ISBN: 978-3-95450-208-0 doi:10.18429/JACoW-IPAC2019-MOPGW114

BAYESIAN APPROACH FOR LINEAR OPTICS CORRECTION

Yongjun Li*, Robert Rainer, Weixing Cheng
Brookhaven National Laboratory, Upton, NY-11973, USA

Abstract

With a Bayesian approach, the linear optics correction algorithm for storage rings is revisited. Starting from the Bayes' theorem, a complete linear optics model is simplified as "likelihood functions" and "prior probability distributions". Under some assumptions, the least square algorithm and then the Jacobian matrix approach can be re-derived. The coherence of the correction algorithm is ensured through specifying a self-consistent regularization coefficient to prevent overfitting. Optimal weights for different correction objectives are obtained based on their measurement noise level. A new technique has been developed to resolve degenerated quadrupole errors when observed at a few select BPMs. A necessary condition of being distinguishable is that their optics response vectors seen at these specific BPMs should be near-orthogonal.

INTRODUCTION

At modern particle accelerator facilities, advanced beam diagnostics instruments with high acquisition rate can generate copious amounts of data within a short time period. A specific example would be obtaining beam turn-by-turn (TbT) data from beam position monitors (BPM) after the beam is disturbed. With a Bayesian approach, the linear optics correction algorithm for storage rings is revisited [1]. The linear optics functions, such as, the envelope function β of betatron oscillation and its phase ϕ [2], can be extracted [3–5]. Due to various measurement noise, accurately identifying quadrupole error sources is important for optics correction. One can average over repetitive measurements, then use the mean values directly. Distributions of measurement noise, which are usually ignored, however, can provide rich information for identifying error sources precisely. Using a Bayesian approach and the information provided by the error analysis, the linear optics correction problem presented by accelerators can be approached from the viewpoint of probability.

Lattice measurement noise and quadrupole excitations errors are usually randomly distributed around their expectation values. Overfitting quadrupole errors must be avoided. Specifically, the optics functions β and ϕ can be measured with BPMs at many locations s_i, where $i = 0, 1, \cdots, N-1$, and N is the total number of BPMs. Given a set of measured data with noise, fitting the actual quadrupole errors ΔK, is a typical nonlinear regression problem since the dependence of β and ϕ on K is nonlinear. In regression problems, overfitting is a modeling error which occurs when a function is too closely fit to a limited set of data points [6, 7]. There are two reasons of revisiting this problem with a Bayesian approach. First, the Bayesian approach is a proven technique in preventing overfitting. Second, several optics distortions caused by quadrupole errors need to be corrected simultaneously, but measured in different units and scales. With the Bayesian approach, the coherence of the correction algorithm, which is capable of dealing with multi-objective regression problems, can be established.

In some scenarios, an optics distortion pattern is indeed caused by a single quadrupole error rather than normally distributed errors. However, the goal of the Bayesian approach is to distribute the error to multiple sources. It can sometimes fail to distinguish the single source from its highly degenerated neighbors. A new technique has been developed where only a few specific BPMs are selected to address the degeneracy. One necessary condition for being distinguishable is that the optics response vectors of those specific BPMs should be near-orthogonal.

BAYESIAN APPROACH

From the viewpoint of probability, identifying quadrupole errors from repetitive and independent measurements can be achieved by computing a posterior conditional probability distribution and determining its maxima. Consider a simple case of β function in the horizontal plane. Based on the Bayes' theorem, the conditional probability of having an error $\Delta \boldsymbol{K}$ with a measured $\boldsymbol{\beta}$ reads as [6]

$$\begin{aligned} p(\Delta \boldsymbol{K}|\boldsymbol{\beta}) &= \frac{p(\boldsymbol{\beta}|\Delta \boldsymbol{K})p(\Delta \boldsymbol{K})}{p(\boldsymbol{\beta})} \\ &\propto p(\boldsymbol{\beta}|\Delta \boldsymbol{K})p(\Delta \boldsymbol{K}). \end{aligned} \tag{1}$$

Eq. (1) can be interpreted as, given a measured optics distortion $\boldsymbol{\beta} = \boldsymbol{\beta_0} + \Delta\boldsymbol{\beta}$, the probability of it being the error source of $\Delta \boldsymbol{K}$ is proportional to the product of a likelihood function $p(\boldsymbol{\beta}|\Delta \boldsymbol{K})$ and a probability distribution of error $\Delta \boldsymbol{K}$. The likelihood function can be recognized as being related to the dependence of β on K, i.e., the Jacobian matrix. $p(\Delta \boldsymbol{K})$ is known as prior probability distribution which will be covered in greater detail later. $p(\boldsymbol{\beta})$ is the normalizing constant.

By maximizing the probability in Eq. (1), the most likely quadrupole error distribution can be obtained. In general, we can assume that both β measurement noise and quadrupole excitation errors are normally distributed. For example, at a particular BPM, repetitive measurement of βs are distributed around an expectation value $\mathbb{E}(\beta) = \bar{\beta}$ with a variance σ_β.

$$\mathcal{N}(\beta|\bar{\beta}, \sigma_\beta^2) = \frac{1}{\sqrt{2\pi}\sigma_\beta} \exp\left[-\frac{(\beta-\bar{\beta})^2}{2\sigma_\beta^2}\right]. \tag{2}$$

Eq. (1) thus can be re-written as

$$p(\Delta \boldsymbol{K}|\boldsymbol{\beta}) \propto \mathcal{N}(\boldsymbol{\beta}|\bar{\boldsymbol{\beta}}(s_i, \Delta \boldsymbol{K}), \sigma_\beta^2) \cdot \mathcal{N}(\Delta \boldsymbol{K}|\boldsymbol{K}_0, \sigma_K^2). \tag{3}$$

* yli@bnl.gov

ISBN: 978-3-95450-208-0 doi:10.18429/JACoW-IPAC2019-MOPGW114

where σ_K is the variance of quadrupole error distribution. Maximizing the probability of Eq. (3) is equivalent to minimizing its negative logarithm,

$$-\ln\left[p(\Delta\boldsymbol{K}|\boldsymbol{\beta})\right] \propto \frac{1}{2\sigma_\beta^2}\sum_i\left[\beta(s_i,\Delta\boldsymbol{K})-\bar{\beta}(s_i)\right]^2+\frac{1}{2\sigma_K^2}\|\Delta\boldsymbol{K}\|^2. \quad (4)$$

Here $\|\bullet\|$ is the Euclidean norm of a vector. Eq. (4) can be recognized as the least-square algorithm but with some well-defined weights. It is important to note that, since we assume a normal distribution for quadrupole errors, the solution to Eq. (4) is intended to allocate errors according to a normal distribution, even if they are not. If there is a systematic calibration error on the quadrupole excitations, the second distribution in Eq. (3) has an non-zero mean. But after the first several iterations, the non-zero mean value should be able to be filtered out. A more detailed discussion on a single outlier of quadrupole error will be addressed the next section.

Now we take a look at the first term on the right-hand side of Eq. (4). By expanding β with respect to quadrupole errors $\Delta\boldsymbol{K}$ at β_0 and keeping the linear components, it reads as

$$\begin{aligned}&\frac{1}{2\sigma_\beta^2}\sum_i\left[\beta_0(s_i)+\frac{\partial\beta(s_i)}{\partial\boldsymbol{K}}\Delta\boldsymbol{K}-\bar{\beta}(s_i)\right]^2\\ =\;&\frac{1}{2\sigma_\beta^2}\sum_i\left[\frac{\partial\beta(s_i)}{\partial\boldsymbol{K}}\Delta\boldsymbol{K}-\Delta\bar{\beta}(s_i)\right]^2, \quad (5)\end{aligned}$$

where $\Delta\bar{\beta}(s_i)=\bar{\beta}_0(s_i)-\beta_0(s_i)$. After differentiating every term with respect to $\Delta\boldsymbol{K}$, it can be expressed in the format of a matrix. The solution is given as

$$\Delta\boldsymbol{K}=\left[\boldsymbol{M}^T\boldsymbol{M}\right]^{-1}\boldsymbol{M}^T\Delta\bar{\boldsymbol{\beta}}. \quad (6)$$

$\left[\boldsymbol{M}^T\boldsymbol{M}\right]^{-1}\boldsymbol{M}^T$ is often known as the pseudo inverse of $\boldsymbol{M}$, because $\boldsymbol{M}$ is usually non-invertible.

Thus far, the measurement noise σ_β has been ignored. The solution to Eq. (6) often overfits quadrupole errors from either noisy BPM data, or even bad BPMs if they are present. The overfitting can be mitigated by taking the second term into account, which is known as regularization technique. By adding an additional penalty term to the sum of squares in Eq. (4), one can prevent the fitted quadrupole errors from deviating from a reasonable normal distribution. In other words, a complete linear optics model provides not only a likelihood function but an informative prior probability distribution of quadrupole errors as well. The solution to the least-squares problem with regularization is

$$\Delta\boldsymbol{K}=\left[\boldsymbol{M}^T\boldsymbol{M}+\lambda\boldsymbol{I}\right]^{-1}\boldsymbol{M}^T\Delta\bar{\boldsymbol{\beta}} \quad (7)$$

It is important to note that the optimal regularization coefficient $\lambda=\frac{\sigma_\beta^2}{\sigma_K^2}$ is well-defined here. More specifically, the variance $\sigma_k(\Delta\bar{\beta})$ of the quadrupole error distribution $p(\Delta\boldsymbol{K})$ should be determined by the measured β-beat level using the designed lattice model. Figure 1 illustrates that the horizontal β-beat is linearly proportional to the variance of quadrupole error distribution at the NSLS-II ring. After averaging repetitive β_x measurements and comparing against the nominal $\beta_{x,0}$, the variance of quadrupole error probability distribution $\sigma_k(\Delta\bar{\beta})$ can be determined with Fig. 1. During an iterative correction, β-beat reduces gradually, as do the corresponding quadrupole errors. Therefore the regularization coefficient should be dynamically adjusted to speed up the convergence as well. The $p(\Delta\boldsymbol{K})$ is named as the prior probability because it can be estimated analytically [2] or numerically in advance. In the previous section, one can still use the regularization technique to avoid overfitting, but the coefficient is not necessarily optimal due to lack of a theoretical basis. Experimentally one can obtain this regularization coefficient based on correction performance on a trial basis [8]. However, the Bayesian approach can explicitly give its statistic and physics interpretation.

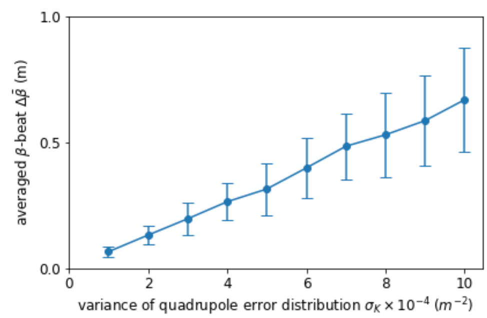

Figure 1: Statistic illustration of the horizontal β-beat due to the random quadrupole errors. This linear correlation with gradually increasing variance are calculated with the NSLS-II ring lattice model in advance. Once an averaged β-beat is measured, its corresponding variance of quadrupole error distribution can be used as the prior probability to prevent overfitting

RESOLVING DEGENERACY

In the previous section, we discussed the case in which the lattice distortion is due to normally distributed quadrupole errors. Once a real error is localized in a particular quadrupole, it may require us to identify which quadrupole is the root cause. This is nontrivial because quadrupoles are closely packed in modern storage rings, the NSLS-II being no exception. Therefore, their lattice response vectors (corresponding columns in $\boldsymbol{M}$) are often highly degenerated, especially between neighboring quadrupoles. The degeneracy between the i^{th} and j^{th} quadrupole is defined by the correlation coefficient [8]

$$C_{i,j}=\frac{\boldsymbol{m}_i\cdot\boldsymbol{m}_j}{\|\boldsymbol{m}_i\|\,\|\boldsymbol{m}_j\|}, \quad (8)$$

where $\boldsymbol{m}_i$ is the i^{th} column of $\boldsymbol{M}$, which has N elements. If $|C_{i,j}|$ approaches 1, it is difficult to distinguish which one is

ISBN: 978-3-95450-208-0 doi:10.18429/JACoW-IPAC2019-MOPGW114

the actual error source with a full Jacobian matrix. It was found that rather than using all BPMs, and instead selecting a few specific BPMs among them, the highly degenerated quadrupoles were distinguishable.

Consider that there are N BPMs. The β-beats seen by these BPMs are N-dimension vectors. Among them, n ($n \ll N$) components can be selected to form two much shorter sub-vectors $\boldsymbol{v}_{i,j}$ in a such way that $\boldsymbol{v}_{i,j}$ have much less correlation between them. This means they should be as near-orthogonal as possible in an n dimensional vector space. There are $N!/(n!(N-n)!)$ different permutations to select from. We found that it is not difficult to distinguish 5-6 BPMs out of 180 BPMs in the NSLS-II ring even if the correlation between some neighboring quadrupoles is above 0.98. Experimentally, we repetitively measure the lattice functions. Then we compute the correlation coefficients between $\boldsymbol{v}_{i,j}$ and the measured lattice distortion patterns, $\boldsymbol{u}$, as seen only at those 5-6 specific BPMs. If the quadrupole error was due to the i^{th} quadrupole, the correlation coefficients $C_{v_i,u} = \frac{\boldsymbol{v}_i \cdot \boldsymbol{u}}{\|\boldsymbol{v}_i\|\,\|\boldsymbol{u}\|}$ should be distributed close to ± 1. Another one, $C_{v_j,u} = \frac{\boldsymbol{v}_j \cdot \boldsymbol{u}}{\|\boldsymbol{v}_j\|\,\|\boldsymbol{u}\|}$ should be around zero, and vice versa.

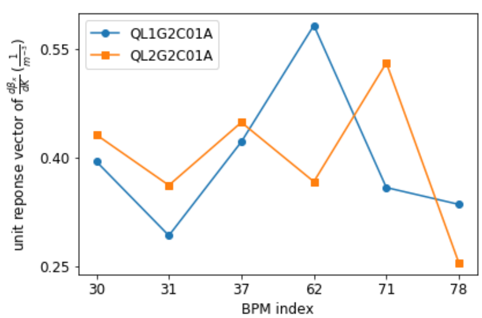

Figure 2: The unit β_x responses vectors of quadrupole 10 and 11 seen at 6 selected BPMs. They are near-orthogonal because their correlation is as low as 0.0612.

To verify this technique, an experiment and a simulation study were carried out on the NSLS-II ring. The excitation current of one quadrupole `QL1G2C01A` (with an index of 10) was changed by 1 Ampere. The bunch-by-bunch feedback system [9] was then used to resonantly drive the beam to perform betatron oscillation at a nearly constant amplitude. Beam TbT data was acquired for 800×1024 turns. For every 1024 turns of data, a set of β_x functions was extracted at 180 BPMs. After averaging them, an error distribution was fitted out with the Bayesian approach. It was found that the maximum error is not `QL1G2C01A` as it should be, but its neighbor `QL2G2C01A` (with an index 11)). It is not surprising because the correlation between the 10^{th} and 11^{th} columns of the Jacobian $\boldsymbol{M}$ is as high as 0.9896.

Among 180 BPMs, we specifically selected 6 of them with their indices as {30, 31, 37, 62, 71, 78}. Observed at these BPMs, the unit β_x functions response to these two quadrupoles is near-orthogonal with a correlation coefficient as low as 0.0612 (see Fig. 2). 800 independently measured β_x-beat patterns at those specific 6 BPMs were compared against these two unit response vectors $\boldsymbol{v}_{10,11}$. The histograms of their correlation coefficients are illustrated in Fig. 3. It becomes clear that the β_x distortion is likely due to the quadrupole `QL1G2C01A` rather than its neighbor `QL2G2C01A`, because the measured optics distortion pattern is highly correlated with its response vector.

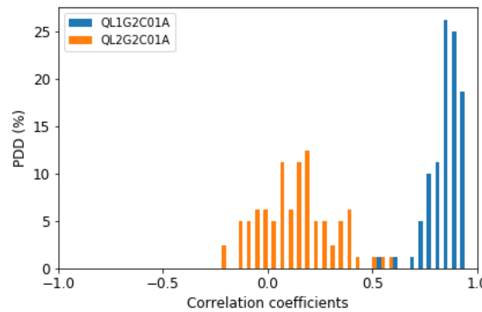

Figure 3: The probability density distribution (PDD) of the correlation coefficients between 800 measured β_x distortion and two unit vectors $\boldsymbol{v}_{10,11}$. The independently and repeatedly measured β-beats are highly correlated with quadrupole 10's pattern, rather than its neighbor. Based on that we can conclude that the actual error source is more likely from the quadrupole 10 (`QL1G2C01A`) instead of quadrupole 11 (`QL2G2C01A`).

ACKNOWLEDGEMENTS

This work was supported by the U.S. Department of Energy under Contract No. DE-SC0012704.

REFERENCES

[1] Y. Li, R. Rainer, and W. Cheng, "Bayesian approach for linear optics correction", *Phys. Rev. Accel. Beams*, vol. 22, p. 012804, 2019.

[2] E. Courant, and H. Snyder, "Theory of the alternating-gradient synchrotron", *Annals Phys.*, vol. 3, pp. 1–58, 1958.

[3] P. Castro-Garcia, "Luminosity and beta function measurement at the electron-positron collider ring LEP", CERN, Geneva, Switzerland, Rep. CERN-SL-96-070-BI, 1996.

[4] J. Irwin *et al.*, "Model-Independent Beam Dynamics Analysis", *Phys. Rev. Lett.*, vol. 82, p. 1684, 1999.

[5] X. Huang *et al.*, "Application of independent component analysis to Fermilab Booster", *Phys. Rev. ST Accel. Beams*, vol. 8, p. 064001, 2005.

[6] C. Bishop, "Pattern Recognition and Machine Learning", Springer, 2006.

[7] Y. Abu-Mostafa *et al.*, "Learning from data", AMLBook, 2012.

[8] X. Huang *et al.*, "LOCO with constraints and improved fitting technique", *ICFA Beam Dyn. Newslett.*, vol. 44, pp. 60–69, 2007.

[9] W. Cheng, B. Bacha, Y. Li, and D. Teytelman, "Developments of bunch by bunch feedback system at NSLS-II storage ring", in *Proc. 9th Int. Particle Accelerator Conf. (IPAC'18)*, Vancouver, BC, Canada, Apr. 2018, pp. 1833–1836, doi:10.18429/JACoW-IPAC2018-WEPAF014