Analysis of Nonlinear Dynamics
by Square Matrix Method

Li Hua Yu

July 2016

National Synchrotron Light Source II
Energy & Photon Sciences Directorate

Brookhaven National Laboratory

U.S. Department of Energy
USDOE Office of Science (SC),
Basic Energy Sciences (BES) (SC-22)

Notice: This manuscript has been authored by employees of Brookhaven Science Associates, LLC under Contract No. DE-SC0012704 with the U.S. Department of Energy. The publisher by accepting the manuscript for publication acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.
DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or any third party’s use or the results of such use of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.
Analysis of Nonlinear Dynamics by Square Matrix Method

Li Hua Yu
Brookhaven National Laboratory, Upton, NY 11973

The nonlinear dynamics of a system with periodic structure can be analyzed using a square matrix. We show that because the special property of the square matrix constructed for nonlinear dynamics, we can reduce the dimension of the matrix from the original large number for high order calculation to low dimension in the first step of the analysis. Then a stable Jordan decomposition is obtained with much lower dimension. The transformation to Jordan form provides an excellent action-angle approximation to the solution of the nonlinear dynamics, in good agreement with trajectories and tunes obtained from tracking. And more importantly, the deviation from constancy of the new action-angle variable provides a measure of the stability of the phase space trajectories and their tunes. Thus the square matrix provides a novel method to optimize the nonlinear dynamic system. The method is illustrated by many examples of comparison between theory and numerical simulation. In particular, we show that the square matrix method can be used for optimization to reduce the nonlinearity of a system.

PACS numbers: 05.45.-a, 45.20.-d, 45.50.-j,95.00.00,46.40.Fj,41.85.-p,29.20.db,29.27.-a

1. INTRODUCTION

The question of the long term behavior of charged particles in storage rings has a long history. To gain understanding, one would like to analyze particle behavior under many iterations of the one turn map. The most reliable numerical approach is the use of a tracking code with appropriate integration methods. For analysis, however, one would like a more compact representation of the one turn map out of which to extract relevant information. Among the many approaches to this issue we may mention canonical perturbation theory, Lie operators, power series, and normal form [1–11], etc.

Here, we would like to look at this problem from a somewhat different perspective: we shall analyze the problem by the method of square matrix. In the following, we organize the paper so that we limit the description of the main body of the work in the first 9 pages from Section 1 to 6, while the most parts of the paper are examples, some details and appendices.

We consider a square matrix [12] constructed out of the power series map [6]. If we use the complex Courant-Snyder variable \( z = x - ip \), its conjugate and powers \( z, z^*, z^2, \ldots \) as a column \( Z \), the one turn map can be represented by a large square matrix \( M \) using \( Z = M Z_0 \). In this paper we find that the analysis of this matrix, after an immediate reduction to a Jordan matrix with much lower dimension, leads to an action-angle approximation \( w_{x0}, w_{y0} \), with its action amplitude and phase advance angle nearly constant up to near the border of the dynamic aperture or resonance lines. In addition, we find that near this border the deviation of these actions from constancy provides a measure of the destruction of invariant tori, or a measure of the stability of trajectories and tunes. Thus the functions \( w_{x0}, w_{y0} \) provide a tool for analysis of nonlinear dynamics.

As an example, for the 7th order problem with 4 variables \( x, p_x, y, p_y \), the dimension of the matrix \( M \) is 330. We find that because the special feature of this matrix that it is upper-triangular with all eigenvalues (i.e., all the diagonal elements) precisely known to be on unit circle, once this matrix is constructed we may immediately reduce the mapping problem into a matrix problem of much smaller dimension. In this example, the dimension is reduced to 10 from 330. In Section 2-3 we show that for this matrix of much smaller dimension we can find a stable Jordan decomposition \( U M \bar{U} = e^{i\mu I + \tau} \), where \( U \) and \( \bar{U} \) are the transformation matrix, \( \mu \) is the tune and \( \tau \) is the Jordan matrix, as given by Eq.3.2, with much lower dimension. In the above example, the dimension \( U \) and \( \bar{U} \) and are 4×330 and 330×4 respectively, and the dimension of \( \tau \) and the identity matrix \( I \) is 4×4. The new variables from the transformation \( W \equiv U Z \) with its rows \( w_0, w_1, w_2, \ldots \) provide several action-angle variable approximations with different lowest power terms, with accuracy from high to low when lowest power increases from low to high. We remark that the Jordan decomposition is carried out in one step to high order, without step by step iteration from low order to high order.

In Section 4 we find that when sufficiently far away from the dynamic aperture the quantity \( i\phi = w_1/w_0 \equiv w_2/w_1 \equiv \ldots \) along the trajectory is nearly constant, represents the amplitude dependent tune. Then, in Section 5 we show that \( \Delta \equiv w_2/w_0 - (w_1/w_0)^2 \) is nearly zero within the dynamic aperture, and its deviation from zero provides information about the loss of stability of the trajectory. The conditions \( \text{Im}\phi \equiv 0, \Delta \equiv 0 \) and the constancy of \( w_0 \) can be used to test the stability of the trajectory and tune variation.

In Section 6, we address the stability of Jordan decomposition, and its uniqueness, and describe the structure of the Jordan basis. We show that by scaling \( z \) we can reduce the range of the singular values of the square matrix by many orders of magnitudes, and assure the stability of the Jordan decomposition. Then we show that there is a way to systematically separate the most accurate chain in the Jordan decomposition from the less accurate shorter chains. This makes the decomposition unique.
In Section 7 we apply the method developed here to several storage ring lattices. We present large amounts of examples to compare the theory with tracking results. Among them, we emphasize the complicated surfaces in the Poincare sections using Courant-Snyder variables, derived from tracking results, become flat planes in the Poincare sections of $w_{x0},w_{p0}$. Section 8 gives an example of the comparison of tune calculated by the square matrix method with the tracking results (see Fig.15). We also study the region where the trajectories are stable, i.e., the "coherence region", and compare its shape with the dynamic aperture, as shown in Fig.21 and 23. Section 8 illustrates how the action-angle approximation can be used to study the formation of islands. Section 9 provides one example of optimization by manipulation of trajectories. In particular, we show in Fig.33 that the complicated motion in $y - y'$ space generated by nonlinear coupling for large amplitude $x$ motion is improved to regular circular motion after optimized by square matrix method. Finally, Section 10 gives a summary.

2. SQUARE MATRIX, ITS STRUCTURE AND ITS INVARIENT SUBSPACE

We consider the equations of motion of a nonlinear dynamic system with periodic structure such as Hills equation, it can be expressed by a square matrix. For example, for a simple Henon map [8],

\[
\begin{align*}
    x &= x_0 \cos \mu + p_0 \sin \mu + e x_0^2 \sin \mu \\
    p &= -x_0 \sin \mu + p_0 \cos \mu + e x_0^2 \cos \mu
\end{align*}
\] (2.1)

we use the variables $z = x - ip$ and $z^* = x + ip$ to write $z, z^*$ and their higher power monomials after one turn of rotation, or after one element in an accelerator lattice, as a truncated power series expansion of the initial $z_0 = x_0 - ip_0$ and $z_0^* = x_0 + ip_0$. For example, up to 3rd order, we have:

\[
\begin{align*}
    z &= e^{i\mu} z_0 - \frac{i}{4} e^{i2\mu} z_0^2 - \frac{i}{2} e^{i\mu} p_0 z_0 + \frac{i}{4} e^{2i\mu} z_0^3 \\
    z^* &= e^{-i\mu} z_0^* + \frac{i}{4} e^{-i2\mu} z_0^2 + \frac{i}{2} e^{-i\mu} p_0 z_0^* + \frac{i}{4} e^{-2i\mu} z_0^3 \\
    z^2 &= e^{2i\mu} z_0^2 - \frac{i}{2} e^{2i2\mu} z_0^3 - \frac{i}{4} e^{i\mu} p_0 z_0^2 + \frac{i}{2} e^{i2\mu} z_0^3 \\
    z^3 &= z_0 z_0^* + \frac{i}{4} z_0^2 - \frac{i}{4} z_0 + \frac{i}{4} z_0^2 - \frac{i}{4} e^{2i\mu} z_0^3 \\
    z^{32} &= e^{-2i\mu} z_0^2 + \frac{i}{2} e^{-2i\mu} z_0^3 - \frac{i}{2} e^{2i\mu} z_0^2 + \frac{i}{2} e^{-2i\mu} z_0^3 \\
    z^3 &= e^{3i\mu} z_0^3 \\
    \ldots \\
    z^3 &= e^{-3i\mu} z_0^3
\end{align*}
\] (2.2)

In general, there are constant terms in the expansion, even though in this example the offset of $x$ is zero, so the constant terms are also zeros. We may now write this in the matrix form:

\[
Z = MZ_0,
\] (2.3)

where to 3rd order, we define the $10 \times 1$ monomial array

\[
\vec{Z} = (1, z, z^*, z^2, z^2 z^*, z^2, z^3, z^2 z^*, z^2 z^2 z^*, z^3 z^3)
\] (2.4)

The row $\vec{Z}$ represents the matrix transposition of the column $Z$. The vector $Z$ spans a 10 dimensional linear space. The matrix $M$, when operated on $Z_0$, represents a rotation to $Z$ in this space. We remark here that even though we mostly use $M$ to represent one turn map for a storage ring, it can be used to represent one element in the storage ring dynamics or other nonlinear dynamics problem.

The square matrix $M$ is upper-triangular, and has the form:

\[
M = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & M_{11} & M_{12} & M_{13} \\
0 & 0 & M_{22} & M_{23} \\
0 & 0 & 0 & M_{33}
\end{bmatrix}
\] (5.2)

The zeros here represent sub-matrixes with all zero elements. $M_{11}, M_{22}$ and $M_{33}$ are diagonal matrixes with dimension $2 \times 2, 3 \times 3$ and $4 \times 4$ respectively, their diagonal elements are $\{e^{i\mu}, e^{-i\mu}\}, \{e^{2i\mu}, 1, e^{-2i\mu}\}, \{e^{3i\mu}, e^{i\mu}, e^{-3i\mu}\}$. They correspond to 1st, 2nd, and 3rd order terms in the series expansion respectively, i.e., length 2, 3 and 4. The eigenvalues of a triangular matrix are its diagonal elements, hence, these 9 numbers and the first diagonal element 1, are the 10 eigenvalues of $M$. The sub-matrixes $M_{12}, M_{13}$ and $M_{23}$ have dimension $2 \times 3, 2 \times 4$, and $3 \times 4$ respectively, represent the cross terms between 1st order and 2nd order, 1st and 3rd order, and between 2nd order and 3rd order respectively. As an example, from equation 2.2, we find

\[
\begin{align*}
M_{11} &= \begin{bmatrix} e^{i\mu} & 0 \\ 0 & e^{-i\mu} \end{bmatrix}, M_{22} &= \begin{bmatrix} e^{2i\mu} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e^{-2i\mu} \end{bmatrix}, \\
M_{23} &= \begin{bmatrix} -\frac{i}{2} e^{2i\mu} & -\frac{i}{2} e^{2i\mu} & -\frac{i}{2} e^{2i\mu} \\ \frac{i}{2} e^{2i\mu} & 0 & 0 \\ 0 & \frac{i}{2} e^{-2i\mu} & -\frac{i}{2} e^{-2i\mu} \end{bmatrix}
\end{align*}
\] (2.6)

This pattern can be generalized to higher order easily. For two variables such as $x$ and $p$, the number of terms of order $n$ is $n+1$. As seen from above, the numbers of terms in order of $1, 2, 3, \ldots, n$ are $2, 3, 4, \ldots, n+1$. The dimension of the square matrix is $(n+1)(n+2)/2$. Thus for 3rd order it is 10, as shown in the example. For 4 variables such as $x, p, x, p$, the numbers of terms in order of $1, 2, 3, \ldots, n$ are $4, 10, 20, \ldots, (n+1)(n+2)(n+3)/6$. The square matrix dimension is $(n+1)(n+2)(n+3)(n+4)/24$. As an example, when truncated to 7th order, the matrix dimension for 4 variables is 330. This rapid increase of matrix dimension as the order increases seems to indicate a very fast increased complexity of the problem.
3. INVARIANT SUBSPACES AND JORDAN DECOMPOSITION

For a \( n \times n \) matrix \( M \), using Jordan decomposition [14](Golub, p.354), we can find a \( n \times n \) non-singular matrix \( U \) such that

\[
UM^{-1} = \begin{bmatrix}
U_1 \\
U_2 \\
\vdots \\
U_k
\end{bmatrix} M \begin{bmatrix} U_1 & U_2 & \ldots & U_k \end{bmatrix} = 
\begin{bmatrix}
N_1 & 0 & \ldots & 0 \\
0 & N_2 & \ldots & 0 \\
0 & 0 & \ldots & 0 \\
0 & 0 & \ldots & N_k
\end{bmatrix} = N
\]  

(3.1)

where the \( m_j \times m_j \) matrix \( N_j = \lambda_j I_j + \tau_j \) with \( j=1,2,\ldots,k \) is the Jordan block with eigenvalue \( \lambda_j \), corresponding to the invariant subspace \( j \) of dimension \( m_j \) in the \( n \) dimensional space of vector \( Z \). \( I_j \) is the identity matrix of dimension \( m_j \), while \( \tau_j \) is a superdiagonal matrix of dimension \( m_j \):

\[
\tau_j = \begin{bmatrix}
0 & 1 & 0 & \ldots & 0 \\
0 & 0 & 1 & \ldots & 0 \\
0 & 0 & 0 & \ldots & 1 \\
0 & 0 & 0 & \ldots & 0
\end{bmatrix}
\]  

(3.2)

\( U_j, U_j \) are \( m_j \times n \) and \( n \times m_j \) submatrixes respectively. Eq. (3.1) leads to the following equations: if \( j \neq l, U_j U_l = 0 \) and \( U_j M U_l = 0 \); if \( j = l, U_j U_l = I_j \) and

\[
U_j M U_j = \lambda_j I_j + \tau_j
\]  

(3.3)

In addition, Eq. (3.1) leads to \( UM = NU \), and hence \( U_j (M - \lambda_j I) = \tau_j U_j \). Similarly, we find \((M - \lambda_j I)U_j = \tau_j I_j \) (Notice the distinction: \( I \) and \( I_j \) are identity matrixes for full space \( Z \) and the subspace respectively.)

It is seen from these equations that when we study the nonlinear dynamics of the system, we can decompose the motion in the full space \( Z \) into the separate motion in many invariant subspaces. In particular, as we shall show, the motion in the eigenspace with eigenvalue \( e^{i\mu} \) provides a wealth of information about the dynamics. Actually, all other eigenvalues also provide this information. However, for the dynamics of the system of \( xy \) motion, we can concentrate on the two simplest invariant subspaces with eigenvalues \( e^{i\mu_x}, e^{i\mu_y} \) only.

A great simplification comes from the fact that we can obtain the transformation matrix \( U_j \) for a specific invariant subspace \( j \) without first having to find the full matrix \( U \). This is based on the fact that we can first find the solution of the left eigenvector equation \( U_j (M - \lambda_j I) = \tau_j U_j \) for a specific \( \lambda_j \) in the following way. As we point out in Section 2, (with details in Appendix A), for a specific eigenvalue \( \lambda_j \), we can first find all the generalized eigenvectors with this eigenvalue without the calculation of the eigenvectors of other eigenvalues. For example, in the case of 4 variables \( x, p_x, y, p_y \) up to 7th order, because we find that the number of diagonal elements with the eigenvalue \( e^{i\mu_x} \) is 10, the total number of generalized eigenvectors is also 10. They form an invariant subspace of dimension 10 even though the full space has dimension 330. If we label these left eigenvectors by \( e_i \), since they form an invariant subspace, when we find them by solving the triangular matrix \((M - \lambda_j I) \) (see Appendix A), we find they satisfy the following equation:

\[
e_i (M - \lambda_j I) = t_{ik} e_k
\]  

(3.4)

Here \( t \) is the matrix derived in Appendix A, we used the Einstein convention: the repeated \( k \) implies a sum over \( k \). The index \( i \) and \( k \) run from 1 to \( m \), where \( m \) is the total number of diagonal elements of \((M - \lambda_j I) \) with eigenvalue equal to zero, i.e., the (generalized) null
space dimension of the matrix \((M - \lambda I)\). In the example above, the dimension of \(t\) matrix is 10. Hence our main issue is greatly simplified to finding the Jordan matrix of \(t\). This involves a much smaller amount of work compared to finding the Jordan form of the matrix \((M - \lambda_j I)\) itself.

Once we find the Jordan decomposition \(T, t = g^{-1}Tg\) with \(T\) in Jordan form, we find \(g_h e_i (M - \lambda_j I) = T_{hj} g_i e_i\). Now we can see that the new basis \(u_k \equiv g_h e_i\) satisfy the left eigenvector equation \(u_h (M - \lambda_j I) = T_{hn} u_n\). Since \(T\) is in Jordan form with eigenvalue zero only, it has the same form as \(N\) in Eq. (3.1), but with \(N_j = \tau_k\) now. Hence the subspace itself is again separated into several invariant subspaces, all of them have the same eigenvalue zero. About how to find these invariant subspaces, please see Appendix B. For the example mentioned above, the 10 dimensional invariant subspace is separated into 4 subspaces again, with dimension 4, 3, 2, 1 respectively, each is spanned by a chain of generalized eigenvectors. Similarly, we can find the invariant subspaces of right generalized eigenvectors. Thus we find the structure of chains in the invariant subspace of one eigenvalue, with each chain corresponds to one Jordan block, serves as the basis of one of the method of Jordan decomposition. There are many programs available for Jordan decomposition, including some of them providing analytical solution. But occasionally the result is unstable. To ensure stable result, we adopt the method by Axel Ruhe (1970), Käström and Ruhe (1980a, 1980b)[15–17], which is referred to by the text book "Matrix Computations, 4th Edition" in page 402 of Golub [14]. For the convenience of the readers, in Appendix B, we outline the crucial steps of the method without the detailed derivation and the proof of the stability of the decomposition, which are given in these papers.

In order to study the dynamics of the system, as we shall show, the most important information is obtained from the Jordan decomposition of \(\ln M\) rather than \(M\) itself. As explained in the Appendix C, we can take a logarithm of Eq. (3.1), then, we can carry out a second Jordan decomposition of \(\ln N\) up to the same order easily and precisely, and arrive at:

\[
U \ln M U^{-1} = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_k \end{bmatrix} \ln M \begin{bmatrix} \bar{U}_1 & \bar{U}_2 & \ldots & \bar{U}_k \end{bmatrix}
\]  
(3.5)

Here we redefined the transformation matrix and Jordan form as \(U\) and \(N\) again to avoid cluttering of notations. After we replace Eq. (3.1) by Eq. (3.5), the formulas following Eq. (3.1) remain the same except the eigenvalue \(e^{\mu_i}\) is replaced by \(i \mu\) so that now \(N_j = i \mu_j I_j + \tau_j\) if we are interested in the Jordan block \(j\) with tune \(\mu_j = m_x \mu_x + m_y \mu_y\).

Since we are mostly only interested in the analysis of the Jordan blocks with tune either \(\mu_x\) or \(\mu_y\), and as we show in the Appendix A, B and this section that all the (generalized) left and right eigenvectors in \(U_j\) and \(\bar{U}_j\) can be calculated without having to find the full matrix \(U\) first, we simply limit to the corresponding invariant subspace, drop the index \(j\) in \(U_j\) and rewrite Eq. (3.3) as

\[
U \ln M U = i \mu I + \tau
\]  
(3.6)

Notice now we use \(U\) and \(\bar{U}\) to represent the submatrix of the transformation matrix. Thus we have

\[
U M U = e^{\mu I + \tau}.
\]  
(3.7)

Or, when we write this in the form of (generalized) left and right eigenvectors,

\[
U M = e^{i \mu I + \tau} U
\]

\[
M \bar{U} = \bar{U} e^{i \mu I + \tau}.
\]  
(3.8)

As mentioned following Eq. (3.2), the rows in \(U\) and columns in \(\bar{U}\) are orthogonal:

\[
U \bar{U} = I
\]  
(3.9)

Where \(I\) is the identity matrix with dimension equal to the dimension of the invariant subspace. For the example of 4 variables \(x, p_x, y, p_y\), at 7th order, the subspace of eigenvalue \(e^{i \mu_x}\) is 4 for the longest chain. For the other 3 shorter chains, the dimension is 3, 2, 1 respectively. We will concentrate on studying the longest chain because it has most detailed information about the nonlinear dynamics.

4. ROTATION IN THE INVARIANT SUBSPACE, TUNE AND AMPLITUDE DEPENDENT TUNE

As \(Z = MZ_0\) (see Eq. (2.3)), Eq. (3.8) gives

\[
U Z = U M Z_0 = e^{i \mu I + \tau} U Z_0.
\]  
(4.1)
Now we define a transformation
\[ W \equiv UZ \]
\[ W_0 \equiv UZ_0 \] (4.2)

Then Eq. (4.1) implies the operation of one turn map
\[ Z = MZ_0, \] corresponds to a rotation in the invariant subspace represented by
\[ W = e^{i\mu l + r}W_0. \] (4.3)

This representation of the rotation in the invariant subspace of lower dimension is much simpler than the matrix \( M \) in the full space of \( Z \).

KAM theory states that the invariant tori are stable under small perturbation (See, for example, [1, 8, 18]). In our examples, for sufficiently small amplitude of oscillation in \( z \), the invariant tori are deformed and survive, i.e., the motion is quasiperiodic. So the system has a nearly stable frequency, and when the amplitude is small, the fluctuation of the frequency is also small. Thus for a specific initial condition described by \( Z_0 \), the rotation in the eigenspace should be represented by a phase factor \( e^{i(\mu + \phi)} \) so that
\[ W = e^{i\mu l + r}W_0 \cong e^{i(\mu + \phi)}W_0. \] (4.4)

This approximation implies that the operation of the matrix \( e^{i\mu l + r} \) on the vector \( W_0 \) in the eigenspace corresponds to replacing the matrix \( \tau \) by a number, i.e., a phase advance \( i\phi \). This means \( W_0 \) is considered to be approximately an eigenvector of \( \tau \):
\[ \tau W_0 \cong i\phi W_0. \] (4.5)

The expression of \( \tau \) given by Eq. (3.2) is well known to have no (proper) eigenvector but only has generalized eigenvectors. However, as the dimension of the eigenspace increases and approaches infinity, the eigenvector of \( \tau \) is defined as a coherent state [19, 20]. Hence Eq. 4.5 means \( W_0 \) is a "coherent state".

We write \( W_0 \) as a column with
\[ \tilde{W}_0 = (w_0, w_1, w_2, \ldots, w_{m-1}) \] (4.6)
where \( m \) is the dimension of the invariant subspace. Based on Eq. (4.2), \( W_0 \equiv UZ_0 \), if we write the rows of \( U \) as \( u_0, u_1, u_2, \ldots, u_{m-1} \), we have the polynomials
\[ w_0 = u_0Z_0, w_1 = u_1Z_0, w_2 = u_2Z_0, \ldots, w_{m-1} = u_{m-1}Z_0. \]
Then, using Eq. (3.2), Eq. (4.5) becomes
\[ \tau \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{m-1} \end{bmatrix} \cong \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ 0 \end{bmatrix} \cong \begin{bmatrix} i\phi w_0 \\ i\phi w_1 \\ \vdots \\ i\phi w_{m-1} \end{bmatrix} \] (4.7)

If the dimension of the subspace of Jordan block is infinity, there is no last row in the column. But when it is finite we see that the last row cannot be satisfied unless \( w_{m-1} \) is zero. However, we find that the lowest power in \( w_m \) always increases by 2 as \( m \) increases, so \( w_{m-1} \) has only high power terms, hence when \( m \) is large it is negligibly small if \(|z| \) is sufficiently small. Thus when the invariant tori survive and there is a stable frequency, we see that Eq. (4.7) requires
\[ i\phi = \frac{w_1}{w_0} = \frac{w_2}{w_1} = \frac{w_3}{w_2} = \ldots = \frac{w_{m-1}}{w_{m-2}} \] (4.8)

Therefore only those vectors \( W_0 \) which satisfy Eq. (4.8) with \( \phi \) a real number represent a motion with a stable frequency given by a phase advance \( \mu + \phi \) every turn. Since \( w_0 = u_0Z_0, \ldots \), we see that \( w_0 = u_0Z_0 \) means \( \phi \) is determined by the initial value \( z_0 \). Hence \( \mu \) represents the zero amplitude tune while \( \phi \) is the amplitude dependent tune shift.

We remark that in practice, only the first few equal signs of Eq. (4.8) are relevant. As \( m \) increases the ratio \( w_{m-1}/w_{m-2} \) becomes the ratio of very small numbers and loses the information about the nonlinearity.

When these conditions are satisfied, the Eq. (4.4) is also satisfied, i.e., \( W \cong e^{i(\mu + \phi)}W_0 \). Hence we have \( w_0(n = 1) = e^{i(\mu + \phi)}w_0(n = 0), w_1(n = 1) = e^{i(\mu + \phi)}w_1(n = 0), \ldots \). Here we use \( n \) as number of turns. That is, after each turn, every row in \( W_0 \) rotates by a factor \( e^{i(\mu + \phi)} \) in their separate complex planes like in a twist map (see, e.g., [1, 8]), or, behaves like an action-angle variable. For example, let \( w_0 = r e^{i\theta} \), then \( r = |w_0| \) remains the same like an action variable, while \( \theta \to \theta + \mu + \phi \) after each turn like the angle variable. As we shall explain later in the section "Stability, Precision, and Uniqueness", even though \( w_0, w_1, w_2, \ldots, w_{m-1} \) all behave like action-angle variables, they have different power orders of monomials of \( z, z^* \), and hence represent different degree of approximation to the action-angle variable. For example, in the case of 7th order square matrix for two variables \( x \) and \( p \), \( w_0 \) has terms of powers from 1 to 7, \( w_1 \) has terms of power of 3 to 7 while \( w_3 \) has only a very small term \( z(z^*)^3 \) of power 7. Thus \( w_3 \) provides very little information about the rotation in the phase space while \( w_0 \) has very detailed information. The detailed discussion of how to extract most detailed information by minimizing high power terms is given in the section "Stability, Precision, and Uniqueness".

Near the border of the survival invariant tori, for example, if the system is near its dynamic aperture or near a major resonance, the condition of Eq. (4.8) and the condition that \( \phi \) is real are violated slightly, and the Eq. (4.4) also has increased errors, for convenience we call this situation as a deviation from a coherent state. Hence these conditions provide information about dynamic aperture and resonances. The next section is concentrated on this discussion.
5. MULTI-TURNS, TUNE AND AMPLITUDE FLUCTUATION

We now consider the map after $n$ turns. Applying Eq. (4.3) $n$ times, we obtain
\[
W(n) = e^{i\pi l + n\tau} W_0 = e^{i\pi \mu e^{n\tau}} W_0, \quad (5.1)
\]
with $W(n=0) \equiv W_0$, and we have moved the constant $e^{i\pi \mu}$ to the front, dropped the identity matrix $I$ to remind us that it is a constant. Before expanding Eq. (5.1), we follow the Dirac notation, let
\[
|0\rangle \equiv \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, |1\rangle \equiv \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, |2\rangle \equiv \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}, \ldots, |m-1\rangle \equiv \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix},
\]
\[
\text{(5.2)}
\]
Using the expression Eq. (3.2) for $\tau$, we find chain $\tau |0\rangle \equiv 0$, $\tau |1\rangle \equiv |0\rangle > 0$, $\tau |2\rangle \equiv |1\rangle > 0$, $\tau |3\rangle \equiv |2\rangle > 0, \ldots$. And, $\tau^2 |0\rangle \equiv 0$, $\tau^2 |1\rangle \equiv 0$, $\tau^2 |2\rangle \equiv |0\rangle > 0$, $\tau^2 |3\rangle \equiv |1\rangle > 0, \ldots$. Hence we have
\[
W_0 = w_0 |0\rangle + w_1 |1\rangle + w_2 |2\rangle + \cdots + w_{m-1} |m-1\rangle + 1 > 0
\]
\[
e^{n\tau} = 1 + n\tau + \frac{n^2}{2}\tau^2 + \cdots + \frac{m-1}{(m-1)!} \tau^{m-1}
\]
\[
\text{(5.3)}
\]
\[
e^{n\tau} W_0 =
\]
\[
w_0 |0\rangle + w_1 |1\rangle + w_2 |2\rangle + \cdots + w_{m-1} |m-1\rangle + 1 > 0
\]
\[
n(w_1 |0\rangle + w_2 |1\rangle + \cdots + w_{m-1} |m-1\rangle + 2 > 0 + \cdots)
\]
\[
\frac{n^2}{2} (w_2 |0\rangle + w_3 |1\rangle + \cdots + w_{m-1} |m-3\rangle + \cdots)
\]
\[
\text{(5.4)}
\]
Thus we find
\[
W(n) = e^{i\pi \mu} e^{n\tau} W_0 =
\]
\[
e^{i\pi \mu} (w_0 + nw_1 + \frac{n^2}{2} w_2 + \cdots) |0\rangle +
\]
\[
e^{i\pi \mu} (w_1 + nw_2 + \frac{n^2}{2} w_3 + \cdots) |1\rangle + \cdots
\]
\[
\text{(5.5)}
\]
\[
\text{Compare this with the expression of } W(n), \text{ we find}
\]
\[
w_0(n) = e^{i\pi \mu} (w_0 + nw_1 + \frac{n^2}{2} w_2 + \cdots) =
\]
\[
e^{i\pi \mu} (1 + \frac{w_1}{w_0} + \frac{n^2}{2} \frac{w_2}{w_0} + \cdots) =
\]
\[
e^{i\pi \mu + \ln(1 + \frac{w_1}{w_0} + \frac{n^2}{2} \frac{w_2}{w_0} + \cdots)} w_0 =
\]
\[
e^{i\pi \mu + \frac{w_1}{w_0} + \frac{n^2}{2} \left( \frac{w_2}{w_0} - \left( \frac{w_1}{w_0} \right)^2 \right) \cdots} w_0 =
\]
\[
e^{i\pi \mu + i\phi + \frac{n^2}{2} \Delta + \cdots} w_0.
\]
\[
\text{(5.6)}
\]
Here we have defined $\phi$ and $\Delta$ ,
\[
\text{i} \phi = \frac{w_1}{w_0}; \Delta = \frac{w_1}{w_0} - \left( \frac{w_1}{w_0} \right)^2.
\]
\[
\text{(5.7)}
\]
To avoid cluttering of symbols, all $w_j$ without specification of $n$ here represent $w_j(n = 0)$. When compare with Eq. (4.4), we identify $\Re \phi$ as the amplitude dependent tune shift. In the region in the phase space where the invariant tori survive with stable frequency, we recognize that $\phi$ is real and remains to be a constant along a circle with radius $r = |w_0|$. In addition, $\Delta$ should be nearly zero in order for the system to have a stable frequency. For convenience, we shall refer these as coherence condition:
\[
\Im \phi \approx 0; \Delta \approx 0.
\]
\[
\text{(5.8)}
\]
Just as pointed out in Section 4, we remark that we pay attention only to the first few terms in the exponent of the right hand side of Eq.5.6 because as $m$ increases the ratio $w_{m-1}/w_{m-2}$ becomes the ratio of very small numbers and lost information.

The analysis given by Eq. (5.6)- Eq. (5.7) paints a physical picture about why and how a chain represents a rotation in the phase space: we need not only $w_0$ to represent a rotation, we also need $w_1$ and $w_2$ to provide information about the phase advance and how stable the frequency is. Thus the phase space is divided into many invariant subspaces, each represents a rotation through the phase shift generated by a chain. For each eigenvalue, the longest chain provides the most detailed information about the rotation while the shorter chains and their sub-chains represent approximation with only high power terms and less information. The invariant subspaces of different eigenvalues represent the rotation of different harmonics of the system.

Clearly $w_0, \phi$, and $\Delta$ are all functions of initial value of $z, z^*$. Or, if we use inverse function of $w_0(z)$ to represent $z$ as function of $w_0$, then $\phi$ and $\Delta$ both are functions of $w_0$. Thus, given initial value of $w_0$, we can examine whether $\phi$ is constant along a circle with radius $r = |w_0|$, whether it is a real function, and whether $\Delta$ is nearly zero, and obtain the information about whether the initial state is close to the border of the survival invariant tori or near resonance. The deviation of the real part of $\phi$ from a constant is the tune fluctuation, while the imaginary part of $\phi$ gives amplitude fluctuation, i.e., the variation of $r = |w_0|$ after many turns. The non-zero $\Delta$ indicates a deviation from coherent state, seems to be related to the Liapunov exponents [1](p.298). All of these can be checked by simulation with calculation, as we shall discuss in detail in Section 7. Actually it is obvious that the condition of $\Delta = 0$ is consistent with the coherence condition Eq. (4.8).

However, before this, we first digress to a discussion of stability, precision, and uniqueness of the square matrix Jordan decomposition, as this is an issue often raised whenever one start to talk about Jordan decomposition.
6. STABILITY, PRECISION, AND UNIQUENESS OF THE SQUARE MATRIX JORDAN DECOMPOSITION

A. Use Scaling to Ensure Stability and Precision of Jordan decomposition

Very often, Jordan decomposition of a matrix is considered to be ill conditioned. However, due to the tridiagonal property and because its eigenvalues are precisely known on the unit circle, the matrix $M$ is already in the form of a stable Schur decomposition[14], and its Jordan decomposition is stable except when we are very close to resonance where some eigenvalues are nearly degenerate. When we are sufficiently far from resonance, all the eigenvalues of the matrix $(M - \lambda I)$ other than zeros are sufficiently far from zero so the null space, i.e., the invariant subspace with eigenvalue $\lambda$ can be solved for its eigenvectors to very high precision.

However, we may lose precision during the Jordan decomposition of the matrix $t$ in the invariant subspace as given in Eq. (3.4) when we use high order square matrices. Eq. (3.4) tells us that when $(M - \lambda I)$ acts on the vectors in the invariant subspace of eigenvalue $\lambda$ from right, it is equivalent to the much lower dimension matrix $t$ acts on the vector from the left. Hence in the following discussion we speak of them as if we were talking about the same thing.

The procedure of Jordan decomposition is based on the fact that the invariant subspace is spanned by several chains of eigenvectors, and the end of each chain contributes to the null space of the matrix: $u_{m-1}(M - \lambda I) = 0$, as explained in Section 3 and in Appendix B in particular. Using Eq. (3.4), we see that the last right eigenvector (the proper right eigenvector) of $t$ for each chain contributes to the null space of $t$. So finding the null space of $t$ gives the space span by the ends of all the chains. Now because the chain relation, we have $u_{m-2}(M - \lambda I)^2 = 0$. Hence for the same reason finding the null space of $t^2$ gives us the space span by the 2nd last vectors of all the chains, in addition to the null space of $t$. This continues until $u_0(M - \lambda I)^{m-1} \neq 0$ while $u_0(M - \lambda I)^m = 0$, where $m$ is the length of the longest chain. In our example of 4 variables at 7th order, there are 4 chains in the invariant subspace of dimension 10, with length $m=4, 3, 2, 1$ respectively. When the null spaces of different powers of $t$ are calculated, it then becomes easy to build the vectors in each chain and establish the Jordan decomposition. More details about this procedure are given in Appendix B, and we refer to reference [15–17]. Here we only point out that our method of Jordan decomposition is based on finding the null spaces of different powers of $t$. To find the null space correctly we need to distinguish small singular values from zero using singular value decomposition. The high power of $t$ may have very small singular values almost reaching the machine precision. In this case we cannot correctly separate the null space anymore, and the Jordan decomposition fails. Hence we need narrow the range of the singular values so that we can distinguish the minimum singular value from zero clearly.

We found that the range of singular values depends on the range of the absolute value of the coefficients of monomial terms at different orders in the square matrix $(M - \lambda I)$. By scaling the variables $z_x = x - ip_x, z_y = y - ip_y$ by a factor $s$, we can reduce the ratio of the maximum and the minimum absolute value of the coefficients above zero. For example, we found for a specific lattice the square matrix has its maximum of the absolute value of the coefficients at 7th order as $1.38 \times 10^{18}$, and the minimum of those other than zeros is found to be the first order as 1 (the coefficients of order of machine precision are excluded). Then let $s^7 = 1.38 \times 10^{18} = 1 \times s$, we find $s = (1.38 \times 10^{18})^{-\frac{1}{7}} = 0.00115$. Now we let $s_{z_x} = x - ip_x, s_{z_y} = y - ip_y$ to make the coefficients of these two terms equal, then we find that the range of the coefficients in the new square matrix using new variables has the coefficients span much smaller range. And the range of singular values is reduced from 18 orders of magnitude to between 0.03 and 35 after the scaling. In Fig.1 we show the spectrum of the singular values before (red) and after the scaling (blue). Therefore, the null space of the invariant subspace is clearly identified; the Jordan decomposition is very stable and accurate.

![FIG. 1: The range of singular value of the square matrix before (red) and after scaling (blue)](image)
the coefficients of the square matrix, we consider the coefficients in the Taylor expansion in Eq. (2.1). The result not only reduces the error of the square matrix, it also reduces the range of the singular values of the matrix \((M - \lambda I)\). The improvement of the scaling before the construction of the square matrix is so significant that in all the cases we studied, there is very little improvement from the second scaling based on the coefficients of the full square matrix. However, since the second scaling is very simple, and does provide some improvement over the maximum-minimum singular value ratio, we always carry out a second scaling.

When on resonance, the eigenvalue \(e^{\mu x}, e^{\mu y}\) is degenerate to the eigenvalue of other harmonic of the tune, in the form of \(e^{(m\mu x + n\mu y)}\). The structure of the Jordan block is different, and the resonance issue should be treated separately, and is not discussed in this paper. However, as our numerical examples show later, in Section 7, our present analysis of nonlinear dynamics by square matrix method is valid even when it is quite close to a resonance.

### B. Non-uniqueness of Jordan decomposition

Another issue of Jordan decomposition is whether it is not unique, and whether we need to impose other conditions based on physics to make it unique. The non-uniqueness of Jordan decomposition is due to its chain structure. A simple example is a chain of two vectors \(u_0\) and \(u_1\), with \(Mu_0 = u_1\) and \(Mu_1 = 0\). Then it is obvious that another two vectors \(u_0' = u_0 + au_1\) and \(u_1' = u_1\) also satisfy the same chain relation: with \(Mu_0' = u_1'\) and \(Mu_1' = 0\), where \(a\) is an arbitrary constant. Hence the basis of Jordan decomposition is not uniquely defined, and the question now is what kind of conditions based on physics will allow us to determine the choice of \(a\). To solve this problem, we must first understand the specific structure of the chains.

### C. The structure of the chains in the Jordan decomposition of nonlinear dynamics square matrix

In the case of two variables \(x\) and \(p\), we find only one chain for one eigenvalue \(e^{i\theta} \). For example, for 7th order, the invariant subspace we found has 4 (generalized) eigenvectors \(u_0, u_1, u_2, u_3\) as the basis of the space. We find that \(u_0 = u_0^Z\) is a polynomial with powers from 1 to 7, \(u_1 = u_1^Z\) has powers from 3 to 7, \(u_2 = u_2^Z\) has powers from 5 to 7, \(u_3 = u_3^Z\) has only a term of power 7, which is proportional to \(z(z^2)^{3}\). When the matrix \((M - i\mu I)\) operates from right on the invariant subspace of \(Z\) spanned by \(u_0, u_1, u_2, u_3\), its operation in the representation using \(u_0, u_1, u_2, u_3\) as basis, is given by the simple matrix \(\tau\) (see Eq. (3.6)). Using the equations of left eigenvectors \(U(\ln M - i\mu I) = \tau U\), we have \(u_0(\ln M - i\mu I) = u_1, u_1(\ln M - i\mu I) = u_2, \ldots, u_3(\ln M - i\mu I) = 0\), i.e., they form a chain. Intuitively, the following helps to understand why when multiplied by the matrix \((\ln M - i\mu I)\), the lowest power terms of the eigenvectors increase progressively by power of 2: the off-diagonal terms in the matrix has at least power of 2.

This is for two variables \(x\) and \(p\). For 4 variables such as \(x, p, y, p_y\), as in the case we studied for a storage ring, there are more independent chains than one for the invariant subspaces of both eigenvalues \(e^{i\mu x}, e^{i\mu y}\). For example, at 7th order, for the eigenvalue \(e^{i\mu x}\), in addition to the chain \(u_0, u_1, u_2, u_3\), there is a linear independent chain \(u_0', u_1', u_2', u_3'\) and a chain with only one element \(u_0''\), forming an invariant subspace of dimension 10. The chains of this eigenvalue span an invariant subspace of dimension 10. There is only one lowest power term in \(u_0 z_x = -ip_x\). We list the lowest power terms of the chains in this example in Table I.

| Terms of | Lowest power | 1 | 3 | 5 | 7 |
|-----------------|-----------------|---|---|---|---|
| \(z_x(z_xz_x^*)\) | \(z_x(z_xz_x^*)^2\) | \(z_x(z_xz_x^*)^3\) | \(z_x(z_yz_y^*)\) | \(z_x(z_yz_y^*)^2\) | \(z_x(z_yz_y^*)^3\) |

Thus the pattern appears: the lowest power terms in the first vector of a chain or sub-chain are always terms with \(z_x\) times the powers of the invariant monomial \(z_xz_x^*\) or \(z_yz_y^*\) so that for small amplitude, it represents a simple rotation as \(z_x\). The first vectors of the shorter chains have terms with \(z_x\) multiplied by higher powers of these invariants. This feature helps us to understand its effects on the non-uniqueness of the Jordan decomposition.

### D. Physical meaning of non-uniqueness

To understand the meaning of the non-uniqueness, let us take \(w_0 = z + z^2 = z(1 + z^2)\), \(w_1 = z(z^2)^{3}\). That is the simplest case where the end of the chain is \(z\) multiplied by an invariant factor of \(z^2\). When \(|z| < < 1\), for a circular motion in \(z\)-plane, both \(w_0\) and \(w_1\) correspond to a circular motion in \(w\)-plane. But as \(z\) increases, \(|w_0|\) no longer remains constant because as the phase of \(z^2\) changes the factor \(1 + z^2\) has interference between its two terms. This modulation of amplitude gives distortion of the trajectory, because a constant \(|w_0|\) does not correspond to a circular trajectory in \(z\)-plane any more. This means \(w_0\) carries information about the distortion.
of the trajectory, while \( w_1 \) does not carry this information. When \( w_1 \) is multiplied by a large number and added to \( w_0 \), it dominates over \( w_0 \), and the information about distortion lost.

It is clear now that those terms of \( z \) times the powers of the invariant of form \( z^3 \) represent pure circular motion in phase space and they do not present any information about nonlinearity. Therefore when they are mixed into the first vector of the longest chain, they blur the distortion generated by the interference between the linear term and terms of other harmonics such as \( z^3 \), and the result is the non-uniqueness.

### E. The need to minimize high power terms from the first vector of the longest chain

In Table I, terms such as \( z^3 \) can appear in \( u_0 \) but it is not in the lowest power terms in \( u_0 \) because if it is then it will create a tune \( 3\mu_x \) rather than \( \mu_x \) even for small amplitude. These terms such as \( z^3 \) interfere with the dominating first order term \( z \) in \( u_0 \) to generate distortion. A mixture of long chain with short chains in Table I such as \( u_0^{(1)} \equiv u_0 + au_1 + bu_0^3, u_1^{(1)} \equiv u_1 + au_3 + bu_1^3, u_2^{(1)} \equiv u_2 + au_3 + bu_2^3, u_3^{(1)} \equiv u_3 \), is still a chain. If \( a \) or \( b \) are very large, the distortion generated by \( z^3 \) will be dominated over by other high power terms and the trajectory becomes more close to a circle with lost information. Hence we need to choose \( a \) or \( b \) to minimize the high power terms. It is obvious that a polynomial with many terms should be able to describe much more detailed complicated curve or surface than a single monomial. Also, it is easy to understand that we need to minimize high power terms to increase the convergence radius of a Taylor series.

Therefore, it becomes clear, to extract more detailed information from \( u_0 \) we need to find a way to minimize higher power terms while maintain a chain satisfying the left eigenvector relation in the invariant subspace. Thus in the Appendix D, we study and find the linear combination of sub-chains such that the higher power terms are minimized. For example, the sub-chain \( u_1, u_2, u_3 \) is used to minimize the 3rd power terms in \( u_0 \), the sub-chain \( u_2, u_3 \) is used to minimize the 5th power terms in \( u_0 \), and the sub-chain \( u_3 \) is used to minimize the 7th power terms in \( u_0 \).

We remark here it is observed that in the Table I the number of chains for each order happen to be equal to the number of lowest power terms, so that the linear combination of these chains can be used to remove these terms from \( u_0 \) completely. Hence when all these higher power terms are minimized, it amounts to removing all the terms of form of \( z^3 \) times the powers of the invariants, this is equivalent to separate the shorter chains from the longest chain. The linear combination is uniquely determined, and hence the Jordan decomposition is stable, accurate, and unique. The procedure is described in Appendix D.

### 7. COMPARISON WITH SIMULATION

#### A. Summary about application of the theory

When we study the nonlinear dynamic equations such as Hill equations, in order to have the square matrix to be tridiagonal, we always first convert the variables \( x, p \) into normalized Courant-Snyder variables \( x, \bar{p} \) using the betatron amplitude matrix \( B^{-1} \) (see, e.g., S.Y. Lee, p.49 [21]), then convert to the scaled variables \( z = \bar{z}/s \equiv (\bar{x} - i\bar{p})/s, z^* = z^*/s \equiv (\bar{x} + i\bar{p})/s \), with the matrix \( K^{-1} \) and the scaling parameter \( s \) as given in Section 6:

\[
\begin{bmatrix}
\bar{x} \\
\bar{z}^*
\end{bmatrix} = K^{-1} \begin{bmatrix}
x \\
p
\end{bmatrix} = K^{-1} B^{-1} \begin{bmatrix}
x \\
p
\end{bmatrix},
\]

\[
B^{-1} = \begin{bmatrix} K_1^{-1} & 0 \\ \sqrt{\gamma} & \sqrt{\beta} \end{bmatrix}, K^{-1} = \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix}
\]

For 4 variables such as \( x, p_x, y, p_y \), it is similar, with both \( B^{-1} \) and \( K^{-1} \) replaced by 4×4 matrixes. But we must first carry out linear decoupling between \( x \) and \( y \). As we point out in Section 6, we select the scaling factor for the variable to minimize the range of the coefficients in the square matrix.

For every element in a storage ring one can construct a corresponding square matrix. For a linear element, there are only linear terms in the Taylor expansion corresponding to Eq. (2.2), hence the matrix corresponding to Eq. (2.3) has nonzero terms only in the diagonal blocks such as \( M_{11}, M_{22} \) and \( M_{33} \). All the off-diagonal blocks such as \( M_{12}, M_{13}, M_{23} \) are zeros. The product of all the square matrixes for all the elements around the ring gives the square matrix \( M \) for the one turn map.

However, during our calculation we use the Taylor expansion provided by the well-known program of TPSA (truncated power series algorithm) [6] as our starting point of the one turn map corresponding to Eq. (2.1). We construct the square matrix \( M \) up to a certain order. Then we apply Jordan decomposition for the invariant subspaces to obtain the transformation matrix \( U \) for the specific eigenvalues \( e^{i\mu_x}, e^{i\mu_y} \). As described in Section 5, we use the first row of \( U \) to obtain the polynomial \( w_0 = u_0 Z \). This serves as an approximation for the action-angle variable (for 4 variables case we have \( w_x = u_0 Z, w_y = u_0 Z \)). Then the 2nd and 3rd row of \( U \) are used to calculate the functions \( \phi \) and \( \Delta \) using...
Eq. (5.7). We remark here that these are no longer polynomial, they are rational functions.

\[w_x \equiv w_{x0}(z_x, z_y), \quad w_y \equiv w_{y0}(z_x, z_y)\]  (7.2)
FIG. 6: The radius r (left) and Δθ (right) after one turn calculated from tracking (solid), and from $i\phi + \Delta/2$ (dots) for various initial x=8,12,16,20,24mm

FIG. 7: plot the difference of maximum and minimum of $|\Delta w/w|$ and $\Delta\theta$ as function of initial x. Tracking (green) is compared with prediction by (red) multiplied by a constant to obtain agreement with the tracking

define the set of variables $w_x, w_y, w_x^*, w_y^*$ as functions of $z_x, z_y, z_x^*, z_y^*$. Notice that because $w_0$ is used very often, we simply use $w$ to represent it, and from now on often we use $w_{x0}$ to represent its initial value when we specify it. Eq. (4.4) and Eq. (5.6) shows that when the coherence condition Eq. (5.8) is satisfied, these two equations provide a transformation from $z_x, z_y$ to the new variables as an approximation to action-angle variables. The inverse function of these functions would be very useful because a set of constant $|w_x|, |w_y|$ describes the motion of the particles. For this purpose we construct a column $\Phi$ with the rows given by $w_x, w_y, w_x^*, w_y^*$ and the monomials constructed from them so its transposition $\tilde{\Phi}$ is similar to the row defined by Eq. (2.4).

Now instead of Eq. (2.3), we can construct a square matrix $V$

$$\Phi = VZ,$$

(7.3)

where $V$ is also a tridiagonal matrix. It is well-known that the inverse matrix of a tridiagonal matrix is easy to calculate as long as its diagonal elements have no zeros, the process does not involve any limiting process, but only addition, multiplication, and division (See Appendix E). The linear part of the polynomial is very simple because as $z_x, z_y$ approach zero, they are proportional to $w_x, w_y$. We can always choose to multiply $U$ by a constant and divide $\overline{U}$ by the same constant (there are 2 transformation matrixes $U$ for x and y separately but here we are not specific about this point) so that $w_x, w_y$ approach $z_x, z_y$ respectively as they approach zero. Thus from Section 2 we find that $V$ has all its diagonal elements equal to 1. Thus it is easy to calculate the inverse
FIG. 8: Poincare surface of section using Courant-Snyder normalized variables $\bar{x}, \bar{y}$ (top row) compared with new variables $w_x, w_y$ (bottom row). Left two plots are for $x$, right are for $y$. Initial values of $|w_x0|, |w_y0|$ for all points are same as a point generated from $x=10\text{mm}, y=2\text{mm}$.

matrix $V^{-1}$ which can now be used to calculate $z_x, z_y$ approximately when $w_x, w_y$ are given. The result can be used as a set of initial trial values for a more accurate solution of the inverse function of the functions Eq. (7.2) and their conjugates. The inverse function solution using $V^{-1}$ for many points on the $w$-plane with $|w_0| = 0.8$ is shown as blue dots in Fig. 2, in agreement with the green contour. The plot shows the triangular inverse matrix $V^{-1}$ gives an excellent solution already, and there is no need to further improve the precision by solving the inverse function more precisely.

In Fig. 3, for the points around a circle in $w$-plane (i.e., a blue circle with $r=\text{constant}$), we use inverse function of $w_x(z_0)$ to find a set of initial $x_0, p_{x0}$ ($y_0, p_{y0}$ are set zero), then after tracking these particles for one turn, calculate $w_x(z_x)$ and plot the red curves. We can see that when $x$ approaches the dynamic aperture at $x=25\text{mm}$, $w_x$ (red) gradually deviates from the circles (blue).

In Fig. 4, for another one of the lattices of NSLSII, we plot on $\bar{z}_x$-plane a circle which passes through the point corresponding to initial $x_0 = 25\text{mm}, p_{x0} = 0$, and plot the tracking result over 512 turns as the magenta curve. In Fig. 4, we also plot all the points calculated according to a constant $|w_x| = r_x$ such that $r_x = |w_x(x = 25\text{mm}, p_x = 0)|$. The calculation is up to 5 (blue), 7 (green), and 9 (red) order, respectively. It is clear the agreement is excellent. To see the errors of different order, we plot the same set of data in Fig. 5 with fine details, showing $|z_x|$ as function of angle $\psi_x$ (the phase angle of $\bar{z}_x = J_x e^{i\psi_x}$). We can see that as the order for the constant $|w_x|$ increases, the agreement with the tracking result (magenta curve) converges slowly, with the 9th order (red) more close to the tracking.

In Fig 6 we show the radius $r$ in $w$-plane and the change of $\theta$ (i.e., phase advance) after one turn as functions of initial phase $\theta_0$ for various initial $r$, which corresponds to different initial $x$. The dotted curves are from tracking, solid curves are calculated from $i\phi + \Delta/2$. It is clear from Fig. 6 that as $r$ increases, both $r$ and $\Delta\theta$ have larger and larger variation. For large amplitude $x=22\text{mm}$, we can see that the two curves do not agree with each other, even though the trend of increased variation is obvious.
The theoretical prediction of $r$ variation is given by $\text{Im}\phi$, while the prediction on $\Delta\theta$ is given by the variation of $\text{Re}\phi + \text{Im}\Delta/2$ as function of $\theta$. These variations are an indication of deviation from coherence, i.e., a violation of the condition given by Eq. (4.8). Thus when this condition is violated, the calculation of $r$ and $\phi$ lost precision, hence they deviate from the tracking results. However, even though the fluctuation of $r$ and $\phi$ do not accurately predict the deviation, they still provide information about the deviation from coherence.

In Fig. 7, we plot the peak to peak deviation of $|\Delta w/w|$ and $\Delta\theta$ as function of initial $x$. We see that $\Delta\phi$ does not give accurate $|\Delta w/w|$ and $\Delta\theta$, so that we need to multiply $\Delta\phi$ by a factor 6 and 2 to obtain agreement with $|\Delta w/w|$ and $\Delta\theta$ found from tracking respectively. This is because $\Delta\phi \neq 0$ itself implies the theory lost its precision. But it is seen from this plot that the deviation from coherence is predicted by $\Delta\phi$ correctly, and it does serve as an index for the proximity to the destruction of invariant tori.

Next, we check the cases of 4 variables $x, p_x, y, p_y$. We study a lattice named nsls2sr_supercell_ch77. In Fig.8, the top row is the Poincare surface of section $[1, 2]$ expressed by the Courant-Snyder variable $\bar{z}_x = x - i p_x$ and $\bar{z}_y = y - i p_y$. The horizontal axes are their phase angles $\psi_x, \psi_y$ respectively. The vertical axes are their amplitude $|\bar{z}_x|, |\bar{z}_y|$. The left plot is for the amplitude $|\bar{z}_x|$, the right one is for $|\bar{z}_y|$. For the case of initial $x=10\text{mm}, y=2\text{mm}$, we track the particle for 512 turns. Every point on these plots is obtained from the coordinates for a specific turn. For the same set of data, when we convert $\bar{z}_x, \bar{z}_y$ to $w_x, w_y$ and plot the Poincare sections for the corresponding variable $\theta_x, \theta_y$ (the phase angle of $w_x = r_x e^{i\theta_x}, w_y = r_y e^{i\theta_y}$) as the transverse axes, and $r_x = |w_x|, r_y = |w_y|$ as the vertical axes, we obtain the bottom row of Fig.8. Clearly, these new variables now move on two separate flat planes in the two Poincare sections, representing two independent rotations. Thus the transformation to new variables $w_x, w_y$ reduces the complicated motion expressed by $z_x, z_y$ to two very simple uniform independent rotations.

In Fig.9 we plot the Poincare section for $w_y$ in the case of all points starting from the same initial $r_{x0}, r_{y0}$ as the point calculated for $x=25\text{mm}$ and $y=5\text{mm}$. That is, for all the points $w_0 = r_0 e^{i\theta_0}$, we take $\theta_0$ from $-\pi$ to $-\pi$ but keep all $r_0 = |r(x = 25\text{mm}, y = 5\text{mm})|$ for both $x$ and $y$. The blue points are for initial values. After one turn, phase angles (green dots) move to new position but $r_y = |w_y|$ deviates from its initial plane in blue. The red dots represent the prediction by $\text{Im}\phi$: $r_y = r_0 e^{-i\text{Im}\phi_0}$. They also deviate from the initial plane in blue, but the deviation is much smaller than the tracking results. Hence, $\text{Im}\phi$ again serves as an indication of the beginning of the destruction of the invariant tori, but it is not an accurate prediction because $\text{Im}\phi \neq 0$ itself serves as a measure the error of the theory. Actually, $x=25\text{mm}$ is near the dynamic aperture, and in particular, it is near a resonance. The analysis of the resonance by square matrix is beyond this article.

If we reduce the initial $r_{x0}, r_{y0}$ so that all points have same $r_x, r_y$ as the point calculated for $x=20\text{mm}, y=5\text{mm}$, we find the Poincare section becomes flat again. This is shown in Fig. 10.

To show this more clearly, in Fig.11 we show $\Delta r_y/r_y$ after one turn in a color scaled plot on the Poincare section for the case of initial $x=25\text{mm}, y=6\text{mm}$. This is to be compared with the prediction by theory, i.e., $-\text{Im}\phi$ in Fig.12. Similar to the observation on Fig.6-7, we see that the theory cannot describe the one turn map deviation from twist map accurately, the theory predict a change much smaller than the tracking result, even though it predicts correctly the trend. Again, this is because the deviation from zero by $\text{Im}\phi$ itself already implies the breakdown of the theory about "coherence" condition.
In Fig.8-12 we tested how well the constancy of $r_y$ is. Now in Fig.13 we test the constancy of the phase advance angles $\Delta \theta_x, \Delta \theta_y$. They are nearly constant. However, on magnified scale, the variation is visible. The tracking results show much larger variation than predicted by theory. Again, this is because the variation of $\Delta \theta_x, \Delta \theta_y$ on the plane itself indicates the precision.

C. Amplitude Dependent Tune shift and Tune footprint

In Fig.14 we plot tune as function of initial x and y position (initial $p_x = p_y = 0$), compare tune from tracking result (green) with tune calculated from $\mu + \text{Re}\phi$ (red) using Eq. (5.7) for x and y respectively. There is an excellent agreement up to near the dynamic aperture. To see the more detailed comparison between tracking and our calculation from the square matrix, instead of the 3D plot of Fig.14, we plot in Fig.15 the comparison in several 2D cross sections of the 3D plot in Fig.14 for x tune. We see that at y=6mm and y=-6mm, when the x passes x=-1mm, there is a resonance. We can see the green curve (tracking) has a discontinuity, and the red curve (the square matrix derived tune) also has a jump. Even though the red curves seem to exaggerate the discontinuity, it does show the resonance clearly. This suggests that the square matrix analysis may provide more detailed understanding about resonances. However, we will not discuss about resonances in general in this paper, except we shall touch this subject about the resonance at x=-1mm, y=6mm with a little more details here.

Near this point $\nu_x \approx 2\nu_y$, we have two frequencies dominate the spectrum of x motion: $\nu_x, 2\nu_y$. The single frequency condition is no longer valid. Hence the coherence condition Eq. (5.8) is violated. The standard deviation of $\text{Im}\phi$, calculated from the particles motion in the phase space, serves as an indication of the violation. We see from Eq. (5.6) that the radius $r = |w|$ is also no longer a constant. Its standard deviation can also serve as an indication of the violation of "coherence" condition.

To check this, for initial y=6mm we scan the initial x near x=-1mm, track the particles, convert their coordinates into the new variables $w_x, w_y$, then calculate the RMS value of $\Delta w_x / w_x$. We also calculate the RMS value of $\text{Im}\phi_x$ for the scanned range of x, and compare the two results. These two quantities are not same, but when we multiply the theoretical value of the RMS value of $\text{Im}\phi_x$ by 4.1 we found there is an excellent agreement in Fig.16. This plot is compared with a set of data extracted from a frequency diagram obtained from tracking using elegant [23] by Yongjun Li[24] in Fig.17.
FIG. 14: 3D plot for $\nu_x$ (top) and $\nu_y$ (bottom) as function of initial $x$ and $y$ (initial $p_x = p_y = 0$.) Compare tracking (green) with theory (red)

We see clearly the resonance behavior and the similarity of the two curves near the resonance. The obvious difference away from resonance is due to the inclusion of $\Delta \nu_y^2$ in Fig.17.

In Fig.18, we plot the tune footprint calculated from tracking (green) and from $\mu + \text{Re} \phi$. Clearly this shows we can calculate tune footprint approximately from square matrix without the time consuming tracking particles for various initial $x$ and $y$.

D. Coherence Region and Dynamic Aperture

We are interested in the range of the region where our coherence condition Eq. (5.8) is valid. We can find this range by tracking particles with different initial conditions and find the tune variation such as in the calculation for a frequency map. However, it is possible to find this range without tracking particles for many turns. For this we need to calculate $\phi$ and $\Delta$ for a set of points where $|w_x|, |w_y|$ are constants. $\phi$ and $\Delta$ are calculated from $w_0, w_1, w_2$ by Eq. (5.7), which in turn are calculated from $z_x, z_y$ according to Eq. (4.2). Hence for a given set of points on two circles in the $w_x, w_y$ planes, we need to find their coordinates $z_x, z_y$ using the inverse function of Eq. (7.2). (Notice the relation to the Courant-Snyder variables $\bar{z} = \bar{z}/s \equiv (\bar{x} - i\bar{p})/s$ has a scaling parameter $s$ included, as mentioned in section 7.1.) Thus we use $V^{-1}$ obtained from Eq. (7.3) to find the initial trial solution. These trial values are often sufficiently close to the inverse solution already when the coherence conditions are satisfied, so there is no need to further calculate the inverse function more precisely. However, to double check the validity of this method, we carry out the inverse function

FIG. 15: Cross sections of Fig.14: plots for $\nu_x$ as function of $x$ for various $y$

FIG. 16: Compare RMS of $\Delta w_x/w_x$ from tracking (red) with theory (green) times 4.1 around a resonance

FIG. 17: Compare tracking (green) with theory (red) times 4.1 around a resonance
calculation and found that the error bar of this calculation itself also provides a criterion for the validity of the coherence conditions. Actually, if a particle motion has stable tunes, it satisfies the coherence condition, and in the w-planes, it moves on two separate circles in the $w_x, w_y$ planes, we can find the inverse solution accurately. Oppositely, if the tune of the particle is not stable, we found that its trajectory on the w-plane deviate from a circle. Thus we can use the error bar, i.e., the standard deviation of $\Delta w/w$ for the inverse function solution as another index for deviation from coherence, in addition to $\text{Im}\phi$ and the standard deviation of $\text{Re}\phi$ and $\Delta$.

In Fig.19 we use color scale to represent the RMS value of $\Delta w_x/w_x$ in xy plane. For every point on this plane, we find the corresponding $|w_x|, |w_y|$ assuming initial $p_x = p_y = 0$. Then, for a set of azimuthal angles $\Delta \theta_x, \Delta \theta_y$ of the corresponding $w_x, w_y$ we find the inverse function solution for Eq. (7.2) and use the result $\bar{z}_x, \bar{z}_y$ to calculate the standard deviation for $\Delta w/w, \text{Re}\phi, \text{Im}\phi$ and $\Delta$ for both x and y motion respectively.

In Fig.20 and 21 we plot the standard deviation for $\text{Re}\phi_x, \text{Im}\phi_x$ and $\Delta_x$ respectively.

In Fig.20 and Fig.21, at $x=-1\text{mm}, y=5.5-9\text{mm}$, we see the resonance behavior discussed regarding to Fig.16.

In Fig.22, for $x$ between 20mm and 25mm and for $y$ from 0 to 8mm, we can see the color changes from dark blue to light blue, passing through yellow to red, reaching dark brown. This is the region we discussed for Fig.9 and Fig.10, where we see $|w_y|$ gradually deviates from being a constant during the motion when initial $x$ increases from 20mm to 25mm.

In Fig.23, we plot the frequency diagram for the same lattice setting. When compared with Fig.19-22, we see that these plots give crude picture about the dynamic
aperture. Even though, without multi-turn tracking, they do not give the detailed structured frequency map, the light blue area gives information about the area of larger tune variation. Fig.19 confirms the expectation that the errors in the inverse function solution of Eq. (7.2) are related to the coherence condition or stability condition.

It is interesting to study what property of the function \( w_x, w_y \) determines the "coherence" condition. Since it is related to the existence of inverse function, we are very naturally led to the calculation of the Jacobian determinant \( |dw/dz| \equiv \left| \frac{\partial(w_x, w_y, w_x^*, w_y^*)}{\partial(z_x, z_y, z_x^*, z_y^*)} \right| \). When the Jacobian determinant is zero, there is no solution for the inverse function. When it is very small, the solution has large errors.

In Fig.24 we plot the logarithm of the Jacobian determinant for a Poincare section scanned for initial \( x=23\text{mm}, y=6\text{mm} \). In most area the variation of the Jacobian is very small (see the fine scale of the color scale of this plot). This type of scan provides information about distribution of zeros of the Jacobian.

In order to study the zeros of the Jacobian determinant, we remark that for a numerical study, the number of zeros is not well defined, so we need to set up a systematical way to count the number of points where the Jacobian determinant is smaller than a certain parameter set as a criterion. In the same way as in the discussion of Fig.19-22, for every initial point (a pair of \( x,y \)) we scan the Poincare section by \( 16 \times 16 = 256 \) points on \( \theta_x, \theta_y \) plane, we use histogram to sort out the 80% of these points around the peak of the distribution, take the average of \( \log(|dw/dz|) \) for this 80% of points (for convenience we call this the core of the distribution), then count the number of points where the Jacobian is smaller than the average by a factor 100. In Fig.25, for a fixed
initial $y=6\text{mm}$, we scan $x$, then plot the count of points as function of $x$. When compared with the frequency diagram at $y=6\text{mm}$ cross section, we see that the number of near zero point significantly increases at the positions where the tune variation is large.

In Fig.25 we plot the RMS width of the $80\%$ core of the distribution of $\log(|dw/dz|)$ for the same case as Fig.25. We see that when the "coherence" condition is satisfied $|dw/dz|$ is nearly a constant from $x=-20\text{mm}$ to $20\text{mm}$, just as it should be if $w$ is a precise action-angle variable instead of an approximation. When it is increased, the "coherence" condition breaks down.

8. ACTION-ANGLE APPROXIMATION AS A TOOL TO STUDY ONE TURN MAP

As we discussed in regarding about Fig.9-12, even though $\text{Im}\phi$ and $\text{Re}\Delta\phi+\text{Im}\Delta/2$ as a function around the trajectory provides the trend of a deviation from "coherence", they cannot provide accurate calculation for the map of one turn change of $w$. However, even though $w_x$ and $w_y$ are not a precise action-angle variables, they serve as a set of new variables. This set of new variables, through Eq. (7.2) and the equations of one turn map for $\bar{z}_x, \bar{z}_y$, provides a new map. This new map, when the system is not far away from "coherence" condition, serves as a twist map with a perturbation. This is similar to the issues discussed by Poincare (about Poincare-Birkhoff theorem, see, e.g., p.85 of [8]) when he studied the destruction of invariant tori and the formation of islands when a twist map is perturbed. Fig.28 provides one example. In this case, the coordinates $\bar{z}_0, \bar{z}_y$ are transformed to $r_x, r_y$ and $\theta_x, \theta_y$. After one turn, the new set of $\bar{z}_x, \bar{z}_y$ are again transformed into the new variables $r_x, r_y$ and $\theta_x, \theta_y$. When not far away from "coherence", $\Delta r_x, \Delta r_y$ are small, the map between $r_{x0}, r_{y0}, \theta_{x0}, \theta_{y0}$ and $r_x, r_y, \theta_x, \theta_y$ is easier to visualize than the map between $\bar{z}_{x0}, \bar{z}_{y0}$ and $\bar{z}_x, \bar{z}_y$, because the map only slightly deviates from the initial plane in the Poincare section. We remark that because this map involves a calculation of inverse function and tracking around one turn, it cannot be expressed in a simple form as a polynomial like Eq. (2.1).

In Fig.28, we plot the movement on the Poincare section, labeled by number of turns following the tracking, starting from $x=20\text{mm}, y=6\text{mm}$. We can see that every 23 turns the point returns to a point nearby. Hence it is interesting to record every 23 turns, and plot the result in Fig.29. In this plot after 943 turns, the point almost returns to the starting point (point 0).

In Fig.30 we plot the motion on the Poincare section in the $\theta_x, r_y$ plane for every 46 turns. It does provide a

**FIG. 25:** For $y=6\text{mm}$, number of points where the Jacobian determinant is smaller than $\log(|dw/dz|)$ by a factor 100

**FIG. 26:** RMS width of $\log(|dw/dz|)$ for $y=6\text{mm}$

**FIG. 27:** RMS width of $\log(|dw/dz|)$ in x-y plane
FIG. 28: Positions on the Poincare section surface using coordinates $\theta_x, \theta_y$ after a number of turns

FIG. 29: Positions on the Poincare section surface using coordinates $\theta_x, \theta_y$ after every 23 turns

FIG. 30: Positions on the Poincare section surface using coordinates $\theta_x, r_y$ after every 46 turns

FIG. 31: Islands to be studied in Fig.32

FIG. 32: Islands to be studied in Fig.32

as initial points and start tracking the motion of these particles for 7 turns. Then we connect the initial point and final point for every point on this net by arrows as a flow diagram.

Then we track 3 particles starting from the 3 red points for 490 turns and plot the result in Fig.32 for every 7 turns as the green dots. We can see that the flow diagram obtained from the 7 turn map provides a clear picture of the formation of the islands. We remark here that the action angle variable $w$ does not provide a simple way to determine the shape of the islands but it does provide a way to visualize their formation.
However, even though the inverse function calculation is made easy and fast by the use of the inverse matrix $V^{-1}$ mentioned in Section 7.2, we would like to carry out this optimization without the inverse function calculation at all. Therefore we remark here that the task of minimizing $\Delta J$ is equivalent to optimize the system so that flat planes in the Poincare sections in $w_x, w_y$ space (as shown in Fig.8) are mapped to approximate flat planes in the Poincare sections in the $\bar{z}_x, \bar{z}_y$ space, and vice versa. Because $w_x, w_y$ have been derived as polynomials of $\bar{z}_x, \bar{z}_y$ already, the optimization can be carried out by minimization of $|\Delta w/w|$ in the Poincare sections in $w_x, w_y$ space instead. This is as shown in Fig.8 but with the rolls of $w_x, w_y$ exchanged with $\bar{z}_x, \bar{z}_y$. Thus for a pair of constants $J_x \equiv |\bar{z}_x|$ and $J_y \equiv |\bar{z}_y|$, and a set of $\phi_x, \phi_y$, we calculate $w_x, w_y$, then minimize $|\Delta w/w|$.

We applied this optimization for the lattice "nls2sr-r-sepercch77" which we have discussed in regard of Fig.8. In Figure 33 we compare the trajectories of several particles in phase space $y - y'$ before (left) and after (right) the optimization. Different color represents different initial $x, x', y, y'$. In these 5 pairs of $x$ and $y$, the initial $y$ is chosen to be proportional to the initial $x$. The maximum initial $x$ is 20mm, so the $x$-motion is nonlinearly coupled into $y$-motion, generating complicated motion in $y - y'$ plane. It is obvious that even though the lattice of the left plot has been optimized for NSLSII operation with nonlinear driving terms minimized already, the further optimization by square matrix method clearly further reduces the nonlinearity of the system significantly.

One possible explanation of this is the significant reduction of $|\Delta w/w|$ on the trajectories. This clearly reduces the radial motion in the $w$-planes and in turn reduces the island widths, the importance of which is discussed in literatures about stability of nonlinear dynamic system (see, e.g., [1]).

For this specific example, 3 sets of Poincare sections are selected to minimize $|\Delta w/w|$. The $J_x, J_y$ for these Poincare sections are derived from the following 3 pairs of initial conditions $x_0, y_0 = \{2.5e-2, 5e-3\}; \{1e-2, 2e-3\}; \{3.5e-2, 3e-3\}$, respectively. This choice is not unique. Actually, the question about how many Poincare sections should be used, and how many points in each section are taken, is open for future exploration of very fast optimization method.

10. CONCLUSIONS AND FUTURE WORK

We have developed a method of analysis of high order non-linear mapping matrix for complicated nonlinear dynamic system. The high order large matrix can be reduced to low dimensional Jordan decomposition. The more accurate long chain in the Jordan decomposition can be separated from the shorter and less accurate chains. The stable Jordan decomposition of very large matrix is made possible mainly because the special
FIG. 33: Trajectories in y-y phase space for 5 particles before (top) and after optimization (bottom) by square matrix feature of the structure of the nonlinear mapping square matrix and the scaling procedure mentioned in Section 6. The more detailed Jordan decomposition steps are outlined in the appendix A, B, C, D. The result is a uniquely defined transformation from Courant-Snyder variables to an action-angle approximation $w_x, w_y$.

We demonstrated that the action-angle approximations remain nearly constant up to near the boundary of the dynamic aperture and resonance lines. They successfully reproduce both the correct phase space structure and tune shift with amplitude. In addition, we have several measures of the stability of the trajectories and their tunes such as the criterions $\text{Im}\phi \equiv 0; \Delta \equiv 0$ and the width of the Jacobian determinant of the transformation to $w_x, w_y$. We compared the region where these criterions satisfied with the dynamic aperture for realistic lattices currently used for NSLSII storage ring in Section 7 showing these measures can be used to find the dynamic aperture approximately.

Preliminary work on manipulation of the trajectories in phase space as described in Section 8 indicates the potential of future application of the square matrix method for nonlinear optimization of lattices. The agreement of the RMS of $\Delta w_x/w_x$ from tracking with theory on resonance in Fig. 16 of Section 7.3 indicates that the square matrix method provides detailed information on resonance. It would be very interesting to be able to understand this resonance behavior in future study.

The examples given here are limited to 2 to 4 variables. However, the method developed here is general. Hence the application to 6 variables or more should be possible. The inclusion of bunch length and energy spread into consideration of this method may be interesting for high energy accelerator physics. The analysis given here is general, and hopefully may be applied to other areas, for example, nonlinear dynamics in physics and astronomy.

ACKNOWLEDGMENTS

The author would like to thank Dr. Yongjun Li for his many comments, suggestions and discussion on this paper, in particular, for his contribution to the optimization result used in Section 9. We also would like to thank Dr. Lingyun Yang for his support on the use of his program of TPSA to construct the square matrixes. We also would like to thank Dr. Yue Hao for discussion and comments on the manuscript.

We thank Dr. Ferdinand Willeke for discussion and support of the work. We also thank Dr. Boaz Nash for the collaboration on the start and early development in the direction of square matrix analysis of nonlinear dynamics.

This work was funded by DOE under Contract No. DE- SC0012704.

Appendix A: Eigen-space of Tridiagonal Matrix

As pointed out in Section 2, for tridiagonal matrix, the generalized eigenvectors can be calculated in a simple straightforward way. For a specific eigenvalue, the eigenvectors span an invariant subspace. Here we give a brief description of the method to solve for a set of basis for this subspace using an example.

For an eigenvalue $\lambda$, we need to find the non-trivial solutions of the equation $MZ = \lambda Z$. As an example, we study the eigenspace of the matrix $M$ of Eq. (2.5) for $\lambda = e^{i\mu}$. Let $m \equiv M - \lambda I$ with $I$ the identity matrix. Given the set of eigenvalues of $M \{1, e^{i\mu}, e^{-i\mu}, e^{2i\mu}, 1, e^{-2i\mu}, e^{3i\mu}, e^{i\mu}, e^{-i\mu}, e^{-3i\mu}\}$, we
see that \( m \) is a tridiagonal matrix with its 2nd and 8th diagonal element equal to zero, the other 8 diagonal elements are non-zeros. Hence there are only two eigenvectors which span the invariant subspace for this eigenvalue.

To save space for this paper and avoid writing large matrix, let us assume that the 2nd and the 5th diagonal element are zero instead of 2nd and 8th. Thus to find the first eigenvector, we write

\[
mX_1 = \begin{bmatrix} a & b & * & * & * & * & * & * \\ 0 & 0 & * & * & * & * & * & * \\ 0 & 0 & c & * & * & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & * & * \\ \end{bmatrix} \begin{bmatrix} x \\ 1 \\ 0 \\ 0 \\ 0 \\ \end{bmatrix} = 0 \quad (A.1)
\]

Here we have chosen \( X_1 \) only has first 2 rows nonzero, thus clearly we only need to choose to satisfy the first row of the matrix equation (we use * to represent a certain number which there is no need to specify). That is, we have \( ax + b = 0 \), and hence \( x = \frac{-b}{a} \). Thus we have solved for the first eigenvector \( X_1 \).

Next we find the 2nd eigenvector \( X_2 \). Since the 2nd zero diagonal element is the 5th, we let only the first 5 rows of \( X_2 \) to be nonzero, and let

\[
mX_2 = \begin{bmatrix} a & b & * & * & * & * & * & * \\ 0 & 0 & * & * & * & * & * & * \\ 0 & 0 & c & * & * & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & * & * \\ \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \end{bmatrix} = t_{21} X_1 = t_{21} \quad (A.2)
\]

\( t_{21} \) is a certain number to be determined. Again we have let the 5th row of \( X_2 \) to be 1. Clearly we only need to find the first 4 rows of \( X_2 \) to satisfy the equation. The 4th row is \( cx_4 + * = 0 \), hence \( x_4 = \frac{-*}{c} \). This process is repeated to find \( x_3 \). When we proceed to the 2nd row, the diagonal element becomes zero, hence the situation is different. And we find the equation \( *x_3 + *x_4 + * = t_{21} \), where \( x_2 \) is absent and can be set to zero. Since \( x_3 \) and \( x_4 \) are already determined, this equation now is used to determine \( t_{21} \). This process also explains why in Eq. (A.2) we cannot set \( t_{21} = 0 \) and hence \( X_2 \) is not a proper eigenvector, but a generalized eigenvector: when it satisfies Eq. (A.2), we have \( mX_2 \neq 0 \), but \( m^2 X_2 = t_{21} mX_1 \).

Once \( t_{21} \) is determined, since \( x_3 \) through \( x_4 \) are already determined, we can proceed to the first row to solve for \( x_1 \): \( ax_1 + *x_2 + *x_3 + *x_4 + * = t_{21} * \) because \( a \) is nonzero.

All the rows in \( X_1, X_2 \) are rational functions of elements in the matrix \( M \), this is just as stated in Section 2, in a more general way.

We can generalize this procedure to the case with more than two zero diagonal elements in \( m \). Without further details, we summarize the result as follows. As long as the system is sufficiently far away from resonance, for example, the minimum value of \( \lambda - e^{in\mu} \) for all \( n \) within the specified order is several orders of magnitude larger than machine zero, there are well-defined zeros in the matrix. Let the number of zeros to be \( n_\lambda \), we can find a set of \( n_\lambda \) (generalized) eigenvectors \( X_j \) such that \( mX_j = \Sigma_{0<k<j} f_{jk} X_k \) with \( 1 \leq j \leq n_\lambda \). Because \( mX_j \) remains to be a linear combination of \( X_k \), these \( X_k \)s serve as basis for the invariant subspace of eigenvalue \( \lambda \) of dimension \( n_\lambda \). Using the Einstein convention (the repeated \( k \) implies a sum over \( k \)), we have \( mX_j = t_{jk} X_k \), where \( t \) is a lower triangular matrix with all diagonal elements equal to zero.

The example here is for the right eigenvectors. To calculate left eigenvectors, we can simply transpose the matrix \( M \) and find its right eigenvectors as columns as discussed here, and transpose them back to rows.

The result is the set of left eigenvectors \( e_i \) such that \( e_i (\lambda - M) = t_{ik} e_k \), as given in Eq. (3.4).

**Appendix B: Outline of a Method of Jordan Decomposition**

As pointed out in Section 3, the structure of chains in the invariant subspace of one eigenvalue, with each chain corresponds to one Jordan block, serves as the basis of the method of Jordan decomposition we outline here. Our main issue has been reduced to the Jordan decomposition of the subspace of matrix \( m \) in Appendix A, which is represented by the much lower dimension matrix \( t \) with eigenvalue zero.

As pointed out in Section 3, this subspace is spanned by vectors of several chains. Each chain has a proper eigenvector at its end. These proper eigenvectors form the null space of \( m \). Hence the dimension of the null space of \( m \) is equal to the number of chains. In the example of Section 3, because there are 4 chains with lengths 4, 3, 2, 1 respectively, when multiplied by \( m \), the 4 proper eigenvectors become zero, hence the null space \( N_1 \) for \( m \) has dimension 4. The chain of length 1 is removed by \( m \), so after multiplied by \( m \) only 3 chains left, and the null space \( N_2 \) for \( m^2 \) has dimension 4+3=7. Continue this way we find the dimension of the null space \( N_k \) for \( m^k \) is \( m_k = 4, 7, 9, 10 \) for \( k = 1, 2, 3, 4 \) respectively. Thus if for every \( k \) we can find the basis of null space \( N_k \) for \( m^k \), we can identify all the eigenvectors as the follows. Since \( m_{4} = 10, m_{3} = 9 \) means there is one vector in \( N_4 \) independent from the basis of \( N_3 \), if we can find this vector \( u \) which satisfies \( m^4 u = 0 \) but \( m^5 u \neq 0 \), we identify this as the first generalized eigenvector in the longest chain because this is the last remaining vector to become zero when we apply the matrix \( m \) to all the basis in the subspace in succession. Thus \( u, mu, m^2 u, m^3 u \) are the basis of the longest chain of length 4. Once we separate these 4 eigenvectors from the subspace of dimension 10, and find the remaining 6 independent 6 vectors, we can
we have

\[
\begin{pmatrix}
0 & -1 & 10 & 29 & -91 & -96 \\
0 & 0 & 0 & 6 & -10 & -45 \\
0 & 0 & 0 & 1 & -3 & 1 \\
0 & 0 & 0 & 0 & 1 & -2 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
\end{pmatrix}
\]  
\begin{equation}
(A.1)
\end{equation}

The singular values are arranged in increasing order, then the first 2 (see Table II: \(m_1 = n_1 = 2\)) singular values are zeros. Then following [15–17], we define \(A^{(2)} = V_1^H U_1 \Sigma_1 V_1^H V_1 = V_1^H U_1 \Sigma_1 V_1^H A^{(1)} V_1\), which has all zero as its first \(m_1 = 2\) columns, corresponding to the null space \(N_1\). To find the null space \(N_2\), we repeat this procedure for the \((n - m_1) \times (n - m_1) = (6 - 2) \times (6 - 2) = 4 \times 4\) submatrix \(A_2\), which is the lower right corner of \(A^{(2)}\), as shown in Fig.34.

\[
\begin{pmatrix}
0 & B_{12} & B_{13} & B_{14} \\
0 & 0 & B_{23} & B_{24} \\
0 & 0 & 0 & B_{34} \\
0 & 0 & 0 & 0 \\
\end{pmatrix}
\]

\begin{equation}
(B.2)
\end{equation}

Here all the zeros represent blocks of zeros with the left column represent \(m_1\) columns of zeros. \(B_{ij}\) is a \(n_i \times n_j\) submatrix. According to the table II, the widths of the blocks are 2,2,1,1. Indeed our calculation result agrees with this form.

\[
\begin{pmatrix}
0 & 0 & -9.99026 & -74.19830 & 15.67463 & 72.29953 \\
0 & 0 & 0.01922 & -23.93292 & 1.71456 & 38.34366 \\
0 & 0 & 0 & -0.40486 & 6.08191 \\
0 & 0 & 0 & -0.00032 & -0.40472 \\
0 & 0 & 0 & 0 & 2.19822 \\
0 & 0 & 0 & 0 & 0 \\
\end{pmatrix}
\]

\begin{equation}
(B.3)
\end{equation}
null space dimension of $A^k$: $m_k = n$-rank = 6-rank
number of chains left after multiplied by $A^{k-1}$: $n_k = m_k - m_{k-1}$
number of chain terminated by $A^k$: $n_k - n_{k+1}$

TABLE II: Use the ranks of the powers of a matrix $A$ to calculate null space size and chain structure

3. Make the Super-Diagonal Blocks of $B$ Upper-Triangular

The next step is to carry out unitary transform such that each of $B_{k,k+1}$ blocks is transformed to upper-triangular form. Let

$$U = \begin{bmatrix} U_1 & 0 & 0 & 0 \\ 0 & U_2 & 0 & 0 \\ 0 & 0 & U_3 & 0 \\ 0 & 0 & 0 & U_4 \end{bmatrix}, U^H = \begin{bmatrix} U_1^H & 0 & 0 & 0 \\ 0 & U_2^H & 0 & 0 \\ 0 & 0 & U_3^H & 0 \\ 0 & 0 & 0 & U_4^H \end{bmatrix}$$

be unitary matrixes with only diagonal blocks nonzero, and the dimension of blocks are $n_k$ as in Table II. Let $T = UBU^H$. Then we find

$$T = UBU^H = \begin{bmatrix} 0 & U_1B_{12}U_2^H & U_1B_{13}U_3^H & U_1B_{14}U_4^H \\ 0 & 0 & U_2B_{23}U_3^H & U_2B_{24}U_4^H \\ 0 & 0 & 0 & U_3B_{34}U_4^H \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We can choose $U_1^H$ to be identity matrix first and carry out QR decomposition $B_{34} = QR$ such that $R$ is upper-triangular and $Q$ is a unitary matrix. Then we choose $U_3 = Q^H$, hence $T_{34} = U_3B_{34}U_4^H = R$ is upper-triangular. We then proceed to $T_{23} = U_2B_{23}U_3^H$. Now since $U_3$ is known, we can carry out another QR decomposition $B_{23}U_3^H = QR$ and choose $U_2 = Q^H$ so that $T_{23} = R$. Here to avoid cluttering of notation we have repeated the use of the notation $Q$ and $R$ for different matrices by redefining them each time we use QR decomposition. This procedure continues until $U_1$ is determined to make $T_{12}$ upper-triangular. Then $T$ has the form in Fig.35 (left), with only upper-triangular blocks nonzero and also with all submatrix $T_{k,k+1}$ upper-triangular.

For the example, the result agrees with this form:

$$T = \begin{bmatrix} 0 & 0 & 0 & 0 & -10.049 & 74.190 & -15.674 & -72.299 \\ 0 & 0 & 0 & 0 & -23.932 & 1.714 & 38.343 \\ 0 & 0 & 0 & 0 & 0.404 & -6.081 \\ 0 & 0 & 0 & 0 & 0 & -0.409 \\ 0 & 0 & 0 & 0 & 0 & 2.198 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Those elements $r_{ij}$ in the $T$ matrix are called coupling elements. If all elements other than the coupling elements are made zero, then the matrix becomes greatly simplified and the chain structure of the subspace will be revealed, as will become clear in the next subsection.

4. Gauss Elimination by Similar Transformation

We now proceed to eliminate those elements in $T$ represented by crosses in Fig.35 by Gauss elimination. The procedure is to eliminate all elements other than coupling elements column by column, start from right to left. Consider the similar transform of $T$

$$V^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -a \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, V^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$
we find

\[
VTV^{-1} = \begin{bmatrix}
0 & T_{12} & T_{13} & T_{14} & T_{15} + aT_{13} & T_{16} \\
0 & 0 & T_{23} & T_{24} & T_{25} + aT_{23} & T_{26} \\
0 & 0 & 0 & T_{34} & T_{35} & T_{36} - aT_{56} \\
0 & 0 & 0 & 0 & T_{45} & T_{46} \\
0 & 0 & 0 & 0 & 0 & T_{56} \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]  

(B.8)

That is, if we choose \( V_{ij} = -a \), then \( V_{ij}^{-1} = a \), and the transform simply add the column \( i \) multiplied by \( a \) to column \( j \), and subtract the row \( j \) multiplied by \( a \) from row \( i \). Thus if we start from \( i=4, j=5 \), and \( a = T_{46}/T_{56} \), because the \( T \) matrix is upper triangular, \( T_{46} \) is eliminated. Column 5 is affected during this transformation, but this does not prevent our elimination process. In particular, the transformed matrix still remains to have the form of the \( T \) matrix. When we repeat this procedure with \( j=5 \) but let \( i=3,2,1 \), the column 6 is eliminated except the coupling element \( T_{56} \).

Next we proceed to column 5, but start from \( i=3, j=4 \). This eliminates column 5 except the coupling element \( T_{45} \). Then, from right to left, the procedure continues to column 4, 3, 2 in the same way. The result is a matrix \( J \) in the form shown in Fig.35(right), with only the coupling elements \( r_{ij} \) nonzero.

For the example discussed from section B1 to B4, we find

\[
J = UAU^{-1} = \begin{bmatrix}
0 & 0 & -10.049 & 0 & 0 & 0 \\
0 & 0 & 0 & -23.932 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.404 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 2.198 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

(B.9)

Here we have multiplied all the transformation matrixes used from section B2 through B4 together into one transformation matrix \( U \). This matrix is not unitary because in the last steps the matrix \( V \) for every step is not unitary, even though in section B2 to B3 all the transformation matrixes are unitary.

5. Permutation and Jordan Form

The matrix \( J \) derived in our example in section B4 indeed only has the 4 coupling elements nonzero: \( J_{56}, J_{35}, J_{24}, J_{13} \). It is already very close to the Jordan form. To see the chain structure we let \( e_j \) represents the column of 6 elements with only element \( j \) equal to 1, and all other elements zero. Then we find

\[
\begin{align*}
J_{e_6} &= J_{56} e_5, J_{e_5} = J_{35} e_3, J_{e_3} = J_{13} e_1, J_{e_1} = 0; \\
J_{e_4} &= J_{24} e_2, J_{e_2} = 0
\end{align*}
\]

(B.10)

This confirms the table in B1 that there are two chains of lengths 4 and 2. To transform to standard Jordan form with all the coupling elements equal to 1, we redefine the norm of the basis \( e_j \), and reorder them according to the chain structure, which amounts to a transformation of renormalization followed by a permutation, and we neglect the details here. The result is the Jordan form (we redefine \( U \) here to include the renormalization and permutation, and choose to have the longest chain first).

\[
J = UAU^{-1} = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

(B.11)

Appendix C: Jordan Decomposition of \( \ln(M) \)

To find the Jordan decomposition of \( \ln(M) \) we first find the transformation for \( M \) itself: \( UMU^{-1} = N \). Use the property of the similarity transformation, we have \( \ln(UMU^{-1}) = \ln(\ln(M)) = \ln(N) \). Since \( N \) is a block diagonal matrix, we have

\[
\begin{align*}
U\ln(MU^{-1}) &= U\lnM \\
\lnN &= \begin{bmatrix}
\lnN_1 & 0 & \cdots & 0 \\
0 & \lnN_2 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & \lnN_k
\end{bmatrix}
\end{align*}
\]

(C.1)

For one block with eigenvalue \( \lambda_j = e^{i\mu} \) we have

\[
U_j \ln(MU_j) = \ln(e^{i\mu} I_j + \tau_j) = \ln(e^{i\mu} (I_j + e^{-i\mu} \tau_j)).
\]

Because we are only interested in this invariant subspace, we drop the index \( j \) when it is not needed, to write

\[
\begin{align*}
U\ln(MU) &= i\mu I + \ln(I + e^{-i\mu} \tau) = i\mu I + e^{-i\mu} \tau \\
&\quad - \frac{1}{2} e^{-2i\mu} \tau^2 + \cdots + \frac{(-1)^{n_j-1}}{n_j!} e^{-(n_j-1)i\mu \tau} n_j-1
\end{align*}
\]

(C.2)

Here now \( U \) represents \( U_j \), i.e., we redefine its submatrix using the same notation to simplify writing. If the dimension of the subspace \( j \) is \( n_j \), the series terminates at \( n_j - 1 \) because \( \tau^{n_j} = 0 \). The right hand side is not in Jordan form, but it is very simple, and can be transformed by another matrix \( V \) (we are excused to redefine the notation \( V \) here, not to be confused with the \( V \) matrix used in Section B4, or Section 7) into Jordan form \( \bar{V}U\ln(M)\bar{U}^{-1} = i\mu I + \tau \). Now we redefine \( VU \) as \( U \), \( \bar{U} \) as \( \bar{U} \), get

\[
\begin{align*}
U\ln(MU) &= i\mu I + \tau
\end{align*}
\]

(C.3)

Notice that since we concentrate only on the subspace of the longest chain in the eigenspace \( e^{i\mu} \), compared with
the discussion in Appendix B, the Jordan decomposition process by $V$ here involves only one chain only.

Appendix D: Minimize Higher Order Terms

As explained in Section 6, we would like to minimize the higher power terms in the first vector of the longest chain by adding linear combination of sub-chains to the longest chain. This is a change of basis in the invariant subspace of a specific eigenvalue, i.e., a linear coordinate transformation. We use an example to illustrate the transformation.

Let us assume the matrix $N$ in Eq. (3.1) has two chains of length 4 and 3. We study a similar transformation $t$ which changes the basis of the eigenspace but keeps the Jordan matrix $N$ invariant. Thus we write

$$N = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}; \quad t = \begin{bmatrix} I_1 + T_1 & T_2 \\ 0 & I_2 \end{bmatrix}; \quad t^{-1} N t = N \quad \text{(D.1)}$$

where $A$ and $B$ are of form Eq. (3.2) with only superdiagonal elements equal to 1 and all other elements equal to zero. $A$ and $I_1$ have dimension 4, $B$ and $I_2$ have dimension 3 respectively. The matrix $t$ is in a general form to represent a change of basis so that the longest chain after the transformation becomes a linear combination of the long chain from the $4 \times 4$ matrix $T_1$ and the short chain from the $4 \times 3$ matrix $T_2$. In Section 6, we mentioned that the Jordan form is not unique. But from Eq. (D.1) we can obtain all the generally possible Jordan basis for the longest chain. Our goal is to choose $T_1$ and $T_2$ to minimize the higher power terms in the first vector in the basis for matrix $N$, as we explained in Section 6.

Eq. (D.1) leads to $N t = t N$, thus we have

$$\begin{bmatrix} A + AT_1 & AT_2 \\ 0 & B \end{bmatrix} = \begin{bmatrix} A + T_1 A & T_2 B \\ 0 & B \end{bmatrix}, \quad \text{(D.2)}$$

i.e., $AT_1 = T_1 A$ and $T_2 B = AT_2$. To determine the form of the matrix $T$, we examine the effect of Jordan form on column and row. Let

$$u = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}; \quad v = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}; \quad A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}; \quad \text{(D.3)}$$

we find

$$Au = \begin{bmatrix} e_2 \\ e_3 \\ e_4 \\ 0 \end{bmatrix}; \quad v A = \begin{bmatrix} 0 & e_1 & e_2 & e_3 \end{bmatrix}. \quad \text{(D.4)}$$

Hence when acted from left by the Jordan matrix, the column shifts up, leaving the last row zero; when acted from right, a row shifts to the right, leaving the left column zero. Thus $AT_1 = T_1 A$ means the matrix $T_1$ after shifted up should be the same as it is shifted to the right. Examine this pattern, we see that $T_1$ must be upper-tridiagonal with all elements on the same super-diagonal same. Also, $T_2$ is of this form. So let

$$t = \begin{bmatrix} 1 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ 0 & 1 & x_2 & 0 & x_4 & x_5 \\ 0 & 0 & 1 & x_1 & 0 & x_4 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \quad u = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \end{bmatrix}. \quad \text{(D.5)}$$

We find $u' = tu = \begin{bmatrix} e_1 + e_2 x_1 + e_3 x_2 + e_4 x_3 + e_5 x_4 + e_6 x_5 + e_7 x_6 \\ e_2 + e_3 x_1 + e_4 x_2 + e_5 x_3 + e_6 x_4 + e_7 x_5 \\ e_3 + e_4 x_1 + e_5 x_2 + e_6 x_3 + e_7 x_4 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \end{bmatrix}$. \quad \text{(D.6)}

$u'$ is the basis of the matrix $N$ after the transform. Our goal is to minimize the high power terms in the first vector of $u'$, i.e., $e_1' = e_1 + e_2 x_1 + e_3 x_2 + e_4 x_3 + e_5 x_4 + e_6 x_5 + e_7 x_6$ in the basis. If the lowest power term in $e_2$ is 3rd power, as explained in Section 6, we need to minimize the 3rd power terms of $e_1'$ by varying $x_1$. We first calculate $f = ((e_1')^* e_1')_3$. The subscript 3 means we take only the 3rd power terms in the scalar product, hence $f$ here is considered to be the norm of the vector $e_1'$ at 3rd power.

The minimization of $f$ requires $\partial f / \partial x_1 = (e_1')^* e_2 = 0$. Taking conjugate, we have

$$(e_1^*(e_1 + e_2 x_1 + e_3 x_2 + e_4 x_3 + e_5 x_4 + e_6 x_5 + e_7 x_6))_3 = 0. \quad \text{(D.7)}$$

This can be written as

$$(e_2^* e_3^*) x_1 + (e_2^* e_3^*) x_2 + (e_2^* e_4^*) x_3 + (e_2^* e_5^*) x_4 + (e_2^* e_6^*) x_5 + (e_2^* e_7^*) x_6 = -(e_2^* e_1^*) \quad \text{(D.8)}$$

If the lowest power term in $e_3$ is 5th power, using similar method we minimize the 5th power terms in $e_1'$ by $x_2$, and find

$$(e_3^* e_2^*) x_1 + (e_3^* e_2^*) x_2 + (e_3^* e_4^*) x_3 + (e_3^* e_5^*) x_4 + (e_3^* e_6^*) x_5 + (e_3^* e_7^*) x_6 = -(e_3^* e_1^*) \quad \text{(D.9)}$$

Assuming the lowest power terms in the basis are power of 1,3,5,7,3,5,7 respectively, we can continue the above
procedure and find the following matrix equation:

\[
\begin{bmatrix}
(e_5^2 e_2)^3 (e_5^2 e_3)^3 (e_5^2 e_4)^3 ... (e_5^2 e_7)^3 \\
(e_5^3 e_5)^5 (e_5^3 e_3)^5 (e_5^3 e_4)^5 ... (e_5^3 e_7)^5 \\
(e_5^4 e_7)^7 (e_5^4 e_4)^7 ... (e_5^4 e_7)^7 \\
... ... ... ... ... ... ...
(e_5^n e_2)^5 (e_5^n e_3)^5 (e_5^n e_4)^5 ... (e_5^n e_7)^5 \\
\end{bmatrix}
\begin{bmatrix}
(x_1) \\
(x_2) \\
(x_3) \\
... \\
(x_n) \\
\end{bmatrix}
\]

By Jordan decomposition the vectors \( e_j \) in u (u is the column of basis as given by Eq.D.5) are known, the Eq. (D.10) can be solved to determine \( x_j \) to establish the new basis \( e_j' \). The basis of Jordan form is uniquely determined now.

**Appendix E: Inverse of an Upper-Triangular Matrix**

Suppose we want to find the inverse of an upper-triangular matrix \( V \) of dimension \( n \). In order to have inverse, all the diagonal elements of \( V \) are nonzero. Let us find a matrix \( L \) consists of column \( y_k \)

\[
L = (y_1, y_2, y_3, ..., y_k, ..., y_n)
\]

And we first find \( y_k \) such that

\[
V y_k = \begin{bmatrix}
* & * & * & * & * & * & * & * & * & * \\
0 & * & * & * & * & * & * & * & * & * \\
0 & 0 & 0 & a & * & * & * & * & * & * \\
0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * \\
\end{bmatrix}
= e_k \equiv 0
\]

Where \( e_k \) has all rows zero except the k'th row equal to 1. Solving the equation starting from the last row, we find \( x_n = x_{n-1} = ... = 0 \) until we reach the k'th row, where we have \( a x_k = 1 \). Hence, \( x_k = 1/a \). Then we can find \( x_{k-1} \) and continue up to \( x_1 \). Thus we find \( y_k \) has only nonzero rows above and include k'th row. Once we find all the \( y_k \), clearly we have the inverse matrix

\[
V^{-1} = L = (y_1, y_2, y_3, ..., y_k, ..., y_n)
\]