

IMIGO: An Optimal Adaptive Nonlinear Program for Accelerator Optics Modeling and Other Applications

Martin J. Lee

SLAC National Accelerator Laboratory, 2575 Sand Hill Road, CA 94025, USA

Abstract:

Model-based electron accelerator control is the maintenance of optimal parameters of an electron beam such as its orbit, size, and shape, as well as machine parameters such as tunes. It works well when the model reflects reality. SLAC pioneered this technique in SPEAR about thirty decades ago. Similar techniques are now employed in particle accelerator and synchrotron laboratories around the world. There is still an inherently complex problem related to the employment of such techniques to manage the operation and analysis of accelerators and storage rings. The problem arises from the use by those techniques of complex numerical algorithms commonly known as nonlinear solvers that are difficult to control and operate. Lessons learned at SLAC have led to the development of a new, simple-to-use, and iterative nonlinear solver that holds much promise not only in advancing the derivation of errors in accelerators and storage rings at SLAC, but also in its ability to tackle a range of complex engineering problems. A CRADA project was recently completed to build on this experience at SLAC, and to collaborate with an outside company GO AI Services to further develop and validate this nonlinear solver for robust SLAC accelerator. In this report, we present the project report submitted to DOE and SLAC.

Work supported in part by US Department of Energy contract DE-AC02-76SF00515.

Table of Contents

	Page
1.0 Introduction.....	2
2.0 Research Objectives Summary.....	3
3.0 Work Performed and Results Obtained.....	5
4.0 Recommendations for Phase II Study.....	12
5.0 Appendix : Background—OASIS for Solving Equations.....	14
6.0 Acknowledgment:.....	19

Introduction

Accelerator models are widely used to control and optimize beam parameters such as orbit, beam size, beam shape, and global machine parameters such as tunes. Since model-based control works well when the model accurately reflects reality, one of the goals in commissioning an accelerator is to develop a well calibrated lattice model. A technique involves the analysis of a specific set of measured orbit data to identify errors in the magnetic elements in the accelerator lattice. The identified errors are then incorporated into the calibration of the lattice model to produce a “true” representation of the physical machine. This technique was pioneered at SLAC on the SPEAR storage ring about three decades ago. Similar techniques are now employed in particle accelerator and synchrotron laboratories around the world. One limitation of these error finding techniques is that it can only be done off-line by experts. This is a result of the fact that these techniques use complex numerical algorithms commonly known as nonlinear solvers that are often slow and difficult to learn and operate even by experts in accelerator control. In the Phase I work, we further developed a new, simple-to-use, nonlinear program to advance the capability for derivation of errors in accelerators and storage rings. The success in using this nonlinear program in solving SLAC problems as identified in the Phase I project has demonstrated its ability in addressing on-line control and optimization needs of accelerator operators who may be unfamiliar with nonlinear programming. Hence, we are ready to use this versatile nonlinear program to accomplish real time, on-line error analysis of accelerator; therefore, real time, on-line beam control and optimization.

This nonlinear program was developed from a nonlinear solver (called OASIS). During Phase I work, it was applied on a number of nonlinear-solver benchmarking problems and a few SLAC accelerator problems. By searching over only two optimization-control parameters, this Phase I work has enabled the development of a systematic procedure to facilitate the finding of a “best-guessed” global minimum solution for various types of problems with different size and complexity. This set of rules has been codified and has led to a versatile nonlinear program IMIGO.

Research Objectives Summary

The significant Phase I accomplishments are detailed in the following “Work Performed” section; here, let us outline the major objectives completed during the Phase I work.

During the time to complete the Phase I project, the above mentioned versatile nonlinear program IMIGO was developed to provide a simplified and easier-to-comprehend visual rendering of how an objective function of n variables is being minimized. The motivation of that work is that, if an objective function has more than two variables, it becomes impossible to graphically and visually render the manner in which the minimization process of the objective function can be shown to gradually “move” into a largest basin of attraction as a result of changes in the variables. Since IMIGO has reduced the optimization of an objective function of n variables (regardless of how large n is) to one in which seeking the “best-guessed” global minimum of the objective function corresponds to seeking the deepest well of convergence of the function based on seeking the best values of just two convergence control parameters (i.e., r and p as will be explained later), an equivalent rendering of the minimization of the objective function can now be graphically and visually presented. This has been verified and tested in Phase I proposal work.

Qualitative description of IMIGO’s characteristics:

1. In the convergent parameter space, there are two notable features: 1) the area of the well of convergence, and 2) the depth of the well.
2. Within the area of the well of convergence, each (r, p) pair corresponds to a path in the n -variable space that connects the start-point to the end-point.

The size of the area of the well of convergence:

The size becomes smaller as the number of variables increases and the complexity of the problem increases. The lowest depth of the well corresponds to the best-guessed global minimum that is connected to the same start-point by a number of (r, p) pairs. In the n -variable space, there are multiple paths connecting the start-point to the same end-point, each path being generated by a given (r, p) pair. Hence, all (r, p) pairs making such a connection forms a bundle between the start-point and the end-point in n -variable space. Such (r, p) pairs connecting the start-point to the same end-point explain why the bottom of the well of convergence is flat.

The depth of the well:

The depth is a measure of how close the objective function is to the “best-guessed” global minimum in n -variable space.

The Phase I project has also developed a search method for selecting the convergence parameters r and p in a systematic way as will be explained later. Such a graphic user interface (GUI) turns out to be very useful to show the effectiveness of the searching algorithm in finding the global minimum.

The Phase I project accomplished the following major objectives:

- A. Explored means to search for a set of “warm” start solutions and associated non-linear paths leading to the global-minimum solution.

There is a significant problem in the use of those popular iterative nonlinear programs—it is known as the basin-of-attraction limit (BOA). A BOA is defined to be the biggest region around a given minimum solution within which the program will converge. The problem with existing iterative nonlinear programs is that they will only find a local solution unless the start solution is inside the BOA corresponding to the global-minimum. In the Phase I project, we have developed a nonlinear program that does not have the inherent BOA limitation of conventional nonlinear programs.

- B. Use the resulting methods to find lattice element errors in SPEAR3.

To reach these goals, the following tasks were accomplished:

1. LOCO and OASIS Pathfinder have been merged into one integrated system.
2. A prototype code for testing this system’s ability to find the global-minimum solution with simulated data from a SPEAR3 model has been produced.
3. A hands-free automated search procedure to find the global optimal solution has been developed for OASIS Pathfinder.
4. Requirements in the use of this lattice error solver as a training tool for operators so that expert analysis can be brought into the control room at all times for finding faults without expert interventions have been investigated.
5. Further research and development on OASIS Pathfinder for other on-line applications have been identified.

- C. Applied the resulting method developed in Phase I to optimize the Accelerator and Free Electron Laser performance at LCLS.

It is known that bunch compression scheme and Free Electron Laser is a highly nonlinear process. So the optimization is not an easy task. Work has been started to merge the OASIS Pathfinder with the existing LINAC model aiming at an on-line model server which can provide on-line real time optimization. Related work on setting up an on-line server has been reported in Kyoto, Japan of the 1st International Particle Accelerator conference¹.

Work Performed and Results Obtained

The inherent difficulty of using an iterative method to find the global-minimum solution is well known. In general, an iterative method requires an initial guess solution. If this start solution is too far from the global-minimum solution, the program will find only a local-minimum solution. As an illustration, a surface plot of the objective function for a minimization problem with two variables is shown in Fig. 1. This figure shows the locations of several local-minimum points and one global-minimum point.

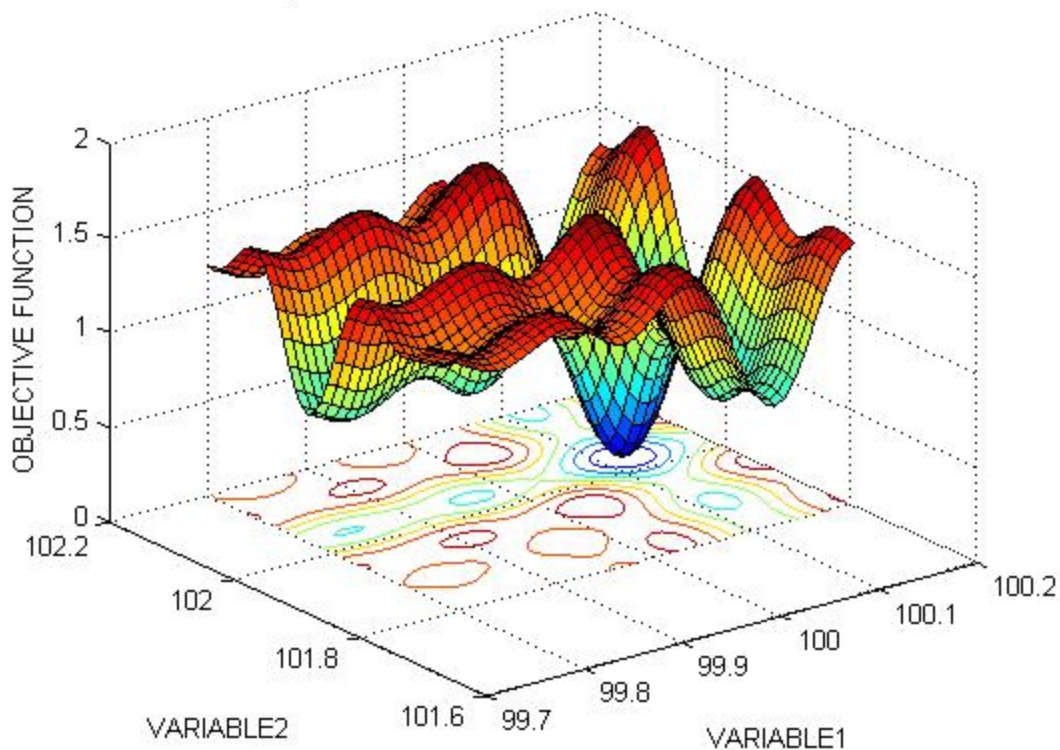


Figure 1, Objective function surface plot.

In general, a nonlinear program is a solver typically employed to find the global minimum of a given objective function subject to certain conditions known as constraints. An objective function of n variables (a_1, a_2, \dots, a_n) is generally written as $f_{\text{obj}}(a_1, a_2, \dots, a_n)$. In the accelerator lattice modeling, the variables are the strengths of a chosen set of magnet components. The complexity of finding the global minimum for the objective function depends on the degree of nonlinearity of the objective function and the number of variables. In the Phase I work, we have developed a nonlinear program, IMIGO that does not have the inherent BOA limitation of conventional nonlinear programs. This new nonlinear program is able to **find a solution path that ends at the global minimum solution even when the start solution is not inside the BOA corresponding to the global-minimum solution.**

IMIGO is composed of two main components. They are named: **Original and Simple Iterative Solution (OASIS)** and **Lee-Wu Way (LWW)**. OASIS is a non-derivative based algorithm that finds the values of the variables iteratively for an initially-guessed start solution. The iterative process in OASIS is shown in Fig. 2. When the values of the variables converge, the set of values of the variables at the end point is a solution that corresponds to a minimum, maximum, or saddle point of the objective function. This unique feature enables LWW—a simple, two-parameters search method working in conjunction with OASIS—to find a path leading to a solution within the BOA corresponding to the global minimum solution developed in the Phase I project.

A main feature of this new search procedure is that it can find the global-minimum solution for any start solution within a given solution space—IMIGO does not have BOA limitation. Another salient feature is that it works equally well regardless of the size and complexity of a given problem, *i.e.*, no matter how many variables an objective function may have and how nonlinear the problem may be.

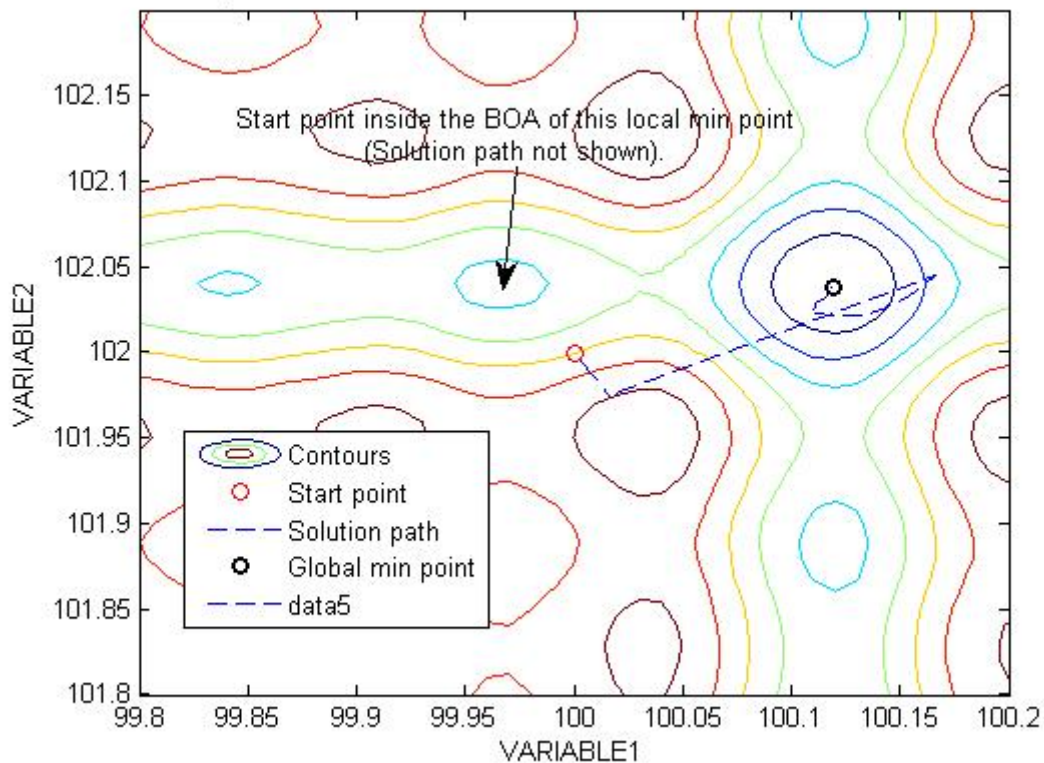


Figure 2, plot of the solution path which ends at the global minimum.

Self-Adaptive Optimization and Examples

One of the unique features of IMIGO is that both the values of the rate and path control parameters are adjusted adaptively along a path: $r_{n=1,2,\dots,N}$ and $p_{n=1,2,\dots,N}$. In order to enhance the convergence, the IMI algorithm adjusts their values using a special recipe developed by us in the Phase I work. This recipe forms the core element inside the automatic method LWW for finding the global optimum solution automatically. In addition, a simple graphical user interface (GUI) was developed to run IMIGO. With this GUI added to IMIGO, users are able to run IMIGO in two different options: hands-free or hands-on mode. The hands-free option was developed to run IMIGO automatically, while the hands-on mode was developed for running IMIGO manually. With the hands-free mode, it is possible to test IMIGO quickly on a variety of optimization problems with different number of variables and different objective functions. In all cases tested, IMIGO is able to find the global minimum solution independent of the size and complexity of the problems for arbitrary chosen start solutions. **With the hands-on mode, it is possible to use IMIGO for training new users as well as for exploring ways to customize IMIGO for solving unresolved optimization problems by expert users.** Three selected cases are presented in the following section as examples.

(A) *The Illustrative Problem in the Phase I Proposal:* As an illustration of how IMIGO works, a typical small-scale minimization problem with two variables a_1 and a_2 is presented. A plot of the objective function to be minimized is shown in Fig. 1. For this case, the start solution is given by $b_1^{start} = 100$ and $b_2^{start} = 102$, and the bounded variables are allowed to vary between: $0.2 > (b_k - b_k^{start}) > -0.2$ for $k = 1, 2$. By running IMIGO in its hands-on mode repeatedly with different initial (r, p) values, IMIGO is able to find a set of solution paths starting at the same point and ending at the global minimum point: $a_1^{end} = 100.13$ and $a_2^{end} = 102.04$. Figure 2 shows the special feature of IMIGO—its unique ability to find the global-minimum solution when the start point is outside of the BOA of the global-minimum point (blue contour-line in the figure). By running IMIGO automatically, IMIGO is able to search over the entire convergence control space for the initial (r, p) values with a path that ends at a point inside the BOA of the global-minimum point. Figure 2 shows one example of a contour plot of the objective function values in the search space for the initial control parameter values. In this case, the paths corresponding to initial values at the points inside the colored (yellow/green) contour all end at solutions that are close to the global minimum solution. These end solutions can be used as warm start solution for the next search using IMIGO. In addition, if the users desire, they also can be used as warm start solution for conventional solvers.

For the purpose of comparing IMIGO with other conventional solvers, one conventional optimization program, *e.g.*, Levenberg–Marquardt (L-M) was used to find the global minimum solution for this case. This program was unable to find the global minimum solution since the starting solution for this case is outside of the BOA corresponding to the global minimum solution. As a test of this BOA-free

property of IMIGO, many other cases with different values of starting solutions were compared. In all cases, whenever L-M failed to find the global minimum solution, IMIGO succeeded.

(B) A Current Accelerator Problem:

As an application of IMIGO to a real accelerator project, we have used IMIGO to optimize the two bunch compressors settings for the LINAC Coherent Light Source (LCLS¹) at SLACⁱⁱ, which is an eight-variable problem. The objective function is formed to set the final electron rms bunch length, the final centroid energy, and the final energy chirp along the electron bunch; and simultaneously minimize the rms bunch length fluctuation and the final energy chirp fluctuation. The objective function is a function of eight variables: the LINAC acceleration phase and total acceleration voltage in the three LINAC sections, and the R_{56} (transport matrix element) of the two bunch compressors. As an example, the objective function as a function of the two LINAC section (L1 and L2) phases is shown in Fig. 3. The IMIGO solves this 8-variable problem and optimize the LINAC setting for LCLS.

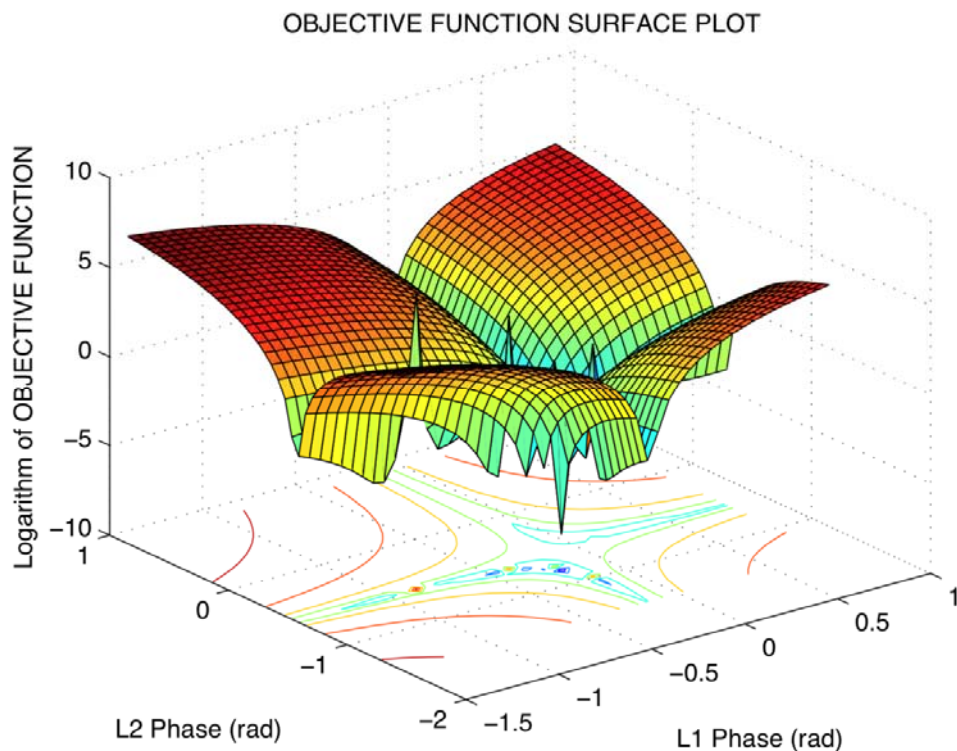


Figure 3: the contour surface plot of the objective function.

¹ LCLS is the world's first x-ray free electron laser recently commissioned at SLAC.

(C) Benchmarking of IMIGO with L-M using LOCO:

One of the recently obtained historical models was used for this work. Three tasks in the Phase I work plan were: (1) integrate IMIGO with the modeling code LOCO, (2) see if IMIGO can find the same solution as L-M, and (3) see if the original solution obtained by LOCO is a global minimum solution. All three tasks have been done. The results are all positive. For example, when the same starting solution in running L-M is used to run IMIGO, IMIGO is able to find the same solution as L-M does. In addition, when the solution in the historical model is used as a start solution to run IMIGO, however, IMIGO is unable to find better solution—one with a lower value of the objective function than the one in the historical model. This observation leads to the conclusion—LOCO has found the “global” minimum solution within the search space as defined by the range limits imposed on the variables. The result will be reported with details of the results for the 2011 Particle Accelerator Conference in New Yorkⁱⁱⁱ.

Here let us outline some features of the work done. Before the Phase I proposal was submitted, a linear solver was used by LOCO to find errors in the strength of the quadrupole magnets in SPEAR 3 ring without imposing constraints on their values. Since then, we have participated in the transition; so that this linear solver was replaced by a nonlinear solver (L-M in this case) in order to impose limits on the range of the variable quadrupole strength errors by SLAC. Substantial improvement in accuracy of the models obtained from L-M has been found as compared to the original ones obtained from using a linear solver. This demonstrated the significance of the concept which we are trying to emphasize, *i.e.*, modern accelerator and light source control systems should be upgraded to use non-linear solver rather than staying on the linear level only. Under this condition, tests were conducted to compare IMIGO and L-M. It was observed that IMIGO and L-M always find the same solution. With the above mentioned success, instead of trying to benchmark IMIGO with L-M on the LOCO modeling problem, attention was redirected to see how the solutions found from using either solver can be improved. This study led to the method to model the magnetic hysteresis effects in the quadrupole field errors (more details will be described in later part of this section).

Phase I derived opportunities

Since future work is a natural continuation and further development of the Phase I work, we view the task accomplished and started in Phase I as dual tasks both for Phase I and future work. Some summary and restatement is in the following.

The work in Phase I can be divided into two main tasks: A. Further develop OASIS Pathfinder and make it easy to learn and use; B. Testing it on modeling and finding a way to improve the existing models.

From working on A, a new automated search method LWW (Fig. 4) to find the global minimum solution was developed and a user friendly GUI was developed for user training (Fig. 5).

As we mentioned above, no matter what the dimension of the optimization problem is, IMIGO is able to convert it into a 2-D search for the global minimum. For example, in Fig. 4, the 8-dimensional problem is visualized as a 2-D search. One nice feature there is that the BOA is widely enlarged, which makes the search not too difficult.

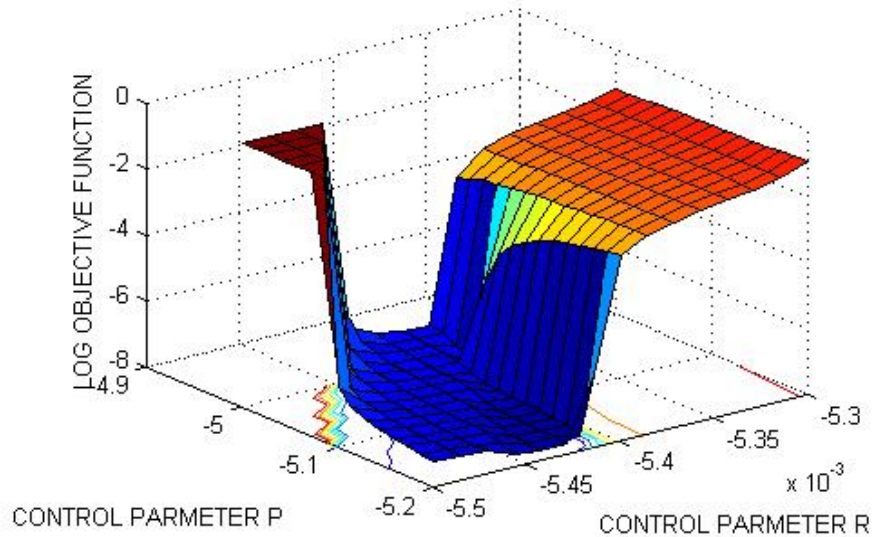


Figure 4: A 8-dimensional optimization is visualized by using the (r, p) pair.

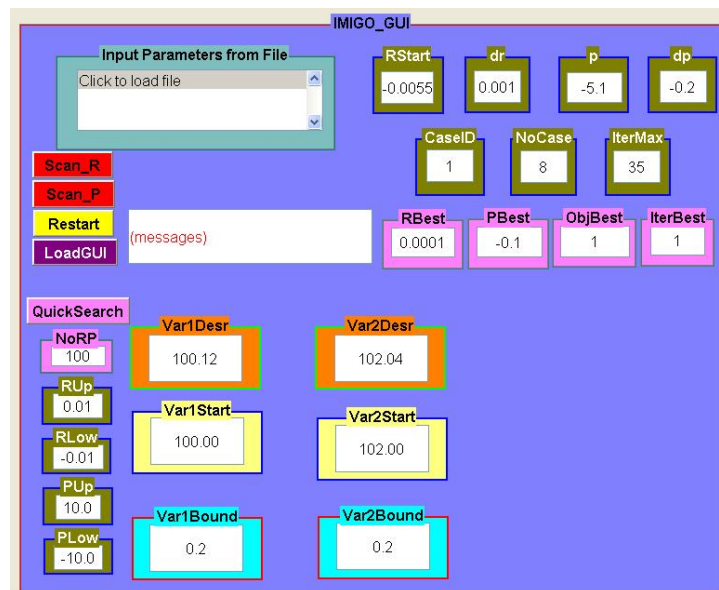


Figure 5: The Graphic User Interface (GUI) developed during Phase I.

To facilitate this 2-D search problem, a GUI is developed as show in Fig. 5. Users can simple type the (r, p) pair and the code will convert the multi-dimensional problem into this simple 2-D search.

From working on B, a new approach based on historical models was conceptualized. This approach can take into account of errors that have not been included in the existing models such as magnetic hysteresis.

Recommendations for Phase II study

During the Phase I work, we identified the following problems which can be conducted in Phase II study.

On-line error identification

A typical example of an off-line model-based analysis is known as the Resolve technique^{iv}. In this method, measured beam trajectories are compared to the set of data generated from the accelerator model in order to diagnose errors in a large set of system parameters such as field strengths of quadrupole magnets and calibrations of beam position monitors. There are two problems with this method: (a) As mentioned above, it can only be done off-line; and (b) It uses a linear model that can only find the first-order lattice errors. Though several methods for finding an accurately calibrated model of an existing ring have been developed, there is another limitation—such a calibrated model can be used to control only the first-order beam parameters accurately such as tunes and beta-functions. To operate next generation accelerators, a well calibrated model will be needed to control even the second-order beam parameters accurately. Two things will be needed to address this problem: First, an off-line error identification method needs to be developed, and then, this method needs to be put on-line.

Therefore, an on-line error finding package that uses a highly nonlinear approach is yet to be invented. As explained above, with the nonlinear solver developed in the Phase I work, it is feasible to complete the development of this on-line nonlinear error identification system in Phase II.

On-line multi-knob control

Model-based control allows a function commonly known as a “multi-knob” to be designed to make changes in several beam parameters at the same time. Currently, the process for making a multi-knob can only be done off-line by experts. To make a multi-knob that changes a set of ‘N’ specific beam parameters linearly, typically the following method is used. First, select a set of N correctors and assign them to a multi-parameter control knob to be adjusted by an operator. As the knob is turned, the values of the corrector strengths change in a specific way such that the beam parameter values change linearly with the change in the position of the knob. Typically, a model is used to compute the N corrector values corresponding to the N beam parameters. Similar to lattice error-finding problems there are two limitations with this control application: (a) a multi-knob can only be designed off-line, and (b) it is difficult to make a multi-knob on demand by operators.

Application of the aforementioned methods is limited to cases where the beam data or desired parameter changes can be computed from a single-particle or first-order optical model. In practice, there are other cases in which such models do not exist. For example, multi-particle tracking codes are often used to compute dynamic aperture. Because of the lack of an analytical model, model-based optimization of dynamic aperture remains

unresolved in non-linear, multi-parameter, on-line optimization problems. Yet, based on the success of Phase I work, we see the hope and possibility to solve this problem as will be detailed in the following sections.

Lack of an on-line control protocol for accelerators described by a complex simulation model

Sometimes, an accelerator system can be described by a complex simulation model that includes space charge effects. However, there is no commensurate on-line control protocol that can adequately deal with on-line system control for problems as such. In practice, numerical simulations are often used to study such effects off-line.

Experts in accelerator control can deal with the unavailability of an on-line protocol for the control of parameters that affect the response of a system described by a complex simulation model. Their success in doing so, often based on heuristic control, depends on an extraordinary level of knowledge about the system, and also on experience and knowledge in the cause and effect of their control actions. Because bad control decisions obviously lead to undesirable outcomes in system behavior, those decision steps are not taken lightly, resulting in a successful control process taking a very long time to go through. Therefore, an easy-to-use, fast nonlinear solver combined with a non-model-based optimization method developed in Phase I is a necessary step towards finding a remedy for this problem. The details of this combined approach will be described in a later section.

Data-mining of past model-calibration files of an accelerator

After an accelerator's operating parameters have been properly calibrated, it will usually run well until: 1) certain uncontrolled changes in the accelerator begin to set in and then lead to undesirable changes in the performance of the system, or 2) reconfigurations of the accelerator operating parameters are required in order to change the performance of the accelerator to some other desired state. In either case, recalibrations of the accelerator model become necessary. Actually, quite a lot of information about the way the accelerator performs is contained in its stored data files, such as values of quadrupole-magnet strength errors. However, prospective use of such data has never been attempted in any systematic way. Therefore, a versatile and flexible-to-use nonlinear program to obtain a parametric model by mining such vast bodies of stored data can be a task for Phase II proposal. The resulting parametric model can replace the first principle model now being used for operation.

Appendix: Background—OASIS for Solving Equations^v

OSASIS uses a special iteratively algorithm to solve for the values of a set of variables (x_1, x_2, \dots, x_m) for a given set of equations:

$$f_i(x_1, x_2, \dots, x_m) = 1 \text{ for } i = 1, 2, \dots, m,$$

where $f_i(x_1, x_2, \dots, x_m)$ for $i = 1, 2, \dots, m$ are positive valued analytic functions of these m variables.

The development of this algorithm started when it was first noticed that this set of equations can be put into a matrix form in terms of a non-zero constant parameter r and the unit vector \hat{u} :

$$M(\bar{x}, r) \bar{x} = \hat{u}$$

where

$$\begin{pmatrix} \frac{1-(m-1)r}{x_1} f_1 & \frac{rf_1}{x_2} & \frac{rf_1}{x_3} \dots & \frac{rf_1}{x_m} \\ \frac{rf_2}{x_1} & \frac{1-(m-1)r}{x_2} f_2 & \frac{rf_2}{x_3} \dots & \frac{rf_2}{x_m} \\ \dots & \dots & \dots & \dots \\ \frac{rf_m}{x_1} & \frac{rf_m}{x_2} & \frac{rf_m}{x_3} \dots & \frac{1-(m-1)r}{x_m} f_m \end{pmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \dots \\ x_m \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}.$$

Since $M(\bar{x}, r)$ can be inverted analytically to obtain

$$M(\bar{x}, r)^{-1} = \frac{1}{1-mr} \begin{pmatrix} \frac{(1-r)x_1}{f_1} & \frac{-rx_1}{f_2} & \frac{-rx_1}{f_3} \dots & \frac{-rx_1}{f_m} \\ \frac{-rx_2}{f_1} & \frac{(1-r)x_2}{f_2} & \frac{-rx_2}{f_3} \dots & \frac{-rx_2}{f_m} \\ \dots & \dots & \dots & \dots \\ \frac{-rx_m}{f_1} & \frac{-rx_m}{f_2} & \frac{-rx_m}{f_3} \dots & \frac{(1-r)x_m}{f_m} \end{pmatrix},$$

the given set of equations to be solved can also be put into an inverse matrix form:

$$\bar{x} = M(\bar{x}, r)^{-1} \hat{u}.$$

By grouping terms, the right-hand-side of this equation can be rearranged to become

$$M(\bar{x}, r)^{-1} \hat{u} \equiv T(\bar{x}, r) \bar{x}$$

where

$$T(\bar{x}, r) \equiv \frac{1}{1 - mr} \begin{pmatrix} \frac{1}{f_1(\bar{x})} - r \sum_{k=1}^m \frac{1}{f_k(\bar{x})} & 0 & 0 \dots & 0 \\ 0 & \frac{1}{f_2(\bar{x})} - r \sum_{k=1}^m \frac{1}{f_k(\bar{x})} & 0 \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \frac{1}{f_m(\bar{x})} - r \sum_{k=1}^m \frac{1}{f_k(\bar{x})} \end{pmatrix}.$$

Since this condition is satisfied for all values of \bar{x} , it is referred to as the Inverse Matrix Identity or IMI.

The IMI Algorithm

By defining

$$g_k = \frac{1}{f_k} \text{ for } k = 1, 2, \dots, m$$

and

$$g_o = \frac{1}{m} \sum_{k=1}^m g_k,$$

the diagonal matrix T can be written as

$$T(\bar{x}, r) = \frac{1}{1 - mr} \begin{pmatrix} g_1 - mrg_o & 0 & 0 \dots & 0 \\ 0 & g_2 - mrg_o & 0 \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & g_m - mrg_o \end{pmatrix}.$$

By replacing $M(\bar{x}, r)^{-1} \bar{I}$ with $T(\bar{x}, r) \bar{x}$, the inverse matrix equation $\bar{x} = M(\bar{x}, r)^{-1} \bar{I}$ becomes

$$\bar{x} = T(\bar{x}, r) \bar{x}.$$

It is the formulation of this diagonal matrix equation that leads to the IMI algorithm for solving a given set of equations. The steps in the IMI algorithm are:

0. Choose the values for \bar{x}^{start} and r .
1. Compute $\bar{x}^{first} = T(\bar{x}^{start}, r) \bar{x}^{start}$.
2. Compute $\bar{x}^{second} = T(\bar{x}^{first}, r) \bar{x}^{first}$.
3. Compute $\bar{x}^{third} = T(\bar{x}^{second}, r) \bar{x}^{second}$.

Stop stepping when the value of \vec{x}^{end} converges.

It can be seen that at the end point: $\vec{x}^{end+1} \approx \vec{x}^{end}$ and

$$T(\vec{x}^{end}, r) = I$$

where I is the identity matrix and

$$T(\vec{x}^{end}, r) = \frac{1}{1 - mr} \begin{pmatrix} \frac{1}{f_1(\vec{x})} - r \sum_{k=1}^m \frac{1}{f_k(\vec{x})} & 0 & 0 \dots & 0 \\ 0 & \frac{1}{f_2(\vec{x})} - r \sum_{k=1}^m \frac{1}{f_k(\vec{x})} & 0 \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \frac{1}{f_m(\vec{x})} - r \sum_{k=1}^m \frac{1}{f_k(\vec{x})} \end{pmatrix}_{\vec{x} = \vec{x}^{end}}.$$

Under this condition, the values of $f_i(\vec{x}^{end}) \cong 1$ for $i = 1, 2, \dots, m$, and the solution $\vec{x}^{solution}$ is given by the value of \vec{x}^{end} , i.e., $\vec{x}^{solution} \cong \vec{x}^{end}$.

The Rate Control Parameters: r

Let $n = 0, 1, 2, \dots, N$ denotes the point at the beginning of each step— $n=0$ refers to the start point and $n=N$ refers to the end point. In OASIS Pathfinder, the set of points is referred to as a path. It computes the values of (x_1, x_2, \dots, x_m) at each point using these formulas:

$$x_i^{(n+1)} = \frac{g_i(\vec{x}^{(n)}) - mr g_o(\vec{x}^{(n)})}{1 - mr} x_i^{(n)} \text{ for } i = 1, 2, \dots, m.$$

The rate of convergence given by

$$\Delta x_i^{(n+1)} \equiv x_i^{(n+1)} - x_i^{(n)} = \frac{(g_i(\vec{x}^{(n)}) - 1) - mr(g_o(\vec{x}^{(n)}) - 1)}{1 - mr} x_i^{(n)} \text{ for } i = 1, 2, \dots, m.$$

This expression shows how the convergent *rate* parameter r affects the rate of change of at each point along the path.

The Unbounded Coordinate Transformation

Before using IMI algorithm, two transformations are performed by OASIS Pathfinder: The first is used to transform the coordinates from a set of bounded variables (b_1, b_2, \dots, b_m) to a set of unbounded variables (a_1, a_2, \dots, a_m) ; the second is used to transform this set of

unbounded variables (a_1, a_2, \dots, a_m) to a set of IMI variables (x_1, x_2, \dots, x_m) . Let the values of the variables be constrained to within given upper and lower limits: $b_{k+} > b_k - b_{ko} > b_{k-}$ where b_{ko} is a constant. Let the set of equations to be solved be given:

$$f_k(b_1, b_2, \dots, b_m) = 1 \text{ for } k = 1, 2, \dots, m.$$

The transformation from the given set of bounded variables (b_1, b_2, \dots, b_m) to a set of unbounded variables (a_1, a_2, \dots, a_m) is given by:

$$a_k - a_{ko} = \tan\left(\frac{\pi(b_k - b_{ko})}{(b_{k+} - b_{k-})}\right)$$

with $b_{ko} = \frac{(b_{k+} + b_{k-})}{2}$ and a_{ko} are constant off-sets for $k = 1, 2, \dots, m$.

By substituting the inverse coordinate transformation

$$b_k = \frac{(b_{k+} - b_{k-}) \tan^{-1}(a_k - a_{ko})}{\pi} + b_{ko}$$

into the set of given equations it becomes:

$$f_k(a_1, a_2, \dots, a_m) = 1 \text{ for } k = 1, 2, \dots, m.$$

The given set of equations to be solved for the value of \vec{a} becomes:

$$M(\vec{a}^{(N)}, r) \vec{a}^{(N)} = \hat{u}$$

where

$$M(\vec{a}, r) = \begin{pmatrix} \frac{1-(m-1)r}{a_1} f_1 & \frac{rf_1}{a_2} & \frac{rf_1}{a_3} \dots & \frac{rf_1}{a_m} \\ \frac{rf_2}{a_1} & \frac{1-(m-1)r}{a_2} f_2 & \frac{rf_2}{a_3} \dots & \frac{rf_2}{a_m} \\ \dots & \dots & \dots & \dots \\ \frac{rf_m}{a_1} & \frac{rf_m}{a_2} & \frac{rf_m}{a_3} \dots & \frac{1-(m-1)r}{a_m} f_m \end{pmatrix}.$$

The Working Coordinates and the Rate Control Parameters: p

Before using the IMI Algorithm, the set of unbounded variables (a_1, \dots, a_m) is transformed into a set of working variables, (x_1, \dots, x_m) . This transformation involves the Hessian matrix of the given objective function evaluated at the start point of the unbounded variables $(a_1, \dots, a_{m-2})^{start}$. Let the extended Hessian matrix evaluated at the start point

$(a_1, \dots, a_m)^{start}$ be denoted as H with its elements given by $h_{jk} = \frac{\partial^2 f_{aof}^{start}}{\partial a_j \partial a_k}$. Eigenvalue

Decomposition of H gives

$$H = V\Lambda V^T,$$

where V is an unitary matrix with $V^{-1} = V^T$, and the elements of the diagonal Λ -matrix are the singular values¹ of H . This coordinate transformation is given by:

$$\bar{x} = V^T \bar{a}$$

where $\bar{x} = (x_1, \dots, x_m)^T$.

Note that for points near the start point

$$\bar{y}_p \approx \bar{J}(\bar{a}) - \bar{J}(\bar{a}^{start}) + H\bar{a}^{start},$$

where \bar{J} is the Jacobian of the objective function with its k^{th} element given by $\frac{\partial f_{aof}(a_1, \dots, a_m)}{\partial a_k}$ for $k = 1, 2, \dots, m$. For points that are not close to the start point, IMIGO

defines a parameter p to account for this approximation:

$$\bar{y}_p \equiv \bar{J}(\bar{a}) - \bar{J}(\bar{a}^{start}) + pH\bar{a}^{start}.$$

Furthermore, since $\bar{x} = V^T \bar{a}$, the condition above can also be expressed as:

$$\bar{y} = V^T \bar{y}_p = V^T (\bar{J}(\bar{x}) - \bar{J}(\bar{x}^{start})) + p\Lambda\bar{x}^{start}.$$

Note that the values of \bar{y}_p at the points on a path depend on the values of p , IMIGO refers to it as the path control parameter.

Acknowledgment

This report is a culmination of many years research into the field of non-linear numerical fitting algorithms and their application to particle accelerators at SLAC. At first the motivation of the work was primarily geared toward lattice diagnostics for both linear and circular accelerators but as the strength and robustness of the algorithms grew the applications were extended into more non-linear problems such as slice-emittance fitting in electron gun beam dynamics, complex electron beam lifetime modeling in storage rings and preliminary considerations for protein crystallography reconstruction. Many of these ventures were born of a sequence of SBIR contracts between SLAC and variety of private companies in industry specializing in a range of topics from neural nets to sophisticated non-linear adaptive algorithms incorporating time dependent concepts along the lines of Kalman filtering. In the later years of the research the SBIR contracts have concentrated the focus more and more on global optimization using an iterative, non-linear adaptive algorithm developed by Dr. Martin Lee. The goal of this more recent research has been to develop an algorithm that could locate the solution or one inside the basin-of-attraction of the 'global minimum' in an otherwise highly non-linear hyperspace of high dimensionality of an optimizing problem.

For demonstration purposes, the entirely new algorithm was benchmarked against standard problems in the field of nonlinear solvers such as the well-known Levenberg-Marquardt algorithm. The years of algorithm development by Martin has resulted in the production of a software package that is relatively easy to configure in terms of identifying variables, constraints, a system model and computational cost-function that runs efficiently on common PC platforms. Of central importance, the algorithm only needs to compute the Hessian once at the beginning of the calculation without the need for repeated calculations during the iterative process and its convergence at the solution depends on only two control parameter values that contain information identifying progress toward attainment of the global minimum as the calculation progresses.

Recognition of assistance and collaboration throughout the many decades of research leading to the present algorithm would produce a list that is long and surprisingly widespread in breadth. Of particular significance are Martin's Professor **Bernard Wisdrow** of Stanford University who for many years taught courses on neural net and non-linear algorithms, his colleague Professor **Richard Blankenbecler** of SLAC and Professor **Gene Golub**, also of Stanford University whose insight into matrix theory and the modern development of Singular Value Decomposition with application to digital computers solidly places his name in history as one of the Greats. Concerning on-line applications and demonstrations, early ground-breaking work was carried out on accelerators at SLAC including SPEAR, PEP, the SLC and associated damping rings and transport lines. Off-shoots of the early developments have successfully applied in particular to second-generation SLAC accelerators PEP-II and SPEAR3 as well as many other machines worldwide that now routinely use response-matrix analysis to model the as-built magnet lattice. It is our hope that this progression of developments continues for many years to come taking advantage of ideas contained in this report, the creativity and ingenuity of young new scientists throughout the world and the increasingly powerful capabilities of modern computers.

We would like to thank Robert Hettel, John Galayda, Ewan Paterson, James Safranek, John Schmerges, and Yiton Yan for their supports and advises on our project.

ⁱ P. Chu, J. Wu, G. Shen, J. Qiang, “*Generic model host system design*”, (submitted to the first International Particle Accelerator Conference, May, 2010, Kyoto, Japan)

ⁱⁱ M.J. Lee and J. Wu, “*Optimal control of electron beam parameters and machine setting with a new nonlinear program*”, Presented at 2009 Particle Accelerator Conference (PAC 09), Vancouver, Canada, 4-8 May, (2009).

ⁱⁱⁱ M.J. Lee, J. Wu, *et al.* (to be submitted to the 2011 Particle Accelerator Conference, March, 2011, New York, USA).

^{iv} W.J. Corbett, M.J. Lee, V. Ziemann, “*A Fast Model Calibration Procedure for Storage Rings*”, SLAC-PUB-6111, May 1993. 3pp. Presented at 1993 Particle Accelerator Conference (PAC 93), Washington, DC, 17-20 May, pp. 108-110 (1993).

^v Martin J. Lee and Richard Blankenbecler, “*Tomographic Reconstruction of Nanoscale Objects with a Nonlineae Optimal Adaptive Simple Iterative Solution*”, SLAC-Report-758, March 2011--
Provisional Patent Application for Stanford University Docket S03-095, filed at USPTO in Nov. (2005).