SLAC-PUB-4091
September 1986
(A)

Some Applications of AI to the Problems of Accelerator Physics*

T. Higo†, H. Shoaee and J. E. Spencer
Stanford Linear Accelerator Center
Stanford University, Stanford, CA 94305

Abstract

Failure of orbit correction schemes to recognize betatron oscillation patterns obvious to any machine operator is a good problem with which to analyze the uses of Artificial Intelligence and the roles and relationships of operators, control systems and machines. Because such error modes are very common, their generalization could provide an efficient machine optimization and control strategy. A set of first-order, unitary transformations connecting canonical variables through measured results are defined which can either be compared to design for commissioning or to past results for 'golden orbit' operation. Because these relate directly to hardware variables, the method is simple, fast and direct. It has implications for machine design, controls, monitoring and feedback. Chronological analysis of such machine signatures can predict or provide a variety of information such as mean time to failure, failure modes and fast feedback or feedforward for optimizing figures of merit such as luminosity or current transmission. The use of theoretical and empirical scaling relations for such problems is discussed in terms of various figures of merit, the variables on which they depend as well as their functional dependences.

Introduction

Many beam diagnostic/correction schemes exist for on-line computer control systems but only with the advent of low-cost, large-storage micros has the potential use of artificial intelligence become practical. At the same time, the size and complexity of accelerator facilities has grown to the extent that both new hardware and software are needed to efficiently commission and operate them. Because one prefers that such techniques not become an end in themselves, we reconsider some old problems and techniques to see to what extent AI can be efficiently applied. Our general attitude toward the man/machine conundrums in AI which implicitly preordain one's approach is that man is not a machine nor should the machine be made to function as a man. Our goal is to provide the man/mind/operator with what they either can't or won't provide for themselves to achieve or maintain certain simple and easily stated goals. Goal setting, modification or codification is then one of the intersections between the control system and the operator. We will begin with the relatively simple problem of maintaining a given system or operating condition since this is easier than initially achieving it just as running an airplane under autopilot is simpler than landings or take-offs. Furthermore, this illustrates most of the basic ideas and variables. Success can be measured by how often one crashes or how hard it is to tell whether the plane is under manual or autopilot. For accelerators a good, practical autopilot should be possible that is significantly better than conventional control systems through consistent and improving hardware characterization which can learn how to prevent crashes and help circumvent aging effects in various ways.

The Paradise or Golden Orbit Problem

Let's assume that we have obtained an ideal operating setup and saved the relevant hardware information as well as such beam related parameters as the orbit $\bar{x}(s), \bar{y}(s)$ and its errors $\sigma_x(s), \sigma_y(s)$[1] in terms of distance along the beam line s. We may then either want to maintain it, extrapolate it or simply turn it off in a way that allows us to reestablish it later.

This is the 'Golden Orbit' problem and although it is an ideal, it repeatedly occurs in one guise or another. For instance, how do you improve an acceptable operating 'point' or keep it from deteriorating so that gold turns into brass. AI is relevant here for dealing with the complex problems of generalized hysteretic and stochastic effects. The way hardware is designed, built, installed, used and even monitored impacts the analysis so one would like as direct a correspondence as possible between the hardware and software or the thing itself and our representations of it. In particular, one wants a close correspondence to those variables that control the figures of merit.

In our example, a direct correlation exists between angle kicks θ_j from air-core correctors at position j and downstream position measurements at i. This is a direct measure of the unitary transformations $R_{12}(s_{ij})$, $R_{32}(s_{ij})$, $R_{34}(s_{ij})$ and $R_{14}(s_{ij})$ where we have used TRANSPORT notation[2]. Such kicks or dipole current errors produce the distinctive betatron oscillation patterns mentioned above and are one of the most probable error modes. Over time, one should be able to track the frequency of occurrence and the signatures of such effects for all j and develop a strategy for dealing with them.

To correct an orbit to the golden orbit, we can normalize orbits and corrector strengths relative to their golden values so one wants to minimize expressions such as

$$Q = \sum_i^L [(\bar{x}_i + \sum_j^N S^x_{ij}\theta_j)^2 \ + \ (\bar{y}_i + \sum_j^N S^y_{ij}\theta_j)^2]$$

where the inner sums are usually taken in lowest order in the kicks θ_j and corrector rotational errors ϕ_j to be

$$\sum_i^N S^x_{ij}\theta_j \ = \ \sum_i^M \widetilde{R}_{12}(s_{ij})x'_j \ + \ \sum_{M+1}^N \widetilde{R}_{14}(s_{ij})y'_j \ .$$

The tilde implies either measured or model derived values e.g.

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

† Visitor from KEK, Japan.

Presented at the Particle Accelerator Conference, Washington, D.C., March 16-19, 1987

$$\widetilde{R}_{12} = R_{12}(s_{ij})cos\phi_j + R_{14}(s_{ij})sin\phi_j \approx \sqrt{\beta_{xi}\beta_{xj}}sin(2\pi\Delta\nu_{xij})$$

in terms of the betatron amplitudes and phase advance between i and j. Such assumptions can be periodically checked and compared to previous results and corrected in a variety of ways.

We can group S^x_{ij} and S^y_{ij} to make two L×N matrices S_x and S_y as well as $\bar{x}_i$ and $\bar{y}_i$ into two vectors $\vec{x}$ and $\vec{y}$ of dimension L. Minimizing Q with respect to θ_j then gives the solution vector

$$\vec{\theta} = -(S_x^T S_x + S_y^T S_y)^{-1}(S_x^T \vec{x} + S_y^T \vec{y}) \ .$$

Again, the S matrices and their transposes S^T can be obtained in a number of ways as described below. The orbit vectors come from beam position monitor(BPM) measurements described elsewhere in this conference[3].

An important point is that the θ_j's can be extended or redefined in a number of ways, some of which have nothing to do with angle kicks. Also, we could use $\vec{x}$ and $\vec{y}$ to represent different beams in a linac[4]. We can include any number of ϕ_j's explicitly as well as orbit errors in quadrupoles which give angle kicks. Neither of these are usually included in on-line analyses even though they are important for obtaining and maintaining golden orbits. Similarly, higher and lower order effects can be included e.g. zeroth order offsets which depend on both orbit and configuration and should therefore be done on-line.

An Example

The transport line of the Stanford Linear Collider SLC which takes positrons from the linac to the damping ring at SLAC is a good example since the beam emittance is comparatively large and transmission is important. Figure 1 shows a typical on-line correlation plot of beam position and current from BPM_i versus one corrector θ_j. The error bars come from averaging over several beam pulses which can vary in energy, position, current, and shape etc. Although the data can be filtered on a pulse-to-pulse basis this wasn't done which explains the larger variations in x which is the predominant bending plane. Nevertheless, the correlations are linear and show a small but determinant amount of x-y coupling.

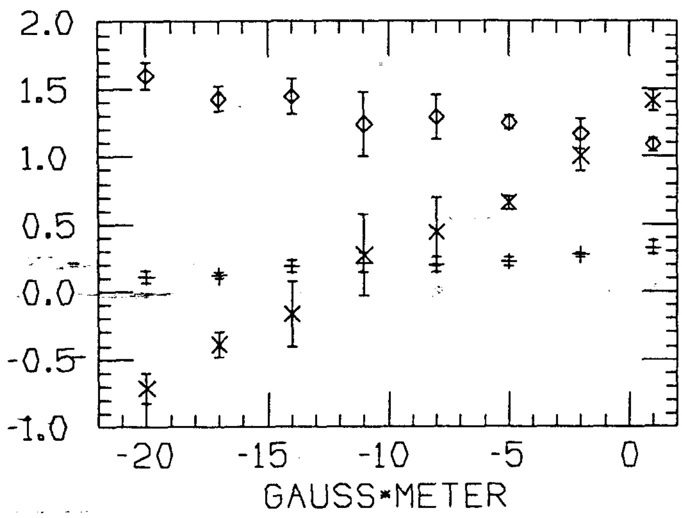

Fig. 1: Typical correlation plot of the mean positions in x(x) and y(+) in mm and the number of positrons per bunch(◇) times 10^{-9} as measured by the BPM at position 605 versus the strength of an x-corrector at position 181 in Fig. 2.

Such measurements, in conjunction with collimators, can also be used to measure profiles and apertures which are the basis for optimizing transmission. They also fix, limit or control initial conditions for measurements such as shown in Fig. 1. Figure 2 shows such correlation data together with various on-line model calculations using COMFORT[5]. Different data points for the same BPM were taken for the same configuration but different runs. All of these R-matrix terms are essentially 'one knob' correlations e.g. R_{16} was obtained using one upstream klystron amplitude.

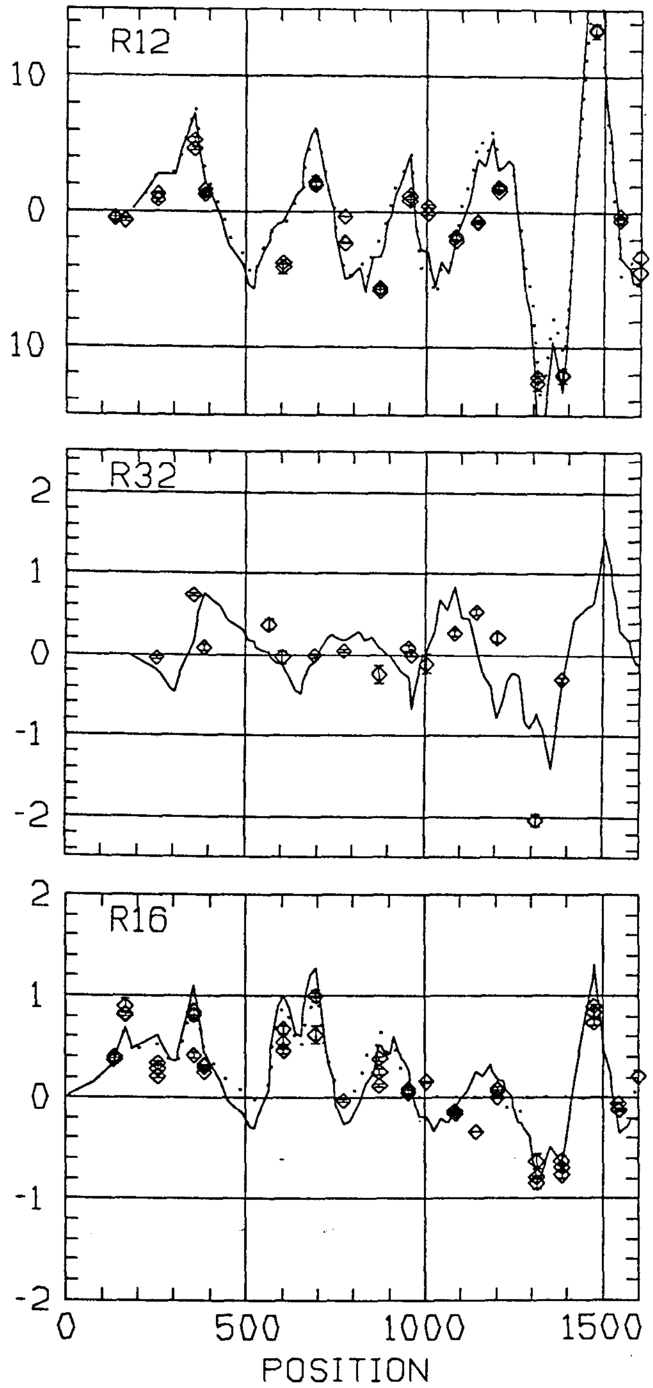

Fig. 2: Typical examples of some measured and calculated R_{ij}'s. All units are in meters. R_{12}'s and R_{32}'s are from the corrector at position 181. R_{16} is the dispersion. In each figure, diamonds show the measured values, solid lines the predictions based on the actual magnet currents and dotted lines the design. The prediction for R_{32} assumes the corrector at position 181 is rotated by $\phi_j = -5°$.

Terms such as R_{11}, R_{13} etc. require two correlated θ_j's related by

$$\frac{\theta_2}{\theta_1} = -\sqrt{\frac{\beta_1}{\beta_2}}[cos(2\pi\Delta\nu_{12}) - \alpha_2 sin(2\pi\Delta\nu_{12})] \ .$$

To do this, we have used a general multi-knob facility which allows one to gang a number of correctors with fixed ratios between their strengths in such a way that they can be varied in unison. Such a capability, when used in conjunction with a flexible correlation plotting and fitting program is a necessary tool for the kinds of AI applications considered here.

Discussion

A general problem for any multiknob program or scheme is the lack of one-to-one correspondence between knobs and the things we are measuring such as the R-matrix elements above. For instance, even if we constrain $\theta_2 = -R_{22}(s_{12})\theta_1$ we will still get some induced dispersion δR_{16} which can introduce errors in our measurements of $R_{11}, R_{13}, etc.$ unless we take it out with software or reduce energy variations. There is also the problem of changing what we are trying to measure by introducing hysteresis errors. This can either be avoided by using air core(or permanent) magnets or else characterized using the beam and AI techiques such as described here.

With due regard for such constraints, one is then ready to compute the rms orbit error as a function of the number of 'correctors' used. Doing this manually, an operator might scan the orbit, subtract it from the golden orbit and then look for what correctors were out of tolerance and tweak the single most likely one. The control program should do very much the same except it would use a relaxation factor based on previous use which is updated and analyzed for each corrector every time it is used. It would also use more than one corrector as well as do a better tolerance analysis. Such an analysis shows how strategic the hardware and it's placement and monitoring can be to overall operation.

With sufficient BPM accuracy, this kind of analysis is easily extended to higher order by successively increasing the magnitude of the excitations θ_j, increasing the dimension of $\vec{\theta}$ and the rank of the matrices. The correlation calculation can give on-line fits to any functional form so one might expect to see nonlinearities appear in plots such as shown in Fig.1 although it is important to guard against such things as bad data from beam loss in the process. A more direct analysis would include measured beam profiles such as the calculations in Ref. 6 can predict. Checks for ill conditioning or whether the resulting system is over or under determined can be used to reduce the rank or determine the most important terms[6].

Some Generalizations

In addition to the specific uses we just described, one can develop a general, multi-knob bump program which can be used in a number of ways e.g. we have used 3 and 4 corrector matched bumps for optimizing injection and extraction past septa, correcting orbit distortions and for aperture studies. If there really was a golden orbit whose measurement was reliable, such a scheme would be a good way to monitor and maintain it since this generalizes what operators already do except it relates errors to their sources more directly and allows constraints such as minimization of corrector strengths to help insure the real sources of error are being corrected.

While the specific optics examples and their refinements are interesting and should be implemented in an on-line facility, our main point is the general utility of such an approach to a broad range of problems which doesn't seem to have been taken advantage of. The R-matrix example and its extensions are specific examples of what can be called a generalized transfer function which links various control variables to the measureable quantities which define the figures of merit. Such an approach appears comparatively simple to implement as an on-line system that could retain its generality and be built-up over time. For instance, the program or operator could select from a menu, the type, order and correlations(ties) to be used based on expert lists of allowed terms or their own prejudices. The calculations would then rank order them with the goal of refining the list to an operating model which provides fast, flexible control with predictive capacity.

One role of operators would be to use such a scheme to monitor and study the functional correlations that improve operation. The generality which is possible here should be sufficient for a number of predilections. Furthermore, the correspondence between hardware and software together with the on-line capability should allow what began as a conventional diagnostic/correction scheme to transform or evolve into a fast-feedback/forward correction system for any disturbances that influence the beam(s) and can be measured. One of many examples, which is particularly important for SLC, would be to characterize those factors which influence the crossings of the two extremely low emittance beams. One possibility would be power supply fluctuations interpreted by a general parametric model of hysteresis whose parameters would be determined by the beam.

Acknowledgments

We should thank many people but particularly Karl Bane and Tom Knight for taking some of the data shown in Fig. 2 as well as Walter Scandale for discussions and work with us on the use of multiknob bumps.

References

1. In some cases one would like higher moments of the beam as a measure of higher order aberrations or stability studies with constraints such as Liapounoff's for a normal distribution in storage rings.

2. K.L. Brown and S.K. Howry, TRANSPORT/360, SLAC-91, 1970.

3. M.C. Ross, Beam Diagnostics and Control for SLC, see proceedings of this conference or SLAC-PUB-4229, 1987.

4. M. J. Lee, et al., Mathematical Models for the Control Program of the SLAC Linear Collider, IEEE Trans. on Nucl. Sci., Vol. NS-28(1981)2155.

5. M.D. Woodley, L. Sanchez-Chopitea and H. Shoaee, Applications of On-line Modelling to the Operation of SLC, see proceedings of this conference or SLAC-PUB-4249, 1987.

6. K.L. Brown and J.E. Spencer, Non-Linear Optics for the Final Focus of the SLC, IEEE Trans. on Nucl. Sci., Vol. NS-28(1981)2568.