Adaptive Method of Closed Orbit Correction

Yao Cheng† and Chen-Shiung Hsue†⋆
†Synchrotron Radiation Research Center
No. 1, R & D Road VI, Science-Based Industrial Park
Hsinchu 300, Taiwan, R.O.C.
⋆ Department of Physics, National Tsing-Hua University
Hsinchu 300, Taiwan, R.O.C.

Abstract

A new adaptive method of closed orbit correction with self-modelling feature is developed for third generation synchrotron light source with many insertion devices. The quasi-Newton method is applied iteratively to minimize the closed orbit distortions and to approach the optimal control. The information from iterative position measurements is used to update the response matrix adaptively. A roughly estimated matrix from optic functions can initialize the correction scheme. The convergence rate is at least superlinear. Hence, the adaptive method may increase the efficiency of fast closed orbit correction. The noise immunity of this method is also considered. Particularly in the case of changing optic functions, this approach provides the searching ability of optimal model.

1 Introduction

The closed orbit correction is one of the basic daily operations in the control routines of every accelerator. The performace of the closed orbit control is important for the beam stability and the flexibility of machine studies. Since the recent progress in the control system supports the online modelling control[1][2], an optimal self-modelling control routine free from the working condition will be the interested subject in the future.

There are many different approachs of the closed orbit correction[3]. No matter which method of correction is used, the basic algorithm is to calculate the new corrector settings from the linear response of the machine and the measured orbit distortions. The orbit measurement starts this closed loop iteratively, untill the orbit distortion converges to a stable minimum. If the machine is linear and the model response is exact, the procedure shall terminate at the first iteration. Usually, we need more than one iteration for the correction. It means that the model response is not exact, since the nonlinear effect near the center orbit is negligible. If a improper model response is used, the iterations could even drive the orbit into the nonlinear region and cause the run-away of machine working point. Since the model response estimated from the theoretical optic functions may be different from the actual values in real operation, many laboratories use the measured response in the model function.

If we find a proper model reponse which leads to the convergence of iterations, the convergence rate and the residual decide whether the model is optimal. Particularly in the fast feedback system, faster convergence rate and smaller residual will improve the efficiency of the system. In section 3 of this paper, we will analyze these criteria for a general approach.

In real machine, the response function is a function of working point and working condition. It may take a long time to find the true response for a certain working point. During the commissioning periode or machine studies, frequent change of working point inhibits the search of the true response. An adaptive self-modelling procedure provides the searching ability of optimal model. It should be helpful for the commissioning and machine studies.

2 Model Function

A general model function to be minimized for the closed orbit distortions by iteration in the least square sense may be written as

$$F = (\widetilde{Y + AX})P(Y + AX) + \gamma\tilde{X}X, \qquad (1)$$

where $\tilde{A}$ denotes the transpose of the matrix A and:

- Y is n-dimensional vector and its element Y_i is the measured closed orbit distortion at station i.
- X is m-dimensional vector and its element X_j is the current setting of corrector j.
- A is the model response matrix $(n \times m)$ between current and distortion.
- P is diagonal matrix of the positive weight factor.
- γ is the positive weight connected to the limit of total corrector current.

The $n \times m$ matrix elements of the response matrix A can be calculated from the optic functions or measured from real machine. After the measurements of the closed orbit distortions $Y^{(k)}$ at kth iteration, the current settings $X^{(k+1)} = X^{(k)} + \sigma\delta X^{(k)}$ are determined uniquely by the Newton step,

$$\delta X^{(k)} = -(\tilde{A}PA + \gamma I)^{-1}\tilde{A}PY^{(k)}. \qquad (2)$$

I indicates the identity matrix. $\delta X^{(k)}$ is multiplied by a selected factor σ to increase the convergence rate and

0-7803-0135-8/91$03.00 ©IEEE

reduce the oscillation or to increase the convergence radius. The positive constant γ is used to avoid the limiting cut of high current settings. It also lets the matrix $\tilde{A}PA+\gamma I$ to be positive definite. Therefore, the inverse of matrix involved always exists. The diagonal positive weight factor P provides the freedom to select suitable stations for extra minimization.

This Newton step will be repeated iteratively with a fixed model response A, until a stable minimum of the orbit distortions is reached. In the next section, we formulate the criterion of convergence and the criterion of optimal orbit control.

3 Convergence Rate

Let us assume that the true linear coefficients between stations and corrector currents are T_{ij}. Here, we neglect the nonlinear part. The residual of initial $\tilde{A}PY^{(0)}$ after kth iteration of correction may be written as

$$\begin{aligned} Y^{(k)} &= Y^{(k-1)}+\sigma T\delta X^{(k-1)}, && (3)\\ \tilde{A}PY^{(k)} &= \left[I-\sigma\tilde{A}PT(\tilde{A}PA+\gamma I)^{-1}\right]^{k}\tilde{A}PY^{(0)}. && (4)\end{aligned}$$

In the case of $A=T$, a similar transformation can always change the symmetric matrix $\tilde{A}PA$ to the diagonal form. The equation (4) becomes

$$\tilde{A}PY^{(k)} = S_a\left[diag\,(1-\sigma\lambda_i^a)^k\right]S_a^{-1}\tilde{A}PY^{(0)}, \quad (5)$$

where $\lambda_i^a=\lambda_i/(\lambda_i+\gamma)$ and λ_i are the semipositive eigenvalues of the matrix $\tilde{A}PA$. The norm of the residuals converges linearly to zero for $0<\sigma<2/max(\lambda_i^a)$. If there are some $\lambda_i^a=0$, we can lower the size of the matrix, or select the most efficient correctors.

In the case of $A\neq T$, we transform the matrix $\tilde{A}PT(\tilde{A}PA+\gamma I)^{-1}$ to Jordan form $J(\lambda_i^t)$ with eigenvalues λ_i^t,

$$\tilde{A}PY^{(k)} = S_t\left[I-\sigma J(\lambda_i^t)\right]^k S_t^{-1}\tilde{A}PY^{(0)}. \quad (6)$$

Note that, only the condition $0<\sigma\lambda_i^t<2$ will guarantee $X^{(k)}$ to converge linearly to the minimizer X^*. Since λ_i^t are close to unity, σ has to be positive. If there is any $\lambda_i^t<0$, no σ exists for the convergence. Such a model response matrix A is certainly improper.

The iterations can always find the minimizer succesfully with a proper model response. Then, the question arises here: "Is this result the optimal correction in the $A\neq T$ case?" The answer is "No". We find only the minimun of the model fuction F. For the optimal correction, we should replace the matrix A by the true linear response T in the model function F and then search for the minimun.

4 Quasi-Newton Method

The quasi-Newton method is a modification of the Newton method[4]. It is almost the only method in a modern approach of unconstrained nonlinear optimization problem. We apply this approach on this adaptive method not because of the nonlinear feature of the orbit correction but for the searching of true response matrix T. Futhermore, if we let $\lim_{k\to\infty}A^{(k)}=T$, the equation (6) transfers to the equation (5) and the eigenvalues λ_i^t approach unity, in the case of $\gamma=0$, after iterations. The convergence rate is increased to be superlinear for $\sigma=1$. That is to say

$$\lim_{k\to\infty}\frac{\|\tilde{A}^{(k+1)}PY^{(k+1)}\|}{\|\tilde{A}^{(k)}PY^{(k)}\|}=0. \quad (7)$$

Hence, we propose $\lim_{k\to\infty}\gamma\to 0$ and $\sigma=1$ in the quasi-Newton approach. The role of σ will be replaced by another constant τ in the later development of this adaptive method.

There is an easy way to understand such an application on the closed orbit correction. We adjust only one corrector current with δX_j. After correction, the vector of position change δY_i divided by δX_j becomes one row vector of the new response matrix. As a matter of fact, we have measured one row vector of matrix. Let the matrix M to be the change,

$$\delta Y_i=(A_{ij}+M_{ij})\delta X_j. \quad (8)$$

This equation (8) is called quasi-Newton condition (the spirit of quasi-Newton method). We say that the feedback information from the measurements is used to correct the response matrix. If several correctors change current simutaneously, then the problem becomes complicated. There are more than one method to update the matrix. However, the quasi-Newton condition must be always hold.

Following, we consider a rank-one matrix $M^{(k)}=u^{(k)}\tilde{v}^{(k)}$ at kth iteration. It is the direct product of a n-dimensional column vector $u^{(k)}$ and a m-dimensional row vector $\tilde{v}^{(k)}$. From the quasi-Newton condition, we have

$$\begin{aligned} u^{(k)} &= \frac{\delta Y^{(k)}-A^{(k)}\delta X^{(k)}}{\tilde{v}^{(k)}\delta X^{(k)}}\\ &= \frac{Y^{(k+1)}-Y^{(k)}-A^{(k)}(X^{k+1}-X^{(k)})}{\tilde{v}^{(k)}(X^{k+1}-X^{(k)})}, && (9)\end{aligned}$$

where $A^{(k)}=A^{(k-1)}+M^{(k-1)}$ is the updated matrix of last iteration. This form takes any vector v of nonzero denominator. To avoid $\tilde{v}^{(k)}\delta X^{(k)}=0$, the vector $v^{(k)}=\delta X^{(k)}$ is chosen. This update is indeed the Broyden's formula[5]

$$M^{(k)} = \frac{(\delta Y^{(k)}-A^{(k)}\delta X^{(k)})\widetilde{\delta X^{(k)}}}{\widetilde{\delta X^{(k)}}\delta X^{(k)}}. \quad (10)$$

This is the most successful update when there are no special feature of matrix A to be reflected[6]. The norm of the matrix $(T-A^{(k)})$ converges to zero with a rate as fast as the rate of convergence of $X^{(k)}$ to the minimizer X^*[4]. It is known that $X^{(k)}$ converge superlinearly with Broyden's update. The convergence radius with the initial matrix $A^{(0)}$ is, in principle, unlimited. If the T matrix is exactly linear, there exists only one minimizer.

Considering the onset of noise in the measurements, the update $M^{(k)}$ may be incorrect. Therefore, we insert a positive constant $\tau \in (0,1)$ to multiply the update and let

$$A^{(k+1)} = A^{(k)} + \tau M^{(k)}. \tag{11}$$

This factor plays a role like a digital filter that damps the noise.

In the next quasi-Newton step,

$$\begin{aligned}\delta X^{(k+1)} &= -(\tilde{A}^{(k+1)} P A^{(k+1)})^{-1} \tilde{A}^{(k+1)} P Y^{(k+1)} \\ &= -H^{(k+1)} \tilde{A}^{(k+1)} P Y^{(k+1)},\end{aligned} \tag{12}$$

it is not necessary to calculate the time-consuming inversion except for the initial inversion $H^{(0)} = \left(\tilde{A}^{(0)} P A^{(0)}\right)^{-1}$. We change the quasi-Newton condition of equation (8) for the symmetrical inverse matrix $H^{(k+1)}$

$$\delta X^{(k)} = H^{(k+1)} \delta Z^{(k)} \tag{13}$$

with

$$\delta Z^{(k)} = (\tilde{A}^{(k)} + \tilde{M}^{(k)}) P \delta Y^{(k)}. \tag{14}$$

The BFGS formula updates the inverse matrix directly[4],

$$\begin{aligned}H^{(k+1)} &= H^{(k)} + \tau \frac{\delta Z^{(k)} \widetilde{\delta Z^{(k)}}}{\widetilde{\delta Z^{(k)}} \delta X^{(k)}} \\ &\quad - \tau \frac{H^{(k)} \delta X^{(k)} \widetilde{\delta X^{(k)}} H^{(k)}}{\widetilde{\delta X^{(k)}} H^{(k)} \delta X^{(k)}}.\end{aligned} \tag{15}$$

The BFGS update preserves the symmetric form of matrix. The updated matrix remains positive definite, if and only if $\widetilde{\delta Z^{(k)}} \delta X^{(k)} > 0$. It means that the positive sequence

$$\widetilde{\delta Z^{(k)}} \delta X^{(k)} = \widetilde{\delta Y^{(k)}} P \delta Y^{(k)} \tag{16}$$

lead to the positive definite sequence of $H^{(k)}$. The inversion of a updated matrix needs $O(m^3)$ floating point computing operation. The BFGS update consumes only $O(m^2)$ floating point computing time. It is the most successful way to update a symmetric inverse matrix[6].

We don't use the same formula uniformly to update $A^{(k)}$ and $H^{(k)}$, because they have the different features. Dennis *et al* showed that both formula are the best update with associated weighting norm in the sense of least change[6]. Since we have the freedom to choose the weight P, there is no contradiction of unequal treatments.

We summarize the whole procedure from the begining as follows

1. Assume that $A^{(0)}, H^{(0)}, Y^{(0)}, X^{(0)}$ are the initial values.
2. From the equation (12), find the $\delta X^{(k)}$ for new current setting $X^{(k+1)}$.
3. After correction, measure the new orbit distortions $Y^{(k+1)}$ and check its convergence for the termination of iterations.
4. use $\delta Y^{(k)} = Y^{(k+1)} - Y^{(k)}$ to calculate the update $A^{(k+1)}$ with formulae (10) and (11).
5. use equation (14) to calculate the update $H^{(k+1)}$ with formula (15) and close the loop here to the step 2.

5 Concluding Remark

The Broyden and BFGS updates have been studied very intensively during the seventies and become slowly the standard algorithms in the nonlinear programming and optimal control. As far as we know, this paper is the first attempt of using this kind of combined algorithms on the closed orbit correction towards an intelligent accelerator control system.

Applying the quasi-Newton method, we can find the true response matrix. The solution of optimizing the model function should be the best correction. Another advantage of this method is to avoid the measurement of A matrix, which is time-consuming and unacceptable for an electron storage ring such as SRRC, if this task has to be executed frequently. We start the correction with the response matrix $A^{(0)}$ estimated from the optic functions. This adaptive method can correct the matrix by itself in the iterations. Even if the optic functions are changed, the method adjusts the response matrix automatically.

The BFGS update keeps the matrix in equation (15) to be positive definite. It is not necessary to pay extra care on the possible sigularity. The BFGS update reduces the time to inverse matrix. The feature of $O(m^2)$ operations instead of $O(m^3)$ operations is benificial and could be essential for the use of the adaptive method in the fast feedback correction.

References

[1] Jan, G.J., "Beam Position Measurement System for SRRC", This conference.

[2] Juan, P., Darpentigny, J., Martin, P.C. and Nghiem, P., "Closed Orbit Measurement and Correction at Super-ACO", Super-ACO/90-27, Orsay, June 1990.

[3] Koutchouk, J.-P. "Trajectory and Closed Orbit Correction" Frontiers of Particle Beams; Observation, Diagnosis and Correction, M. Month, S. Turner(Eds), Springer-Verlag Berlin Heidelberg 1989.

[4] Dennis, J.E. JR, and More, J.J., "Quasi-Newton Methods, Motivation and Theory", SIAM Review, vol 19, No. 1, January 1977.

[5] Broyden, C.G., "A class of methods for solving nonlinear simutaneous equations' Math. Comp., 19, pp. 577-593, 1965.

[6] Dennis, J.E. JR, and Schnabel, R. B., "Least Change Secant Updates for Quasi-Newton Method", SIAM Review, vol 21, No. 4, October 1979.