Beam-based correction methods for Accelerator Middle Layer

06/20/2024

Python Accelerator Middle Layer Workshop (DESY)

Presented remotely

Xiaobiao Huang

SLAC National Accelerator Laboratory

https://indico.desy.de/event/43233/

U.S. DEPARTMENT OF ENERGY Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

Outline

- Beam-based correction for accelerators
- Closed-orbit based linear optics correction methods
  - LOCO and LOCOM
  - Mitigation of over-fitting

Not discussing methods based on turn-by-turn BPM data due to time consideration.

- Implementation considerations
- Summary

Errors in real accelerators

- Real accelerators differ from the designs

Accelerator design ← Physics processes may be omitted, or modeled with simplified approach

↓

Build the machine according to design ← Mechanical (machining and alignment) and magnetic errors.

↓

Set set-points to design values ← Calibration errors. Measurement errors

↓

Machine in reality — Machine as built does not perform exactly as designed.

Beam-based methods: Correction vs. Optimization

- Beam-based methods are used to restore accelerator performance
  - Two categories: correction or optimization
- Beam-based correction:

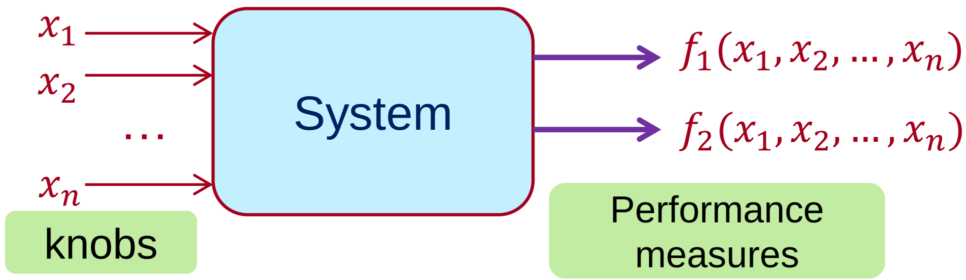
Correct the operating condition of a subsystem toward a target condition through beam based measurements and a deterministic procedure.

	Actuators (knobs)	Diagnostics (monitors)	Deterministic method	Target
Orbit correction	Orbit correctors	BPMs	Orbit response matrix	Ideal orbit
Optics correction	Quadrupole correctors	Beta, phase advance, orbit response matrix	Response (Jacobian) matrix	Design optics

Beam-based optimization (online tuning)

- When correction methods are not available
  - E.g., target condition is not determined, not enough diagnostics, or no correction method is established
- Beam-based optimization

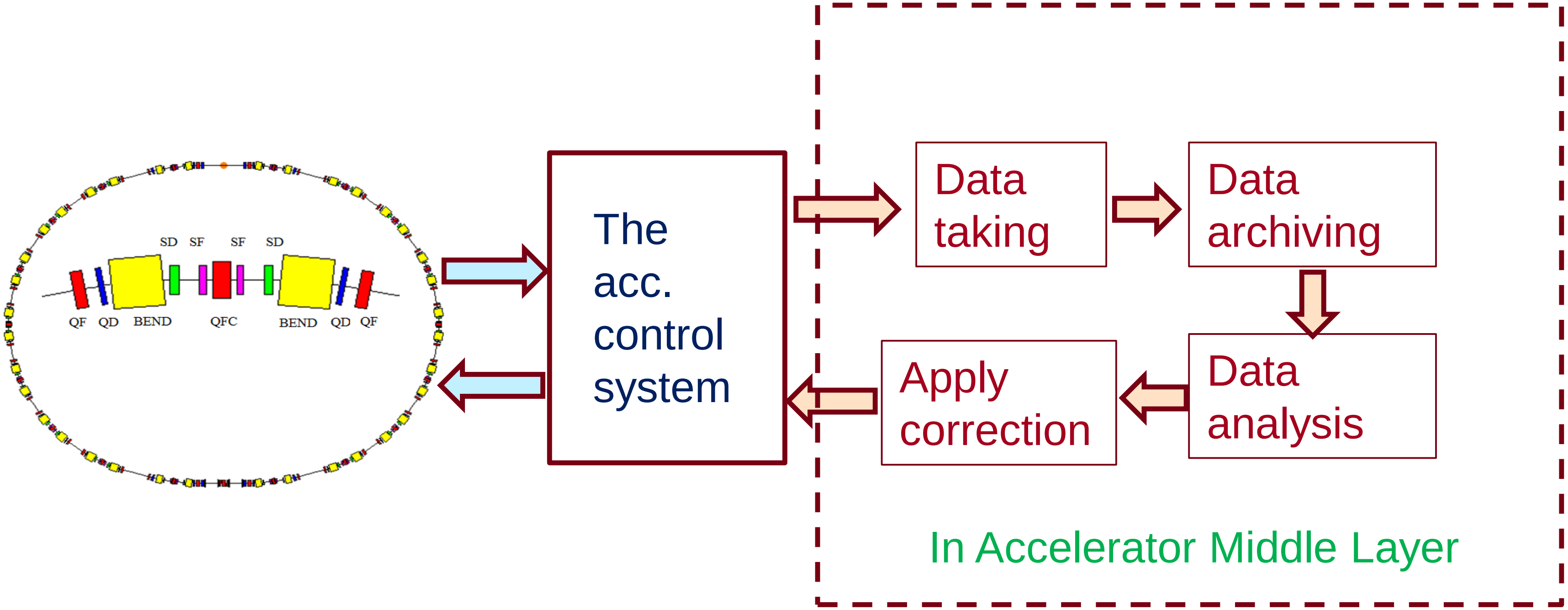
Adjust the operating condition to optimize machine performance directly.



We know the system works – changing input leads to performance responses.
But we don't (need to) know **exactly** how it works – the functions are unknown.
(But system knowledge could be used to speed up the tuning)

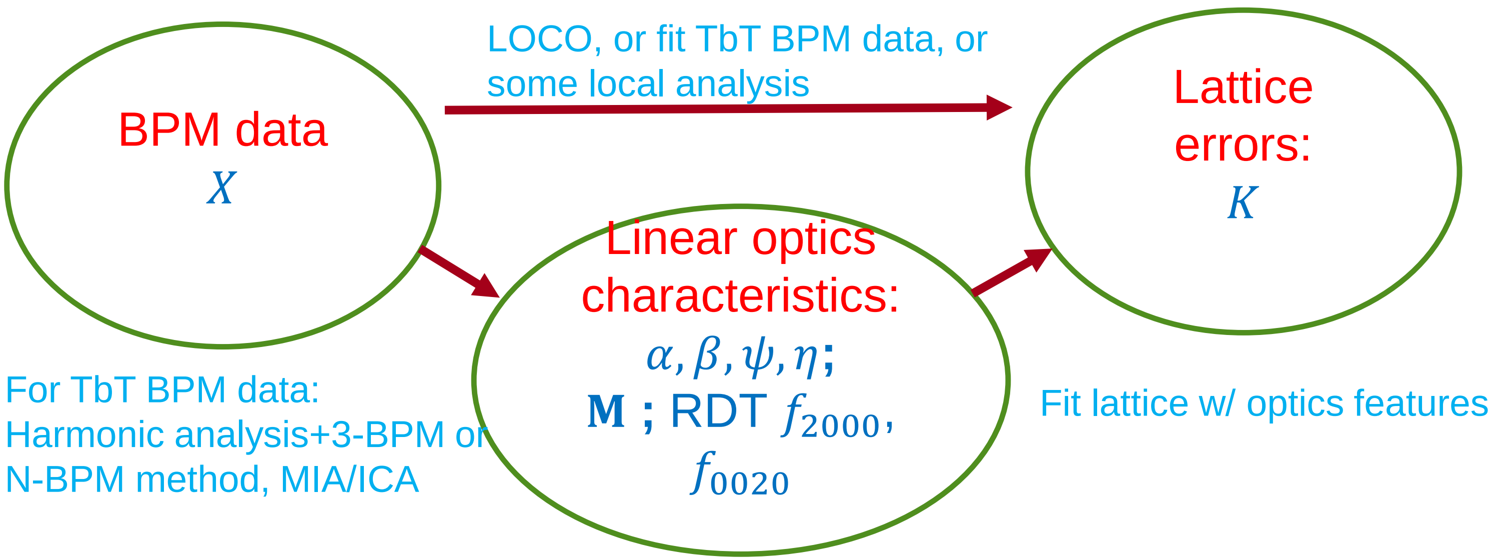
The usual data flow for beam-based correction

- The process involves data taking, data analysis, application of correction, followed by verification or more iterations
  - The Accelerator Middle Layer can help with all three steps



Linear optics correction overview

- Lattice errors affect linear optics characteristics, which can be used to determine and correct linear optics.
  - It is possible to measure beta functions with quadrupole modulation for optics correction.



Linear Optics from Closed Orbit (LOCO)

SLAC

- Orbit response matrix contains optics information

$$\begin{pmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta}_x \\ \boldsymbol{\theta}_y \end{pmatrix} = \begin{pmatrix} \Delta \boldsymbol{x} \\ \Delta \boldsymbol{y} \end{pmatrix}, \quad \text{or } \mathbf{R\theta} = \mathbf{\Delta X}$$

Calculation of response matrix

$M\Delta X + \Delta\theta_{xj} = \Delta X, \quad \rightarrow \quad \Delta X = (I - M)^{-1}\Delta\boldsymbol{\theta}_{xj}$

With one turn transfer matrix M at the corrector location, and $\Delta\boldsymbol{\theta}_{xj} = [0, \Delta\theta_j, 0,0,0,0]'$ for a horizontal kick. Without coupling

$$\Delta \mathrm{x}\,(s) = \frac{\sqrt{\beta(s)\beta_0}\Delta\theta_j}{2\sin\pi\nu}\cos(|\psi(s) - \psi(s_0)| - \pi\nu)$$

The orbit response matrix is an indirect representation of the optics of the machine lattice (including coupling). Therefore fitting orbit response matrix to the lattice model can effectively recover the machine optics into the model.

J. Safranek, M. Lee, SLAC-PUB-6442 (1994)
J. Safranek, NIMA, 388, 27 (1997)

Fitting ORM data to lattice model

- A least-square problem
  - Data include: measured orbit response matrix and dispersion functions.
  - Fitting parameters include: quadrupole strengths (gradients) in model, BPM and correction calibration parameters (gains, rolls and crunch).

$$\begin{pmatrix} x_{meas} \\ y_{meas} \end{pmatrix} = \begin{pmatrix} g_x & c_x \\ c_y & g_y \end{pmatrix} \begin{pmatrix} x_{beam} \\ y_{beam} \end{pmatrix}$$

Note the difference between c_x and c_y accounts for BPM "crunch" (deformation from ideal configuration).

  - Objective function:

$$f(\boldsymbol{p}) = \chi^2 = \sum_{i,j} \frac{\left(\left(R_{ij}^{beam} - R_{ij}^{model}\right)^2\right)}{\sigma_i^2} + \sum_i \frac{\left(D_{xi}^{beam} - D_{xi}^{model}\right)^2}{\sigma_{xi}^2} + \frac{\left(D_{yi}^{beam} - D_{yi}^{model}\right)^2}{\sigma_{yi}^2}$$

where $\boldsymbol{p}$ includes all fitting parameters.

  - Solving the least-square problem with an iterative method:

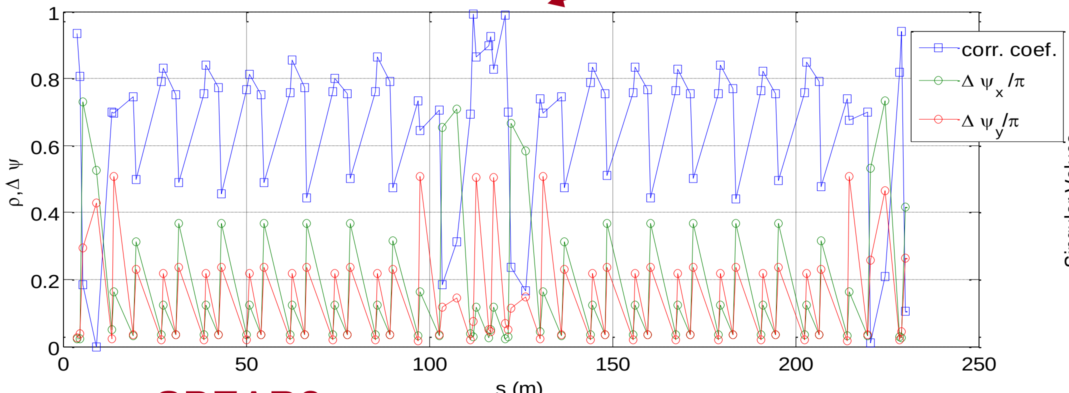
$f(\boldsymbol{p}) = \boldsymbol{r}^{\boldsymbol{T}}\boldsymbol{r}$ with residual vector $\boldsymbol{r}$. Calculate Jacobian matrix $\boldsymbol{J}$ with $J_{ij} = \frac{\partial r_i}{\partial p_j}$.

Solve $\boldsymbol{J}\boldsymbol{\Delta}\mathbf{p} = -\mathbf{r_n}$ at each iteration and move to $\mathbf{p}_{n+1} = \mathbf{p}_n + \boldsymbol{\Delta}\mathbf{p}$

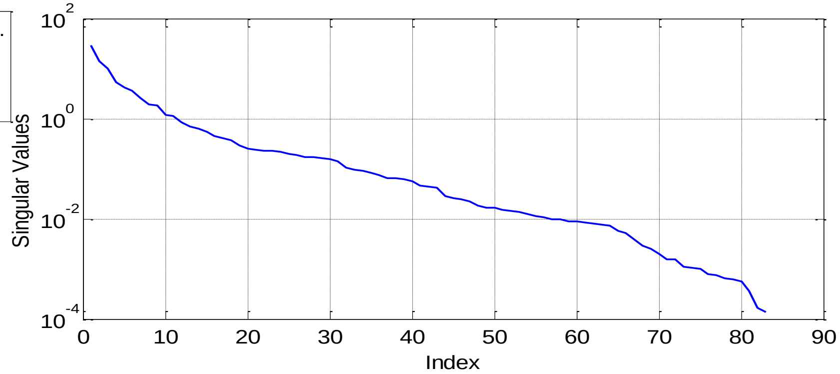
The over-fitting issue

- The linear optics fitting problem is prone to over-fitting
  - Neighboring quadrupoles have similar impact to linear optics (high correlation between columns of Jacobian matrix)
  - Resulting in large quadrupole errors $\frac{\Delta K}{K}$, mostly in the unconstrained directions (w/ small singular values)

$$r \equiv \frac{\mathbf{v}_1^T \mathbf{v}_2}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|}$$



SPEAR3



X. Huang, et al, ICFA Newsletter 44, 60 (2007)

Adding constraints to curtail degeneracy

- Removing singular values or fitting quadrupole parameters is commonly used to reduce over-fitting
- Adding constraints through penalty terms to χ^2 function is more flexible

Penalty terms

$$\chi_c^2 = \sum_{ij} \frac{1}{\sigma_{ij}^2} (R_{ij}^{\text{meas}} - R_{ij}^{\text{model}})^2 + \frac{1}{\sigma_K^2} \sum_{i=1}^{N_q} w_i^2 \Delta K_i^2,$$

Without constraints (Gauss-Newton method):

$$\Delta \mathbf{p} = -(\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T \mathbf{r}_0$$

$$\mathbf{J}_c = \begin{pmatrix} \mathbf{J} \\ \mathbf{W} \end{pmatrix}$$

W is diagonal with

$$W_{K,ii} = \frac{w_i}{\sigma_K}$$

With constraints:

$$\Delta \mathbf{p} = -(\mathbf{J}_c^T \mathbf{J}_c)^{-1} \mathbf{J}_c^T \mathbf{r}_{c0} = -(\mathbf{J}^T \mathbf{J} + \boldsymbol{W}^T \boldsymbol{W})^{-1} \mathbf{J}^T \mathbf{r}_0$$

Other ways to add constraints

- Adding penalty to certain patterns, e.g., parameter pairs

$$\frac{w^2}{\sigma_K^2}(L_i\Delta K_i - L_j\Delta K_j)^2$$

- Adding penalty through the Levenberg-Marquardt method

$$\lambda \sum_{i=1}^{N_q} \mathbf{J}_i^T \mathbf{J}_i \Delta p_i^2, \quad \longrightarrow \quad \Delta\mathbf{p} = -(\mathbf{J^T J} + \lambda \mathrm{diag}(\mathbf{J^T J}))^{-\mathbf{1}} \mathbf{J^T r_0}$$

- Adding penalty to the singular-value patterns

$$\chi_c^2 = \chi^2 + \sum_{i=1}^{P} \lambda_i^2 (\mathbf{v}_i^T \Delta\mathbf{p})^2$$

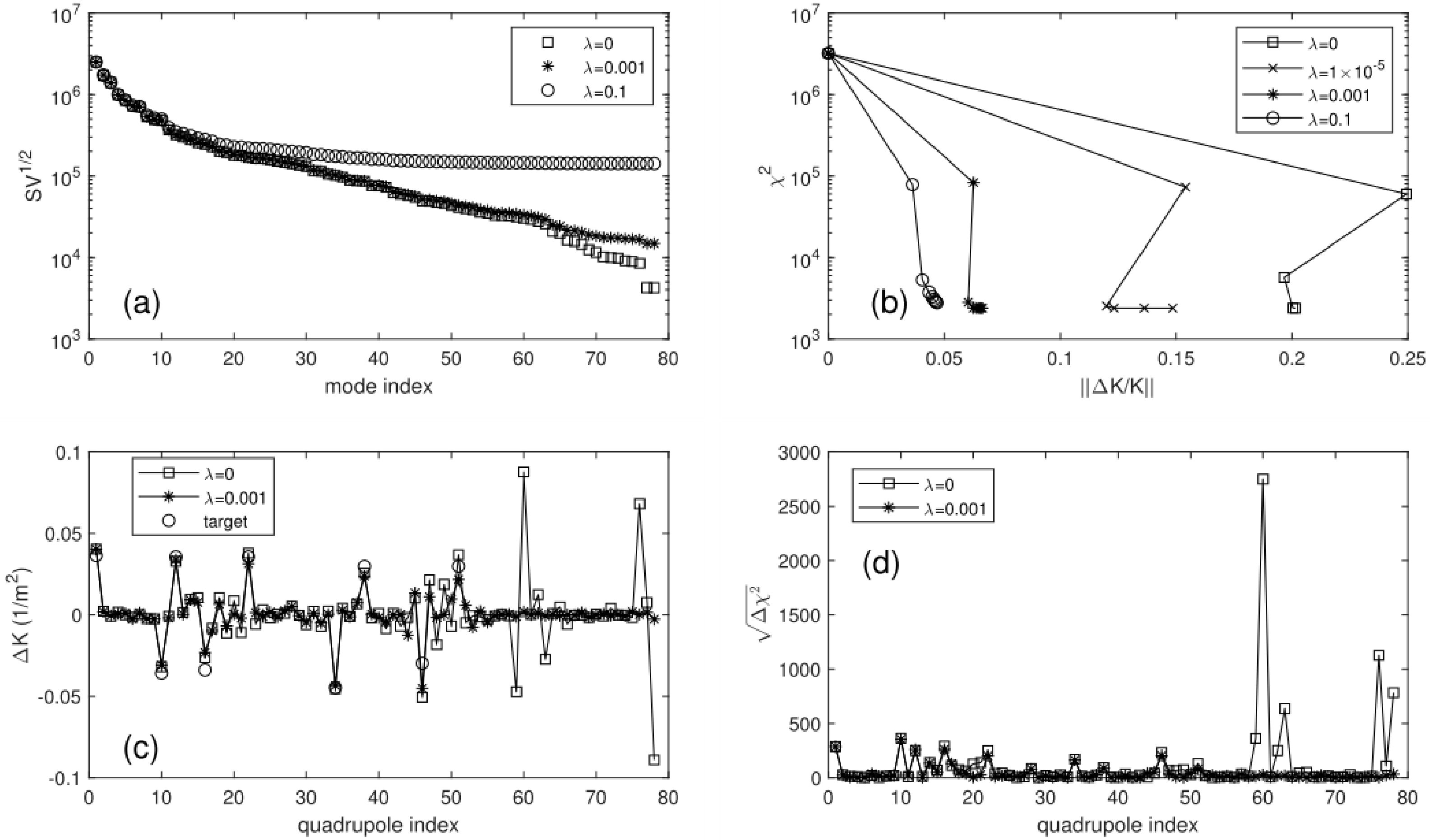
$$\longrightarrow \quad \Delta\mathbf{p} = -\mathbf{V}(\mathbf{S}^2 + \mathbf{\Lambda}^2)^{-1}\mathbf{S}\mathbf{U}^T\mathbf{r}_0 = -\sum_{i=1}^{P} \frac{s_i}{s_i^2 + \lambda_i^2} \mathbf{v}_i (\mathbf{u}_i^T \mathbf{r}_0)$$

X. Huang, Beam-based correction and optimization for accelerators (2019)

Fitting w/ constraints – an example

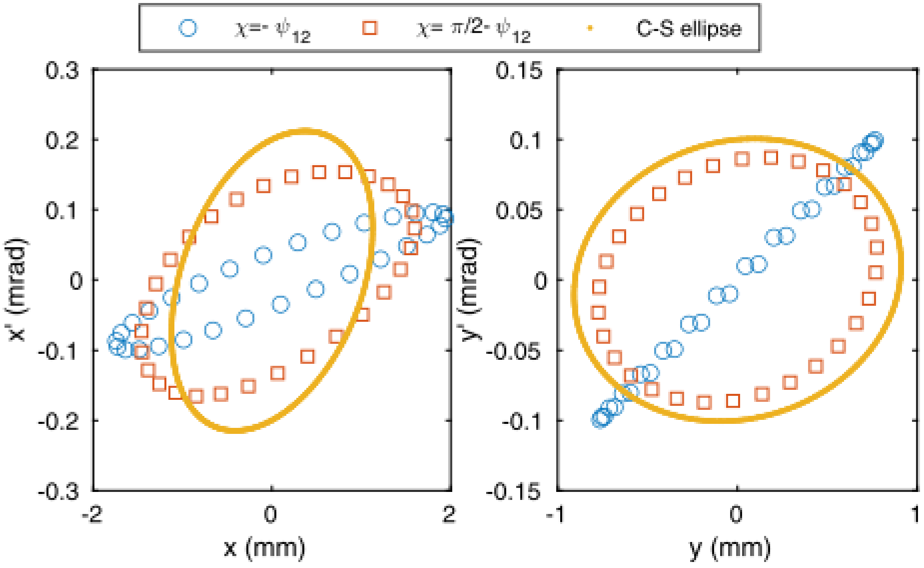
- The constraints shape the convergence behavior

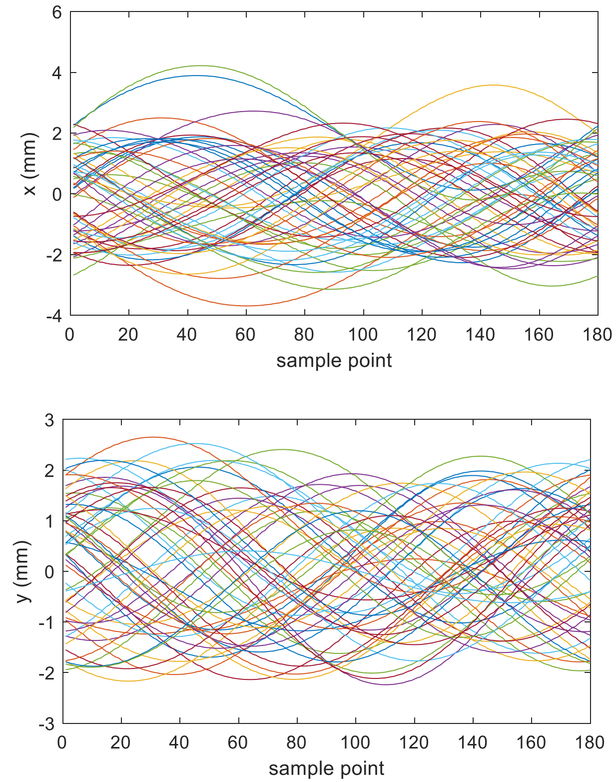
SPEAR optics fitting w/ L-M method (fixed λ)



The LOCOM method

- The LOCOM method is similar to LOCO in using closed-orbit to probe the linear optics and coupling
- But it uses the combination of **two** correctors, not many





X. Huang, PRAB 24, 072805 (2021)

A condensed form of LOCOM data

- The LOCOM orbit waveforms can be reduced to mode amplitudes

$$x_i(n) = A_i \sin(2\pi\nu n + \phi_i) \\ = A_{si} \sin 2\pi\nu n + A_{ci} \cos 2\pi\nu n,$$

If we choose modulation frequency, $\nu = \frac{1}{N}$

$$A_{si} = \frac{2}{N}\sum_{n=1}^{N} x_i(n) \sin\frac{2\pi n}{N},$$

$$A_{ci} = \frac{2}{N}\sum_{n=1}^{N} x_i(n) \cos\frac{2\pi n}{N}.$$

- The mode amplitudes are connected to the transfer matrices and corrector waveforms

Corrector waveforms

$$\theta_1(n) = \theta_{1m} \sin(2\pi\nu n + \phi_1),$$
$$\theta_2(n) = \theta_{2m} \sin(2\pi\nu n + \phi_2),$$

The closed orbit at a BPM downstream of corrector 1

$$\mathbf{Y}_P(n) = \mathbf{M}_{P1}\left(\mathbf{I} - \mathbf{M}_1\right)^{-1}\mathbf{Q}_1 \begin{pmatrix} \cos 2\pi\nu n \\ \sin 2\pi\nu n \end{pmatrix}$$

$$Q_{1,11} = m_{12}\theta_{2m} \sin\phi_2,$$
$$Q_{1,12} = m_{12}\theta_{2m} \cos\phi_2,$$
$$Q_{1,21} = m_{22}\theta_{2m} \sin\phi_2 + \theta_{1m} \sin\phi_1,$$
$$Q_{1,22} = m_{22}\theta_{2m} \cos\phi_2 + \theta_{1m} \cos\phi_1.$$

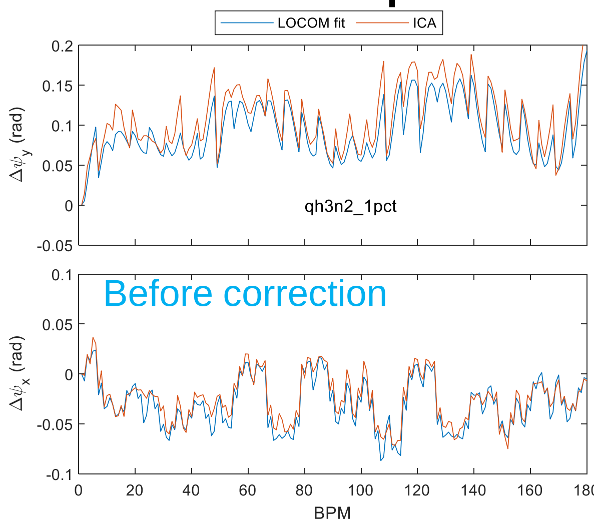
X. Huang, X. Yang, 26, 052802 (2023)

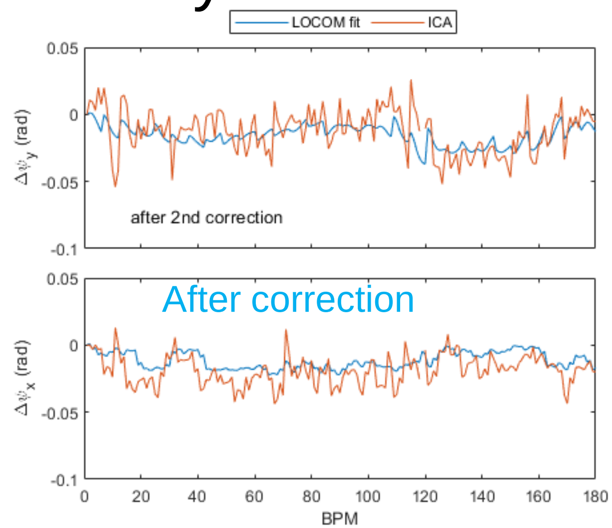
Lattice fitting, with coupling included

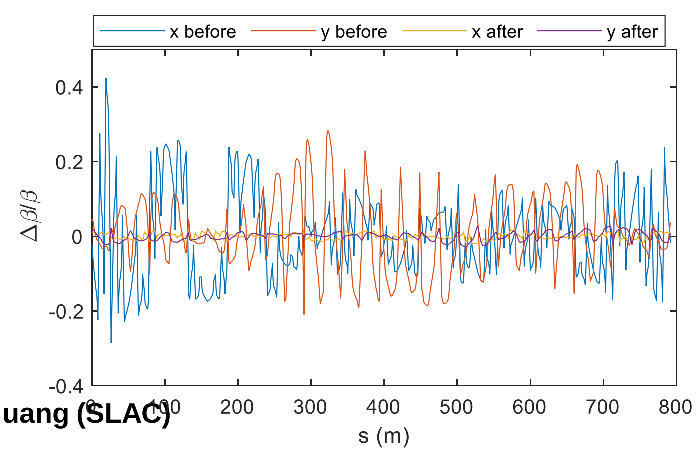
- Fitting data: 8 mode amplitude per BPM, plus H/V dispersion
  - 4 in-plane (2 H + 2 V) amplitudes, 4 cross-plane amplitudes
  - $10\ N_P$ data points total
- Fitting parameters:
  - N_Q, Quadrupoles in lattice
  - N_{SQ}, Skew quadrupoles
  - $4N_P$, BPM gains and coupling coefficients
  - 8, Corrector gains and coupling coefficients
- Fitting method: same as LOCO and other optics correction methods
  - Levenberg-Marquardt with penalty term to slow down divergence on under-constrained directions*

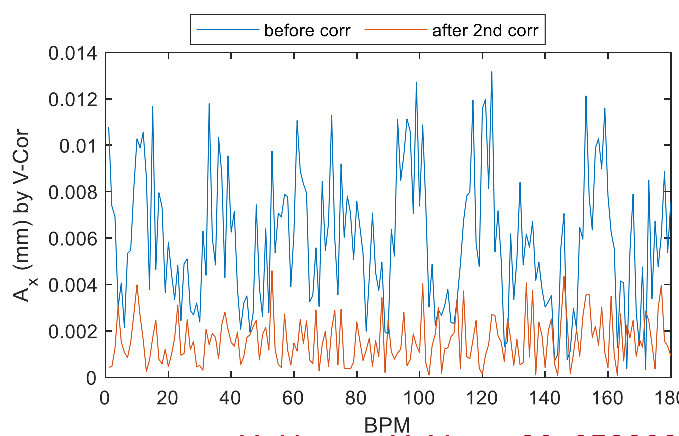
LOCOM experiments at NSLS-II

- Optics and coupling correction with LOCOM are tested at NSLS-II and compared to direct turn-by-turn BPM data









X. Huang, X. Yang, 26, 052802 (2023)

Implementation considerations for the middle layer

- It easy to implement an algorithm to be used by the developer himself/herself
- It takes a lot more effort to develop a software for an non-expert to use
- The Matlab Middle Layer is a great success from the software standing point
  - Many facilities, many users
  - Experience from one facility is transferrable to another

The strengths of MML come from its well thought-out design and meticulous implementation.

Desirable features for Accelerator Middle Layer high-level applications

- Using a common interface to interact with different accelerators
  - Consisting of functions for data taking and accelerator control
  - Hiding the details specific to a machine, through configuration file(s)
  - This interface is core to the Accelerator Middle Layer
- Integrating with an accelerator modeling code
  - Providing easy access to accelerator data (e.g., momentum compaction factor, beta functions, transfer matrix, etc)
  - Evaluating impact of parameters (e.g., Jacobian matrix)
- Decoupling data taking and data analysis
- Consistent data management
  - Careful design at the beginning, avoid frequent modifications to data structure
- Providing a graphic user interface

Summary

- Beam-based correction is the preferred approach if available
  - Combine capability of modern diagnostics and our understanding of physics
  - Often faster and more reliable
- But not without weakness
  - Target might not correspond to optimal beam performance
  - Demanding in terms of diagnostics
  - Degeneracy and over-fitting
  - It is hard to extend to nonlinear optics correction
- Implementation of high-level applications for beam-based correction benefits greatly from a powerful middle layer library