TRAJECTORY AND CLOSED ORBIT CORRECTION

J.P. Koutchouk

CERN, Geneva, Switzerland

ABSTRACT

After a reminder of the linear beam dynamics, this course reviews various trajectory and closed orbit correction methods.

Lecture given at the joint US-CERN Accelerator School
on Beam observation, diagnosis and correction
Capri, October 1988

G E N E V A
January 1989

TRAJECTORY AND CLOSED ORBIT CORRECTION

Jean-Pierre KOUTCHOUK

CERN

CH-1211 Geneva 23

The control of the closed orbit (and of the betatron tune) is one of the basic ingredients of any accelerator control system. It is thus a subject which has been studied over a long period. However the significant increase in the size of machines and their exploitation in collider mode impose tighter tolerances and cost optimization. The quality of the closed orbit control is an essential component towards efficiency in operation, flexibility in machine studies and low background in the experimental detectors. It is thus not surprising that the closed orbit correction is still the subject of many publications.

The aim of this lecture is to review the orbit correction techniques developed in the accelerator laboratories around the world; some emphasis will be given to the new field opened by the possibility of turn-by-turn trajectory measurement.

1 THEORY OF ORBITAL MOTION

The aim of the trajectory and closed orbit correction is to approach as close as possible the reference trajectories and closed orbit of a machine. In most cases, the latter defines the magnetic axis of the lenses. As a consequence, only the linear particle dynamics is relevant to orbit correction.

Indeed, the kicks experienced by a particle displaced by x_{co} in a sextupole S of length l_s is $S l_s x_{co}^2$. The sextupole strength S is chosen to cancel the chromatic focusing aberrations : $S l_s D_x \approx K l_Q$, where D_x is the dispersion and $K l_Q$ the integrated strength of the focusing. Thus

$$S l_s x_{co}^2 \approx K l_Q \left(\frac{x_{co}}{D_x} \right) x_{co} \approx \theta_{foc} \left(\frac{x_{co}}{D_x} \right) \quad (1)$$

where $\theta_{foc} = K l_Q x_{co}$ is the quadrupole focusing kick. But $x_{co}/D_x \ll 1$, showing the hopeful predominance of the linear focusing over the sextupolar aberrations. The non-linearities may however require an iterative linear correction if the orbit deviations are large.

1.1 COORDINATE SYSTEM

Let us define the coordinate system for the transverse motion of a particle (figure 1). The origin of the coordinate system lies on the ideal or reference orbit. The corresponding bending radius ρ is the design bending radius.

Figure 1: Coordinate system

1.2 EQUATIONS OF LINEAR MOTION

The linearized transverse particle motion is well documented [1,2,3],... Only the essentials are recalled here. The two degree of freedom (4-dimensional) transverse linear motion is described by the differential equations:

$$x'' + [K + \frac{1}{\rho^2}]x = +\frac{\delta}{\rho} + \frac{\Delta B_y}{B\rho} - [K_s + \frac{1}{2}M']y - My' \quad (2)$$

$$y'' - Ky = +\frac{\Delta B_x}{B\rho} - [K_s - \frac{1}{2}M']x + Mx' \quad (3)$$

All derivatives are with respect to s . δ is the relative momentum deviation; $\Delta B_x/B\rho, \Delta B_y/B\rho$ are the normalized magnetic field imperfections, arising either from dipole field inhomogeneities or quadrupole misalignments; K_s, M_s are the skew quadrupole and solenoid normalized strengths. We assume here that the machine is plane, i.e. the vertical bending radius is infinite.

In most cases, the coupling gradients K_s arise from random tilts of the focusing quadrupoles and from vertical orbit deviations in the chromaticity sextupoles, whilst the longitudinal fields M_s vanish. The coupling terms on the r.h.s. of (2,3) are thus to second order in small quantities and are neglected in the orbit correction techniques. Purposely coupled sections including experimental solenoids, like large orbit deviations in non-linear lenses (e.g. injection bumps) require a specific treatment which may be derived straightforwardly from the analysis of the linear uncoupled case.

Following these simplifications, the motion is reduced to two independent forced oscillations. Examination of the r.h.s. of (2,3) shows that the forcing terms are periodic with the machine circumference. The solutions of (2,3) are thus known to be a superposition of a periodic forced solution (the closed orbit) with the free oscillation (the betatron oscillation).

1.2.1 The betatron oscillation

The free oscillation characterizes the motion of either a particle or a closed orbit between two kicks. Its well known solution can be written:

$$x_\beta(s) = a_x \sqrt{\beta_x(s)} \cos[\mu_x(s) + \mu_{x0}] \quad (4)$$

$$y_\beta(s) = a_y \sqrt{\beta_y(s)} \cos[\mu_y(s) + \mu_{y0}] \quad (5)$$

where β and μ are the Twiss parameters [1], and a and μ_0 are defined by the initial conditions.

1.2.2 The off-momentum orbit

If the beam momentum is slightly different from the design momentum by δ_0 in relative value, the horizontal closed orbit is displaced to first order by the quantity:

$$x_{\delta 0}(s) = D_x(s) \delta_0 \quad (6)$$

where D_x is the periodic dispersion function. Its differential equation may be found by substituting equation (6) into equation (2).

In large electron machines like LEP, the particles lose a significant fraction of their energy due to synchrotron radiation in the dipoles. The compensating acceleration only occurs at a few points (the RF cavities). The instantaneous momentum therefore varies along the circumference C , even for the ideal particle [4]. The equation of the off-momentum orbit may then be shown to take the form [5]:

$$x_\delta(s) = x_{\delta 0}(s) + x_\pm(s) \quad (7)$$

$$x_\pm(s) = D_x(s) \delta(s) - \frac{\sqrt{\beta(s)}}{2 \sin \pi Q} \int_s^{s+C} \sqrt{\beta(\sigma)} [2D'\delta' + D\delta'']_\sigma \cos[\pi Q - |\mu(s) - \mu(\sigma)|] d\sigma \quad (8)$$

If the dispersion and its derivative vanish at the RF cavities, the integral in equation (8) is small and the first term $D_x(s) \delta(s)$ dominates. It changes sign for electrons and positrons, so that the two types of particles do not follow the same trajectories. This splitting of the trajectories grows very rapidly with the energy E :

$$\delta_\pm(s) \propto \frac{E^3}{\rho} \quad (9)$$

For LEP at 55 GeV (figure 2), the peak electron-positron horizontal separation is as large as 6 mm, and thus must be taken into account when correcting an orbit. The corresponding vertical phenomenon is only excited by coupling and the separation is more than an order of magnitude lower.

Figure 2: LEP energy sawtooth

1.2.3 The on-momentum imperfect closed orbit

If the motion equations (2,3) are solved with only field errors on the r.h.s., the solution obtained describes the quantity which is usually referred to as the closed orbit. Whilst it is the full solution

for the vertical plane, it ignores any momentum effect in the horizontal plane. The solution has the form (e.g. in the vertical plane):

$$y_{co}(s) = \frac{\sqrt{\beta}}{2 \sin \pi Q} \int_s^{s+C} \sqrt{\beta(\sigma)} \frac{\Delta B(\sigma)}{B\rho} \cos[\pi Q - |\mu(s) - \mu(\sigma)|] d\sigma \quad (10)$$

The correcting elements may be considered as field errors so producing a corresponding closed orbit deviation. Expression (11) provides the effect of a kick θ_i in the i^{th} orbit corrector:

$$y_{co}(s) = \frac{\sqrt{\beta}}{2 \sin \pi Q} \sqrt{\beta_i} \theta_i \cos[\pi Q - |\mu(s) - \mu_i|] \quad \theta_i = \frac{\Delta B_i l}{B\rho} \quad (11)$$

It should be noted that this expression is exact for a pure, thick, dipole corrector and a good approximation if a field gradient is superimposed, the correction factor being $(1 - Kl^2/24)$ [6].

Another useful solution may be found by solving (2,3) in the frequency domain:

$$y_{co}(s) = \sqrt{\beta} Q^2 \sum_{k=-\infty}^{+\infty} \frac{f_k}{Q^2 - k^2} e^{ik\mu/Q} \quad (12)$$

with

$$f_k = \frac{1}{2\pi Q} \sum_j \sqrt{\beta_j} \theta_j e^{-ik\mu_j/Q} \quad (13)$$

f_k is the k^{th} Fourier coefficient of the superposition of field errors (index j). Equation (12) shows that the accelerator optics acts like a very sharp filter resonating around the frequency Q . This property is often put to work both for analytical work and closed orbit correction.

1.2.4 Summary

The beam position measured by position monitors can be conveniently decomposed in a periodic and aperiodic part:

$$x = [x_{co} + D_x \delta_0 + D_x \delta_{\pm}] + x_{\beta} \quad (14)$$

Closed orbit/trajectory correction can be applied to x_{co} and x_{β} , y_{co} and y_{β} using the dipole correctors. δ_0 may be corrected by changing the beam energy ; nothing can be done for $\delta_{\pm}$.

2 MEASUREMENT OF TRAJECTORIES AND CLOSED ORBITS

In most modern accelerators, whenever the revolution frequency is sufficiently low, the beam position is measured on each passage by beam monitors (PU's) and stored in a local memory. The availability of such a wealth of information opens up a wide spectrum for data processing beyond the traditional averaging. Trajectory and closed orbit correction requires first the various components of the beam position (14) to be disentangled.

2.1 EXTRACTION OF THE CLOSED ORBIT

Except after a kick, each particle performs aperiodic betatron oscillations, but their center of gravity should steadily remain on the closed orbit. However, there are various reasons for the beam to perform small coherent oscillations. This aperiodic x_{β} component may be suppressed by averaging each PU data over n passages. If A_{coh} is the expected amplitude of the coherent motion and Δx the requested resolution on the closed orbit, n must follow the rule:

$$n \geq \frac{\sin n\pi Q}{\sin \pi Q} \frac{A_{coh}}{\Delta x} \quad (15)$$

2.2 EXTRACTION OF THE BETATRON OSCILLATION

It is simply obtained by subtracting the closed orbit from the measured trajectory.

2.3 EXTRACTION OF THE AVERAGE MOMENTUM DEVIATION δ_0

Following (14), one may have access to δ_0 by averaging the closed orbit data obtained as in section 2.1 over all the PU's. The first closed orbit component tends to vanish because it is centered, whilst the last one vanishes for symmetry reasons. It is more precise to average over the normalized closed orbit:

$$\delta_0 = \sum_i \frac{x_i \Delta \mu_i}{\sqrt{\beta_i}} / \sum_i \frac{D_i \Delta \mu_i}{\sqrt{\beta_i}} \quad (16)$$

2.4 EXTRACTION OF THE ENERGY SAWTOOTH $\delta_{\pm}$

This may readily be obtained by subtracting two closed orbits corrected for δ_0 measured for electrons and positrons. Alternatively, the energy sawtooth may be accurately predicted using the classical part of the synchrotron radiation.

2.5 EXTRACTION OF THE IMPERFECT CLOSED ORBIT

It is found by subtracting the energy-dependent components calculated in sections 2.3 and 2.4 from the closed orbit obtained as in 2.1.

2.6 A FAST MEASUREMENT OF THE CLOSED ORBIT (and of the tune)

Let us assume that the beam is performing small oscillations x_{β} around a closed orbit X which is not restricted to being linear. At the n^{th} passage, a beam monitor will record a position x_n :

$$x_n = X + (C_n + \alpha S_n)x_0 + \beta S_n x'_0 \quad (17)$$

with

$$C_n = \cos n\mu \quad S_n = \sin n\mu \quad (18)$$

α , β and μ are the usual Twiss parameters, X is the unknown closed orbit and x_0 , x'_0 are the unknown initial conditions. Three such equations (three turns) are necessary to eliminate x_0 and x'_0 :

$$4X \sin^2 \frac{\mu}{2} = -2x_{n+1} \cos \mu + (x_n + x_{n+2}) \quad (19)$$

Two such equations (four turns altogether) yield:

$$\cos \mu = \frac{x_n - x_{n+1} + x_{n+2} - x_{n+3}}{2(x_{n+1} - x_{n+2})}$$

$$X = \frac{1}{4 \sin^2 \mu/2} [x_n + x_{n+2} - 2x_{n+1} \cos \mu]$$

After four turns, each PU provides a value for the tune which can be averaged before calculating the closed orbit X , thus providing a reasonable accuracy. This technique may be advantageous when commissioning new machines or new optics as it provides a way of measuring and correcting the closed orbit and the tune provided the beam circulates at least four turns. The method is derived from a closed-orbit-finding algorithm successfully used in optics simulation [7].

3 CORRECTION METHODS

3.1 CLASSIFICATION OF THE METHODS

The correction methods are numerous. To clarify the subject, a classification can be helpful. Let us distinguish the correction of the trajectories, essentially relevant to open sections, from the correction of the imperfect closed orbit. In the first class, we will analyse:

- the classical beam steering,
- the beam ‘threader’, which is a more ambitious steering,

and, in the second class :

- the methods which minimize the closed orbit distortions with no attempt at finding their sources,
- the methods which select the smallest set of correctors to obtain a given orbit quality; these methods may locate imperfection sources with an ambiguity,
- and finally the methods which attempt to localize the azimuth of field imperfections.

This classification also corresponds to an increasing complexity of the algorithms.

3.2 BEAM STEERING

3.2.1 Minimization of trajectory deviations in beam transfer lines

Figure 3: Beam steering

Solving the equations (2,3) yields the well-known transport matrix for the phase-space coordinates (x, x') :

$$\begin{pmatrix} x_{PU} \\ x'_{PU} \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} x_{cor} \\ x'_{cor} \end{pmatrix} \quad (20)$$

This matrix may either be expressed in terms of the strengths and lengths of the elements between the corrector and the PU or in terms of the Twiss parameters. The m_{12} coefficient which relates a displacement x to a kick x' depends upon $\sin(\mu_{PU} - \mu_{cor})$. The best position for a corrector is thus $\pi/2$ upstream of the PU in betatron phase advance (figure 3).

In spite of its primitiveness, a manual steering of beam transfer lines is quite popular. For long beam transfer lines, the method may be straightforwardly parallelized:

$$\begin{pmatrix} x_{PU1} \\ x_{PU2} \\ \vdots \\ x_{PU_n} \end{pmatrix} = \begin{pmatrix} a_{11} & 0 & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & . & . & . & a_{nn} \end{pmatrix} \begin{pmatrix} x'_{cor1} \\ x'_{cor2} \\ \vdots \\ x'_{corn} \end{pmatrix} \quad (21)$$

where a_{ij} corresponds to the m_{12} coefficient of the transport between corrector j and PU i . The problem is simply solved by substitution.

3.2.2 Minimization of the betatron oscillations at injection

Figure 4: Injection steering

An injection system can put a beam onto orbit provided it enters the injection system on the design trajectory (figure 4). Due to imperfections, this is not so and one needs to adjust the four transverse coordinates with four corrector magnets. Due to the oscillatory nature of x and y , the best corrector arrangement is obtained when using two correctors per plane, $\pi/2$ apart in betatron phase.

If only the envelope of the injection oscillation is available, a minimization process is required. At the ISR for instance, fast algorithms for 4-D search were developed and used for many years until the clear superiority of the 'threader' technique was recognized.

3.3 BEAM THREADER

The beam threader algorithm may be used to flatten a trajectory in a beam transfer line, to adjust the injection in a circular machine or to cancel closed orbit perturbations. Its principle is to recognize a pattern of free oscillation following a kick or an injection with imperfect initial coordinates (figure 5).

The algorithm is as follows:

- Subtract the closed orbit from the measured trajectory (if the closed orbit is known),
- Define the fitting range (pure oscillation),

Figure 5: Beam threader

- Fit $a \sin [\mu(s) + \Phi] + D\delta_0$, where a , Φ , and δ_0 are the unknowns,
- Select two orthogonal correctors; if the azimuth of the perturbation is known, this step is trivial; otherwise guess the origin of the oscillation,
- Flatten the oscillation by exciting the correctors to:

$$\theta_1 = -\frac{a \sin [\mu_2 - \Phi]}{\sqrt{\beta_1} \sin [\mu_2 - \mu_1]} \quad \theta_2 = \frac{a \sin [\mu_1 - \Phi]}{\sqrt{\beta_2} \sin [\mu_2 - \mu_1]}$$

This procedure has been successfully used to tune the ISR injection, switch-on a superconducting low- β in the same machine without quench and to inject the beam in the first octant of LEP (figure 6).

Figure 6: Threading the beam in the first LEP octant

3.4 MINIMIZATION OF THE CLOSED ORBIT DEVIATIONS

3.4.1 Closed orbit flattening using bumps

- *Definition of orbit bumps*

An orbit bump is a local deformation of the closed orbit: Let $\mathcal{L}$ be the length of the bump. Its amplitude $y(s)$ is such that:

$$\forall s \notin \mathcal{L}, y(s) = y'(s) = 0$$

$$\text{for } s = s_0 \in \mathcal{L}, y(s_0) = A \text{ [and } y'(s_0) = B]$$

In order to fulfill these three or four conditions, one needs an equivalent number of corrector magnets in general. An explicit bump equation may be calculated by superposing the deformation due to each magnet (equation (11)) and imposing the above-mentioned conditions.

- *Popular orbit bumps*

One may take advantage of certain particular phase advances between corrector magnets or of the optics symmetry to design simpler bumps. The most usual bumps are represented on figure 7.

- *Beam bump correction à la Collins*

A serious worry at the beginning of the use of strong focusing was that the closed orbit might partly lie outside of the vacuum chamber. This was considered to be an intrinsic limitation of strong focusing until Collins, in the 60's, proposed an orbit correction scheme which is still widely used.

Let us assume a machine made of regular FODO cells, with a beam monitor and an orbit corrector at each focusing quadrupole in each plane (figure 8). If the phase advance per cell is $\pi/2$, a simple two-magnet π bump may be created with the first and third magnet to correct the beam displacement measured at the symmetry point of the bump. If the phase advance is not $\pi/2$, the magnet at the bump center is used to create a symmetric three-magnet bump which is completely general. A beam displacement y_i at magnet/PU i is corrected by the three kicks:

$$\theta_{i-1} = \theta_{i+1} = \frac{y_i}{\beta \sin \mu} \quad \theta_i = -2\theta_{i-1} \cos \mu \quad (22)$$

with μ phase advance per cell. The process is repeated with each set of three corrector magnets all around the machine.

Each magnet is shared by three bumps. If the orbit distortion is a free oscillation over some length, it is easy to verify that the superposition of the three bumps will cancel out the excitation θ of the corrector magnets over that range. There remains only a non-vanishing excitation close to the field defect. This property may, unfortunately, be confused by the limited accuracy of the beam position measurements.

- *Generalization 1*

In practice, corrector magnets and PU's are not positioned at the same azimuthal position and differ in betatron focusing function β_{PU} , β_{cor} and phase advance (ϵ). The first equation (22) must be modified as follows:

$$\theta_{i-1} = \theta_{i+1} = \frac{y_i}{\sqrt{\beta_{PU}} \sqrt{\beta_{cor}} \sin(\mu - \epsilon)}$$

A second consequence is that, at the closure, the last bump distorts the already corrected orbit at the first PU. To alleviate this difficulty, two possibilities are opened: one may iterate

Figure 7: Popular orbit bumps

Figure 8: Beam bump technique

the correction a few times around the machine. Alternatively one may proceed as suggested in Ref. [8], i.e. find directly the correction by the inversion of the tri-diagonal matrix relating the corrector strengths θ to the orbit deviations y :

$$\begin{pmatrix} y_{PU1} \\ y_{PU2} \\ y_{PU3} \\ \vdots \\ y_{PU_n} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & 0 & 0 & . & . & a_{1n} \\ a_{21} & a_{22} & a_{23} & 0 & . & . & 0 \\ 0 & a_{32} & a_{33} & a_{34} & 0 & . & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & 0 & . & . & . & a_{nn-1} & a_{nn} \end{pmatrix} \begin{pmatrix} \theta_{cor1} \\ \theta_{cor2} \\ \theta_{cor3} \\ \vdots \\ \theta_{corn} \end{pmatrix} \quad (23)$$

• Generalization 2

The above-mentioned beam-bump technique is well adapted to synchrotrons but requires one orbit corrector per cell. In colliders, the optics is much less regular; it is, moreover, possible to demonstrate that an efficient orbit correction may be obtained with significantly less correctors [9]. The beam-bump technique may still be used with two possible implementations.

In the first one, all possible consecutive three-magnet bumps are considered sequentially. Within each bump, the closed orbit oscillation is minimized in the least-squares sense, using the only free parameter of the bump:

$$\frac{\partial}{\partial \theta_{j-1}} \sum_i (y_i - \alpha_i \theta_{j-1})^2 = 0$$

where j runs over the bumps, i runs over the PU's within a bump, and α_i is the normalized bump amplitude at PU i . As in the previous technique, a perfect closure cannot be obtained after one turn and iteration is required. This method is available in PETROC [10], the CERN version of the DESY PETROS optics program.

Alternatively, the problem can be solved in one go if expressed as a least-squares problem relating n observables (PU's) to m parameters (the bumps' amplitudes). If $\mathbf{A}$ is the (n, m) matrix defining the bump j amplitude at position i , the solution is given by

$$\vec{\theta}_m = -(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \vec{y}_n$$

This problem is in general simpler than relating the observables to the individual orbit correctors. The matrix $(\mathbf{A}^T \mathbf{A})$ is indeed sparsely filled and may be inverted more easily [11].

3.4.2 Harmonic correction

From equation (12), the Fourier transform $\mathcal{F}$ of the normalized orbit may be simply deduced:

$$\mathcal{F}\left[\frac{y}{\sqrt{\beta}}\right] = \frac{Q^2}{Q^2 - k^2} \mathcal{F}\left[\beta^{3/2} \frac{\Delta B}{B\rho}\right] \quad (24)$$

If the field errors ΔB are random, the orbit spectrum is the accelerator transfer function, i.e. the first term of the r.h.s. of (24). The spectrum pattern is simple, with only a few strong harmonics around the tune (figure 9). The ISR implementation of a harmonic correction technique was

Figure 9: Orbit spectrum before correction

straightforward: the contributions of all correctors ($j = 1 \dots m$) to the orbit harmonics $i = 1 \dots h$ are computed and arranged in a square matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{pmatrix} (a_{kj}) \\ (b_{kj}) \end{pmatrix}, \quad h = \frac{m}{2} \quad (25)$$

a and b being the sine and cosine component of each harmonic. In the transformation, the azimuthal angle must be the betatron phase advance and not the geometrical angle. $\mathbf{A}^{-1}$ is computed once and stored with the optics data base. The closed orbit is then Fourier transformed and filtered by retaining only a few harmonics around Q , and the corrector excitations calculated using:

$$\begin{pmatrix} \theta_1 \\ \vdots \\ \theta_n \end{pmatrix} = \mathbf{A}^{-1} \begin{pmatrix} (0) \\ (a_i) \\ (0) \\ (b_i) \\ (0) \end{pmatrix} \quad (26)$$

Retaining only four to five harmonics is an efficient protection against measurement errors.

Another approach is to exploit the natural machine symmetries when they exist. This was used for instance in PETRA, which has a four-fold symmetry (figure 10). Four corrector magnets are positioned in equivalent positions, while four others are mirror-symmetric. This arrangement allows the independent control of four consecutive harmonics. The excitation table for the correctors is

Figure 10: Corrector grouping in a 4-fold symmetric ring

given in [12]. For another scheme with other symmetries, it may easily be recalculated by expressing the Fourier components and the conditions put onto them:

$$\begin{pmatrix} a_n \\ b_n \end{pmatrix} = \oint \sum_i y_i(\mu) \begin{pmatrix} \sin n\mu \\ \cos n\mu \end{pmatrix} d\mu \quad (27)$$

where the y_i 's are the distortions associated with each corrector magnet. In this case, $n = 1 \dots 4$ and only one of the a_n, b_n must be non-zero.

This technique provided an efficient way of improving the beam polarization by an empirical orbit optimization based on the orthogonal control of detrimental harmonics. In this particular case, the relevant parameter in the Fourier transformation is the spin precession angle for a unit spin tune.

3.4.3 Methods to find the most efficient correctors

Another family of closed orbit correction methods attempts to select the smallest set of orbit correctors which minimizes a measure of the orbit 'badness'. The physical justification relies on the fact that there is often a small number of large field errors emerging from a background of small errors. A mechanism which selects a small number of efficient correctors is thus bound to more or less localize the large field errors and correct them at the source with a minimal amount of resources.

- *The measure of the badness*

The closed orbit deviations can be represented by an n -vector:

$$\vec{y} = (y_1, y_2, \dots, y_n)^T$$

where n is the number of PU's. Although an n -vector is a classical mathematical entity, defining a measure (or a norm) to be minimized is a pragmatic step. An extension of the

usual length concept is given by the class of l_p norms:

$$\|\vec{y}\| = \left(\sum_{i=1}^n |y_i|^p \right)^{1/p}, \quad p \in \mathcal{N}$$

The l_2 norm is the Euclidian length and leads to the least-squares minimization. Although it seems natural, it does not disqualify the other norms.

The l_1 norm penalizes errors independently of their magnitude and thus does not overweigh wrong measurements. The implementation (Simplex) moreover allows to set a limit to the corrector magnet excitation.

The l_∞ (Tchebychev) norm allows the maximum beam excursion to be controlled. It is thus the most practical norm provided the beam position readings are reliable.

Minimization of these various norms has been tested at the SPS [13].

- *The least-squares problem*

Minimizing the Euclidian norm is the usual approach that we will use from now on. The least-squares problem has already often been mentioned and is well documented in all books on statistics and physics. We will nevertheless present it here with a different point of view.

Let us assume that the orbit measurement $\vec{y}$ can be predicted from kicks $\vec{\theta}$ through a known model $\mathbf{A}$; however an unknown noise $\vec{\epsilon}$ is superimposed:

$$\vec{y} = \mathbf{A}_{nm} \vec{\theta}_m + \vec{\epsilon}_n$$

with n the number of PU's, m the number of correctors and $\mathbf{A} = (a_{ij})$.

$$a_{ij} = \frac{\sqrt{\beta_i \beta_j}}{2 \sin \pi Q} \cos [\pi Q - |\mu_i - \mu_j|]$$

Clearly, there are $(n + m)$ unknowns for n equations and m other relations must be provided. The deterministic problem being solvable, one may only state a property of the noise. A natural assumption is that any arrangement of the m correctors shall not be able to reproduce the noise. Mathematically, this is expressed by a vanishing cross-correlation between the orbit distortion created by any corrector j and the noise. Using the symbol of mathematical expectancy:

$$\forall j \quad E[\vec{a}_j \cdot \vec{\epsilon}] = 0$$

which can be approximated by

$$\mathbf{A}^T \vec{\epsilon} \approx \vec{0}$$

The system becomes

$$\begin{pmatrix} \vec{y} \\ \vec{0} \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{I} \\ \mathbf{0} & \mathbf{A}^T \end{pmatrix} \begin{pmatrix} \vec{\theta} \\ \vec{\epsilon} \end{pmatrix} \quad (28)$$

with the solutions:

$$\begin{aligned} \vec{\theta} &= [(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T] \vec{y} \\ \vec{\epsilon} &= \vec{y} [\mathbf{I} - \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T] \end{aligned}$$

In other words, the least-squares technique cannot fit any type of noise. The solution relies on the non-correlation of the 'noise' with the optical model. The method is thus biased by optical model errors.

- *Minimization strategies*

The aim being to select the smallest set of correctors which minimizes the orbit deviations, the algorithm is bound to be iterative in two stages: selection of a sample of correctors, computation of the corrector strengths. Various strategies are possible: Let m be the total number of available correctors.

1. *Brute force.* For a subset p of correctors ($p = 1 \cdots m$), select all combinations one after the other. For each combination, solve the least-squares problem to compute the orbit residue. If it is below some threshold, retain the corrector strengths and exit.

This technique is extremely heavy indeed; however, if the number of correctors involved is small, it guarantees that the corrector subset selected is the closest to the field imperfections. This has been of interest for the understanding of machine imperfections.

2. *More Selective.* A straightforward simplification consists in retaining the already selected correctors. An iteration is thus only concerned with the selection of the next most efficient corrector magnet. After p iterations amongst m correctors, only about $p \cdot m$ corrector sets will have been examined instead of $m!/(m-p)!p!$ in the previous case. To avoid a bias, it is best to solve the whole least-squares problem i.e. to recompute the excitations of all the correctors of the selected subset at each iteration.

The price to be paid for this simplification is a fair chance not to find the optimal solution: if there are two magnetic defects at distinct phases of about equal amplitudes, the algorithm selects a corrector which best corrects both; it becomes thus impossible, after another iteration, to come to a state where each defect is corrected by the closest corrector [14].

The control room approach often proceeds in steps: measurement, correction, measurement, correction, ... The danger mentioned above is even worse because there is no chance that the strength of the first selected corrector in the example given above be ever reduced. This may result in orbit bumps produced by the correction algorithm. However, the execution speed makes this approach attractive and efficient provided its results are checked and reviewed periodically.

3. *MICADO.* This algorithm, initially developed for the ISR [15], is now widely used [16,10,17]. It basically follows the selective approach presented above with an important improvement: the selection of the next most effective corrector is not based on solving the least-squares problem. Instead, a cross-correlation is formed between the orbit residues of the previous step and the remaining correctors. The highest correlation, called the pivot, defines the next most efficient corrector. The calculation is simpler and allows a faster selection. The least-squares problem is only solved for the selected corrector set at each iteration.

Both calculations are combined and facilitated by use of the Householder transformation. It is an orthogonal transformation and thus preserves the norm of the vectors operated upon. The selection of the best correctors based on a test of the norm of the residuals is thus unaffected. Under this transformation, all the vector components except one vanish. This eventually allows the least-squares problem to be expressed as a triangular set of linear equations which can be solved by back-substitution. The detailed mathematical technique is involved and may be found in Refs. [18,15].

- *Spectral properties of the best-corrector techniques*

In comparing the various correction techniques, it is of interest to understand the spectral properties of an orbit corrected with one of the above 'spatial' methods. The following analysis was developed to clarify whether the best-corrector techniques can increase the harmonic amplitudes away from Q , which is detrimental for beam polarization [19].

A reasonable assumption is that the best-corrector technique selects mostly correctors which either are close to the defects or shifted in phase by π . The spectrum of a corrected orbit may still be expressed by (24) provided the r.h.s. is replaced by the sum of the magnetic defects and of the kicks due to the orbit correctors:

$$\mathcal{F} \left[\frac{y}{\sqrt{\beta}} \right] = \frac{Q^2}{Q^2 - k^2} \mathcal{F} \left[\beta^{3/2} \left\{ \frac{\Delta B}{B\rho} + \frac{\Delta B}{B\rho} \otimes \delta\left(\phi - \frac{\pi}{Q}\right) \right\} \right] \quad (29)$$

where the convolution with the shifted δ operator expresses the defects corrected at π (the defects corrected by a close-by corrector have indeed disappeared). The Fourier transformation yields:

$$\mathcal{F} \left[\frac{y}{\sqrt{\beta}} \right] = \frac{Q^2(1 + e^{-ik\pi/Q})}{Q^2 - k^2} \mathcal{F} \left[\beta^{3/2} \frac{\Delta B}{B\rho} \right] \quad (30)$$

If the imperfections are randomly distributed, the Fourier transformation on the r.h.s. of (30) is a constant f_0 . Expression (30) shows thus that the enhancement of the spectrum away from Q is periodic and at most a factor of two. For $k \approx Q$, expression (30) may be developed:

$$\mathcal{F} \left[\frac{y}{\sqrt{\beta}} \right] \approx -i\frac{\pi}{2} \left[1 - \frac{\epsilon}{2}(1 + i\pi) \right] f_0 \quad (31)$$

with $k = Q(1 + \epsilon)$. This expression shows that the resonant properties of the accelerator transfer function have been suppressed by the orbit correction algorithm using the best-corrector technique. The result obtained in practice is close to this simplified analysis (figure 11).

Figure 11: Orbit spectrum after correction

- *Sensitivity to errors*

Unlike the harmonic correction, the best-corrector technique is very sensitive to large isolated measurement errors, which must somehow be removed beforehand. One may either perform local fitting or filter out the orbit and suppress the measurements larger than some threshold. The smaller statistical errors from the PU's set a limit to the orbit residuals. A more important consequence is that they increase the rms corrector strength. By applying the law of the

propagation of the variance to the least-squares problem, it is possible to derive the following approximate relation for LEP [20]:

$$\langle \theta \rangle \approx \sqrt{\frac{\langle x_{co} \rangle^2}{p\beta} + \frac{2p \langle \epsilon \rangle^2}{n\beta_{PU}}} \quad (32)$$

with p being the number of correctors, n the number of PU's, $\langle x_{co} \rangle$ the rms initial orbit deviation, $\langle \epsilon \rangle$ the rms PU noise. This expression shows that a reduction of the rms strength by increasing the number of correctors is limited by the noise.

3.4.4 Looking for imperfections

All correction techniques so far aim at minimizing the closed orbit distortions. They in general do not allow a reliable localization of the magnetic imperfections. Local methods are more appropriate.

- *Statistical approach*

The principle is to displace a small window around the machine and test a free oscillation within this window. The fit residue, properly scaled, is a probability of magnetic defect. Left and right fitting provide the position of the probable kick. This method was developed and successfully applied to the ISR [21].

- *Search for signatures*

The approach is basically identical, but very different in implementation. Let us consider a sliding window encompassing three PU's. If there is only one faulty PU, three consecutive high fit residues will be observed when displacing the window, yielding a first rule. If there is only one magnetic defect, the same technique gives one or two consecutive high fit residues, yielding a second rule. Additional rules can generate suspicions to narrow the search, provide means of disentangling combined defaults and allow a diagnosis to be verified.

This rule-based system is implemented in PROLOG on an IBM PC [22] and has already been used successfully [23]. The 'fashionable' implementation does not change the nature of the problem nor does it find miraculous solutions. It does however allow a clear presentation of the methodology which incorporates not only deterministic but equally heuristic rules to compensate for the lack of data.

4 CONCLUSION

There is a wide spectrum of closed orbit correction techniques and a reasonable question would be which one is the best. Implementing and experimenting with one of these is a non-trivial effort. There is thus little opportunity to compare them. From simulation work or actual closed orbit correction known to the author, all methods are approximately equivalent from the point of view of the final orbit residues, i.e. they are all equally efficient, provided the PU data have a reasonable signal/noise ratio.

The harmonic correction is however significantly more immune to the noise, which can be caused, for example, by a large burst of radiation striking a PU. In the ISR, the best-corrector technique was often preceded by a harmonic analysis to reject such readings.

If the orbit is corrected by moving quadrupoles, or if the machine is exploited in long sessions of beam coasting which demand a very high reliability, the best-corrector techniques are certainly well adapted. They, moreover, allow a cheaper design of the closed orbit corrector scheme.

The beam bump techniques have the advantage of simplest design and implementation. Simulations have shown a high sensitivity to the noise.

Given the large number of published work on this topic; the references are necessarily somewhat arbitrary; it is possible that some interesting correction techniques are not reported here. The author would like to apologize for this and would appreciate being informed about them.

References

- [1] E. Courant and H. Snyder, Theory of alternating-gradient synchrotron, Ann.Phys. 3(1958).
- [2] K. Steffen, Basic course on accelerator optics, Proceedings of the 1984 CERN Accelerator School, CERN 85-19(1985).
- [3] E. Courant, Introduction to accelerator theory, Fermilab Summer School 1981, AIP Conf. Proc. 87(1982).
- [4] M. Bassetti, Effects due to the discontinuous replacement of radiated energy in an electron storage ring, Proc. 11th Int. Conf. High En. Acc., Geneva, 1980(Birkhauser, Basle, 1980), p. 650.
- [5] J.P. Koutchouk, The horizontal closed orbit of a radiating beam, (CERN internal) LEP Note 578 (1987).
- [6] A. Verdier, Private communication.
- [7] T. Risselada and A. Verdier, Preparation of the operation of the superconducting low- β insertion in the ISR, CERN/ISR-BOM-OP/82-19(1982).
- [8] J.R. Maidment and C.W. Planner, An analytic method for closed orbit correction in high energy synchrotrons, Nuclear Instr. Meth. 1971.
- [9] G. Guignard, Effets des champs magnétiques perturbateurs sur l'orbite fermée ainsi que leur compensation, CERN 70-24(1970).
- [10] G. Guignard and Y. Marti, PETROC users' guide, CERN/ISR-BOM-TII/81-32(1981).
- [11] S. Peggs, Ph.D. Thesis, Cornell 1981.
- [12] R. Rossmanith, R. Schmidt, Compensation of depolarizing effects in electron positron storage rings, DESY 84-105(1984).
- [13] J. Miles, Preliminary note on closed orbit correction in the l_p norms, (CERN Internal) SPS/AOP/Note-83-3(1983).
- [14] T. Risselada, An improved version of the orbit program ORBCOR, (CERN Internal) CERN/PS 87-90(1987).
- [15] B. Autin and Y. Marti, Closed orbit correction of A.G. machines using a small number of magnets, CERN ISR-MA/73-17(1973).
- [16] F.C. Iselin, J. Niederer, The MAD program (Methodical Accelerator Design), CERN/LEP-TII/88-38(1988).
- [17] E. Close et al., A proposed orbit and vertical dispersion correction system for PEP, SLAC-PEP Note 271(1978).

- [18] B. Autin, Lattice perturbation, CERN/PS/84-22 (AA) (1984) and Brookhaven Summer School on HEPA 1983.
- [19] J.P. Koutchouk, Systematic integer spin resonances, 8th Symposium on high energy spin physics, Minneapolis 1988.
- [20] J.P. Koutchouk, About the strength of LEP orbit correctors, (CERN Internal) LEP Note 511 (1984).
- [21] G. Guignard, The closed orbit measurement as a diagnostic tool for localisation and correction of the misalignments in the ISR, CERN-ISR-BOM/80-21(1980) and 9th Int. Conf. on High En. Acc. CERN July 1980.
- [22] D. Brandt, F. Varlot, A. Verdier, (CERN internal) LEP note to be published.
- [23] D. Brandt and A. Verdier, Search for magnetic defects in EPA, 1st EPAC, Rome 1988.