SVD BASED ORBIT CORRECTION INCLUDING CORRECTOR LIMITATIONS AT DELTA

Marc Grewe*, Peter Hartmann, Gerald Schmidt, Klaus Wille

University of Dortmund, Institute of Accelerator Physics and Synchrotron Radiation
Maria-Goeppert-Mayer Str. 2, 44221 Dortmund, Germany

Abstract

Singular Value Decomposition (SVD) of the orbit response matrix has become an invaluable tool for orbit correction at storage rings worldwide. SVD based orbit correction has now been realised at DELTA, a 1.5 GeV synchrotron light source[1]. However, due to certain orbit demands at DELTA, we frequently have to face corrector limitations during the process of orbit correction. This work focuses on presenting an analytic algorithm on how to treat physical current limitations when seeking for a SVD based orbit correction. In contrast to previously published methods, this approach is fairly easy to implement and does not afford a numerical solver. Concepts and results are presented.

INTRODUCTION

An applied dipole kick ϑ_j at a position j to a beam in a storage ring will yield in a stationary orbit distortion u_i at another position i with $u \in [x, z]$. Neglecting a change of twiss functions or tune for moderate ϑ_j as well as higher order multipole fields in the vicinity of the beam, the orbit response matrix (ORM) $\mathbf{R}$ composed of columns u_i for each corrector j is to be considered constant and independent of actual orbit offsets $\vec{u}$. Thus, the resulting orbit offset $\vec{u}_0$ due to a set of dipole kicks $\vec{\vartheta}$ is given by the linear superposition

$$\vec{u}_0 = \mathbf{R}\vec{\vartheta}. \tag{1}$$

The task of orbit correction is to find a suitable set of corrector kicks $\vec{\vartheta}$ which will result the inverse of a measured orbit deviation to a sought orbit reference $\vec{u}_0 = \vec{u}_{\mathrm{ref}} - \vec{u}$. For the special case of a non-singular, square matrix $\mathbf{R}$ this is accomplished by multiplying equation (1) with $\mathbf{R}^{-1}$ from the left to receive the solution $\vec{\vartheta} = \mathbf{R}^{-1}\vec{u}_0$. However, in general $\mathbf{R}$ is neither squared nor 'well-conditioned', as will be explained later. For those cases SVD helps to design a 'pseudo-inverse' $\mathbf{R}^\sharp$ so that the sought solution vector takes the form

$$\vec{\vartheta}' = \mathbf{R}^\sharp\vec{u}_0, \tag{2}$$

with $\chi^2 = (\vec{u}_0 - \vec{u}'_0)^2$ becoming a minimum for $\vec{u}'_0 = \mathbf{R}\vec{\vartheta}'$ within the range($\mathbf{R}$). Yet, under practical circumstances, this may not be the whole story. In the need of local orbit bumps with sufficiently large orbit offsets and small phase advances between the dipole correctors to be used, the desired orbit kick afforded by some dipole correctors $\{k\}$ may exceed their physical current limitation. However, the above solution does not include these side conditions, thus rendering the gained solution useless.

A common technique to overcome the stated problem is to successively pick any corrector k, which is to be set beyond its limit, truncate its share to the solution to its limiting value and subtract the resulting orbit offset from the sought orbit difference. Then, equation (2) is solved again for the remaining correctors by setting the corresponding column k of $\mathbf{R}$ to zero. This procedure is iterated until no more correctors exceed their limits. Although this method will yield a physically feasible corrector setting, we observed, that it is not unusual to end up with a solution that might deteriorate the present orbit, rather than improving it. This may be the consequence of the algorithm itself, mostly due to a lack of an appropriate criterion to decide upon the correct order of corrector limitations. However, even for a prediction to improve the orbit, the application of the corresponding corrector setting may in fact worsen the orbit or even result in beam loss. This phenomenon was identified as being a result of the inherent disregard of a potentially error dominated null space (see below).

An alternative approach to overcome these problems was to suggest a 'numerical hybrid method' [2]. This method employs the analysis of SVD in a first step to clear the set of linear equations of possibly existent singularities. The restated problem is then fed into a numerical solver, capable of respecting the side conditions stated. While this method will certainly lead to a satisfying solution, it remains to be of an opaque numerical nature.

Here, we propose a transparent approach yielding a solution with analytic accuracy according to the requirements stated below.

SVD ANALYSIS

Singular value decomposition states a theorem of linear algebra, according to which any $M \times N$ matrix $\mathbf{A}$ may be decomposed into the product of three matrices:

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{W} \cdot \mathbf{V}^T.$$

The matrix $\mathbf{U}$ is of the same geometry as $\mathbf{A}$, consisting of orthonormal column vectors $\vec{U}_j$. $\mathbf{W}$ is a square $N \times N$ diagonal matrix, whose diagonal elements w_j are called 'singular values'. The $N \times N$ square Matrix $\mathbf{V}$ is orthonormal, such that $\mathbf{V} \cdot \mathbf{V}^T = 1$. Taking the ORM $\mathbf{R}$ to be decomposed this way, it is easy to see that the N columns $\vec{V}_j$ of $\mathbf{V}$ form an orthonormal basis V of the N-dimensional corrector space, which, when multiplied to $\mathbf{R}$, result in an orbit

* marc.grewe@delta.uni-dortmund.de

$w_j\vec{U}_j$. Thus, the columns $\vec{U}_j$ describe an orthonormal N-dimensional orbit base U, linked to V by the metric tensor $\mathbf{W}$. Obviously, those base vectors $\vec{V}_j$, whose singular values w_j are close to zero, represent corrector combinations, whose effect on the monitored orbit is very small. If such singular values exist, the matrix is referred to as being 'ill-conditioned' or 'singular' if $w_j = 0$ for any j. Now, given an orbit difference $\vec{u}_0$ to be corrected, we may project this vector into U to receive $\vec{U}_0 = \vec{U}_0'$. By multiplying each component j of $\vec{U}_0$ by either its reciprocal metric factor w_j^{-1} for $w_j > w_{\rm cut}$ or zero otherwise,we get the corresponding representation $\vec{V}'$ in V, excluding the chosen null space by definition of $w_{\rm cut}$. Using the basis representation $\vec{V}_j$, $\vec{V}'$ is then translated into real corrector settings $\vec{\vartheta}'$ required to reproduce $\vec{u}_0'$ in real orbit space. This procedure is easily represented in Matrix form in a more general way:

$$\begin{aligned}\vec{\vartheta}' &= \mathbf{V}\cdot\vec{V}' \\ &= \mathbf{V}\cdot[\mathrm{diag}(1/w_j)]_{w_{\rm cut}}\cdot\vec{U}_0 \\ &= \mathbf{V}\cdot[\mathrm{diag}(1/w_j)]_{w_{\rm cut}}\cdot\mathbf{U}^T\cdot\vec{u}_0 \\ &= \mathbf{R}^\sharp\cdot\vec{u}_0\end{aligned}$$

The last line is to be understood as a definition for $\mathbf{R}^\sharp$. '$[\mathrm{diag}(1/w_j)]_{w_{\rm cut}}$' denotes a $N\times N$ diagonal matrix with diagonal elements $1/w_j$ for $w_j > w_{\rm cut}$ or zero otherwise.

CORRECTOR LIMITATIONS

Obviously, corrector limitations restrict the available corrector space, thereby the accessible orbit range. For the purpose of visualisation of the geometries involved, we shall have a look at a two dimensional example assuming two correctors to be used for orbit correction, see figure 1. Their orbit responses $\vec{R}_{1,2}$ may have a higher dimension-

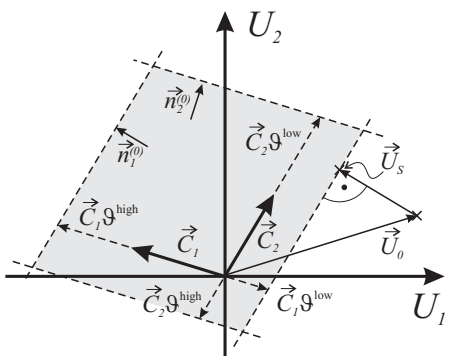

Figure 1: Geometry of corrector limitations in the orthonormal range U of $\mathbf{R}$

ality than two, but their representation $\vec{C}_{1,2}$ in U is again two dimensional. Given their allowed span of kicks within $\vartheta_{1,2}^{\rm low} \le \vartheta_{1,2} \le \vartheta_{1,2}^{\rm high}$, the physically accessible domain forms a diamond in two dimensions (shaded in figure 1), and a hyper dimensional parallelepiped in general. Note, that the origin of U is defined by the actual corrector setting, not by the actual orbit. The stated problem arises, once $\vec{U}_0$ comes to lie outside this diamond, so that one or both correctors are incapable to access this point. The best one can do instead is to find the point closest to $\vec{U}_0$ on the surface of the diamond – or the parallelepiped in general.

The Algorithm

We shall call the sought surface point $\vec{U}_S$. The set of limiting hyper planes, which are exceeded by an orbit point $\vec{U}^{(m)}$ shall be called $K(m)$ for a given step of iteration m. The proposed algorithm assumes

1. $\vec{U}_S$ to lie within at least one plane in $K(m)$ for any m, unless $K(m)$ is the empty set;
2. that the plane $k(m)\in K(m)$ with the largest Euclidian distance $\Delta_{k(m)}^{(m)}$ to $\vec{U}^{(m)}$ is such a plane:

$$\left|\Delta_{k(m)}^{(m)}\right| \ge \left|\Delta_j^{(m)}\right| \qquad \forall j\in K(m). \tag{3}$$

The iteration starts off using $\vec{U}^{(0)}$ as a starting point (see below). Each step m of the iteration will project $\vec{U}^{(m)}$ perpendicularly into the plane $k(m)$. The problem is then restated within this reduced dimensionality until $\vec{U}^{(m)} = \vec{U}_S$. Figure 2 illustrates this reasoning in two dimensions: for

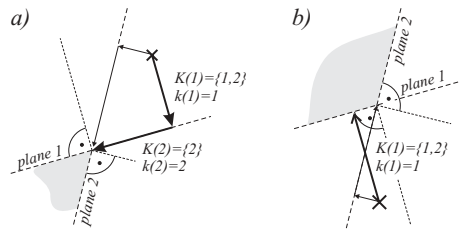

Figure 2: Two dimensional examples to illustrate the path of iteration at sharp and obtuse angles. Bold arrows denote the path chosen by the iteration, semi-bold arrows that of a possibly mistaken path, yielding either to the same or an objectionable point. See text.

the case of a sharp angle (a), the order of projections is irrelevant, whereas (b) reveals areas for an obtuse angle, where the order of projection is indeed of relevance: starting off beyond both planes 1 and 2, projection into plane 1 immediately yields the desired surface point, whereas a projection into plane 2 with a successive projection into plane 1 yields in an objectionable point.

This algorithm may also be applied to a weighted set of equations:

$$\mathbf{G}\cdot\mathbf{R}\vec{\vartheta} = \mathbf{G}\vec{u}_0, \tag{4}$$

with a diagonal $M\times M$ weight matrix $\mathbf{G}$. U is still isometric as required by the geometric methods to follow.

Implementation

The following approach assumes that

$$\mathrm{rank}(\mathbf{R}) = N, \tag{5}$$

hence all corrector representations $\vec{C}_j$ are linearly independent for all $j = 1 \ldots N$. From computational considerations this is typically given for $w_j > 10^{-18}\,\forall j$. Each limiting plane j is then uniquely defined by its normal vector $\vec{n}_j^{(0)}$ and its distance $d_j^{(0)}$ to the origin. $\vec{n}_j^{(0)}$ is characterized by being orthogonal to all other corrector representations $\vec{C}_{i\neq j}$, and enclosing some angle $\neq \pm\pi/2$ with $\vec{C}_j$. Because of (5), this is a squared set of non-singular linear equations with a unique solution, which is to be solved for a vector $\vec{n}_j'^{(0)}$ pointing into the same direction as $\vec{n}_j^{(0)}$. We then receive the sought normal vector by normalisation:

$$\begin{aligned} \vec{n}_j'^{(0)} &= \left(\mathbf{R}^T\mathbf{U}\right)^{-1}(0,\ldots,0,1,0,\ldots,0)^T \\ \rightarrow \vec{n}_j^{(0)} &= \vec{n}_j'^{(0)} \cdot \left|\vec{n}_j'^{(0)}\right|^{-1}. \end{aligned}$$

Using the known space point $\vartheta_j^{\lim}\vec{C}_j$ within this plane, we obtain $d_j^{(0)} = \vartheta_j^{\lim}\vec{n}_j^{(0)}\vec{C}_j$, with the index 'lim' corresponding to the upper or lower limit under consideration. In order to obtain a starting point within the *chosen range* of $\mathbf{R}$ which is spanned by the vector set $\vec{U}_i$ with $w_i > w_{\rm cut}$, we simply zero those components i of $\vec{U}_0$ for which $w_i < w_{\rm cut}$. Let $\vec{U}^{(1)}$ be the result of this action. For each step $m \geq 1$ of the iteration, $k(m)$ is determined by evaluation and comparison of the distances $\Delta_j^{(m)}$ *within the intersection of the chosen range with previously projected planes*. To do so, we need to reduce the normal vectors of planes under investigation by the collateral dimensions $N^{(m)}$ spanned by the vector sets $\{\vec{U}_i | w_i < w_{\rm cut}\} \otimes \{\vec{n}_{k(n)} | 1 \leq n < m\}$. Since these vectors are generally skewed, we need to create an orthogonal basis of $N^{(m)}$ first. Again, SVD is used to decompose a matrix $\mathbf{N}^{(m)}$ made up of the column vectors spanning $N^{(m)}$. The resulting matrix $\mathbf{U}_N^{(m)}$ is then used to project the normal vectors into the remaining subspace, so that the sought distances are found to be (see figure 3):

$$\begin{aligned} \vec{n}_j^{(m)} &= \left(\vec{n}_j^{(0)} - \mathbf{U}_N^{(m)}\left(\mathbf{U}_N^{(m)}\right)^T\vec{n}_j^{(0)}\right)\frac{1}{|\ldots|} \\ \Delta_j^{(m)} &= \frac{\Delta_j^{(0)}}{\cos\theta_j^{(m)}} = \frac{\vec{U}^{(m)}\vec{n}_j^{(0)} - d_j^{(0)}}{\vec{n}_j^{(m)}\vec{n}_j^{(0)}} \end{aligned} \qquad (6)$$

The periods $(\ldots)$ within the bars replace the parenthesized term left to it. By equation (5), $\cos\theta_j^{(m)}$ is guaranteed to be non-zero. $k(m)$ is then determined according to equations (3) and (6). Ultimately, $\vec{U}^{(m)}$ is projected orthogonally into $k(m)$:

$$\vec{U}^{(m+1)} = \vec{U}^{(m)} - \vec{n}_{k(m)}^{(m)}\Delta_{k(m)}^{(m)}$$

The iteration ends, once $K(m)$ is the empty set.

While it is seldom to occur that an ORM becomes truly singular[1], certain situations exist, such that the matrix under consideration will become numerically singular[2]. Instead of rendering the above procedure useless, it proved an easy procedure to simply add some noise to the matrix elements; just enough to guarantee the necessary precision of the result but lifting the smallest singular values above $\sim 10^{-18}$, depending on the numerical stability of the computational implementation.

[1] since usually the number of evaluated monitor readings M is equal or greater than the number N of correctors to be used

[2] E.g., the columns of the matrix might be constructed such, that they already fulfill side conditions, such as the closeness of a local orbit bump.

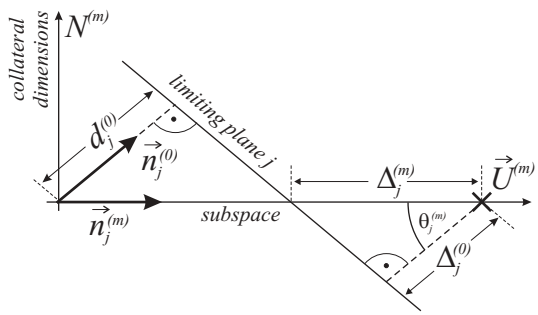

Figure 3: Determination of the distance $\Delta_j^{(m)}$ of $\vec{U}^{(m)}$ to a plane j within a subspace of U.

Note, that for the special case of nullity$(\mathbf{R}) = 0$ ($\Leftrightarrow w_{\rm cut} < w_j, \forall j$) *and* a homegeneous weight matrix $\mathbf{G} = \lambda\mathbf{1}$ (λ being some real number), this algorithm may also be carried out in the real orbit space u instead of U.

RESULTS

The above procedure has been implemented into the SVD based orbit correction at DELTA with great success. Since at DELTA the ORM is overdetermined for each plane, it is reasonable to add weights to relevant monitors. Thus, the condition number as the ratio of the largest to the smallest singular value is easily raised to about 10^6, such that the epiped in U becomes highly distorted due to the metric of singular values. In contrast to the method of successive corrector limitations as stated in the introduction, the use of the presented algorithm predicts a reduction of the orbit deviation $\vec{u}_0$ for the vast majority of tested circumstances, while strictly obeying the restriction of the null space. As a result, orbit correction has become very stable, even with up to ~ 10 correctors at their limits under test conditions (usually up to ~ 3 under real conditions). However, by successive increase of the chosen rank($\mathbf{R}$), the distance of $\vec{U}^{(0)}$ to $\vec{U}_S$ does not always decrease for a given $\vec{u}_0$, as indeed it should. At the moment it is not understood, wether these observations are a matter of the implementation or the stated assumptions regarding $\vec{U}_S$ not applying to the closest point on the epiped to $\vec{U}^{(0)}$.

REFERENCES

[1] D. Schirmer et al., "Status of The Synchrotron Light Source DELTA", EPAC'04, Lucerne, 2004

[2] G. White et al., "A Hybrid Numerical Method for Orbit Correction", PAC'97, Vancouver, 1997