = a_j \cdot \begin{pmatrix}
\cos(\phi_j) & -\sin(\phi_j) \\
\sin(\phi_j) & \cos(\phi_j)
\end{pmatrix} \cdot \begin{pmatrix} t^+ \\ t^-
\end{pmatrix},
\]

where the angles depend only on the rotationally invariant quantities \((t_1^+)^2 + (t_1^-)^2\) and possible system parameters, and thus the angle represent the tunes of the motion. This allows for a direct and accurate computation of the dependence of the tune on amplitude. As such, it offers a **fast and powerful** alternative to the conventional numerical computation of tune dependence on amplitude, which is based on tracking a seed of particles for a sufficient number of turns and then computing the tune for each as the average angle advance.
We want to study the practical aspects of the computation of the tunes using the DA normal form algorithm implemented in COSY INFINITY. As a first example, we calculate the tune shifts of the Saskatoon EROS storage ring with energy used as a parameter using both the DA normal form algorithm and the method to calculate high-order resonances. Table III shows the results. The chromaticity column shows only terms depending on energy, the exponents of which appear as the fifth entry in the exponent column. On the other hand, the normal form column shows terms depending on both energy and $x^2$, $\alpha^2$ and $y^2$, $b^2$, which describe amplitude-dependent tune shifts. As can be seen, the energy-dependent terms that are calculated by both methods agree to nearly machine precision. However, since the two algorithms for their computation are different, this can be used to assess the overall accuracy of the methods.

An independent test was performed using the code DIMAD which calculates tunes numerically by applying the fast Fourier transform method to tracking data. The various orders of chromaticities were compared to those found by COSY, and the results are listed in Table IV. Agreement within the accuracy of the numerical methods used in DIMAD was obtained.

Next, we want to address some applications for the computation of resonances. By virtue of Eq. (7.64), the coefficient of $\alpha$ necessary to obtain removal of terms in $S$ in the normal form process is given by

$$\left(\mathcal{T}_{m_j}^{\pm} | \vec{k}_{}^+, \vec{k}_-^{+}\right) = \frac{(S_{m_j}^{\pm}) (\vec{k}_{}^{+}, \vec{k}_-^{+})}{(\prod_{l=1}^{n} r_j^{k_l^+ + k_l^-}) \cdot e^{i\theta_j^{k_l^++k_l^-}} - r_j \cdot e^{\pm i\mu_j}}. \tag{7.72}$$

Even if the system is not on a resonance and thus the denominator does not vanish, these terms can become large if the system is near a resonance and the term $(S_{m_j}^{\pm}) (\vec{k}_{}^{+}, \vec{k}_-^{+})$ is large. If this is the case, then the normal form transformation map $A_m$ will have large coefficients, which leads to large coefficients in the partially transformed map $A_m \circ M \circ A_m^{-1}$ and entails that subsequent transformation attempts have to confront larger coefficients. The ultimate consequence is that the map in normal form coordinates, which is circular to order $n$, will have large coefficients beyond order $n$ that induce deviations from circularity.

Therefore, the coefficients $(\mathcal{T}_{m_j}^{\pm} | \vec{k}_{}^{+}, \vec{k}_-^{+})$ that appear in the normal form transformation process have very practical significance in that they quantify the total effect of the proximity to the resonance [via $(\prod_{l=1}^{n} r_j^{k_l^+ + k_l^-}) \cdot e^{i\theta_j^{k_l^++k_l^-}} - r_j \cdot e^{\pm i\mu_j}$] and the system’s sensitivity to it [via $(S_{m_j}^{\pm}) (\vec{k}_{}^{+}, \vec{k}_-^{+})$]. As such, the study and analysis of the resonance strengths $(\mathcal{T}_{m_j}^{\pm} | \vec{k}_{}^{+}, \vec{k}_-^{+})$ is an important characteristic of the system. Figure 7.15 shows a graphical representation of the resonance strengths for the Large Hadron Collider, which allows for a fast identification of the dominating resonances that would benefit from correction.

Having assessed the strengths of the various resonances, a system can be optimized to suppress them. Since in the normal form picture, the resonance described
TABLE III
THE VERTICAL TUNES OF THE EROS RING, COMPUTED WITH DA NORMAL FORM THEORY AND THE CHROMATICITY TOOLS

<table>
<thead>
<tr>
<th>Total Tune Shift (normal form)</th>
<th>Energy tune shift (chromaticity)</th>
<th>Order</th>
<th>Exponent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8800260865565617</td>
<td>0.8800260865565616</td>
<td>0</td>
<td>0 0 0 0 0</td>
</tr>
<tr>
<td>0.4044489434279511</td>
<td>0.4044489434279519</td>
<td>1</td>
<td>0 0 0 0 1</td>
</tr>
<tr>
<td>20.27300549111636</td>
<td></td>
<td>2</td>
<td>0 0 0 0 0</td>
</tr>
<tr>
<td>-12.7052391367074</td>
<td></td>
<td>2</td>
<td>0 0 2 0 0</td>
</tr>
<tr>
<td>19.93950369004059</td>
<td>19.93950369004062</td>
<td>2</td>
<td>0 0 0 0 0</td>
</tr>
<tr>
<td>511.199179412686</td>
<td></td>
<td>3</td>
<td>2 0 0 0 1</td>
</tr>
<tr>
<td>-272.9888390984160</td>
<td></td>
<td>3</td>
<td>0 0 2 0 1</td>
</tr>
<tr>
<td>353.1162640647963</td>
<td>353.1162640647955</td>
<td>3</td>
<td>0 0 0 0 3</td>
</tr>
<tr>
<td>13502.07274250955</td>
<td></td>
<td>4</td>
<td>4 0 0 0 0</td>
</tr>
<tr>
<td>-4936.135382965938</td>
<td></td>
<td>4</td>
<td>2 0 2 0 0</td>
</tr>
<tr>
<td>1072.710823056354</td>
<td></td>
<td>4</td>
<td>0 0 4 0 0</td>
</tr>
<tr>
<td>10058.89904831771</td>
<td></td>
<td>4</td>
<td>2 0 0 0 2</td>
</tr>
<tr>
<td>-5699.740274508591</td>
<td></td>
<td>4</td>
<td>0 0 2 0 2</td>
</tr>
<tr>
<td>12684.92656366016</td>
<td>12684.92656366014</td>
<td>4</td>
<td>0 0 0 0 4</td>
</tr>
<tr>
<td>-5111660.969538658</td>
<td>5 4 0 0 0 1</td>
<td>5</td>
<td>2 0 0 2 1</td>
</tr>
<tr>
<td>-111233.3887058197</td>
<td>5 2 0 0 2 1</td>
<td>5</td>
<td>0 0 4 0 1</td>
</tr>
<tr>
<td>32316.40175202939</td>
<td>5 2 0 0 3 3</td>
<td>5</td>
<td>0 0 4 0 1</td>
</tr>
<tr>
<td>321102.2916307596</td>
<td>5 2 0 0 0 3</td>
<td>5</td>
<td>0 0 2 0 3</td>
</tr>
<tr>
<td>-313130.5638738061</td>
<td>5 0 0 2 0 3</td>
<td>5</td>
<td>0 0 2 0 3</td>
</tr>
<tr>
<td>487095.4522760212</td>
<td>487095.4522760201</td>
<td>5</td>
<td>0 0 0 0 5</td>
</tr>
<tr>
<td>10165111.83496411</td>
<td></td>
<td>6</td>
<td>6 0 0 0 0</td>
</tr>
<tr>
<td>7921291.158179961</td>
<td></td>
<td>6</td>
<td>4 0 2 0 0</td>
</tr>
<tr>
<td>753492.5011067157</td>
<td></td>
<td>6</td>
<td>2 0 4 0 0</td>
</tr>
<tr>
<td>-117199.2118569492</td>
<td></td>
<td>6</td>
<td>0 0 6 0 0</td>
</tr>
<tr>
<td>2005266230.252126</td>
<td></td>
<td>6</td>
<td>4 0 0 0 2</td>
</tr>
<tr>
<td>-497920.0057811320</td>
<td></td>
<td>6</td>
<td>2 0 2 0 2</td>
</tr>
<tr>
<td>1629044.258487686</td>
<td></td>
<td>6</td>
<td>0 0 4 0 2</td>
</tr>
<tr>
<td>4141105.646457314</td>
<td></td>
<td>6</td>
<td>2 0 0 0 4</td>
</tr>
<tr>
<td>-21315430.11959259</td>
<td></td>
<td>6</td>
<td>0 0 2 0 4</td>
</tr>
<tr>
<td>19692776.49371354</td>
<td></td>
<td>6</td>
<td>0 0 0 0 6</td>
</tr>
</tbody>
</table>

*While the chromaticity column contains only the energy-dependent tune shifts, the normal form column also shows amplitude-dependent terms.

TABLE IV
COMPARISON BETWEEN NUMERICAL AND DA COMPUTATION OF TUNE DEPENDENCE ON ENERGY FOR THE SASKATOON EROS RING

<table>
<thead>
<tr>
<th>Order</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$1 \times 10^{-5}$</td>
</tr>
<tr>
<td>1</td>
<td>$1 \times 10^{-4}$</td>
</tr>
<tr>
<td>2</td>
<td>$5 \times 5 \times 10^{-3}$</td>
</tr>
<tr>
<td>3</td>
<td>$1 \times 10^{-2}$</td>
</tr>
<tr>
<td>4</td>
<td>$4 \times 4 \times 10^{-2}$</td>
</tr>
</tbody>
</table>
Resonances of Order 7

by the term \( \left( T_{m,j}^{\pm} \right) \tilde{k}^+ \), \( \tilde{k}^- \) is driven by terms of order \( \sum k_i^+ + k_i^- \), by suitably placing correction elements of that order in the lattice, it is possible to reduce the magnitude of \( \left( S_{m,j}^{\pm} \right) \tilde{k}^+ \), \( \tilde{k}^- \) and hence \( \left( T_{m,j}^{\pm} \right) \tilde{k}^+ \), \( \tilde{k}^- \). In the context of such numerical optimization, both the speed of the normal form approach and the high computational accuracy are of crucial significance.

Figure 7.16 shows the effect of such a resonance correction procedure for the case of the Superconducting Supercollider low-energy booster. Tracking pictures as well as numerically obtained resonance strengths are shown before and after the resonance correction with COSY. As expected, the tracking picture shows a more regular behavior, and the main offending resonance decreased in magnitude and is now in line with others (note the change in scale).

7.3.5 Invariants and Stability Estimates

Because of the clean appearance of motion in normal form coordinates, a detailed study of the dynamics is possible. For example, if the normal form motion would produce perfect circles, we know we have found an invariant of the system for each degree of freedom and hence the system is integrable. Considering that the transformation of normal form circles to conventional variables is continuous, the
resulting set is bounded and we can conclude that the motion is stable forever and particles can never get lost.

However, because most systems are not integrable, the order-by-order normal form transformations do not actually converge but rather approach a certain saturation beyond which a new order does not bring any improvement, or sometimes even worsens the situation. The mathematical reasons for this phenomenon lie in the resonance-driving terms and the associated resonance denominators...
In most practical cases, the normal form transformation leads to near-circular motion with varying amounts of deviation from circularity. To demonstrate the effect in more detail, we show the invariant defects, i.e., the residual deviation from perfect invariance, for several cases.
We begin with the invariant defects as a function of normal form radius and angle for a simple one-dimensional pendulum. In the case of this nonlinear motion, there is an invariant (the energy). Thus, in principle, the normal form algorithm could converge. Figure 7.17 shows the invariant defect for amplitudes of 1/10 rad using a normal form map of order 16. In this case the scale is approximately $10^{-17}$, and all the errors are of a computational nature. In Fig. 7.18 the amplitude of the pendulum is increased to 1/2 rad into a nonlinear regime. Again, a normal form transformation map of order 16 was used. Now the scale of the invariant defects is $10^{-13}$, and some systematic effects due to the limited order become apparent.

In the next example, we study the PSR II, a ring accelerator proposed by Los Alamos National Laboratory. In this case, an invariant does not have to exist a
priori. It is likely that the motion is indeed nonintegrable, preventing convergence of the normal form map. Figure 7.19 shows the invariant defect, i.e., the deviation from perfectly circular structure, in this case with a normal form transformation map of order 6. The defect is shown as a function of normal form radius and angle, and the increase with radius due to the limited convergence is clearly visible. The scale is approximately $10^{-6}$. 

**Figure 7.19.** Normal form defect for the Los Alamos PSR.
While the existence of invariant defects makes statements about general stability impossible, it allows an estimate of stability for finite but long times (Berz and Hoffstätter 1994; Hoffstätter 1994; Hoffstätter and Berz 1996a). This approach has been used repeatedly in theoretical studies including Nekhoroshev (?) and was later studied for several cases (Warnock, Ruth, Gabella, and Ecklund 1988; Warnock and Ruth 1991; Warnock and Ruth 1992; Turchetti 1990; Bazzani, Marmi, and Turchetti) by trying to fit models for approximate invariants to numerical data. Since in practice the invariant defects are very small, we can make estimates of how long a particle takes to traverse a certain region of normal form coordinate space for all the $n$ subspaces in which the motion follows approximately circular shape. This method is illustrated in Fig. 7.20 for one of these subspaces. Let us assume that the whole region of normal form coordinates up to the maximum radius $r_{\text{max}}$ corresponds to coordinates within the area which the accelerator can accept in its beam pipes. Let us further assume that nowhere in the $r - \phi$ diagram is the invariant defect larger than $\Delta r$. If we launch particles within the normal form region below $r_{\text{min}}$, then all these particles require at least

$$N = \frac{r_{\text{max}} - r_{\text{min}}}{\Delta r} \quad (7.73)$$

turns before they reach $r_{\text{max}}$. Considering the small size of $\Delta r$ in practical cases, this can often ensure stability for a large number of turns.

In most cases, the invariant defects grow quickly with increasing values of $r$, as shown in Fig. 7.17 and 7.18. Therefore, this estimate can be refined in the following obvious way. Suppose the region of $r$ values between $r_{\text{min}}$ and $r_{\text{max}}$ is subdivided in the following manner:

$$r_{\text{min}} = r_1 < r_2 < \cdots < r_l = r_{\text{max}}. \quad (7.74)$$

Let us assume that in each of these regions the maximum invariant defect is bounded by $\Delta r_i$. Then we can predict stability for

$$N = \sum_{i=1}^{l-1} \frac{r_{i+1} - r_i}{\Delta r_i} \quad (7.75)$$

turns. Since in practice, at least for the first values of $i$, $\Delta r_i$ can be substantially less than $\Delta r$, this lower bound can be much greater than $N$ in Eq. (7.73).

To determine the survival time of a particle, one can determine the corresponding numbers of $N$ for all the $n$ normal form subspaces and then take the smallest $N$ as a lower bound.

In practice, the appeal of the mathematically rigorous method outlined here hinges critically on the ability to determine rigorous bounds for the $\Delta r_i$, and its practical usefulness is directly connected to the sharpness of these bounds. However, in practice these functions have a large number of local maxima, and a computation of their bounds requires much care. For all the $l - 1$ regions in phase
space, we are faced with the task of finding $n$ bounds for the maxima $\Delta r^{(j)}$ of deviation functions:

$$\Delta r^{(j)} \geq \max[ r^{(j)}(\tilde{M}(\bar{x})) - r^{(j)}(\bar{x})], \quad (7.76)$$

where $r^{(j)}(\bar{x})$ is the normal form radius in the $j$th normal form subspace of a particle at position $\bar{x}$. The regions in which the bounds for the maxima have to be found are the regions where $r^{(j)}(\bar{x}) \in [r_{\min}^{(j)}, r_{\max}^{(j)}]$. As shown previously, these functions exhibit several local maxima in a six-dimensional space. To be useful, the bounds for the maxima have to be sharp to about $10^{-6}$, and for some applications to $10^{-12}$. Methods are under development that allow both the rigorous bounding of the deviation function (Berz and Hoffstätter 1994; Hoffstätter 1994; Makino 1998; Makino and Berz 1996b; Berz and Hoffstätter 1998; Makino and Berz 1999b) and the computation of rigorous bounds for the remainders of maps (Makino 1998; Berz and Makino 1998; Berz 1997b).

### 7.3.6 Spin Normal Forms

Similar to the case of the orbit motion, for spin dynamics many quantities of interest can be obtained directly within a normal form framework. In particular,
these include the so-called invariant polarization axis as well as the spin tunes. Recalling that according to Eq. (5.60), the motion assumes the special form

\[
\begin{align*}
\vec{z}_f &= \hat{\mathcal{M}}(\vec{z}_i, s), \\
\vec{s}_f &= \hat{A}(\vec{z}_i, s) \cdot \vec{s}_i.
\end{align*}
\]

In particular, the orbit motion is unaffected by the spin, and the spin motion is linear with an orbit-dependent orthogonal matrix. This entails that compared to the situation in the previous sections, there are certain differences.

As a first step of the normal form analysis, we want to perform an orbit-dependent transformation of the spin variables to coordinates that preserves one component of the spin vector at every value of \( \vec{z} \). To this end, we will determine an orbit-dependent axis \( \hat{n}(\vec{z}) \), the invariant polarization axis, such that if the spin is described in a coordinate system in which one axis is \( \hat{n}(\vec{z}) \), the spin’s component in this direction is constant from turn to turn. In the first step we do not concern ourselves with the specific orientation of the coordinate system around the axis \( \hat{n}(\vec{z}) \). This question will be addressed in the second step in connection with the spin tune.

As in the case of orbit motion, the first step is similar to the transformation to the fixed point. In the orbit case, particles on or near that fixed point stay on or near it, and for practical purposes it is thus advantageous to inject near the fixed point to limit the regions visited by the beam particles during further revolutions. In the spin case, spin vectors on or near the invariant polarization axis will stay on or near it, and for practical purposes it is advantageous to inject near the invariant polarization axis to preserve overall polarization of the beam for further revolutions.

In addition to the significance of the invariant polarization axis for the preservation of polarization of the beam, it is also important in connection with electron or positron machines subject to synchrotron radiation damping. In this case, the damping leads to the buildup of polarization due to the non-random spin of emitted photons, the amount of which can be estimated via semiempirical arguments derived by Derbenev and Kondratenko (1973) from the knowledge of the motion of the axis \( \hat{n}(\vec{z}) \) around the ring. Methods for the computation of \( \hat{n} \) were developed to first order in spin and orbit by Chao (1981), to third order by Eidelmann and Yakimenko (1991), to arbitrary order by Mane (1987) and Balandin and Golubeva (1992), and to arbitrary order including \( s \)-dependent elements by Berz (1995b). There are also various approaches to determining \( \hat{n} \) and the equilibrium polarization from numerical tracking tools (Yokoya 1992; R. Rossmanith 1989).

If the spin matrix \( \hat{A} \) is independent of \( \vec{z} \), the axis \( \hat{n} \) is apparently merely the rotation axis of \( \hat{A} \). Also, since the matrix \( \hat{A} \) has a constant part \( \hat{A}_0 \) that is independent of \( \vec{z} \), the constant part \( \hat{n}_0 \) of \( \hat{n} \) is just the rotation axis of \( \hat{A}_0 \). However, for a \( \vec{z} \)-dependent \( \hat{A} \), the axis \( \hat{n} \) depends on \( \vec{z} \), and the requirement is different.
Indeed, to preserve the projection of the spin on \( \vec{n}(\vec{z}) \), we have to satisfy

\[
S^i_t(\vec{z}_t) \cdot \vec{n}(\vec{z}_t) = S^f_t(\vec{z}_f) \cdot \vec{n}(\vec{z}_f) = \left( \hat{A}(\vec{z}_t) \cdot S^i_t(\vec{z}_t) \right)^t \cdot \vec{n}(\mathcal{M}(\vec{z}_t)) = \hat{S}^i_t(\vec{z}_t) \cdot \hat{A}^t(\vec{z}_t) \cdot \vec{n}(\mathcal{M}(\vec{z}_t)).
\]

Since this must hold for every vector \( S^i_t(\vec{z}_t) \), we infer that \( \vec{n} \) must satisfy \( \vec{n}(\vec{z}_t) = \hat{A}^t(\vec{z}_t) \cdot \vec{n}(\mathcal{M}(\vec{z}_t)) \), or

\[
\hat{A}(\vec{z}) \cdot \vec{n}(\vec{z}) = \vec{n}(\mathcal{M}(\vec{z})). \quad (7.77)
\]

In the following, we will discuss an algorithm for the determination of the equilibrium polarization \( \vec{n}(\vec{z}) \) similar to the DA normal form approach. We assume the orbital map is already in normal form and the linear spin map is diagonalized. We proceed in an iterative way.

For 0th order, observe that \( \mathcal{M}(\vec{z}) = 0 \), and thus the equation reads \( \hat{A}_0 \cdot \vec{n}_0 = \vec{n}_0 \). Since \( \hat{A}(\vec{z}, s) \) and thus \( \hat{A}_0 \) are in \( SO(3) \), the eigenvalue spectrum of \( \hat{A}_0 \) contains the value 1 as well as a complex conjugate pair of unity modulus. The eigenvector corresponding to the eigenvalue 1 represents the axis \( \vec{n}_0 \), and the phase of the complex conjugate pair represents the linear spin tune \( \mu_s \).

For higher orders, assume we already know \( \vec{n} \) to order \( m - 1 \) and want to determine it to order \( m \). Assume \( \mathcal{M}(\vec{z}) \) is in normal form, i.e., \( \mathcal{M}(\vec{z}) = \mathcal{R} + \mathcal{N} \). Write \( \hat{A} = \hat{A}_0 + \hat{A}_{\geq 1}, \vec{n} = \vec{n}_{< m} + \vec{n}_m \), and obtain

\[
(\hat{A}_0 + \hat{A}_{\geq 1}) \cdot (\vec{n}_{< m} + \vec{n}_m) = (\vec{n}_{< m} + \vec{n}_m) \circ (\mathcal{R} + \mathcal{N})
\]

To order \( m \), this can be rewritten as

\[
\hat{A} \cdot \vec{n}_{< m} + \hat{A}_0 \cdot \vec{n}_m = m \vec{n}_{< m} \circ (\mathcal{R} + \mathcal{N}) + \vec{n}_m \circ \mathcal{R} \quad (7.78)
\]

or

\[
\hat{A}_0 \cdot \vec{n}_m - \vec{n}_m \circ \mathcal{R} = m \vec{n}_{< m} \circ (\mathcal{R} + \mathcal{N}) - \hat{A} \cdot \vec{n}_{< m}.
\]

The right hand side must be balanced by choosing \( \vec{n}_m \) appropriately. However, as in Eq. (7.64) for the orbit case, the coefficients of \( \hat{A}_0 \vec{n}_m - \vec{n}_m \circ \mathcal{R} \) differ from those of \( \vec{n}_m \) only by resonance denominators, and so the task can be achieved as soon as these resonance denominators do not vanish. This requires the absence of spin-amplitude resonances in which, because of the purely linear structure of \( \hat{A}_0 \cdot \vec{n} \), the spin tune \( \nu_s \) appears only linearly. Therefore, spin-amplitude terms are characterized by the condition

\[
\mu_s + \vec{n} \cdot \vec{\mu} = 0 \, \text{mod} \, 2\pi. \quad (7.79)
\]
As a result of the iterative scheme, we obtain the value of the invariant polarization axis $\vec{n}(\vec{z})$ for all values of $\vec{z}$, probably the single most important nonlinear characteristic of spin motion.

In order to proceed further and determine a nonlinear normal form as well as the associated tune shifts for symplectic orbit motion, we perform a coordinate rotation of the spin variables that is linear in spin and nonlinear in orbit, such that at each point $\vec{e}_z$ is parallel to $\vec{n}(\vec{z})$ and $\vec{e}_x$ is in the horizontal plane. Expressed in these coordinates, the spin matrix then has the form

$$\hat{A}(\vec{z}) = \begin{pmatrix} a_{11}(\vec{z}) & a_{12}(\vec{z}) & 0 \\ a_{21}(\vec{z}) & a_{22}(\vec{z}) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (7.80)$$

Furthermore, $a_{11}(\vec{z}) \cdot a_{22}(\vec{z}) - a_{12}(\vec{z}) \cdot a_{21}(\vec{z}) = 1$.

We now drop the invariant third component of the spin coordinates and introduce

$$\hat{A}_r(\vec{z}) = \begin{pmatrix} a_{11}(\vec{z}) & a_{12}(\vec{z}) \\ a_{21}(\vec{z}) & a_{22}(\vec{z}) \end{pmatrix}, \quad (7.81)$$

the reduced matrix for the remaining spin coordinates. The resulting spin–orbit motion in these eight coordinates $(\vec{S}_r, \vec{z})$ has the form

$$\vec{S}_{r,f} = \hat{A}_r(\vec{z}) \cdot \vec{S}_{r,i}$$
$$\vec{z}_f = \mathcal{M}(\vec{z}_i). \quad (7.82)$$

where for every $\vec{z}$, the matrix $\hat{A}_r(\vec{z})$ has a complex conjugate pair of eigenvalues with product 1. For the particular case of $\vec{z} = \vec{0}$, these two eigenvalues have the form $\exp(\pm i \cdot \mu_s)$.

The resulting linearization of the system is apparently

$$\begin{pmatrix} \vec{S}_{r,f} \\ \vec{z}_f \end{pmatrix} = \begin{pmatrix} \hat{A}_r & \mathbf{0} \\ \mathbf{0} & \hat{M} \end{pmatrix} \cdot \begin{pmatrix} \vec{S}_{r,i} \\ \vec{z}_i \end{pmatrix}. \quad (7.84)$$

In the absence of spin-orbit resonances, this eight-dimensional linearization is of the form in Eq. (7.57) required for the conventional DA normal form algorithm.

In the absence of spin–orbit resonances, the linear eigenvalues are all disjoint and it is thus possible to execute the conventional DA normal form algorithm discussed in Section 7.3.1. In this scenario, it is most useful that this algorithm is not restricted to merely symplectic systems but readily adapts to handle this case in which one part of the map satisfies the symplectic symmetry and one part the orthogonal symmetry.
Because of the fact that in each step of the transformation, the spin coordinates occur only linearly, the resonance-driving terms for all spin resonances of higher order in spin vanish, and the only requirement is that

$$\mu_s + \vec{n} \cdot \vec{\mu} = 0 \mod 2\pi$$

(7.85)
cannot be satisfied for any choice of \(\vec{n}\). As a result of the overall transformation, we obtain that the motion in spin–orbit normal form variables has the form

$$\mathcal{M}_j^\pm = \begin{pmatrix} \cos(\phi_j) & -\sin(\phi_j) \\ \sin(\phi_j) & \cos(\phi_j) \end{pmatrix} \cdot \begin{pmatrix} t_j^+ \\ t_j^- \end{pmatrix}$$

(7.86)

$$\dot{\mathcal{A}}_j^\pm = \begin{pmatrix} \cos(\phi_s) & -\sin(\phi_s) \\ \sin(\phi_s) & \cos(\phi_s) \end{pmatrix} \cdot \begin{pmatrix} t_s^+ \\ t_s^- \end{pmatrix}$$

(7.87)

As a result, the motion in both the orbit and spin blocks now follows \textbf{plain rotations}. Since the entire orbit block was independent of the spin block, the \(\phi_j\) are again only dependent on orbit amplitudes. On the other hand, since the spin motion is linear and dependent on the orbit motion, the \(\phi_s\) are only dependent on orbital amplitudes. Therefore, the newly obtained values \(\phi_s\) describe \textbf{amplitude spin tune shifts}.

### 7.4 Symplectic Tracking

The potential buildup of inaccuracies in the dynamical system characterizing an accelerator is a fundamental and unavoidable problem that prevents exact predictions of the motion. While from a purist’s point of view this fact represents the point of retreat, in many practical situations it is nevertheless often possible to obtain a qualitative understanding of the phenomena that can occur by enforcing \textbf{symmetries} that the system is known to have. An example that is particularly clearly understandable is demonstrated by a system that conserves \textbf{energy} (or any other function on phase space such as the Hamiltonian). If during the simulation one enforces that despite the presence of errors the energy of the system is indeed preserved, then this at least prevents the system from drifting off into regions with the wrong energy. Depending on the shape of the energy function, this sometimes ensures that the system stays in the right territory; of course, if the energy surfaces have a complicated topology, even this does not have to be the case.

Another symmetry of systems that are Hamiltonian that is often favored for artificial enforcement is the \textbf{symplectic symmetry} discussed in Section 4.3.2, which entails that if \(\mathcal{M}\) is the Jacobian of the map, then

$$\tilde{\mathcal{M}} \cdot \tilde{J} \cdot \tilde{\mathcal{M}}^t = \tilde{J}.$$
The geometric ramifications of enforcing this symmetry are more difficult to understand than those of energy conservation; in fact, one of the very unfortunate theoretical results is that for most systems, it is indeed impossible to simultaneously preserve symplecticity and energy, unless the simulation by accident tracks the completely correct system.

7.4.1 Generating Functions

Enforcing the symplectic symmetry (Eq. 7.88) directly is not straightforward since the symplectic condition represents a complicated nonlinear implicit partial differential equation for the map. The question is whether for a map that is nearly symplectic, be it by truncation or because of errors, there is a symplectic function nearby that can be used instead. A particularly robust method for this purpose is to fabricate such a function through the use of generating functions in mixed variables. The four most commonly used generators are

\[
    F_1(\tilde{q}_i, \tilde{q}_f), \quad F_2(\tilde{q}_i, \tilde{p}_f), \\
    F_3(\tilde{p}_i, \tilde{q}_f), \quad \text{and} \quad F_4(\tilde{p}_i, \tilde{p}_f);
\]

which are known to represent symplectic maps via the implicit equations

\[
    (\tilde{q}_i, \tilde{q}_f) = (\nabla_{\tilde{q}_i} F_1, -\nabla_{\tilde{q}_f} F_1) \\
    (\tilde{p}_i, \tilde{q}_f) = (\nabla_{\tilde{p}_i} F_2, \nabla_{\tilde{p}_f} F_2) \\
    (\tilde{q}_i, \tilde{p}_f) = (-\nabla_{\tilde{q}_i} F_3, -\nabla_{\tilde{q}_f} F_3) \\
    (\tilde{p}_i, \tilde{p}_f) = (-\nabla_{\tilde{p}_i} F_4, \nabla_{\tilde{p}_f} F_4).
\]  

However, to assert universal existence of a generator, as shown in Chapter 1, Section 1.4.6, a wider class of generators is necessary.

Furthermore, the map represented by any generating function, be it the right one or not, is always symplectic. An approximative generating function can be used to perform symplectic tracking in the following way: Use the underlying approximative map to compute first values of the final coordinates \((\tilde{q}_f, \tilde{p}_f)\). Depending on the accuracy of the map, the quadruple \((\tilde{q}_i, \tilde{p}_i, \tilde{q}_f, \tilde{p}_f)\) is already very close to a solution of the implicit Eq. (7.90). It is used as a starting point of a numerical solution of the implicit equations, and \(\tilde{q}_f\) and/or \(\tilde{p}_f\) are varied to determine an exact solution. This can be done by Newton’s method, and usually one iterative step is enough to obtain machine accuracy.

To illustrate the practical use of the method, we apply it to a nonlinear dynamics problem in which the motion is well understood. We utilize the map of a homogeneous sector magnet with a bending magnet of deflection angle \(2\pi/n\), where \(n\) is an integer. The resulting map is nonlinear, but since the motion inside a homogeneous bending magnet is purely circular, after \(2\pi\) and hence \(n\) iterations,
the overall map will be unity. In particular, while nonlinear, the system does not exhibit any amplitude dependent tune shifts. Figure 7.21 shows map tracking for large amplitudes obtained with a nonsymplectified Taylor map of order 20. After many turns, individual separate points are reproduced faithfully, as should be the case because of the complete lack of amplitude-dependent tune shifts. Figure 7.22 shows the same tracking with a fifth-order map.

Because of the inaccuracy of the fifth-order approximation, errors in both tune and amplitude result. Figure 7.23 shows the motion using two different generating function symplectifications of the Taylor map. In both cases, after a suitable number of points, the overall deviation from the correct position is similar to that for
the case of the nonsymplectified tracking; on the left, the accuracy is higher because the generating function is less nonlinear as it is on the right. It is noteworthy that the resulting corrections due to symplectification lead mainly to deviations in angular direction, whereas the radial direction is apparently well preserved. This phenomenon is mostly due to the fact that symplectification preserves phase space volume, and so in this two-dimensional case no shrinking in phase space is possible.

The observations in this particular example are found in a similar way as those in other cases: symplectification has a tendency to preserve radial information but does not help in the reproduction of angular information; in fact, it sometimes leads to larger errors in angular direction. While tracking the preservation of radial information perhaps leads to a more realistic looking picture, for analytical causes such as the normal form methods, not much is necessarily gained. In particular, the errors in angular direction are usually approximately as strong in amplitude tune shifts as they are without symplectification. Similarly, since the resonance strengths in Eq. (7.64) are a measure of the proximity to resonances for various amplitudes, the errors in resonance strengths are usually not affected beneficially by symplectification. For the previous example, the structure of tune shifts and resonance strengths after symplectification is not reproduced faithfully, and although symplectification indeed preserves the symplectic symmetry of the system, it is not able to preserve the underlying resonance symmetry properly.

7.4.2 Prefactorization and Symplectic Extension

For all practical applications, the behavior of the nonlinear generating function is very important; in particular, it is important that its nonlinearities are not too large. While the nonlinearities of the transfer map are a more or less direct measure of
the nonlinearity of the system, this is not the case for generating functions. By inspecting the algorithm for the computation of the generating function, it becomes apparent that the new results are concatenated over and over with the inverse of the linear relationship in mixed variables. Although, as shown in Chapter 1, Section 1.4.6, a set of mixed variables can be found such that the linearization is nonsingular, it may be the case that the determinant is close enough to zero that the inverse has large terms.

Therefore, while symplectic matrices are always restricted to unity determinant, this is not the case for the matrices that have to be inverted here, and so the resulting generating functions sometimes have stronger nonlinearities than the underlying maps.

This problem can be alleviated largely by factoring out the linear part of the map

\[ \mathcal{M}_n = \mathcal{M}_1 \circ \mathcal{M}_n^* \]

and then treating \( \mathcal{M}_1 \) and \( \mathcal{M}_n^* \) separately. For the map \( \mathcal{M}_n^* \), the linear part is now unity, which implies that two of the four generating functions can be computed without the risk of small determinants. The symplectification of the linear map \( \mathcal{M}_1 \) represented by the matrix \( \bar{\mathcal{M}}_1 \), which is also its Jacobian, is a linear algebra problem. It suffices to satisfy \( \bar{\mathcal{M}}_1 \cdot \bar{J} \cdot \bar{\mathcal{M}}_1^t = \bar{J} \) which using the scalar product (Eq. 1.173) with metric \( \bar{J} \) defined by \( \langle \bar{x}, \bar{y} \rangle = x \cdot \bar{J} \cdot y^t \) can be enforced by procedures such as the Gram–Schmidt.

Using the Taylor series expansion terms of the generating function as described previously, it is straightforward to compute a transfer map that agrees with the given transfer map to order \( n \) yet is symplectic to higher orders compared to the old transfer map. Even if the old transfer map violates symplecticity noticeably because of truncation errors, it is possible to have the extended map satisfy the symplectic condition to increasingly higher orders. Depending on the case, it is often possible to obtain machine accuracy symplecticity for the phase space regions of interest by the method of symplectic extension.

To this end, one first computes a suitable generating function to the same order as that of the original map, following the algorithm discussed in the last section. While the result is not the proper generating function for the true map, it has the same Taylor expansion as the proper one and agrees with it increasingly better the higher the order. One now approximates the real generating function by its Taylor series and computes the map that is generated from the approximate generating function using the previous algorithm. Up to order \( n \), the old map is reproduced; continuing beyond order \( n \) produces increasingly higher orders extending the original transfer map, and the map can be made symplectic to any order \( k > n \).

If the missing higher order terms are sufficiently small, what we have produced is an explicit symplectic integrator. This algorithm is particularly useful for systems in which the computation of the map is expensive to high orders, but
whose inherent nonlinearity is not too high. In particular, this could be the case for machines consisting of many different elements or compact machines with very complicated fields requiring detailed integrations.

7.4.3 Superposition of Local Generators

Here, we discuss the symplectification of a map that is not given by one Taylor map but rather by a group of Taylor maps around different expansion points. The goal is to find a global generating function that locally reduces to the respective Taylor maps but that is globally symplectic.

When using generating functions for symplectic tracking, it is of course not mandatory that they actually have the same Taylor expansion as that of the true generating function. Indeed, in the case of nonlinear problems in which the function $\mathcal{N}_n$ is not well behaved, it may be advantageous to produce generating functions that are smoother overall. This can be achieved by a superposition of local generating functions.

To this end, a representative ensemble of nodes in phase space is chosen, preferably in a regular way as in Fig. 7.24. For each of these nodes, a transfer map is computed to a certain order. Then, for each of the nodes the respective generating function is computed. Each of these generating functions is uniquely determined except for its constant value $c_i$.

A total generating function can now be determined by a smooth interpolation of the local polynomial-type generating functions in a nonequidistant mesh. This has the form

$$F(\bar{q}_i, \bar{p}_j) = \sum_{j=1}^{n} F_j(\bar{q}_i, \bar{p}_j) \cdot w_j(\bar{q}_i, \bar{p}_j),$$

(7.91)

where the $w_j$ are smooth weighting factor functions that ensure that the influence of $F_i$ only extends near the respective next nodes and not far beyond. For example, they can be Gaussians centered at $(\bar{q}_i, \bar{p}_j)$, with widths determined by the distances to the next nodes and a height chosen accordingly.
While in the case of a single generating function, the unknown constant term was irrelevant, here it is significant since it is multiplied by the position dependent weighting function and thus appears in the implicit solution \((\vec{q}_f, \vec{F}_f)\). Therefore, it is necessary to choose the \(c_i\) in a self-consistent way.

One solution to this problem is to demand that at each node, the predictions of all the neighboring nodes are as close as possible (which is similar to the finite element potential-solving problem). This yields a least squares problem for the \(c_i\) which can be solved using conventional techniques. Naturally, the higher the orders of the individual \(F_i\), the better will be their prediction at the neighboring nodes and the smaller will be the resulting sum of squares.

Therefore, one obtains a generating function that is not Taylor series like, and one can cover large and rather nonlinear areas of phase space.

### 7.4.4 Factorizations in Integrable Symplectic Maps

In this section, we discuss symplectification techniques of other methods that allow the representation of a map by a symplectic map. The idea is to first generate a large pool of symplectic maps \(S_i\) that depend on certain parameters and that can be evaluated explicitly: such maps are sometimes called Cremona maps. Important examples of such a pool are drifts, rotations in phase space, and kick maps that affect only the momenta of particles as proposed by Irwin and Dragt. Other important classes of such explicitly integrable functions were proposed, including those by Shi (Shi and Yan 1993; Shi 1994).

Then this pool is used to model the original function as a composition of such maps as

\[
M_n = n \prod_{i=1}^{k} S_i,
\]

where the parameters are chosen such that the overall map \(M_n\) is reproduced. In all cases, it is important and not easily achieved that the map pool is rather exhaustive and allows the representation of the desired map through maps from the pool with well-behaved parameter settings. Furthermore, since the number of parameters that have to be adjusted is very large and usually equal to the number of free coefficients in a symplectic map, it is very advantageous if the system of equations that has to be solved is linear.

In the case of kick factorizations, a frequently used approach is to attempt to represent the map by kicks that affect only one particular order each and that are separated by rotations of fixed angles, all of which are fractions of \(2\pi\). The kick strengths are chosen as the free parameters \(\vec{p}\), and they are determined in an order-by-order way. Specifically, for order \(n\), all kicks of orders \(<n\) are already known and only parameters \(\vec{p}_n\) of order \(n\) are adjusted. Denoting by the vector \(\vec{m}_n\) the \(n\)th order map properties that have to be adjusted, this leads to a linear system of
equations of the form

\[ \ddot{m}_{n} = \hat{A}_{n} \cdot \vec{p}_{n}. \]  

(7.92)

Here, the information of the specific map is entirely contained in \( \ddot{m}_{n} \), whereas the coupling matrix \( \hat{A} \) depends only on the placement of the kicks within the fixed rotations. In order to limit spurious higher order nonlinearities introduced by the approach, it is highly desirable that the matrix \( \hat{A} \) that couples the coefficients to the nonlinearities of the map has a well-behaved inverse. A measure of this is the spectrum of the \( k \) eigenvalues \( \lambda_i \) of \( \hat{A} \). The spacing and angles of rotations are adjusted in such a way as to maximize the minimum eigenvalue modulus

\[ \lambda_{\min} = \min(|\lambda_i|); \]  

(7.93)

obviously, a vanishing eigenvalue prevents inversion of the matrix, and it is expected that the coefficients of the inverse are more well behaved if \( \lambda_{\min} \) is as large as possible. Likewise, it is also desirable that the average

\[ \lambda_{\alpha} = \frac{\sum_{i=1}^{k} \lambda_i}{k}, \]  

(7.94)

be as large as possible. To satisfy both of these demands is rather nontrivial, and optimal and robust solutions for these questions have not been found for the general case.

### 7.4.5 Spin Tracking

We address another type of tracking that, although not preserving the symplectic symmetry, conceptually can be treated in a very similar way: the tracking of spin. As discussed in Chapter 5, Section 5.2, the map of a system consisting of orbit motion and spin motion is described by

\[ \begin{cases} \vec{z}_f = \mathcal{M}(\vec{z}_i), \\ \vec{S}_f = \hat{A}(\vec{z}_i) \cdot \vec{S}_i \end{cases}, \]  

(7.95)

where \( \mathcal{M} \) is the symplectic orbit map, and \( \hat{A}(\vec{z}_i) \in SO(3) \) is the spin matrix. For the purpose of tracking of spin, it is necessary to successively obtain \( (\vec{z}_n, \vec{S}_n) \) from \( (\vec{z}_{n-1}, \vec{S}_{n-1}) \). This requires the determination of \( \vec{z}_n \) via

\[ \vec{z}_n = \mathcal{M}(\vec{z}_{n-1}), \]  

(7.96)

for which symplectic integration techniques can be used. However, to compute \( \vec{S}_n \) requires that first the orbit-dependent spin map \( \hat{A}(\vec{z}_{n-1}) \) be evaluated. Since the
full dependence of $\hat{A}$ on $\vec{z}$ is not known, but only its polynomial representation to a given order, the evaluation will in general result in a matrix that is almost, but not exactly, orthogonal. For the same reason, symplectification may be desirable; one may thus want to make $\hat{A}$ orthogonal before computing the new spin.

While the preservation of the symplectic symmetry required the determination of a sufficiently close-by symplectic map, here it is necessary to find a sufficiently close-by orthogonal matrix. This, however, is straightforward, because it can be achieved with orthonormalization techniques such as Gram-Schmidt. Furthermore, if the starting map is nearly orthogonal, the procedure will not result in large deviations and it will produce a nearby matrix. After the matrix $\hat{A}(\vec{z}_n)$ is orthogonalized to become the matrix $\hat{A}^*$, the new spin coordinates are determined via

$$\vec{S}_n = \hat{A}^* \cdot \vec{S}_{n-1}. \quad (7.97)$$