A HYBRID NUMERICAL METHOD FOR ORBIT CORRECTION

Greg White, Tom Himel, Hamid Shoaee *
Stanford Linear Accelerator Center, Stanford University, Stanford CA 94305

Abstract

We describe a simple hybrid numerical method for beam orbit correction in particle accelerators. The method overcomes both degeneracy in the linear system being solved and respects boundaries on the solution. It uses the Singular Value Decomposition (SVD) to find and remove the null-space in the system, followed by a bounded Linear Least Squares analysis of the remaining recast problem. It was developed for correcting orbit and dispersion in the B-factory rings.

1 INTRODUCTION AND PROBLEM STATEMENT

The main objective in accelerator steering is to minimize deviations of the beam from the center of the beam pipe, that is, to minimize the rms of the orbit. There may be other objectives such as minimizing the corrector strengths, or dispersion, but these are generally secondary. What follows can be generalized very easily to include these secondary objectives, but for illustration only orbit correction will be discussed.

For practical purposes of online orbit correction in a linear accelerator, its reasonable to assume that there is only a linear relationship between a corrector magnet's strength (the extent to which it bends the beam) and the beam's position when measured horizontally or vertically by any "down-stream" Beam Position Monitor (BPM). The magnitude of this influence can be computed or measured, and recorded, for all magnets to all BPMs in the accelerator. This coefficient is sometimes called, informally, T_{12}, the subscripts refer to the position of this coefficient in the larger "Transport matrix" discussed in [5] and elsewhere, which describes to the first-order the action-response relationship between control elements of a beam-line.

The problem of accelerator steering can then be posed as a system of simultaneous linear equations relating M BPMs to N magnets used for orbit correction.

$$\begin{aligned}
T_{12}^{11}\Delta\theta_1 + T_{12}^{12}\Delta\theta_2 + \cdots + T_{12}^{1N}\Delta\theta_N &= \Delta BPM_1 \\
T_{12}^{21}\Delta\theta_1 + T_{12}^{22}\Delta\theta_2 + \cdots + T_{12}^{2N}\Delta\theta_N &= \Delta BPM_2 \\
&\vdots \\
T_{12}^{M1}\Delta\theta_1 + T_{12}^{M2}\Delta\theta_2 + \cdots + T_{12}^{MN}\Delta\theta_N &= \Delta BPM_M
\end{aligned}$$

where $\Delta BPM_i,\ i = 1 \ldots M$, is the desired change in the ith BPM, and $\Delta\theta_j,\ j = 1 \ldots N$, is the sought change in the bend angle of the jth corrector magnet to achieve that change. A suitable vector of the N $\Delta\theta$ values which solves this system would then constitute a solution to the steering problem for the M BPMs.

* Work supported by the Department of Energy, contract DE-AC03-76SF00515

The equations are often written in matrix form, which lends itself to solution by numerical methods:

$$\begin{bmatrix} T_{12}^{11} & T_{12}^{12} & T_{12}^{13} & \cdots & T_{12}^{1N} \\ T_{12}^{21} & T_{12}^{22} & T_{12}^{23} & \cdots & T_{12}^{2N} \\ & & \vdots & & \\ T_{12}^{M1} & T_{12}^{M2} & T_{12}^{M3} & \cdots & T_{12}^{MN} \end{bmatrix} \cdot \begin{bmatrix} \Delta\theta_1 \\ \Delta\theta_2 \\ \vdots \\ \Delta\theta_N \end{bmatrix} = \begin{bmatrix} \Delta b_1 \\ \Delta b_2 \\ \vdots \\ \Delta b_M \end{bmatrix}$$

In practice though, there is a barrier constraint; that each $\Delta\theta$ does not result in exceeding the practical maximum strength for that magnet. That is, each element in the unknown vector of $\boldsymbol{\Delta\theta}$ has an upper limit.

So, we could characterize this as a linear least squares problem

$$\|\mathbf{Ax} - \mathbf{b}\|_2 \tag{1}$$

$$\text{subject to } \mathbf{x}_j \le \mathbf{x}_j^{\max}$$

for which a solution $\mathbf{x}$ can be found by least-squares inversion of $\mathbf{A}$ giving $\mathbf{x} = \mathbf{A}^{\dagger}\mathbf{b}$. $\mathbf{b}$ would be the vector of all the desired changes to the BPMs values, $\mathbf{x}$ is the vector of necessary changes in corrector magnet strengths as bend angles, and $\mathbf{A}$ is the matrix of all the ratios between a corrector strength and its concomitant beam monitor value. Specifically, $\mathbf{A}$ is the matrix of the T_{12}s in the system, where each column of $\mathbf{A}$ represents a single corrector.

Posed as a linear system then, the problem lends itself to questions such as "is there a solution $\mathbf{x}$, if so is there a family of solutions, and if there is, which is the "best" solution according to some criteria, if there is no solution, is there at least some optimum compromise?"

These questions are all related to the property of *rank*, the maximal number of linearly independent columns in $\mathbf{A}$, and the ratio of this rank to the number of unknowns in $\mathbf{x}$. The term *singularity* is also often used particularly for square matrices. There are a number of separate theoretical methods to answer these questions, but one, the SVD has become popular because its very robust and is easily applicable to all of these questions.

The SVD though does not respect barrier conditions so its not possible to include the magnet limits in the problem posed to classical SVD algorithms, and it is this drawback we address later.

2 THE SINGULAR VALUE DECOMPOSITION

The SVD technique is based on a factorization of $\mathbf{A}$ which we shall summarize as follows: $\mathbf{A} = \mathbf{U\Sigma V^T}$ where $\mathbf{A}$ is M rows $\times$ N columns, $\mathbf{U}$ is an $M \times M$ column-orthogonal matrix, $\mathbf{\Sigma}$ is an $M \times N$ diagonal matrix of positive or 0 values, and $\mathbf{V}$ is an $N \times N$ column-orthogonal matrix, i.e. $\mathbf{U^TU} = \mathbf{V^TV} = \mathbf{I}$. Since V is square it is also row-

0-7803-4376-X/98/$10.00 © 1998 IEEE

orthogonal.

$$(\mathbf{A}) = (\mathbf{U}) \cdot \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_N \end{pmatrix} \cdot \left(\mathbf{V^T}\right) \quad (2)$$

The elements σ_j are the *singular values* of $\mathbf{A}$. The condition-number of $\mathbf{A}$ is given by the ratio of the largest of the singular values to the smallest. The rank of $\mathbf{A}$ is given by the number of non-0 singular values, and so the nullity (explained more fully below) of $\mathbf{A}$ is given by the number of 0 valued singular values. Given the above properties of orthogonality, the pseudo-inverse of A is given by $\mathbf{A}^\dagger = \mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{U^T}$. $\mathbf{\Sigma}$ is a diagonal matrix, so its inverse is the diagonal matrix of reciprocals of its elements. This then is the simple method of solving a linear system given in eq 1: using the formalism used in Press et al[2]:

$$\mathbf{A}^{-1} = \mathbf{V} \cdot diag(1/\sigma_j) \cdot \mathbf{U}^T \quad (3)$$

The properties of these matrices for diagnosing algebraic problems are well explored, particularly by Golub and Riensch [1]. The interpretation of results under the conditions $M < N, M = N$ and $M > N$ are very practically explained in Press et al, and Strang[3] describes Linear Algebra in general, and its geometric interpretation in particular.

3 THE SOLUTION SPACE

To illustrate our method of finding solutions which respect barrier conditions lets look at the case of the underdetermined system, one in which there are simply less equations than unknowns, $M < N$, or the rank of $\mathbf{A}$ is insufficient to find a single perfect solution. This may have been due simply to there being more correctors than BPMs, or more subtly because of correctors and BPMs being poorly separated in phase space, or one sub-set of correctors having roughly the same influence on the BPMs as some other sub-set. All of these conditions would cause degeneracy in the transformation matrix $\mathbf{A}$.

The SVD will return at least $N - M$ 0 or small σ_j's. There may also be additional 0 σ_j due to rank deficiency. Call the number of 0 singular values k.

One must also set σ_j that are very close to 0 to 0, since those are probably dominated by numerical error. If these values are allowed to remain they will tend to attract the computation in 3 toward a null-space vector. Specifically what constitutes "close to 0", is related to computational precision and accuracy of original data. Guidelines for deciding appropriate cut-off values are given in the literature.

The SVD can easily be used to select a 'particular solution' and it will be the solution which is smallest in the least squares sense: one simply sets all the 0 valued σ (after editing for computational precision) to 0 rather than $1/\sigma_j$.

$$\mathbf{x} = \mathbf{V} \cdot [diag(1/\sigma_j)] \cdot \left(\mathbf{U}^T \cdot \mathbf{b}\right) \quad (4)$$

If we now wanted to look at alternative solutions, perhaps because the particular solution involved exceeding the limit of some corrector, we can look more closely at the geometric interpretations of $\mathbf{U}$ and $\mathbf{V}$. To do this lets make explicit the ideas of range and null-space. If there are alternative solutions $\mathbf{x}$, $\mathbf{A}$ must be singular or $M < N$, and then there must be some sub-space $\mathbf{x}$ for which $\mathbf{A} \cdot \mathbf{x} = \mathbf{0}$. This is the null-space of $\mathbf{A}$, and its dimension is called the "nullity" of $\mathbf{A}$. The space that can be reached by $\mathbf{A}$ multiplied by any $\mathbf{x}$ at all is called the "range" of $\mathbf{A}$. The rank of $\mathbf{A}$ is equivalent to the dimension of this range.

The non-0 singular values define the dimensions of these sub-spaces. Specifically, the columns of $\mathbf{U}$ that correspond to same-numbered non-0 valued elements of $\mathbf{\Sigma}$ form a set of orthonormal basis functions for the range of $\mathbf{A}$, call this matrix $\mathbf{U_1}$. A basis is a set of vectors which spans the same sub-space as the original matrix. It is orthonormal in the sense that it is a set of mutually orthogonal unit vectors, and so makes up a necessary and sufficient description of the sub-space.

The columns of $\mathbf{V}$ that correspond to 0 valued elements of $\mathbf{\Sigma}$ form a set of orthonormal basis functions for the null-space of $\mathbf{A}$, call this matrix $\mathbf{V_0}$.

Using $\mathbf{V_0}$ one can compute alternative solutions by adding to the particular solution $\mathbf{x}$ linear multiples (or "combinations") of columns v_j drawn from $\mathbf{V_0}$, giving $\mathbf{x}'$ and the overall value $\mathbf{Ax}'$ won't be different from $\mathbf{Ax}$.

Taking this process further, how can we find those specific alternative solutions which do not exceed some specific barriers?

4 BOUNDED LEAST SQUARES SEARCH

Recall that the objective is to find a vector $\mathbf{x}$ which minimizes $\|\mathbf{Ax} - \mathbf{b}\|_2$ - a linear least squares problem. This is equivalent to minimizing the RMS of the beam orbit when $\mathbf{A}$ is the "transport matrix" of an accelerator. But the problem is made difficult to solve by the classical linear least squares algorithm implementations such as LSSOL[4] if there is degeneracy in the equations $\mathbf{A}$. Degeneracy significantly compounds the effects of rounding error in numerical computations.

One way to overcome the problem would be to remove the trouble-some null-space from $\mathbf{A}$, and search for solutions $\mathbf{x}$ in the remaining sub-space. The constraints on the solution would also have to be transformed in to the coordinates of that, re-cast, problem. When a solution is found it would be transformed back again into the original coordinates.

To delineate these transformations, define the basis functions that will be used to generate them:

$\mathbf{U_1} \stackrel{\text{def}}{=}$ columns $\mathbf{u_j}$, for which $\sigma_\mathbf{j} \neq \mathbf{0}$: An orthonormal basis for the range of $\mathbf{A}$. $\mathbf{U_1}$ is $M \times N - k$.

$\mathbf{U_0} \stackrel{\text{def}}{=}$ columns $\mathbf{u_j}$, for which $\sigma_\mathbf{j} = \mathbf{0}$: An orthonormal basis for the orthogonal complement of the range of $\mathbf{A}$. $\mathbf{U_0}$ is $M \times k$.

$\mathbf{V_0} \overset{\text{def}}{=}$ columns $\mathbf{v_j}$, for which $\sigma_\mathbf{j} = \mathbf{0}$: An orthonormal basis for the null-space of $\mathbf{A}$. $\mathbf{V_0}$ is $N \times k$.

$\mathbf{V_1} \overset{\text{def}}{=}$ columns $\mathbf{v_j}$, for which $\sigma_\mathbf{j} \neq \mathbf{0}$: An orthonormal basis for the orthogonal complement of the null-space of $\mathbf{A}$. $\mathbf{V_1}$ is $N \times N - k$.

$\mathbf{\Sigma_1} \overset{\text{def}}{=} \sigma_\mathbf{j}$, for which $\sigma_\mathbf{j} \neq \mathbf{0}$: The extent of each dimension of the range of $\mathbf{A}$. $\mathbf{\Sigma_1}$ is a square diagonal matrix of non-0 singular values and is $N - k \times N - k$.

Then define the orthogonal sub-space of $\mathbf{A}$:

$$\mathbf{A_1} \overset{\text{def}}{=} \mathbf{U_1 \Sigma_1} \tag{5}$$

The solution vector being sought must similarly be interpreted in the coordinates of orthogonal complement of the null-space of $\mathbf{A}$:

$$\mathbf{x_1} \overset{\text{def}}{=} \mathbf{V_1^T x} \tag{6}$$

$\mathbf{x_1}$ will be $N - k \times 1$.

Then the minimization can be re-written to exclude the null-space:

$$\begin{aligned} & \|\mathbf{Ax} - \mathbf{b}\|_2 \\ \iff & \|\mathbf{U_1 \Sigma_1 V_1^T x} - \mathbf{b}\|_2 \\ \iff & \|\mathbf{A_1 x_1} - \mathbf{b}\|_2 \end{aligned} \tag{7}$$

This is then a least squares problem of smaller dimension - those dimensions in the null-space of $\mathbf{A}$ have been removed.

To incorporate the barriers on the original problem, $\mathbf{x_j} \leq \mathbf{x}^{\max}$, where $\mathbf{x}^{\max}$ are the limits on changes to the corrector magnet settings, we need to pose those barriers also in the recast coordinates. From the interpretation of $\mathbf{V_1}$, and the definition of $\mathbf{x_1}$ given in 6, and since the inverse of an orthogonal matrix is its transpose, then

$$\mathbf{x} = \mathbf{V_1 x_1} \tag{8}$$

$\mathbf{V_1}$ then is the matrix whose rows each define a general linear constraint on $\mathbf{x_1}$. Each corresponding row vector multiplication $\mathbf{V_{1j} x_1}$ must not exceed $x_j^{\max}$.

Altogether, the linear least squares problem in the recast coordinate space is to minimize:

$$\|\mathbf{A_1 x_1} - \mathbf{b}\| \tag{9}$$

$$\text{subject to } \mathbf{V_1 x_1} \leq \mathbf{x}^{\max}$$

This can be solved by any linear least squares solver that accepts a linear constraints matrix as part of the problem parameters, such as LSSOL. When some $\mathbf{x_1}$ is found, it can be transformed back into regular coordinates by 8.

5 PERFORMANCE

Both the SVD and the Linear Least Squares method for non-singular matrices are known to be very robust.

Although the decomposition operation itself is fairly expensive, it need at least only be computed once and then all orbit corrections using the same transport matrix can proceed from it.

We do not submit the recast problem to a constrained linear least squares solver in the case that the SVD solution, by 4, finds solutions that are in bounds, since the minimum solution has already been found. Also, the linear least squares technique in the case of negligible degeneracy is equivalent to the SVD when all of the singular values are used.

The hybrid method has been tested in simulation and been in operation in the SLC linac for some time with success. It was developed for the B-factory, which will be commissioned in the summer of '97, when heavier requirements will be made of its ability to deal with degenerate matrices. It is also part of a larger project in which dispersion is corrected simultaneously with orbit [6].

6 THANKS

Our very warm thanks are extended to Prof. SVD, Gene Golub for pointing us down the hybrid path, and to Jeff Corbet and Martin Lee. And Haitham Hindi, who is to formal argument what a buzz-saw is to a match-stick.

7 REFERENCES

[1] Golub G., Reinsch C. Singular Value Decomposition and Least Squares Solutions, in Wilkinson J.H. and Reinsch C. 9 (editors) Handbook for Automatic Computation, vol II. Springer. 1971.

[2] Press W.H., Flannery B.P, Teukolsky S.A., Vetterling W.T. Numerical Recipes, The Art of Scientific Computing. Cambridge University Press. 1989.

[3] Strang, G Linear Algebra and its Applications. Harcourt Brace Jovanovich. 1988.

[4] Gill P. E., Hammarling S.J., Murray W., Saunders M.A., Wright M.H., LSSOL, Systems Optimization Laboratory, Stanford University. 1986.

[5] Sheppard J. C., Lee M. J., Ross C., Seeman J. T., Stiening R. F., Woodley M. D. Beam Steering in the SLC Linac, Proceedings of PAC 1985.

[6] Chi Y., Donald M., Shoaee H., White G., Yasukawa L.A. An Orbit and Dispersion Correction Scheme for the PEP-II. Proceedings of PAC 97.