METHODS OF ORBIT CORRECTION SYSTEM OPTIMIZATION*

Yu-Chiu Chao, Thomas Jefferson National Accelerator Facility

Abstract

Extracting optimal performance out of an orbit correction system is an important component of accelerator design and evaluation. The question of effectiveness vs. economy, however, is not always easily tractable. This is especially true in cases where betatron function magnitude and phase advance do not have smooth or periodic dependencies on the physical distance. In this report a program is presented using linear algebraic techniques to address this problem. A systematic recipe is given, supported with quantitative criteria, for arriving at an orbit correction system design with the optimal balance between performance and economy. The orbit referred to in this context can be generalized to include angle, path length, orbit effects on the optical transfer matrix, and simultaneous effects on multiple pass orbits.

1 INTRODUCTION

In designing orbit correction systems simple rules such as phase advance counting are often followed in placing monitors and correctors. While such rules work well with smooth, periodic betatron functions and phase advances, one may encounter difficulty using them in areas where large betatron variations contribute significantly to response matrix elements compared to pure phase contributions, or in areas where smooth periodicity is absent. When these problems are present, it is difficult to arrive at a *global and quantitative* design criterion for the orbit correction system based on phase advance counting. Here we present a self consistent program based on response matrices. It ensures the *globally consistent* application of the same quantitative criteria for observability, controllability and non-degeneracy defined by the designer, independent of the smoothness of the local lattice.

The program starts by evaluating the observability of the monitor system to ensure knowledge of the orbit to the same level everywhere. An algorithm for adding monitors is introduced in case of deficiency in observability. The redundancy of the monitor system is then evaluated and an algorithm for monitor minimization introduced to ensure that a minimally necessary set is obtained that would not place unjustified demands on the corrector system. We follow by evaluating the controllability of the corrector system to ensure control of the orbit to the same level everywhere. An algorithm for adding correctors is introduced in case of deficiency in controllability. The overall redundancy of the corrector system is then evaluated and an algorithm for corrector minimization introduced to ensure that a minimally necessary set is obtained that would not lead to large local orbits due to correction using near-singular response matrices. The application of this program to the CEBAF accelerator, where localized deficiencies in monitors and overall redundancies in correctors have been identified and corrected, will be described.

* Work supported by the U.S. Department of Energy, contract DE-AC05-84ER40150.

2 NOMENCLATURE

For simplicity we limit our discussion to the x-plane only with the usual index assignments of 1, 2, 6 for position, angle and momentum. Generalization is straightforward.

2.1 Error-to-monitor response matrix M^{EM}

The error-to-monitor response matrix M^{EM} summarizes the orbit disturbance at any monitor caused by any physical error which can affect any of the beam orbit coordinates:

$$O_i = \sum_j M_{ij}^{EM} \bullet E_j$$

where O_i is the orbit disturbance at the i-th monitor and E_j the magnitude of the j-th physical error, including injection errors, magnetic field errors, misalignments etc.. The elements of M^{EM} consist of optical transfer elements M_{11}, M_{12}, and M_{16} from the sources of error to the monitors. In constructing M^{EM} for subsequent analysis, one must identify all the major potential sources of errors that the entire orbit correction system is designed to correct. This usually includes quadrupole offsets, large dipole field errors, suspected misalignments etc.. Any estimate on the relative magnitude of such errors can be incorporated into M^{EM} by properly scaling individual columns.

2.2 Error-to-all-location response matrix M^{EA}

The error-to-all-location response matrix M^{EA} summarizes the orbit disturbance at all representative locations caused by the physical errors described above. These locations, not tied to any physical elements, should effect a dense coverage of the entire beam line and will be collectively denoted by a set C_A.

$$O_i = \sum_j M_{ij}^{EA} \bullet E_j$$

where O_i is the orbit disturbance at the i-th location.

2.3 Corrector-montior response matrix M^{CM}

The corrector-monitor response matrix M^{CM} summarizes the orbit disturbance at any monitor by any corrector.

$$O_i = \sum_j M_{ij}^{CM} \bullet K_j$$

where K_j is the magnitude of the j-th corrector kick.

0-7803-4376-X/98/$10.00 © 1998 IEEE

2.4 All-location-to-montior response matrix M^{AM}

The all-location-to-monitor response matrix M^{AM} summarizes the orbit disturbance at any monitor caused by coordinate error at any representative location in the set C_A.

$$O_i = \sum_j M_{ij}{}^{AM} \bullet K_j$$

where K_j is the magnitude of the error at the j-th location.

2.5 Corrector-to-all-location response matrix M^{CA}

The corrector-to-all-location response matrix M^{CA} summarizes the orbit disturbance at all representative locations caused by any corrector.

$$O_i = \sum_j M_{ij}{}^{CA} \bullet K_j$$

2.6 Singular value decomposition (SVD)

SVD is the process of decomposing a matrix M into the product of three matrices U, W and V:

$$M = U^T \bullet W \bullet V$$

U and V afford useful physical interpretation when applied to response matrices. The rows of V represents orthonormal combinations of the "actuators", either errors or correctors, whose effects are magnified by the diagonal elements of W before being realized as orthonormal orbit patterns represented by the rows of U. SVD allows us to decompose the response matrix into decoupled cause-effect relations between linear combinations of the actuators and monitors, with the magnification factors contained in W. The diagonal elements of W are called singular values and the condition number N_M^{svd} of M is the ratio between the largest and the smallest singular values.

2.7 Null space vectors

The null space vectors E_M for a given matrix M are the vectors which are projected into 0 by M:

$$M \bullet E_M = 0, \qquad |E_M| = 1$$

Notice we choose to have all E_M's normalized.

2.8 Pseudoinverse and projected components

The pseudoinverse $M^{\dagger}$ of a given matrix M is defined as

$$M^{\dagger} = (M^T \bullet M)^{-1} \bullet M^T$$

The pseudoinverse is related to the projection operator Π_M, which decomposes any vector X into components X^M and $\overline{X}^M$ respectively inside and outside the subspace spanned by the column vectors of M, through

$$\Pi_M = M \bullet M^{\dagger},$$

$$V^M = \Pi_M \bullet X, \qquad P_M^X = |X^M| / |X|$$

$$\overline{X}^M = X - \Pi_M \bullet X, \qquad Q_M^X = |\overline{X}^M| / |X|$$

where we have also defined the fractional components $P^X{}_M$ and $Q^X{}_M$ of X inside and outside the subspace spanned by M.

2.9 Gram determinant and orthogonality

The Gram determinant G_M of a matrix M is given by

$$G_M = \mathrm{Det}(M^T \bullet M) \qquad \text{row dim.} > \text{column dim.}$$

$$= \mathrm{Det}(M \bullet M^T) \qquad \text{column dim.} > \text{row dim.}$$

$$G_M = \left(\prod_j S^M{}_j\right)^2$$

where $S^M{}_j$ is the j-th singular value of M. The normalized Gram determinant $\overline{G}_M$ is defined as

$$\overline{G}_M = G_M / L_M, \qquad 0 \le \overline{G}_M \le 1$$

$$L_M = \prod_j \left(\sum_i M_{ij}{}^2\right) \qquad \text{row dim.} > \text{column dim.}$$

$$= \prod_i \left(\sum_j M_{ij}{}^2\right) \qquad \text{column dim.} > \text{row dim.}$$

3 THE OPTIMIZATION PROGRAM

In the following we outline the entire optimization program using the quantitative measures defined in the previous section. It should be noted that the general philosophy of accelerator design demands the following numerology to hold: $N_E > N_M > N_C$, where N_E, N_M and N_C are the total number of potential errors, monitors and correctors respectively. Thus the matrix M^{EM} always has more columns than rows and the opposite is true for M^{CM}. One can start the program with an arbitrary initial monitor-corrector configuration and iterate until all criteria are satisfied. A set of candidate locations for monitors and correctors should be identified, for example at all quadrupole locations, in case additional monitors or correctors are demanded in an iteration. These sets will be denoted C_M and C_C in the following. We will also denote by C_A the set of all representative locations used for establishing M^{EA}. Various cutoff numbers will be used for terminating iterations. Their physical meaning will be briefly described, but not quantitatively elaborated.

3.1 Eliminating monitor deficiency

1. Determine cutoff number R in units of orbit displacement, a measure of the error-induced orbit anywhere that is undetectable by existing monitors.

2. Obtain null space vectors $E_{M^{EM}}$ of M^{EM}, calculate $V^A = M^{EA} \bullet E_{M^{EM}}$ for all $E_{M^{EM}}$.

3. If any element of any V^A is greater than R, identify index j of the largest such element in V^A.

4. Add to monitor list the candidate monitor in C_M closest to the location represented by the j-th location in C_A.

5. Iterate steps 2-4 until all elements of V^A are less than R.

6. Perform SVD on M^{EM}, obtain the row vector v of V with the smallest singular value, calculate $V^A = M^{EA} \bullet v$.
7. Iterate steps 3, 4 and 6 until all elements of V^A are less than R.

3.2 Minimizing monitor redundancy

1. Determine cutoff numbers R and S with 0 < R < 1, 0 < S < 1. R is a measure of the extent to which all error-induced orbits contribute to a single monitor, and S is a measure of the orthogonality of the monitors.
2. Calculate all N_M fractional components $Q^{X_i}_{M^{EM}}$, with X_i the vector representing unit orbit peak at the i-th monitor.
3. Eliminate all monitors whose corresponding $Q^{X_i}_{M^{EM}}$ exceed (1-R).
4. Calculate the normalized Gram determinants $\bar{G}_{M^{EM}}$, continue if it is less than S to the N_M-th power.
5. Perform SVD on M^{EM}, obtain the row vector u of U with the smallest singular value, identify the largest component of u and its index i.
6. Eliminate the i-th monitor.
7. Iterate steps 4-6 until $\bar{G}_{M^{EM}}$ is greater than S to the N_M-th power.

3.3 Eliminating corrector deficiency

1. Determine cutoff number R and S, 0 < R < 1. R measures the fraction of an error-induced orbit pattern uncorrectable by the correctors. S measures corrector limits.
2. Perform SVD on M^{EM}, obtain all row vectors u of U.
3. Calculate all N_M fractional components $Q^{u_i}_{M^{CM}}$, and all N_M pseudo-inverted vectors $K_i = M^{CM\dagger} \bullet u_i$ with u_i the i-th row vector of U. Identify the maximum K_i^{max} of each K_i.
4. Continue if any $Q^{u_i}_{M^{CM}}$ is greater than R, or any K_i^{max} is greater than S. In the former case identify the u_i with the largest $Q^{u_i}_{M^{CM}}$, calculate $T_i = u_i - \Pi_{M^{CM}} \bullet u_i$. In the latter identify the u_i with the largest K_i^{max} and set $T_i = u_i$.
5. Calculate the normalized inner product between Ti and all the column vectors of M^{AM}. Identify index j with the largest inner product.
6. Add the candidate corrector in C_C closest to the location represented by the j-th location in C_A.
7. Iterate steps 2-6 until all $Q^{u_i}_{M^{CM}}$ are less than R and all K_i^{max} are less than S.

3.4 Minimizing corrector redundancy

1. Determine cutoff numbers R and S, R measures the evenness in the corrector effect distribution among monitors. S with 0 < S < 1 measures the orthogonality of the corrector effects on the monitors.
2. Identify correctors forbidden from removal.
3. Perform SVD on M^{CM}, if the condition number $N^{svd}_{M^{CM}}$ is greater than R, or the normalized Gram determinants $\bar{G}_{M^{CM}}$ is less than S to the N_C-th power, continue.
4. Identify the row vector v of V with the smallest singular value and the index j of the largest element in v.
5. If the j-th corrector is not forbidden, remove it. If it is, remove the largest non-forbidden corrector in v.
6. Monitor $Q^{u_i}_{M^{CM}}$ and K_i^{max} defined in the previous program relative to their respective cutoff numbers to ensure freedom from deficiency.
7. Iterate steps 2-6 until $N^{svd}_{M^{CM}}$ is less than R and $\bar{G}_{M^{CM}}$ is greater than S to the N_C-th power.

3.5 Alternative to corrector redundancy

A more advantageous alternative to the last corrector reduction program aimed at eliminating excessive orbit correction caused by near-degeneracy is to introduce “virtual monitors” which automatically keep the correction result confined and free of singular behavior. The algorithm for adding virtual monitors is discussed in [1].

4 APPLICATION TO CEBAF ACCELERATOR

4.1 Monitor deficiency

The program 3.1 for eliminating monitor deficiency was applied to the existing set of BPM’s in CEBAF. It was discovered that all elements V_A are within a limit of 3 mm, with the exception in the East Extraction Region with the elements of V_A exceeding 15 mm for all beam passes, causing orbit excursions undetectable from available data. This is supported by simulation and operation data. It was determined that additional BPM’s be installed according to this program as the potential orbit error can cause emittance distortion on the order of 10% due to suspected higher order field in nearby dipoles.

4.2 Corrector redundancy

A corrector reduction program at CEBAF was performed based on the program 3.4 above. There have been operation and simulation evidences that an overly dense coverage of the beam line by correctors led to excessive correction in the lower arcs and poor reproducibility in the spreaders and recombiners. 66 correctors were removed from a total of about 860 while the corrector deficiency criteria were monitored at each step to prevent over-reduction. The machine has been operating with this reduced corrector set and no compromise in orbit correctability has been observed.

5 REFERENCES

[1] ‘Orbit Correction Using Virtual Monitors at Jefferson Lab’, Y. Chao et al, these proceedings.