Improving fitting technique for global optics correction

Presentation at Advanced Optics Control Workshop, CERN

Xiaobiao Huang

SSRL/SLAC

2/6/2015

Outline

- Global optics correction by fitting lattice model to data
  - Orbit response matrix data: LOCO
  - Extraction of optics functions from turn-by-turn BPM data
  - Fitting turn-by-turn BPM data directly to model
- Difficulty with the fitting approach
  - Quadrupole errors predicted by fitting are too big
- Understanding the problem
- Solution to the difficulty
- A method for optics and coupling correction with TBT BPM data.

Linear Optics from Closed Orbit (LOCO)

SLAC

- Orbit response matrix contains optics information

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

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, but significant 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.

References: 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:

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

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 $\mathbf{p}_{n+1} = \mathbf{p}_n + \boldsymbol{\Delta}\mathbf{p}$

J. Safranek, NIMA, 388, 27 (1997);

Turn-by-turn BPM data for optics calibration

- Turn-by-turn (TBT) BPM data sample the machine optics

  The beam coordinates at turn n, $X_n = (x, x', y, y')_n^T$, is given by
  $X_{n+1} = MX_n$

- Beta functions and betatron phase advances can be derived from simultaneous TBT BPM data with beam oscillation.
  - P. Castro's method [1]   [1] P. Castro, et al, PAC 93
  - MIA [2] and ICA [3]   [2] Chun-xi Wang, et al. PRSTAB 6, 104001 (2003)

- Measured beta and phase can be used to fit lattice model [3].

$$f(\mathbf{q}) = \chi^2 = \frac{1}{2}\sum_{i=1}^{240} r_i^2, \qquad r_i = \frac{y_i(\mathbf{q}) - y_i^d}{\sigma_i},$$

$$\mathbf{y} = (w_1\boldsymbol{\beta}_x, w_2\Delta\psi_x, w_3\boldsymbol{\beta}_z, w_4\boldsymbol{\Delta\psi}_z, w_5\mathbf{D}_x),$$

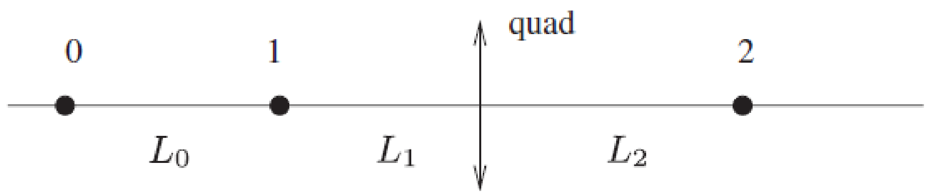
This approach can be extended to include coupling (see details below).

[3] X. Huang, et al, PRSTAB, 8, 064001, (2005)

Another way of using TBT BPM data for optics correction

SLAC

- The raw measured TBT data can be fitted against model prediction.
  - Simple case: one quadrupole



The angle coordinate can be determined from two BPMs with known optics in between (such as a drift).

$$x_1' = \frac{x_1 - x_0}{L_0},\ y_1' = \frac{y_1 - y_0}{L_0}.$$

TBT data at BPM 2 can be compared to predictions based on (x_1, x_1', y_1, y_1') and the model between BPM 1, 2.

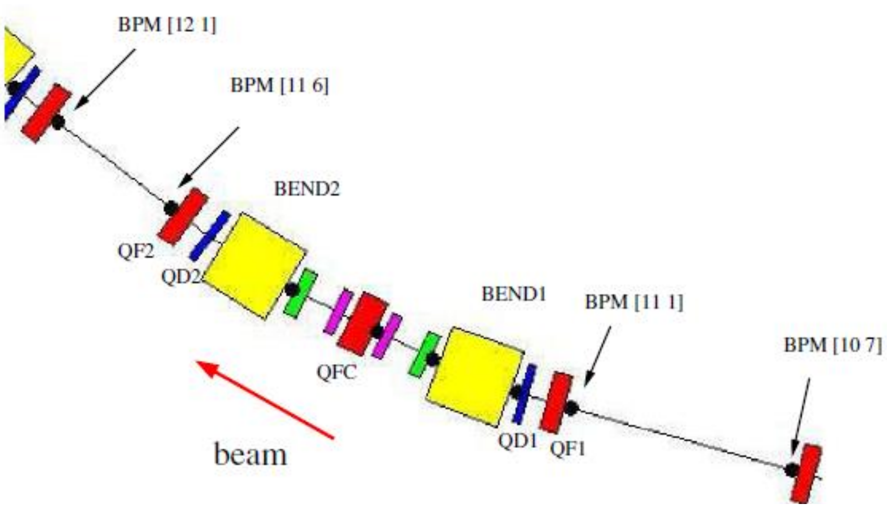
  - Transport line and storage ring cases are similar.

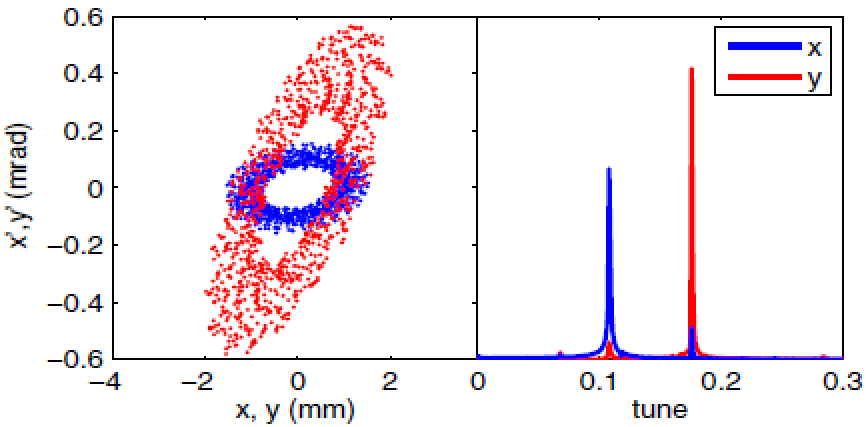
$$\chi^2 = \sum_{n=1}^{N} \sum_{i=2}^{M+1} \left[\left(\frac{x_i(n) - \tilde{x}_i[\mathbf{p}; \mathbf{X_1(n)}]}{\sigma_{xi}} \right)^2 + \left(\frac{y_i(n) - \tilde{y}_i[\mathbf{p}; \mathbf{X_1(n)}]}{\sigma_{yi}} \right)^2 \right],$$

Fitting parameters include quadrupole strengths and BPM gains and rolls.

X. Huang, et al, PRSTAB, 13, 114002, (2010)

Experiment with TBT data for a section of SPEAR3





8 BPMs on SPEAR3 were converted to have turn-by-turn capability for the experiment.

X and y motion are simultaneously excited.

Fitting 5 quadrupole parameters. Errors were added to initial values.

Results: the 3 upstream quads are very well determined. The 2 downstream quads are not well constrained.

TABLE I. Quadrupole parameters.

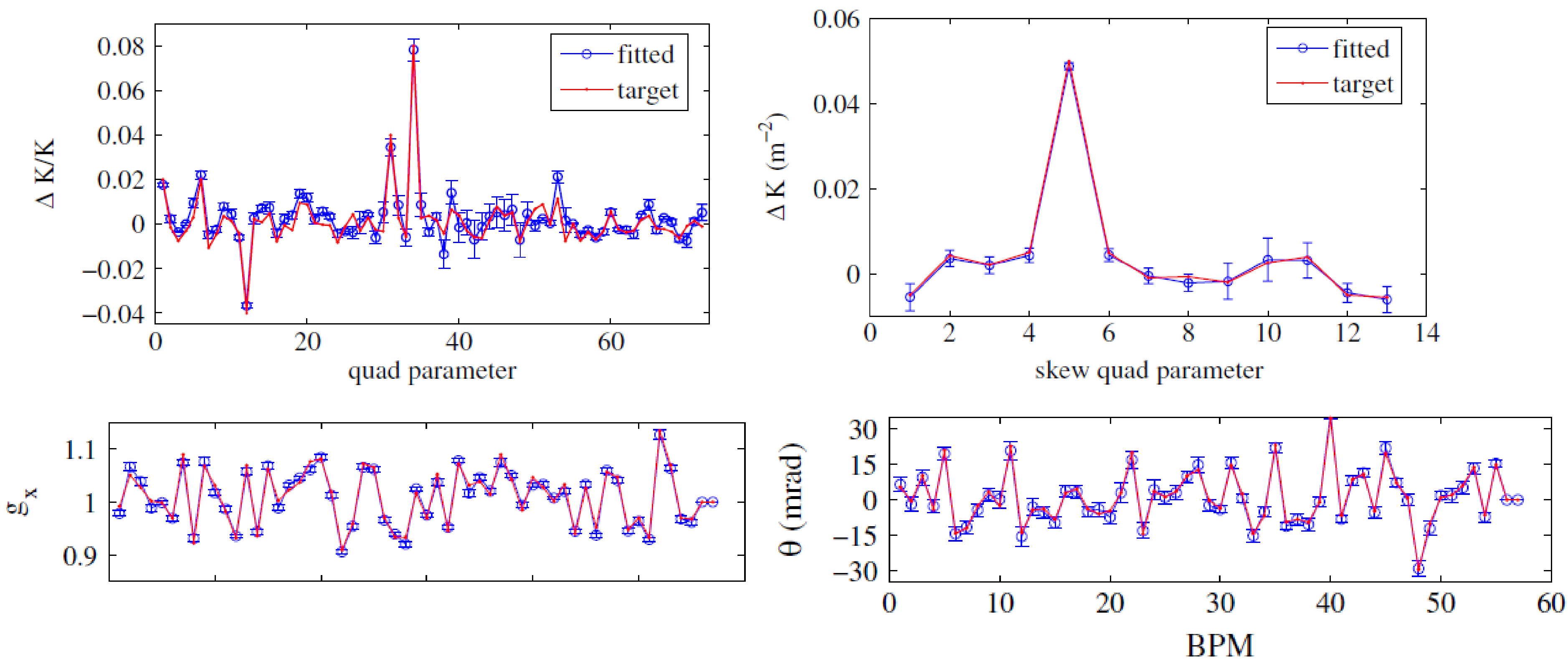
Quad	Design	LOCO	Fitted	rms	Initial
QF1	1.823	1.824	1.810	0.001	1.883
QD1	−1.920	−1.922	−1.911	0.001	−1.890
QFC	1.683	1.683	1.694	0.002	1.653
QD2	−1.347	−1.331	−1.344	0.003	−1.347
QF2	1.691	1.686	1.691	0.002	1.691

LOCO results may be better in this case because it uses data from full ring. It is used here as a reference.

Simulation with full ring for SPEAR3

Simulated data: 0.5% of random quad error (rms) were added to model. Skew quad error of rms 0.001 m^{-2}. Gaussian errors of 50 microns added to data. Initial excitation of 2 mm for both planes. Using 200 turns of data for fitting.

Fitting parameters: 72 quads and 13 skew quads.



The seeded errors to quads, skew quads and BPM parameters are all successfully recovered.

Difficulty with global optics fitting

- When fitting optics model to orbit response matrix data, it often happens the predicted ΔK is too large.
  - It can be too large to be reasonable. Sometimes iterative fitting won't work since the second iteration lattice has no closed orbit.
  - It can happen even in simulation without random noise in data (Fermilab Booster [1]).
  - More common for other rings is that optics correction with the fitted ΔK fails.
- Causes of the difficulty: correlation between fitting parameters.

Correlation of two parameters

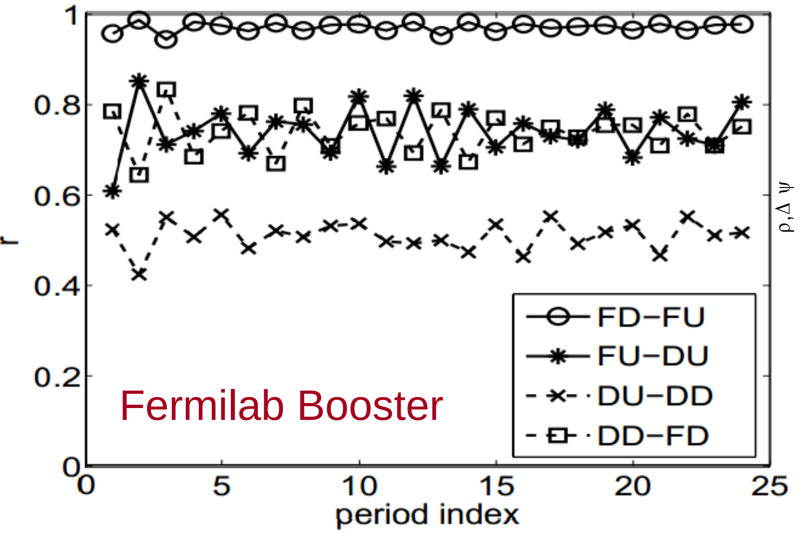
$$\rho_{12} = \frac{\mathbf{J}_1^T \mathbf{J}_2}{\|\mathbf{J}_1\| \|\mathbf{J}_2\|}$$

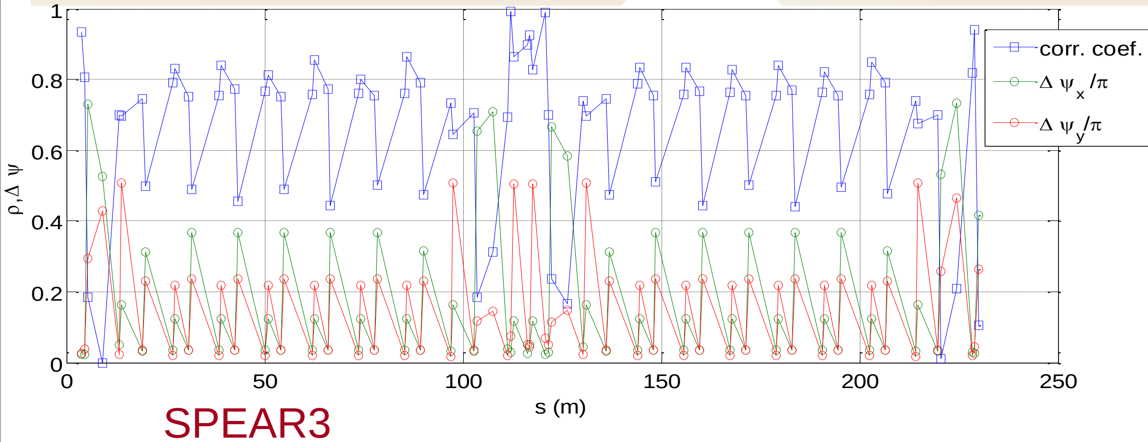
$\mathbf{J}_i$ is the column in the Jacobian matrix for parameter i.

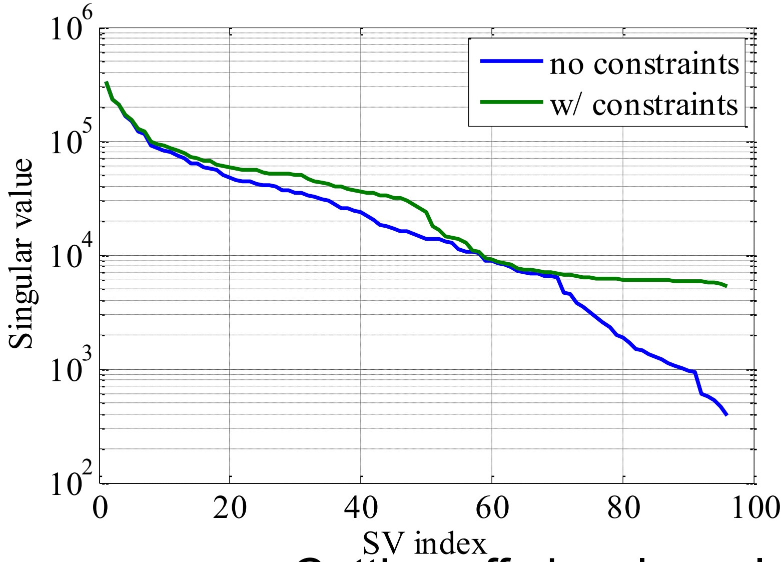
Large correlation between two parameters indicates that their effects on the objective function are similar and thus hard to resolve.

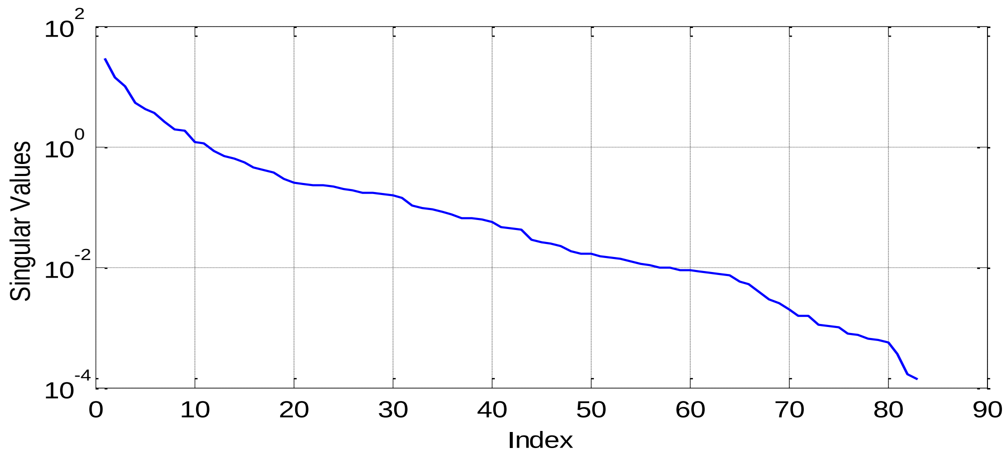
[1] X. Huang, et al, PAC'05, (2005)

Correlation between quadrupole parameters





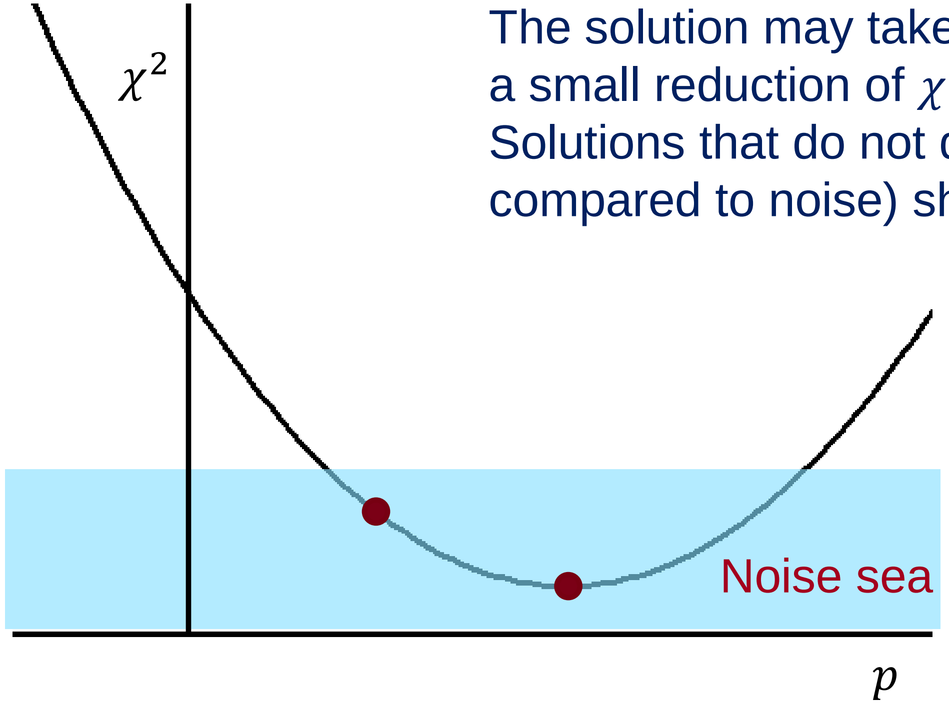




Cutting off singular values in matrix inversion doesn't always help.

Symptoms with coupled fitting parameters

- Large excursion of the fitting solution
- Large error bars to the fitting solution



The solution may take a large step in $\Delta\boldsymbol{p}$ to chase a small reduction of χ^2.
Solutions that do not differ significantly in χ^2 (as compared to noise) should be deemed equivalent.

Solutions to the coupled-parameter problem

- Cutting off singular values in matrix inversion
- Removing fitting parameters
- Adding constraints to fitting parameters[1-2]
  - This is the preferred approach. It works better, probably because it is more “gentle”, allowing the solution $\Delta\boldsymbol{p}$ to develop along the otherwise prohibited directions.

At every iteration, solve for $\Delta\boldsymbol{p}$ with extra constraints, such as

$$\Delta K_k = 0.$$

These constraints are not strict, but with certain pre-specified weights. Equivalently, for the iteration, the objective is modified to

$$\chi^2 = \chi_0^2 + \frac{1}{\sigma_{\Delta K}^2}\sum_k w_k^2 \Delta K_k^2$$

Note the global minimum is not changed by the constraints, only the convergence path.

[1] X. Huang, et al, PAC’05, (2005)
[2] X. Huang, et al, ICFA beam dynamics newsletter, 44, (2007).

Constraints in math terms

The least-square problem: to minimize

$$f(\boldsymbol{p}) = \chi^2 = \sum_i \frac{1}{\sigma_i^2}(y_i - y(x_i;\boldsymbol{p}))^2 = \boldsymbol{r}^T\boldsymbol{r}$$

With residual vector $\boldsymbol{r}$, $r_i = (y_i - y(x_i;\boldsymbol{p}))/\sigma_i$.

Solving the LS with iteration:

$$f(\boldsymbol{p}) = f(\boldsymbol{p}_0 + \Delta\boldsymbol{p}) \approx \boldsymbol{r}_0^T\boldsymbol{r}_0 + 2\Delta\boldsymbol{p}\boldsymbol{J}^T\boldsymbol{r}_0 + \Delta\boldsymbol{p}^T\boldsymbol{J}^T\boldsymbol{J}\Delta\boldsymbol{p}$$

Condition $\frac{\partial f}{\partial \boldsymbol{p}} = 0$ leads to (the normal equation)

$$\boldsymbol{J}^T\boldsymbol{J}\Delta\boldsymbol{p} = -\boldsymbol{J}^T\boldsymbol{r}_0$$

(This is equivalent to solving $\boldsymbol{J}\Delta\boldsymbol{p} = -\boldsymbol{r}_0$ with SVD, but computationally easier)

The constraints are additional conditions that extends the Jacobian matrix and the residual vector:

$$\boldsymbol{J}_n = \begin{pmatrix}\boldsymbol{J}\\ \boldsymbol{D}\end{pmatrix},\ \ \boldsymbol{r}_n = \begin{pmatrix}\boldsymbol{r}_0\\ \boldsymbol{0}\end{pmatrix}$$

where elements of matrix D is zero except $D_{kk} = \frac{w_k}{\sigma_{\Delta K}}$ for parameter k. The new normal equation (with constraints) is essentially modified to

$$(\boldsymbol{J}^T\boldsymbol{J} + \boldsymbol{D}^{\mathrm{T}}\boldsymbol{D})\Delta\boldsymbol{p} = -\boldsymbol{J}^T\boldsymbol{r}_0$$

The matrix $\boldsymbol{D}$ can be different if other types of constraints are imposed.

The Levenberg-Marquadt algorithm for fitting

The L-M method for fitting

$$(\boldsymbol{J}^T\boldsymbol{J} + \lambda\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D})\Delta\boldsymbol{p} = -\boldsymbol{J}^T\boldsymbol{r}_0$$

With diagonal matrix D, $D_{ii} = \|\boldsymbol{J}_i\|$. Parameter λ is adjusted after every iteration. It tends to zero when the solution approaches the minimum.

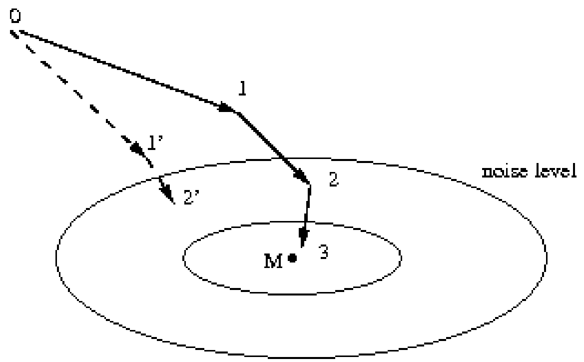
So with λ at fixed value, the L-M method is another way of applying constraints. Changing λ is to change the weight.

This is more convenient since one needs not to manually choose the constraints.

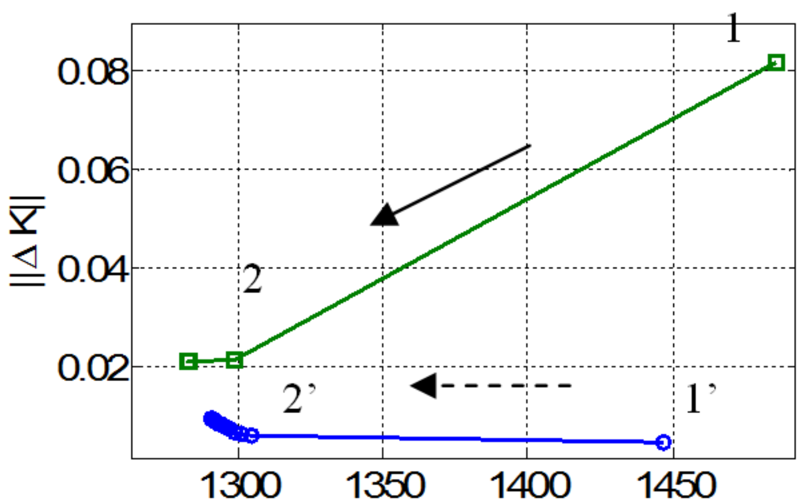
Both the Gauss-Newton with manual constraints method and the L-M method (with fixed, changeable weight factor λ) are implemented in the Matlab LOCO code[1].

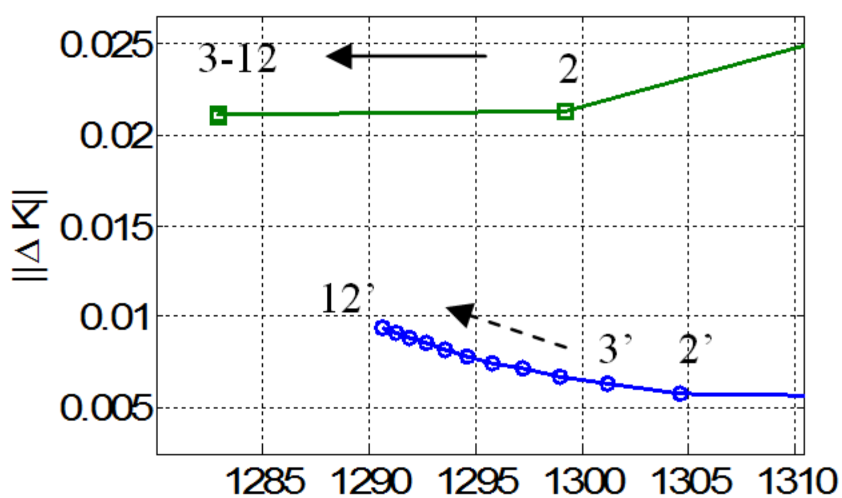
[1] G. Portmann, et al, ICFA newsletter, 44 (2007)

Illustration of the effects by the constraints



Equivalent solutions with small ΔK are more reasonable and are preferred for optics correction.

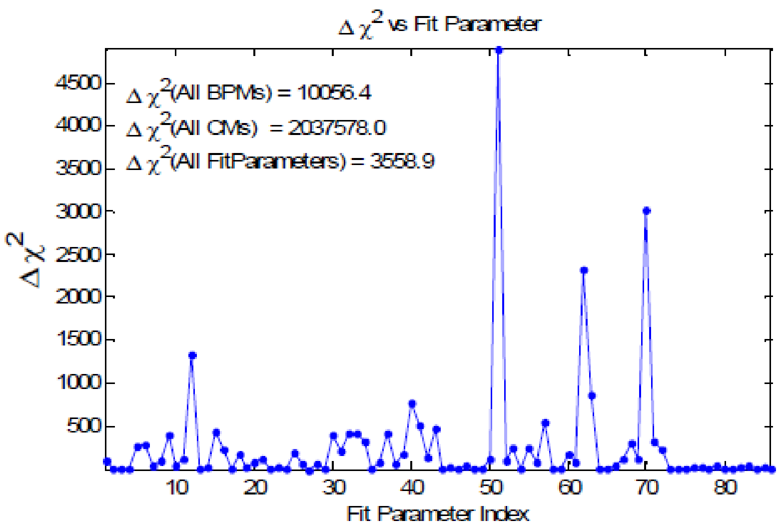




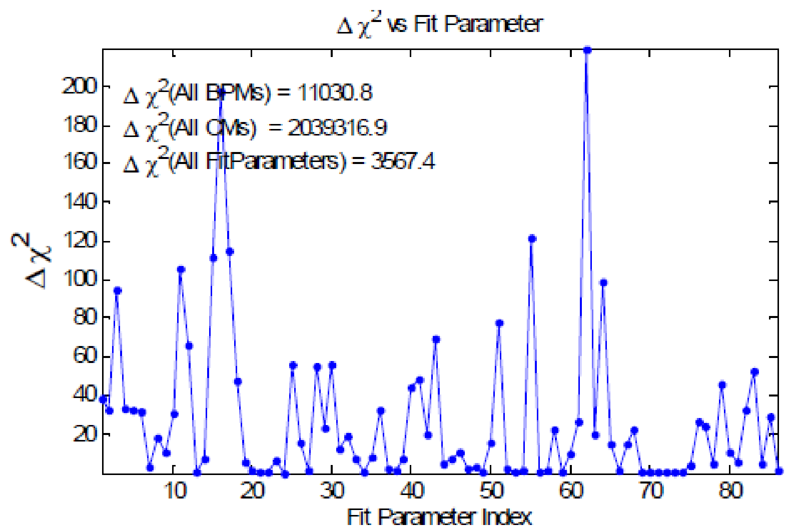
The rms relative change of gradients vs. the residual χ^2 for real SPEAR3 data set. Green: no constraints; Blue: with constraints. Point 0 is located at (2×10^6, 0).

χ^2 contribution of parameters

Calculation of χ^2 contribution for an individual parameter or a group of parameters: put all other parameters at the fitted values except the parameter(s) in question, which is set to the initial values; then find the change to χ^2.



Solution 3 (no constraint)



Solution 3’ (w/ constraint)

Without constraints (left), the sum of contributions of all quads is much larger than the total contribution of these quads. This indicates cancellation (fighting) between the quadrupoles.

Practical considerations for optics correction

- For optics correction, fit data only with knobs that you can or plan to use in correction.
  - These knobs may include quads on group power supplies (combined knob).
  - For optics "measurement", one may use more quadrupole knobs in fitting to more fully extract information from data.
- Adjust the weight of constraints (λ in L-M method) so that quadrupole parameters result in significant χ^2 reduction, but with reasonable predicted $\Delta K/K$ (typically $< 1\%$).
  - Stop fitting iteration when χ^2 change becomes small, then apply correction and take new data.

Do not wait until fitting converges. The global minimum in the fitting problem may not be suitable for optics correction because of large, unrealistic errors.

A new method to simultaneously correct optics and coupling with TBT BPM data

Turn by turn BPM data are decomposed to normal modes with Independent Component Analysis (ICA)

$$x(n) = A\cos(2\pi\nu_1 n + \psi_1) + B\sin(2\pi\nu_1 n + \psi_1) + a\cos(2\pi\nu_2 n + \psi_2) + b\sin(2\pi\nu_2 n + \psi_2)$$
$$y(n) = c\cos(2\pi\nu_1 n + \psi_1) + d\sin(2\pi\nu_1 n + \psi_1) + C\cos(2\pi\nu_2 n + \psi_2) + D\sin(2\pi\nu_2 n + \psi_2)$$

Amplitude and phase advances of the normal mode components in x and y can be constructed with the ICA mode coefficients.

Normal modes can also be obtained from the transfer matrix (obtained with model)

$$X = (x, x', y, y')^T,$$
$$X_{n+1} = TX_n$$
$$X = P\Theta,$$
$$\Theta_{n+1} = R\Theta_n.$$

$$\Theta = \begin{pmatrix} \sqrt{2J_1}\cos\Phi_1 \\ -\sqrt{2J_1}\sin\Phi_1 \\ \sqrt{2J_2}\cos\Phi_2 \\ -\sqrt{2J_2}\sin\Phi_2 \end{pmatrix}$$

$$R_{1,2} = \begin{pmatrix} \cos\phi_{1,2} & \sin\phi_{1,2} \\ -\sin\phi_{1,2} & \cos\phi_{1,2} \end{pmatrix}$$

$$R = \begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix}$$

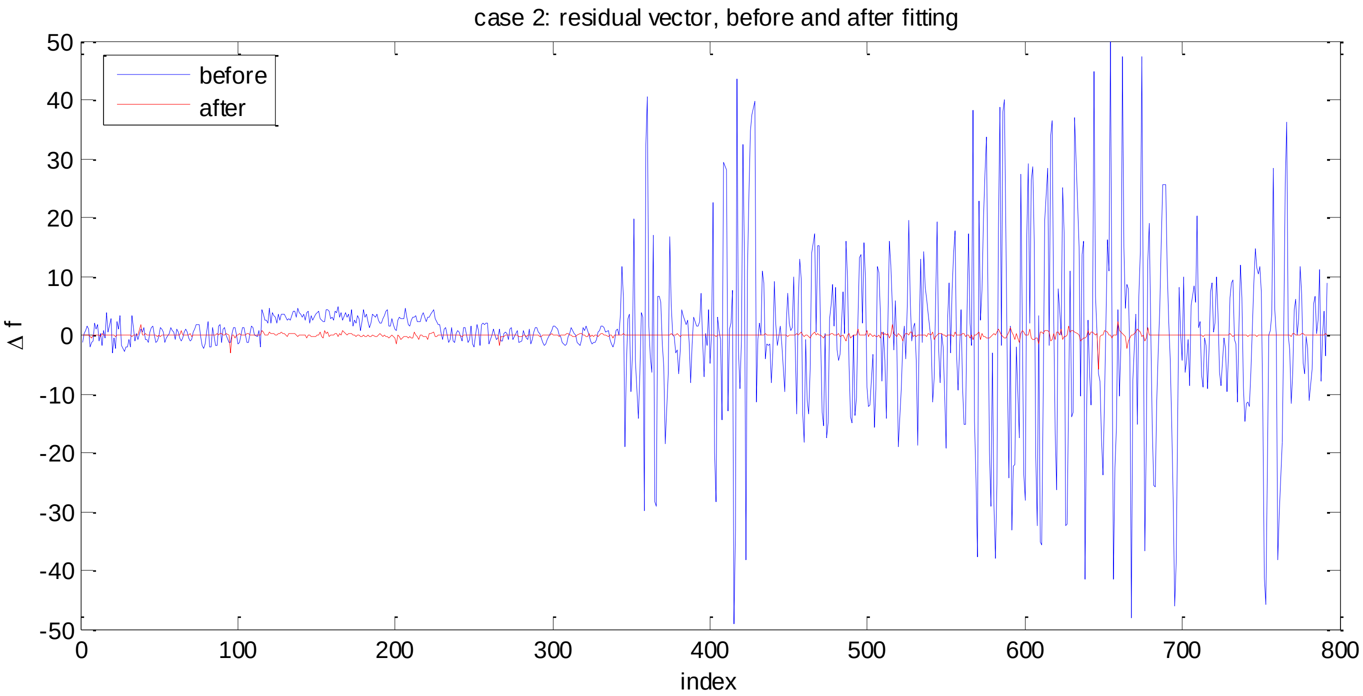
Hence

$$x = p_{11}\sqrt{2J_1}\cos\Phi_1 + p_{13}\sqrt{2J_2}\cos\Phi_2 - p_{14}\sqrt{2J_2}\sin\Phi_2$$
$$y = p_{31}\sqrt{2J_1}\cos\Phi_1 - p_{32}\sqrt{2J_1}\sin\Phi_1 + p_{33}\sqrt{2J_2}\cos\Phi_2$$

Fitting the lattice model with quadrupole and skew quadrupole parameters to reproduce the amplitude and phase advances of ICA.

Matrix *P* can be found with Y. Luo's approach (PRSTAB 7, 124001, 2004) or with Sagan-Rubin approach (PRSTAB 2, 074001, 1999).

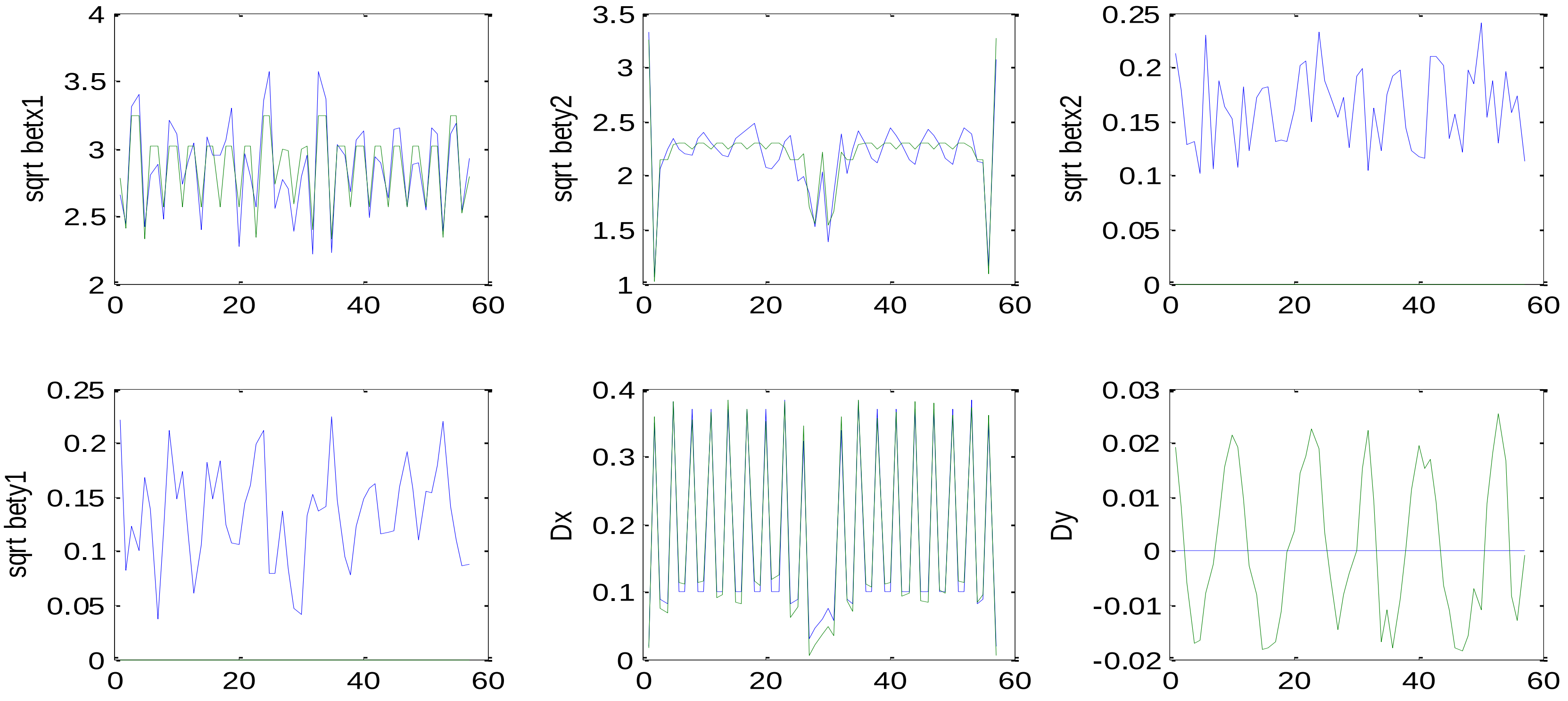
Simulation with SPEAR3 model



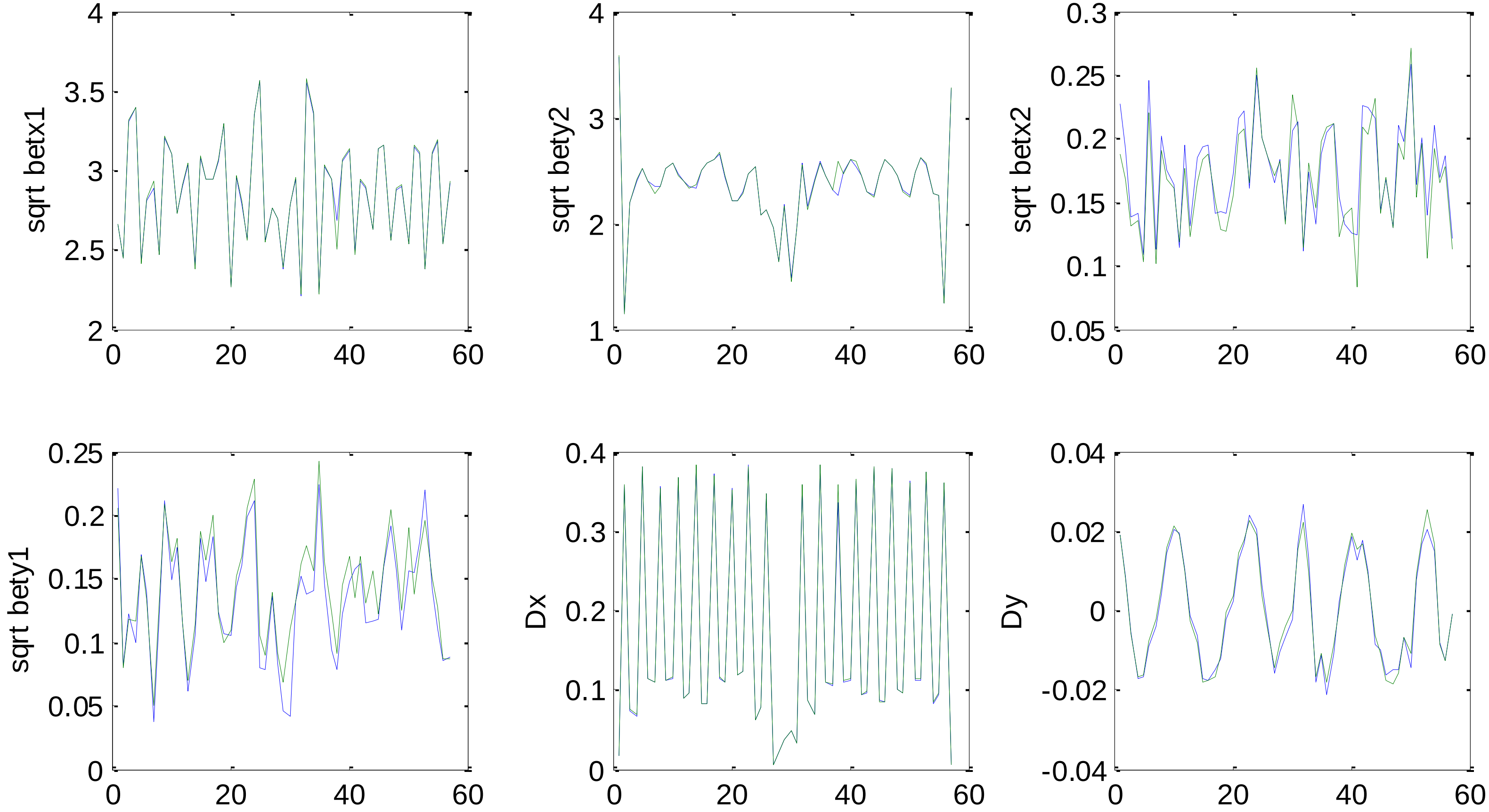
For SPEAR3, a test case is generated by adding quadrupole and skew quad errors to the ideal lattice. BPM rolls are also added. Tracking one particle for 1100 turns, with initial offset of x=2 mm and y=1mm, respectively.

Dispersion functions are also included in fitting.

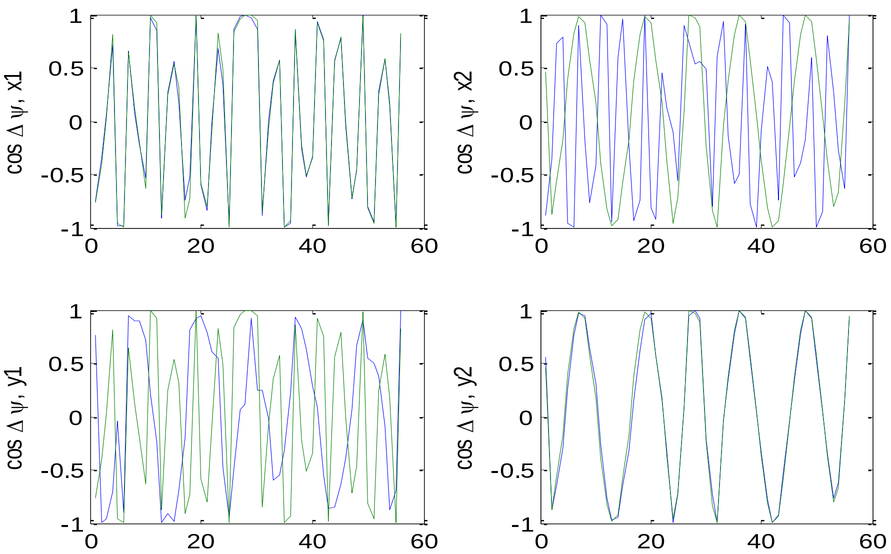
Before fitting



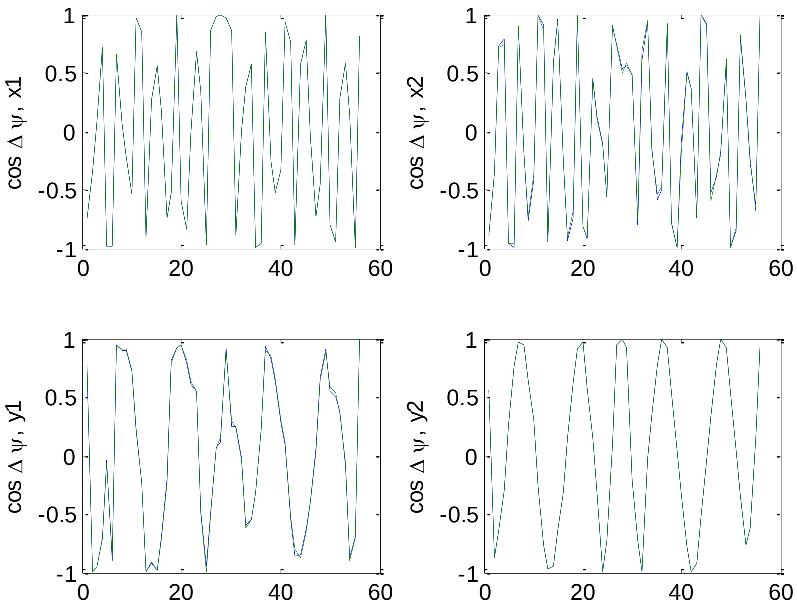
After fitting



Comparison of phases

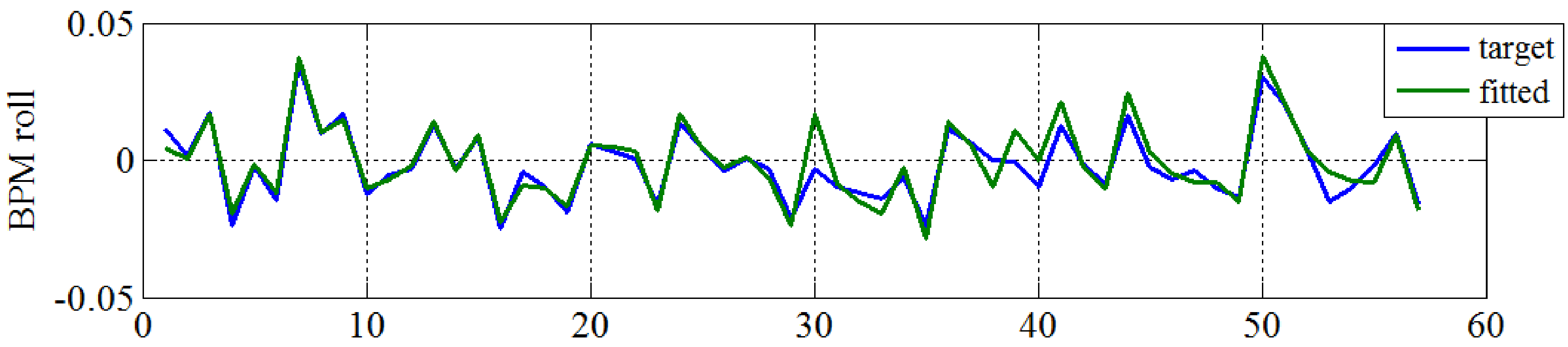


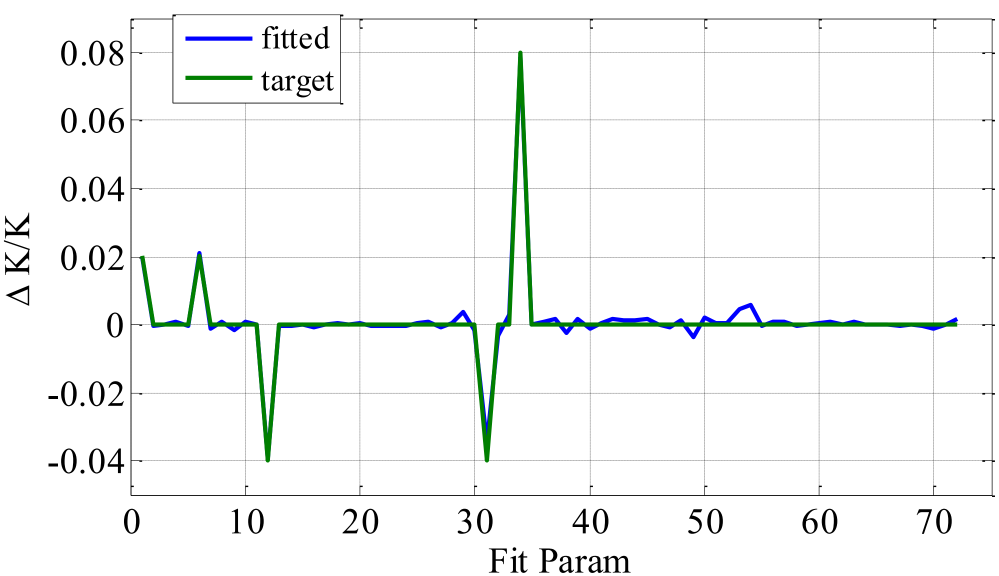
Before fitting

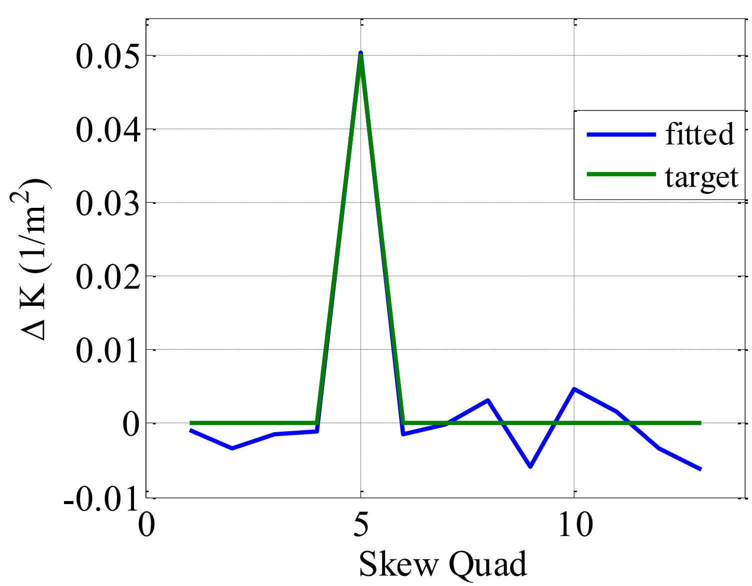


After fitting

Fitting parameters







Summary

- Global optics can be obtained by fitting lattice model to data that sample the machine optics.
- For many machines correlation between quadrupole parameters makes the fitting problem degenerate or poorly constrained.
- Imposing constraints on parameter deviation in each iteration helps finding more reasonable equivalent solutions that are suitable for optics correction.
  - This method was implemented in Matlab LOCO since 2007 and has benefitted many storage rings.
- A new method that uses ICA to obtain normal modes can simultaneously correct optics and coupling with in TBT BPM data.