EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH
CERN - SL DIVISION

OPTIMIZATION OF ORBIT CORRECTION SYSTEMS USING GENERALIZED RESPONSE MATRICES AND ITS APPLICATION TO THE LHC INJECTION TRANSFER LINES

Yu-Chiu Chao[†], Volker Mertens

Abstract

The LHC injection transfer lines will transport intense high-energy beams over considerable distances. Their relatively tight apertures require precise control of the trajectory. An analytic method has been developed to optimize the performance, reliability and cost of these lines. This method invokes a comprehensive set of performance criteria, makes detailed predictions on orbit correction effectiveness, configuration defects and critical elements, and gives indications for improvement. A path approaching an optimized configuration can thus be established based on well-defined quantitative criteria.

Geneva, Switzerland
August 2001

OPTIMIZATION OF ORBIT CORRECTION SYSTEMS USING GENERALIZED RESPONSE MATRICES AND ITS APPLICATION TO THE LHC INJECTION TRANSFER LINES

Yu-Chiu Chao[†], Volker Mertens, CERN, Geneva, Switzerland

Abstract

The LHC injection transfer lines will transport intense high-energy beams over considerable distances. Their relatively tight apertures require precise control of the trajectory. An analytic method has been developed to optimize the performance, reliability and cost of these lines. This method invokes a comprehensive set of performance criteria, makes detailed predictions on orbit correction effectiveness, configuration defects and critical elements, and gives indications for improvement. A path approaching an optimized configuration can thus be established based on well-defined quantitative criteria.

1 OVERVIEW

This report describes an analytic optimization program developed for the LHC transfer lines TI 2 and TI 8, currently under construction at CERN [1]. By establishing response matrices for generalized actuator and responder spaces that completely characterize an orbit correction system, most performance issues can be cast in the form of mapping properties between such spaces with a global and probabilistic significance. Analyzing the structure of these matrices further provides a systematic path for optimization. All methods developed are implemented through rigorous linear recipes, as this is the only way to realize intuitive pictures on performance criteria with efficient and robust numerical methods. In the following the construction of the generalized response matrices and various error distributions, the recipes for mapping and solutions, and structural analyses of the response matrices will be given, followed by a description of the optimization program and its application to the LHC transfer lines. Further details can be found in [2].

2 GENERALIZED ACTUATORS AND RESPONDERS

Besides physical actuators (correctors) and responders (BPM's), the same concept is extendable to elements such as errors and unmonitored locations [3]. Analyzing the structure of the response matrices linking this extended set of actuators and responders, as given in Table 1, can provide quantitative predictions on the impact of errors, a higher level of error interpretation and better control through configuration changes. The matrix $\mathbf{M^{MM}}$ formalizes monitor errors as actuators in orbit correction.

With errors treated as generalized actuators, one can formulate most of the performance criteria on a probabilistic basis, by defining overall error distributions according to machine and operation parameters. The

Table 1. Generalized response matrices.

	Generalised actuator	Generalised responder	Response coefficients
CM	$\mathbf{A^C}$: correctors, dipoles & strings	$\mathbf{R^M}$: position & angle at monitors	$M^{11}, M^{12}, M^{21}, M^{22}$ and linear comb.
EM	$\mathbf{A^E}$: alignment type errors	$\mathbf{R^M}$: position & angle at monitors	$M^{11}, M^{12}, M^{21}, M^{22}$
CA	$\mathbf{A^C}$: correctors, dipoles, & strings	$\mathbf{R^A}$: pos. & angle at all elements	$M^{11}, M^{12}, M^{21}, M^{22}$ and linear comb.
EA	$\mathbf{A^E}$: alignment type errors	$\mathbf{R^A}$: pos. & angle at all elements.	$M^{11}, M^{12}, M^{21}, M^{22}$
AM	$\mathbf{A^A}$: angle at all elements	$\mathbf{R^M}$: position & angle at monitors	M^{12}, M^{22}
MM	$\mathbf{A^M}$: monitor offset error	$\mathbf{R^M}$: orbit error at monitor	δij

recipes discussed here are applicable to such distributions if the constant probability density contours are convex ellipsoidal surfaces [2]. In other words,

$$\mathbf{P_E} = \mathbf{P_E}(\mathbf{Q})$$
$$\mathbf{Q} = \mathbf{A^{E^T}} \cdot \mathbf{E^E} \cdot \mathbf{A^E} \tag{1}$$

where $\mathbf{P_E}$ is the probability density function of the error represented by $\mathbf{A^E}$ of Table 1, and $\mathbf{E^E}$ a square symmetric matrix with non-negative eigenvalues. This includes the special case of multi-dimensional Gaussian distributions.

3 SECONDARY RESPONSE MATRICES

In order to further quantify physical concepts about the performance of configurations in terms of probabilistic envelopes, mapping of the distributions of (1) by various "secondary response matrices" needs to be studied. These are transformations derived from linear algebraic operations on the response matrices of Table 1. Different performance criteria require different operations and combinations. An example is the response matrix

$$\mathbf{M^{UOE}_{EMM}} = \left(\Pi^{\perp}_{\mathbf{CM}} \cdot \mathbf{M^{EM}} \right) \oplus \left(-\Pi^{\parallel}_{\mathbf{CM}} \cdot \mathbf{M^{MM}} \right), \tag{2}$$

with projection operators defined as

$$\Pi^{\parallel}_{\mathbf{CM}} = \mathbf{M^{CM}} \cdot \mathbf{M^{CM}}^{\dagger} = \mathbf{M^{CM}} \cdot \left(\mathbf{M^{CM}_T} \cdot \mathbf{M^{CM}} \right)^{-1} \cdot \mathbf{M^{CM}_T}$$
$$\Pi^{\perp}_{\mathbf{CM}} = \mathbf{I} - \Pi^{\parallel}_{\mathbf{CM}}, \tag{3}$$

and the direct sum ($\oplus$) being a column-wise concatenation. This matrix maps the concatenated vector

$$\mathbf{A^{EM}} = \mathbf{A^E} \oplus \mathbf{A^M} = \left(A^E_1, \dots A^E_{N_E}, A^M_1, \dots A^M_{N_M} \right)^T,$$

representing the combined alignment-type error (i.e., injection, alignment, field, etc.) and monitor error, to the underlying orbit error after orbit correction at monitors. More secondary response matrices can be constructed to realize different physical processes and criteria [2].

4 MATHEMATICAL RECIPES

In order to understand the impact of error distributions such as (1) on the performance criteria, through orbit monitoring and correction processes such as (2), it

remains to develop mathematical recipes to perform mapping, intersection, extremum determination, and other operations based on (1) and (2). The apparent quadratic form of (1) implies that direct application of off-the-shelf numerical optimization tools is not an option, especially if efficiency and robustness are desired when the number of error sources and monitored points run to the hundreds or even thousands. Thus all calculations must be rigorously cast in linear form for which efficient and robust tools exist. A list of such linear recipes is given below:

- Orthonormal transformation and decomposition.
- Projection of ellipsoids onto lower/higher dimensions.
- Inverse projection of point(s) onto ellipsoids.
- Tangent point between ellipsoids and hyper-planes.
- Extreme values of operators on a constrained surface.
- Curvature of ellipsoids through the Hessian operator.
- Recipes for multi-dimensional Gaussian distribution.

Different recipes have to be devised depending on whether the map is under-, over-, or critically-constrained. With the involved high dimensionality in mind, measures against rank deficiency also need to be built in. A complete delineation of the recipes can be found in [2].

5 MEASURES OF PERFORMANCE

5.1 Fundamental Observability, Monitor Coverage, and Effects of Residual Orbit

Unobservable Error
The unobservable subspace of $\mathbf{A^E}$ is defined by the error projection operator orthogonal to $\mathbf{M^{EM}}$. Taking the 3σ ellipsoid in this space and the boundary of its image by $\mathbf{M^{EA}}$, we get the unobservable error effect at all elements.

Effect of Finite Residual RMS Orbit
The observable subspace of $\mathbf{A^E}$ is defined by the error projection operator parallel to $\mathbf{M^{EM}}$. Taking the 1σ ellipsoid in $\mathbf{A^M}$, enhanced by $\sqrt{N_M}$, and the boundary of its image by the inverse of this projection followed by $\mathbf{M^{EA}}$, one obtains the maximal orbit error at all elements if the RMS orbit observed at the monitors is equal to the assumed monitor offsets. This is further scaled so that the final error is within a criterion determined by a cutoff.

Combined Effects of Finite Residual RMS Orbit
The error envelopes from the two previous sections come from complementary orthogonal subspaces. Their quadratic sum gives the real underlying orbit envelope at the same combined probability density as individual ones.

An alternative view can be obtained by treating alignment and observed orbit errors on an equal basis [2].

5.2 Monitor Efficiency and Redundancy

Monitor Efficiency
The efficiency of a monitor configuration is measured by the projection of $\mathbf{E^E}$ in (1) along all axes in the monitor space. Inspection of such envelopes can reveal strategic locations and gaps in the monitor configuration.

Redundant Monitor Combination
An SVD analysis on the response matrix $\mathbf{M^{EM}}$ can be used to reveal excessive monitors or their combinations.

5.3 Fundamental Correctability, Correction Range and Residual Orbit

Corrector Range
By taking the projection of $\mathbf{E^E}$ onto the corrector space by $\mathbf{M^{EM}}$ followed by the pseudo-inverse of $\mathbf{M^{CM}}$, and applying the technique for tangent points in Section 4, one can derive the maximum error, in units of the global error distribution σ, that can be handled by each corrector. Increasingly more vulnerable correctors can be tabulated.

If $\mathbf{M^{EA}}$ and $\mathbf{M^{CA}}$ are used instead, the evaluation will be independent of the monitor configuration and reflect "fundamental" correction range for each of the correctors.

Uncorrectable Orbit
By mapping $\mathbf{E^E}$ onto the corrector space by $\mathbf{M^{EM}}$ followed by the projection orthogonal to $\mathbf{M^{CM}}$, one gets the uncorrectable orbit at all monitors. If $\mathbf{M^{EA}}$ and $\mathbf{M^{CA}}$ are used, the result will be independent of the monitors.

Monitor Offset Induced Orbit Error
The normalized monitor error ellipsoid can be projected onto the element space via $\mathbf{M^{MM}}$, the pseudo-inverse of $\mathbf{M^{CM}}$, and $\mathbf{M^{CA}}$, providing evaluation on the effect of monitor offsets on the global corrected orbit.

Orbit Error Implied by Observed Orbit at Monitors
The implication of an observed, corrected orbit on the underlying errors and correctors, and in turn on its effect at all elements, can be calculated in a least-square sense by projecting the 1σ ellipsoid in $\mathbf{A^M}$ via a complicated secondary response matrix [2]. Proper scaling is needed such that it amounts to no more than 3σ in the errors.

5.4 Actual Underlying Corrected Orbit Error

Similar to (2), be the secondary response matrix defined

$$\mathbf{M_{EMA}^{UOE}} = \left(\mathbf{M^{EA}} \oplus \mathbf{Z^{MM}}\right) - \mathbf{M^{CA}} \cdot \mathbf{M_{CM}^{\dagger}} \cdot \left(\mathbf{M^{EM}} \oplus \mathbf{M^{MM}}\right),$$
$$\mathbf{R^{UA}} = \mathbf{M_{EMA}^{UOE}} \cdot \mathbf{A^{EM}}, \qquad\qquad (4)$$
$$\mathbf{Z^{MM}} = Z_{ij}^{MM} = 0, \; i=1,2,\ldots N_A, j=1,2,\ldots N_M.$$

where $\mathbf{R^{UA}}$ is a responder representing the real underlying orbit and angle after correction at all elements, $\mathbf{A^{EM}}$ an actuator of combined alignment-type and monitor errors, and $\mathbf{Z^{MM}}$ an all-zero matrix. The projection of an error ellipsoid of all such errors under a global distribution is mapped by $\mathbf{M_{EMA}^{UOE}}$, its image boundary solved, and the maximum along each axis of $\mathbf{R^{UA}}$ obtained. This is an important measure of the correction configuration, and can be directly compared with simulation.

5.5 Near-Degeneracy in Response Matrices

Null-space and SVD analyses of $\mathbf{M^{CM}}$ measure degeneracy in a correction configuration and identify offending corrector combinations that can lead to excessive correction and orbit error in unobservable locations. A well-defined procedure can be iteratively applied to eliminate degeneracy in a configuration.

6 THE OPTIMIZATION PROGRAM

Optimizing an orbit correction configuration involves

the interplay between applying the performance criteria discussed above and invoking well-defined optimization recipes. The latter will be discussed in the following.

6.1 Configuration Optimization

Two approaches are adopted to improve the performance:

Structural Improvement

For structural defects in a configuration, the following algorithms are applied to identify weak points and provide solutions to bring the performance to acceptable levels:

- Adding monitors by unobservable error-induced orbit.
- Adding correctors by uncorrectable residual orbit.
- Adding correctors by principal axes of error ellipsoid.
- Removing monitors by null space/SVD analysis.
- Removing correctors by null space/SVD analysis.

Exhaustive Fine Tuning

The need for fine-tuning arises where the effect of configuration parameters is algebraically intractable. For example, locations of correctors may have considerable impact on the residual orbit envelope, but the interplay between competing factors prevents an analytic path to the optimal solution. In this case scanning of the parameter space is done with one performance criterion as a merit function. This can be very efficient because the analytic performance criteria can be calculated quickly for a large number of candidate configurations. It would be difficult to do the same scanning based on simulation.

6.2 Failure Mode Analysis

Using the mathematical recipes one can also decompose an observed performance defect into contributing errors. Examination of the combination provides insight leading to improved configurations. As example, Figure 1 shows the alignment, monitor and field errors, and corrector responses that led to a peak in the actual underlying orbit (abscissa: beam line in m, ordinate: offsets in m / kicks in rad; full line: underlying orbit, bars tagged "MCIAxxxxxx": corrector kicks*100, bars tagged "BPMIxxxxx": monitor offsets*100, other larger bars: quadrupole offset kicks*100; other smaller bars: dipole tilt kicks*100, bars clipped at boundary).

6.3 Critical Elements

Study of the most vulnerable elements due to various failure modes can also be efficiently carried out, using the analytic performance criteria as merit functions.
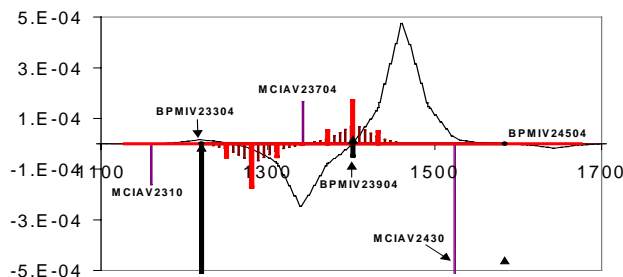

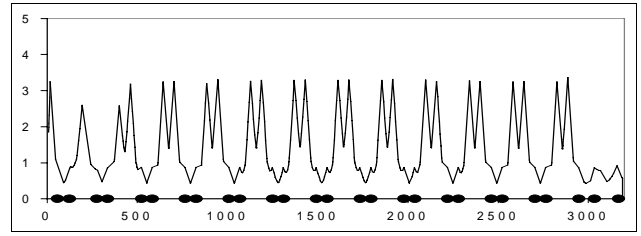Figure 1: Decomposition of Contributing Errors.

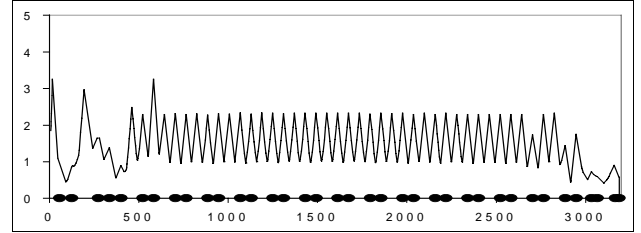Figure 2: TI 2 (H), 2-in-4 monitor and corrector.

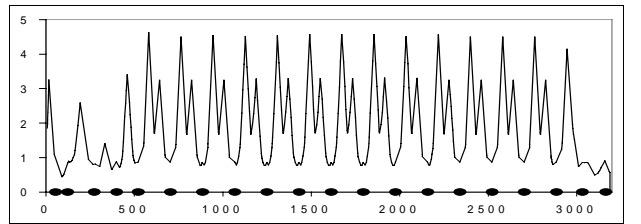Figure 3: TI 2 (H), 2-in-3 monitor, 1-in-3 corrector.

Figure 4: TI 2 (H), 1-in-3 monitor and corrector.

7 APPLICATION TO TI 2 AND TI 8

The method described has been applied to the LHC injection transfer lines TI 2 and TI 8 for evaluation and optimization. Figure 2 shows the 3σ envelope of the corrected orbit (horizontal plane) for the reference configuration with 2 out of every 4 consecutive focusing quadrupoles in the periodic section complemented by monitors (marked by dots) and correctors (abscissa in m, ordinate in mm). Figure 3 shows that for an alternative configuration with monitors in 2 out of every 3 and correctors in 1 of every 3 such locations. The latter implies an over-constrained steering assisted by computers, but provides a much tighter underlying orbit envelope than the "2-in-4" scheme (2.3 mm vs. 3.3 mm in the periodic section). A variation of the over-constrained scheme has every other monitor disabled to allow a 1-to-1 steering without resorting to computers; its corrected orbit envelope is shown in Figure 4. Similar schemes have been developed for the vertical plane. These schemes have been subjected to all other performance criteria discussed and found satisfactory.

8 ACKNOWLEDGEMENTS

REFERENCES

[1] A. Hilaire, V. Mertens, E. Weisse, Proc. EPAC'98, Stockholm (1998), p2117ff, and LHC Project Report 208.
[2] Y.-C. Chao, V. Mertens, LHC Project Report 470 (2001).
[3] Y.-C. Chao, Proc. ABS Workshop, ed. M. Lindroos, CERN 1998, and Proc. PAC'97, Vancouver, Canada, p2262ff.