Analysis and Optimization of Singular Value Decomposition Orbit Correction Schemes

Arman Ballado
Department of Physics, University of Florida, Gainesville, Florida, 32611
(Dated: August 14, 2010)

The proposed Cornell Energy Recovery Linac (ERL) requires precise control of particle trajectories especially in the undulators where highly brilliant and coherent hard X-rays will be produced. A configuration of beam position monitors and correctors ensures sufficient error correction, and this report focuses on evaluating the performance of this given orbit correction scheme. Previously developed analytic methods using linear algebraic tools such as the singular value decomposition were applied to the ERL lattice to determine its effectiveness and detect any defects. These methods detects deficiencies and redundancies in the monitor and corrector configurations, and the optimization of the configuration then follows directly from the results of the analyses.

I. INTRODUCTION

The Cornell Energy Recovery Linac will extend the pre-existing electron storage ring with two linacs. The electrons will pass through the accelerator twice. In the first run, the particles are accelerated to full energy and pass through the undulators to produce the hard X-rays. In the second pass, they are decelerated, so electromagnetic fields in the cavities can recover and recycle the energy. This drastically reduces the machine's power consumption. Furthermore, the maximum quality of the X-ray beams depends entirely on the injector's ability to produce very narrow and short electron pulses, and as a result, the beam can degrade as it traverses the accelerator due to field errors and misalignments and from synchrotron radiation.

These errors must be corrected to ensure that particles are not lost, so the correction of these errors remain one of the important studies done in designing an accelerator. The basic correction scheme requires beam position monitors (BPM) and correctors. The BPMs feed information about the beam to the correctors that provide a kick to fix the beam's trajectory. The proper placement of BPMs and correctors then becomes a main concern to ensure that beam diagnostics have sufficient coverage and control over the particle trajectory throughout the entire accelerator. This report focuses on applying previously developed methods [1–3] to evaluate the performance of the ERL lattice and then find a more optimal configuration.

These analytic methods frame the problems in linear terms, and this choice allows for more robust calculations using numerical tools, specifically *Mathematica*, without sacrificing mathematical rigour. In fact, with the number of error sources, monitors, and correctors encountered in the study, an approach through linear solvers distinguishes itself as a reliable and efficient method. To verity the results from the analysis, the configurations are simulated through the *Tool for Accelerator Optics (Tao)* , a general purpose program for simulating beams in particle accelerators and storage rings [4] which uses the software library *Bmad* [5]. This program can easily simulate the particle orbit after changes to the machine such as the removal or addition of monitors and correctors.

II. ERROR SOURCES

Before any analysis can begin, potential error sources must first be identified. For the purposes of this report, three main errors are considered: quadrupole misalignments, RF cavity pitches, and dipole field errors.

Quadrupoles magnets provide beam focusing in one axis and defocusing in another. In either case, the quadrupole field is directly proportional to the orbit displacement from the quadrupole center which should ideally be at the reference trajectory. Therefore, a misaligned quadrupole introduces unwanted dispersion from

FIG. 1: Quadrupole Misalignment

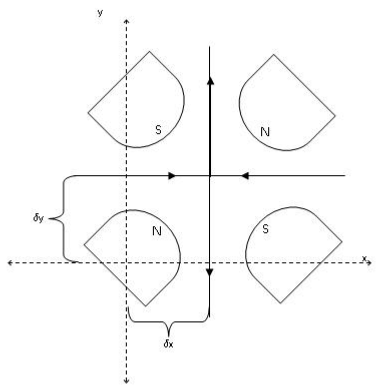
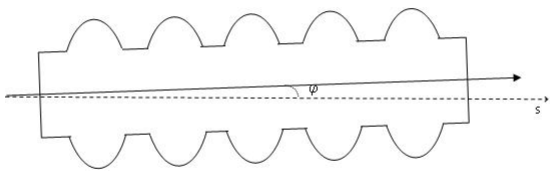


FIG. 2: RF cavity pitch



an effective dipole field. For a quadrupole magnet with a center shifted from the origin to $(\delta x, \delta y)$, it effectively steers a particle beam by

$$\Delta\theta_x = k_l \delta x \tag{1}$$

$$\Delta\theta_y = -k_l \delta y \tag{2}$$

where k_l is the integrated quadrupole strength [6].

An RF cavity accelerates a particle beam longitudinally, but if it has a pitch angle, it introduces a transverse acceleration leading to deflections and emittance growth. A cavity with a pitch angle φ generates an effective steering

$$\Delta\theta \approx \frac{\Delta P}{P}\varphi, \tag{3}$$

where P is the final on-axis momentum and ΔP is the momentum gain [6].
The dipole fields bend the beam around the designed trajectory, but errors in this field can easily result in a different bending radius, deflecting the beam away from the reference orbit.

III. GENERALIZED RESPONSE MATRICES

After the identification of these error sources, the right coefficients map them to the appropriate spaces such as the orbit monitor space, and the mapping is done through generalized response matrices. Response matrices relate a specific action imparted by an actuator A with its result to a responder R defined by the equation $R = M^{RA}A$, where M^{RA} is the response coeffecient [3]. This relation may be further expanded to include m responders and n actuators represented by column vectors.
There are several generalized response matrices crucial for the analyses. The error-to-monitor response matrix $\mathbf{M^{EM}}$ relates all potential error sources to the orbit disturbance at the monitors. A similar matrix, $\mathbf{M^{EA}}$, summarizes the orbit errors at all representative elements. These matrices must include all major potential error sources that the system is designed to correct. Another pair of response matrices are the corrector-to-monitor $\mathbf{M^{CM}}$ and corrector-to-all-location $\mathbf{M^{CA}}$ matrices. These describe the orbit disturbances caused by the angle kicks from the correctors. For the purpose of corrector optimization, it helps to construct two other response matrices. The first is the response matrix $\mathbf{M^{AM}}$ which represents the orbit error at the monitors due to angle kicks at all representative elements (i.e. magnetic elements, mainly the quadrupoles). It provides a list of potential new correctors and its contribution to orbits at the monitors. Finally, the matrix $\mathbf{M^{AA}}$ takes the same actuators but propagates its effects to all locations. These response matrices with their respective actuators and responders are summarized in Table I.
To generate these response matrices, *Tao* has a command that immediately calculates the derivative $dModel\_Value/dVariable$ [4] which determines how a change in a variable affect a data. Each derivative becomes the coefficient M^{ij} between the i-th responder and the j-th actuator. One only needs to use the appropriate data (responders) and variable (actuators) in TAO and construct the matrix from the derivative. For example, M^{CM} is formed from the

TABLE I: Summary of Generalized Response Matrices

Matrix	Actuator	Responder
$\mathbf{M^{CM}}$	Corrector Kicks	Orbit at monitors
$\mathbf{M^{CA}}$	Corrector Kicks	Orbit at all elements
$\mathbf{M^{EM}}$	Potential Error Sources	Orbit at monitors
$\mathbf{M^{EA}}$	Potential Error Sources	Orbit at all elements
$\mathbf{M^{AM}}$	Quadrupole Kicks	Orbit at monitors
$\mathbf{M^{AA}}$	Quadrupole Kicks	Orbit at all elements

calculation of the derivative matrix between the variable kicks at the correctors and the orbit at the beam position monitors.

IV. SINGULAR VALUE DECOMPOSITION ORBIT CORRECTION

A system of linear equation $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$ can be solved exactly using the inverse $\mathbf{A^{-1}}$ of the coefficient matrix, so that $\mathbf{x} = \mathbf{A^{-1}} \cdot \mathbf{b}$ However this presents an obvious problem for singular matrices whose inverses do not exist. For example, the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \tag{4}$$

does not have an inverse. In such cases, an exact solution to the system also does not exist, but it may suffice to have a linear least square solution. These solutions are calculated in different ways, one being the singular value decomposition. Singular value decomposition, or SVD, decomposes an $M \times N$ matrix $\mathbf{A}$ into a product of three matrices $\mathbf{U}$, $\mathbf{W}$, and $\mathbf{V}$:

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{W} \cdot \mathbf{V^T}, \tag{5}$$

where $\mathbf{U}$ and $\mathbf{V}$ are $M \times M$ and $N \times N$ column-orthonormal matrices, respectively. The $M \times N$ matrix $\mathbf{W}$ is a diagonal matrix with positive or zero elements called the singular values. The previous example is decomposed as

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{6}$$

One of the SVD's usefulness comes from its construction of the orthonormal bases for the nullspace and range of a matrix. The columns of $\mathbf{V}$ that correspond to very small or zero singular values in $\mathbf{W}$ form the orthonormal set of basis vectors for the nullspace while those of $\mathbf{U}$ that have non-zero singular values are the orthonormal basis for the range [7]. A closer inspection of the decomposition of the example above easily reveals this property. The optimization methods discussed later will exploit this property of the SVD.

The pseudoinverse of $\mathbf{A}$ is the defined as :

$$\mathbf{A}^{\dagger} = \mathbf{V} \cdot [\mathbf{diag(1/w_j)}] \cdot \mathbf{U^T} \tag{7}$$

replacing $1/w_j$ by zero if $w_j = 0$. For a set of simultaneous linear equations $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$, the solution $\mathbf{x} = \mathbf{A}^{\dagger} \cdot \mathbf{b}$ does not exactly solve the original set of equation but will do so in the least-squares sense. It finds the $\mathbf{x}$ which minimizes the residual $r \equiv |\mathbf{A} \cdot \mathbf{x} - \mathbf{b}|$.

In terms of orbit correction, the linear equation of the orbit is

$$\mathbf{R} = \mathbf{M^{EM}} \cdot \mathbf{A^E} + \mathbf{M^{CM}} \cdot \mathbf{A^C}, \tag{8}$$

and the goal is to find the corrector kicks $\mathbf{A^C}$ that minimizes the residual orbit after correction. From SVD, the best corrector kicks $\mathbf{A^C_{best}}$ that minimizes the residue $|\mathbf{R}|$ is $\mathbf{A^C_{best}} = \mathbf{M^{\dagger}_{CM}} \cdot \mathbf{M^{EM}} \cdot \mathbf{A^E}$.

Our *Tao* simulation also uses SVD to correct the orbit. It internally generates the response matrices and applies the SVD on the BPM readings. The corrector kicks are modified and the step is repeated until enough particles pass through the entire lattice without being lost to the chamber walls.

V. PERFORMANCE ANALYSIS

A. Secondary Response Matrices

The following secondary response matrices are derived by linear algebraic operations on the general response matrices. They help further quantify the performances of the configuration allowing a more intuitive comparison of the performance before and after optimization. These matrices are originally from [3].

- Unobservable Error

$$\mathbf{M^{unobs}} = \mathbf{M^{EA}} \cdot (\mathbf{I} - \mathbf{M^{\dagger}_{EM}} \cdot \mathbf{M^{EM}}) \tag{9}$$

  This matrix maps from the error space to its null space then to all monitor space (i.e. For a given error configuration, it finds the part of this configuration that is undetectable at the monitors then propagates it at all the elements).

- Corrector Range

$$\mathbf{M^{resp}_{CME}} = \mathbf{M^{\dagger}_{CM}} \cdot \mathbf{M^{EM}} \tag{10}$$

  Given an error configuration, the matrix maps it to the corrector space after an SVD type correction, providing the necessary kicks to minimize the orbits at the correctors. It was already encountered in the section about SVD in the equation $\mathbf{A^{C}_{best}} = \mathbf{M^{\dagger}_{CM}} \cdot \mathbf{M^{EM}} \cdot \mathbf{A^{E}}$.

- Residual Orbit

$$\mathbf{M^{resid}_{CMA}} = \mathbf{M^{EA}} - \mathbf{M^{CA}} \cdot \mathbf{M^{resp}_{CME}} \tag{11}$$

  With a given corrector configuration, this response matrix maps from the errors to the orbit after steering.

B. Mapping Errors

All of these response matrices map the error space to the relevant responder space, thus providing, for instance, the corrected orbit at all the elements given a specific error combination. This mapping is a straightforward multiplication but it also presents an obvious limitation because errors are hidden from direct measurement in beam diagnostics. Instead, errors are estimated through probability distributions, so these error distributions must be mapped to the appropriate distributions in the responder space. Recalling that for random variables R_i and A_j related by a constant coefficient

$$R_i = \sum_{j=1}^{n} M_{ij} A_j, \tag{12}$$

their variances are

$$Var(R_i) = \sum_{j=1}^{n} M_{ij}^2 Var(A_j) + 2\sum_{j<k} M_{ij} M_{ik} Cov(A_j, A_k). \tag{13}$$

Assuming that the actuators are uncorrelated and independent, then $Cov(A_j, A_k) = 0$. So the above equation simplifies to

$$Var(R_i) = \sum_{j=1}^{n} M_{ij}^2 Var(A_j). \tag{14}$$

Since it is more interesting to describe the performance of a configuration in terms of standard deviations, then

$$\sigma_{R_i} = \sqrt{\sum_{j=1}^{n} M_{ij}^2 \sigma_{A_j}^2}. \tag{15}$$

C. Scaling Response Matrices

Certain elements in the lattice have more emphasis placed on them during the orbit correction, and this is exactly the case in the ERL where the orbit at the undulators must be minimized as best as possible. *Tao* handles this by attaching weights to the data and the variables. The analytic methods here must also take this into account. This is done by scaling the relevant general response matrices; each row of $\mathbf{M^{EM}}$, $\mathbf{M^{CM}}$ and $\mathbf{M^{AM}}$ is multiplied by $\sqrt{w_i}$, where w_i is the weight given to the i-th monitor. The undulator monitors are assigned a weight stronger by a factor of 10 compared to the default monitors.

Additionally, Tao sacrifices some orbit correction to ensure that there are no excessively strong kicks coming from the correctors. In terms of the matrices, this minimization can be incorporated by vertically concatenating a diagonal matrix to the bottom of $\mathbf{M^{CM}}$, and the diagonal elements are $\sqrt{w_i}$ for the weight w_i of the corresponding i-th corrector. All correctors have the same weight. An $N_C \times N_E$ zero matrix should also be concatenated to the bottom of $\mathbf{M^{EM}}$ to keep proper dimensions for matrix multiplication. The scaled matrices become

$$\overline{\mathbf{M^{CM}}} = \begin{pmatrix} \mathbf{M^{CM}} \\ \mathbf{W} \end{pmatrix}, \tag{16}$$

$$W_{ii} = \sqrt{w_i} \tag{17}$$

$$\overline{\mathbf{M^{EM}}} = \begin{pmatrix} \mathbf{M^{EM}} \\ \mathbf{Z} \end{pmatrix}, \tag{18}$$

$$Z_{ij} = 0, i = 1, 2, \cdots, N_C; \tag{19}$$

$$j = 1, 2, \cdots, N_E \tag{20}$$

VI. LATTICE OPTIMIZATION PROCEDURE

The following section describes the optimization algorithms developed in [1, 2]. They rely on the general response matrices $\mathbf{M^{EM}}$, $\mathbf{M^{CM}}$ and $\mathbf{M^{AM}}$ and their corresponding all-element counterparts to determine the candidate monitors and correctors for addition and removal. To add monitors, one simply appends the desired monitor represented by a row from the all-element matrix to the original matrix, and new correctors are added by including columns from $\mathbf{M^{AM}}$. Monitors (correctors) are deleted by eliminating the respective rows (columns). The algorithms are summarized in Tables II,III,IV,and V.

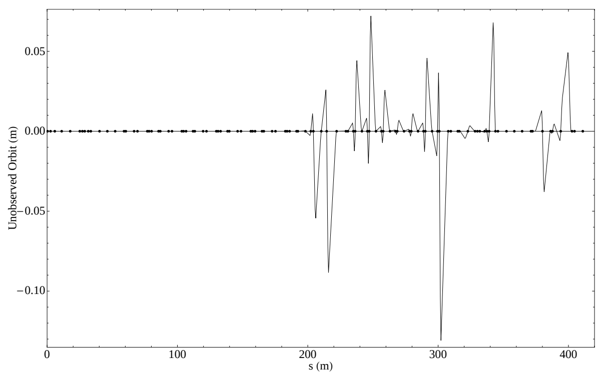
A. Monitor Deficiency

The only information about the orbit is gathered from the readings of the beam position monitors, so most of the orbit's behavior in between monitors remains largely hidden from direct observation. As a result, large undetectable orbits may be present although the readings are well behaved.Then, monitors should be added at locations where this large orbits occur to eliminate the monitor deficiency.
The unobservability of a configuration may be studied through the null space and small singular values of $\mathbf{M^{EM}}$. The null space of $\mathbf{M^{EM}}$ forms an orthonormal basis of all $\mathbf{A^E}$ that satisfy

$$\mathbf{M^{EM}} \cdot \mathbf{A^E} = \mathbf{0}. \tag{21}$$

Each vector is then propagated to all elements using $\mathbf{M^{EA}}$. The largest orbit error lying outside a numerical criterion is noted, and a monitor is placed at its location. Similarly, very small singular values of $\mathbf{M^{EM}}$ correspond to error combinations that are unobservable at the monitors. As discussed earlier, the columns of $\mathbf{V}$ in the decomposition that correspond to very small or zero singular values form an orthonormal basis for the null space, so these basis vectors result in unobservable orbits at the monitors. These errors are again propagated to all elements to find locations of the new monitors.
Figure 3 illustrates an example of a large undetectable orbit. The black dots represent the monitors. Note that the orbit is well-behaved at these locations but have large errors once propagated at all the elements. In this example, a monitor should be placed at the orbit bump near the 300m mark.

FIG. 3: Large Unobservable Orbit



B. Monitor Redundancy

The previous procedure can possibly add too many monitors that have a minimal contribution to improving observability, and excessive monitoring may even constrain the correctors [1]. The following algorithm detects and eliminates these redundancies. There are two parts in the program to minimize the monitor redundancy. The first looks at the coupling between the monitors and errors through the projection operator of $\mathbf{M^{EM}}$. This operator $\mathbf{\Pi^{EM}} = \mathbf{M^{EM}} \cdot \mathbf{M_{EM}}^{\dagger}$

TABLE II: Eliminating Monitor Deficiencies

1 :	Find the null space vectors E of $\mathbf{M^{EM}}$.
2 :	Calculate $V^A = \mathbf{M^{EA}} \cdot \mathbf{E}$, for every E.
3 :	For each V^A add a monitor at the index of its largest element greater than R.
4 :	Repeat steps 1-3 until all elements of V^A are less than R.
5 :	Decompose $\mathbf{M^{EM}}$ by SVD.
6 :	Take the column vector v of $\mathbf{V}$ that corresponds to the smallest singular value.
7 :	Calculate $V^A = \mathbf{M^{EA}} \cdot \mathbf{v}$.
8 :	Repeat steps 3,5-7 until all elements of V^A are less than R.

divides the orbit vectors into parts that are spanned by the errors, and it is applied to unit orbit peaks X_i at each monitor. The vector X_i contains zeros except for the i-th row representing a single reading at i-th monitor. For each monitor, the fractional part $|\bar{X}_i|/|X_i|$ is calculated, where $\bar{X}_i = X_i - \Pi^{EM} \cdot X_i$. This fraction reflects the amount of the unit orbit peak that is not caused by the errors, and the i-th monitor is removed if its fractional component exceeds a numerical criteria $(1 - R)$. Thus,error-induced orbits must contribute to at least a fraction R of a monitor reading if this monitor is to be kept.
The second part of this optimization algorithm compares the Gram determinant of $\mathbf{M^{EM}}$ and the orthogonality of the monitors. For the smallest singular value of $\mathbf{M^{EM}}$, the index of the largest component of the corresponding column vector of $\mathbf{U}$ represents a redundant monitor that should be removed. This program is iterated until the Gram determinant exceeds S^{N_M}.
The Gram determinant measures the orthogonality of the row vectors of a matrix. If the Gram determinant is non-zero then the vectors are linearly independent, and hence orthogonal. By deleting rows of the matrix, the row vectors become more independent. The criterion S is the required Gram determinant between two monitors, and by calculating S^{N_M}, one finds the required orthogonality measure for all the remaining monitors.
The Gram determinant G_M of the matrix $\mathbf{M}$ is

$$G_M = (\prod_j S_j^M)^2 \tag{22}$$

where S_j^M is the j-th singular value. The Gram determinant can be normalized as follows

$$\bar{G_M} = G_M / L_M \tag{23}$$

$$L_M = \prod_j (\sum_i M_{ij}{}^2), rows > columns \tag{24}$$

$$L_M = \prod_i (\sum_j M_{ij}{}^2), columns > rows \tag{25}$$

To illustrate, consider a basic configuration with three error sources and monitors. The configuration has the following simple response matrix and singular value decomposition:

$$\mathbf{M^{EM}} = \begin{pmatrix} 0.99 & 1.99 & 2.99 \\ 2.00 & 3.00 & 4.00 \\ 1.01 & 2.01 & 3.01 \end{pmatrix} \tag{26}$$

$$\mathbf{U} = \begin{pmatrix} -0.493381 & -0.516403 & -0.699931 \\ -0.713325 & 0.700691 & -0.01414 \\ -0.497737 & -0.492302 & 0.714071 \end{pmatrix} \tag{27}$$

$$\mathbf{W} = \begin{pmatrix} 7.53586 & 0. & 0. \\ 0. & 0.459751 & 0. \\ 0. & 0. & 3.63 \cdot 10^{-17} \end{pmatrix} \tag{28}$$

$$\mathbf{V} = \begin{pmatrix} -0.320841 & 0.854631 & 0.408248 \\ -0.547018 & 0.184673 & -0.816497 \\ -0.773196 & -0.485285 & 0.408248 \end{pmatrix} \tag{29}$$

The configuration clearly has a redundancy in the first and last monitors. The smallest singular value of 0. corresponds to the third column of $\mathbf{U}$, and the largest element of this column vector indicates that the last monitor should be removed. The first part of the algorithm also points to this same redundancy. Recalling that $\bar{X}_i = X_i - \Pi^{EM} \cdot X_i$,

$$X_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \bar{X}_1 = \begin{pmatrix} 0.489903 \\ 0.00989703 \\ -0.4998 \end{pmatrix}, \quad (30)$$

$$\frac{|\bar{X}_1|}{|X_1|} = 0.699931 \quad (31)$$

$$X_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \bar{X}_2 = \begin{pmatrix} 0.00989703 \\ 0.00019994 \\ -0.010097 \end{pmatrix}, \quad (32)$$

$$\frac{|\bar{X}_2|}{|X_2|} = 0.01414 \quad (33)$$

$$X_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \bar{X}_3 = \begin{pmatrix} -0.4998 \\ -0.010097 \\ 0.509897 \end{pmatrix}, \quad (34)$$

$$\frac{|\bar{X}_3|}{|X_3|} = 0.714071 \quad (35)$$

C. Corrector Deficiency

Take the fraction of the error-induced orbit that is uncorrectable by the correctors. This fraction is calculated by applying the projection operator $\mathbf{\Pi^{CM}} = \mathbf{M^{CM}} \cdot \mathbf{M_{CM}}^{\dagger}$ on certain column vectors u_i of $\mathbf{U}$ from the SVD of the $\mathbf{M^{EM}}$. Only column vectors with non-zero singular values or above the cutoff defined before are taken to ensure that only the basis for the range is considered for the optimization program. The projection finds the part u_i that is uncorrectable, represented by $Q^{u_i}_{M^{CM}} = |\bar{u_i}|/|u_i|$, where $\bar{u_i} = u_i - \Pi^{CM} \cdot u_i$. For the largest $Q^{u_i}_{M^{CM}}$ greater than the criterion R, the inner product of the corresponding $\bar{u_i}$ with all the column vectors of M^{AM} is calculated to determine the location of the new corrector that best compensates for this deficiency. By taking this inner product, one essentially quantifies the amount by which the uncorrectable orbit represented by $\bar{u_i}$ is projected to the column vectors of $\mathbf{M^{AM}}$, so the procedure detects the location where a corrector would have the best contribution to this uncorrectable orbit.
The required corrector kicking strengths also provides a pathway for eliminating corrector deficiencies. The pseudoinverse $\mathbf{M^{\dagger}_{CM}}$ transforms from the orbit monitor space to the corrector space, so once applied to the u_i defined above, the matrix finds the necessary corrector kicks $K_i = \mathbf{M^{\dagger}_{CM}} \cdot \mathbf{u_i}$. The largest kick K^{max}_i for all K_i greater than the corrector limit S indicates that the i-th corrector requires extremely strong kicks. Again, the inner product of its corresponding u_i with all the column vectors of $\mathbf{M^{AM}}$ is calculated where the largest inner product locates the best new corrector that can decrease the required kick from the i-th corrector.

D. Corrector Redundancy

The response matrix $\mathbf{M^{CM}}$ is singular value decomposed to to find its singular values to find the column vector of V with the smallest singular values. This column vector represents the corrector combination that has minimal to no effect on the monitor orbit. The index j of its largest element is the j-th monitor that should be removed. The procedure is iterated until two numerical criteria are achieved.
The first criterion is the matrix condition number defined as the ratio of the largest singular value to the smallest. This number measures the behavior of the response matrix; a matrix that has an infinite or extremely large condition number is singular or ill-conditioned while one with a smaller condition number is well-behaved. It is important to consider this behavior because an ill-conditioned matrix means that a small change in the input results in a large changes in the output, leading to unreliable solutions. While a definitive cutoff between an ill-conditioned and well-behaved matrix does not exist, it is possible to pick out a sensible cutoff based on the problem's requirements.
The final criterion is S the measure of the orthogonality of the corrector effects on the monitors. The normalized Gram determinant of $\mathbf{M^{CM}}$ is compared with S^{N_C}, where N_C is the number of remaining correctors. The significance of this criteria has already been discussed in the monitor redundancy section.

TABLE III: Reducing Monitor Redundancies

1 :	For every unit orbit peak X_i, calculate $Q^{X_i}_{M^{EM}} = \frac{\lvert X_i - \mathbf{M^{EM}} \cdot \mathbf{M^{\dagger}_{EM}} \cdot \mathbf{X_i} \rvert}{\lvert X_i \rvert}$.
2 :	Delete the i-th monitor with $Q^{X_i}_{M^{EM}} > (1 - R)$.
3 :	Decompose $\mathbf{M^{EM}}$ by SVD.
4 :	Take the column vector u of $\mathbf{U}$ with the smallest singular value.
5 :	Eliminate the i-th monitor corresponding to the index of the largest element of u.
6 :	Repeat steps 3-5 until $\bar{G}_{M^{EM}} > S^{N_M}$.
7 :	Stop at any point when a minimally required number of monitors is achieved.

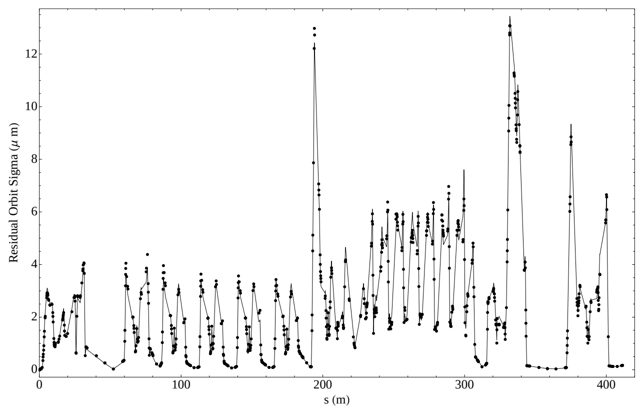
TABLE IV: Eliminating Corrector Deficiencies

1 :	Take column vectors u_i of $\mathbf{U}$ with non-zero singular values from the SVD of $\mathbf{M^{EM}}$.
2 :	For each u_i, calculate $Q^{u_i}_{M^{CM}} = \frac{\lvert \bar{u}_i \rvert}{\lvert u_i \rvert}$, where $\bar{u}_i = u_i - M^{CM} \cdot M^{\dagger}_{CM} \cdot u_i$.
3 :	For $Q^{u_i}_{M^{CM}} > R$, calculate the inner product of $\bar{u}_i$ with all the column vectors of $\mathbf{M^{AM}}$.
4 :	Identify the index of the column of $\mathbf{M^{AM}}$ with the largest inner product.
5 :	Add a corrector at this index.
6 :	Repeat steps 2 - 5 until all $Q^{u_i}_{M^{CM}} < R$.
7 :	Calculate $K_i = \mathbf{M^{\dagger}_{CM}} \cdot \mathbf{u_i}$ for every u_i.
8 :	Identify the largest element K^{max}_i for all K_i.
9 :	Find the largest of all K^{max}_i and take its corresponding u_i.
10 :	Calculate the inner product of u_i with all column vectors of $\mathbf{M^{AM}}$.
11 :	Identify the index of the column of $\mathbf{M^{AM}}$ with the largest inner product.
12 :	Add a corrector at this index.
13 :	Repeat steps 7-12 until all $K^{max}_i < S$.

VII. APPLICATION TO THE SOUTH ARC SECTION OF THE ERL

To test these algorithms, they are first applied to the South Arc. The South Arc is a smaller section of the lattice that contains the majority of the undulators, thus it represents the primary needs of the machine. Since the ERL aims to generate X-ray beams, the orbit at the undulator remains a high priority that must be monitored to ensure that the optimization program does not compromise the orbit at these locations. The section also do not have the added complication of multipass elements, so the generalized response matrices are much easier to calculate.

FIG. 4: SA: Residual Orbit before Optimization



The current configuration of the South Arc contains 103 BPMs and 62 correctors. Its baseline performance is in Figure 4. The residual orbit resulted from including quadrupole x-offset errors at all 111 quadrupole elements. These errors were given a standard deviation of $10\mu m$. The results from the matrix calculations are depicted

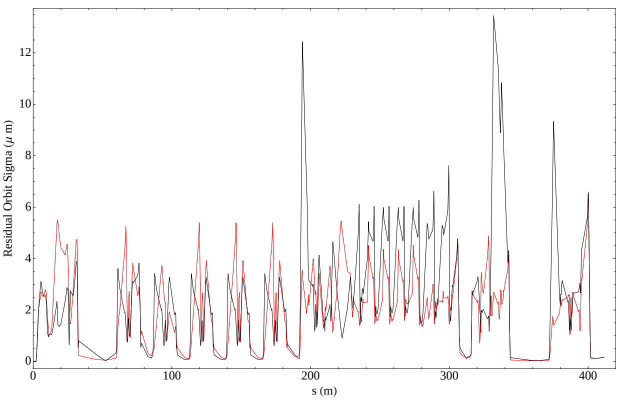
TABLE V: Minimizing Corrector Redundancies

1 :	Decompose $\mathbf{M^{CM}}$ by SVD.
2 :	Obtain the column vector v of $\mathbf{V}$ with the smallest singular value.
3 :	Remove the corrector corresponding to the largest element of v.
4 :	Repeat steps 1-3 until the condition number $N < S$ and $\bar{G}_{M^{CM}} > S^{N_C}$.

as solid lines. To verify their accuracy, they are compared with the results from the Tao simulation, depicted here as the black dots, and the two agree very well.
The nine undulators are located where there are local minima around the 50, 80, 110, 135, 160, 190, 315, 360, 400 meter marks. The orbits here should be kept at a minimum while optimizing and eliminating huge orbit peaks.

A. Optimization of the Current Configuration

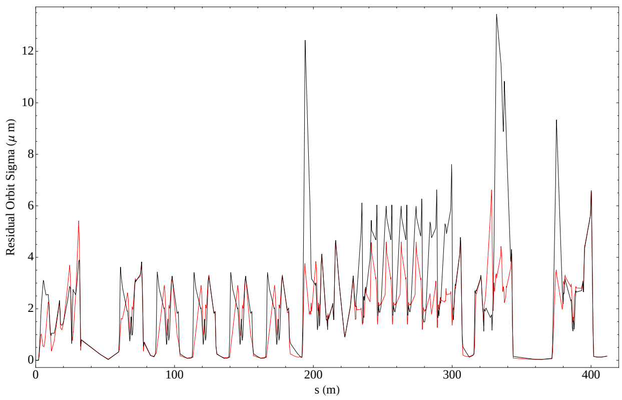
FIG. 5: SA: Residual Orbit After Optimization



The residual orbit after applying both the monitor and corrector optimization programs is shown in Figure 5. A new constraint was placed on the monitor redundancy algorithm due to its sensitivity. Monitors are placed before and after the undulators to increase the monitoring power at these elements, but the program deletes the first of these monitors. Therefore, an exception list and a new criterion indicating the minimum number of monitors required are introduced.

The program simply rearranges the 103 monitors but reduces the number of correctors to 58. The figure represents the new residues as the red envelope. The residual orbits here are more evenly distributed and eliminated the major peaks from the original configuration, cutting the maximum peaks by more than a half, while having a minimal effect on the undulator orbits.

FIG. 6: SA: Residual Orbit After Corrector Optimization



Finally, the original correction scheme was optimized again using only the corrector program. This procedure assumes that the current lattice lack any significant monitor deficiencies and redundancies. It removed nine correctors while still improving the residual orbits, as depicted in Figure 6.

VIII. CONCLUSION

The program written in *Mathematica* takes as input the derivative data from Tao to construct the generalized response matrices. It separates the four optimization algorithms so they can be

applied according to the needs and assumptions behind a given lattice though a full optimization involves the application of all four in the order that they are presented here.
Its application to the South Arc of the ERL lattice has promising results indicating the effectiveness of these algorithms. The next step is to apply the study to the entire ERL lattice. The generality of the code facilitates this study since it only requires the data to generate the appropriate matrices then the optimization immediately follows.

IX. ACKNOWLEDGMENTS

Special thanks goes to my mentor, Chris Mayes, for his support and guidance throughout summer. I would also like to thank David Sagan for his help in solving technical problems with Tao. This project was support by the NSF Grant PHY-0849885 and DMR-0937466.

---

[1] Y. Chao, *Orbit Correction Methods - Basic Formulation, Current Application at Jefferson Lab, and Future Possibilities*, Proc. Workshop on Automated Beam Steering and Shaping (ABS), Edited by Lindroos, CERN, (1998).

[2] Y. Chao, *Methods of Orbit Correction System Optimization*, Proceedings of 1997 Particle Accelerator Conference, Vancouver, Canada.

[3] Y. Chao & V. Mertens, *Analysis and Optimisation of Oribt Correction Configurations Using Generalized Response Matrices and its Application to the LHC Injection Transfer Lines TI2 and TI8*, LHC Project Report 470, May 2001.

[4] D. Sagan, J. Smith, *The Tao Manual*.

[5] D. Sagan, *Bmad: A relativistic charged particle simulation,* Nuc. Instrum. Methods Phys. Res. A, 558 (2006).

[6] C. Song, G. Hoffstaetter, *Particle Orbit Correction in an Energy Recovery Linac*, July 31, 2007.

[7] W. Press, B. Flannery, S. Teukolsky, W. Vetterling, *Numerical Recipies in C*, Cambridge University Press, 1998.