Beam based optimization

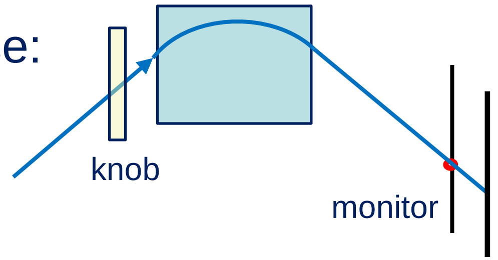
X. Huang

USPAS 2015 Summer – Beam based diagnostics

How to make a machine work?

- For a complex machine such as an accelerator to work well, typically it requires
  - A model of the machine
    - Specify the ideal working conditions
      - At the physics model level, working conditions may include magnet strengths, ideal orbit, RF voltage and phase, etc.
    - Predict performance
    - Predict response of performance w.r.t. working conditions.
  - Diagnostics
    - Monitor the working conditions and detect deviations.
  - Adjustable parameters – knobs
    - Restore or compensate the deviations in the working conditions

A simple case:



Scenarios when things don't go well

- Model is not accurate
  - Effects not included the model
    - Example: fringe fields, insertion device effects
  - Imperfections in the machine
  - →Model specified working conditions are not ideal.
  - Examples:
    - Linear optics error due to misalignment of magnets, magnet calibration errors, magnet fringe field effect, insertion device imperfections.
    - Similarly for nonlinear optics errors
  - Solution: calibrate the model with beam based measurements (e.g. LOCO)
- Lack of diagnostics
  - Example: transport line steering and optics
- Lack of effective correction methods
  - Example: optics correction before LOCO

Automatic tuning (online optimization)

- When model specified solution doesn't give the optimal performance, "tuning" of the machine is necessary.
- Manual tuning
  - Not scalable to large problems (with # of knobs ≥ 4?)
  - Not efficient (evaluation of solution is slow; multi-variable search algorithm not efficient)
- Automatic tuning*
  - Essentially an optimization problem
  - Requirements:
    - dealing with noise in functions.
    - efficient
  - Questions: what algorithms are suitable for automatic tuning.

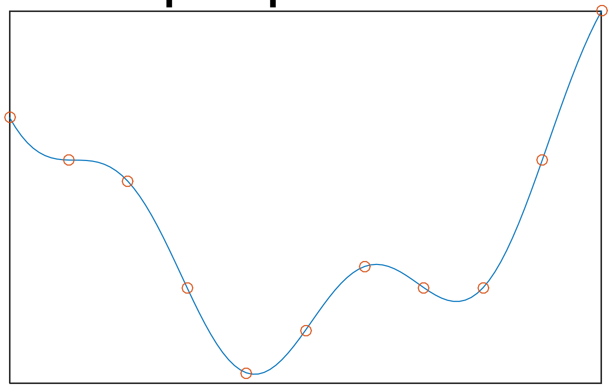
*First attempt of automatic tuning in accelerator is by L. Emery (APS), PAC03. Algorithms tried were 1D scan and downhill simplex.

Candidate optimization algorithms

- Parameter scans
- Gradient based algorithms
- Downhill simplex algorithm
- Powell’s method (conjugate direction set method)
- Robust conjugate direction search (RCDS)
- Stochastic methods (more assurance for finding global optimum)
  - Simulated annealing
  - Genetic algorithms
  - Particle swarm optimization

Parameter scan

- 1D scan: evaluate the function at parameter values equally distributed in a given range.
- Iterate through multiple parameters.



- Advantage
  - Samples the 1D parameter range globally.
- Disadvantages
  - Inefficient: evaluate solutions in areas of no interest
  - Inefficient: the search directions are in general not independent.
  - Scan setup (range, number of steps) is problem/parameter dependent.

Gradient-based methods

Expanding the objective function at the present solution:

Hessian matrix

$$f(\boldsymbol{p}) \approx f(\boldsymbol{p}_0) + \frac{\partial f}{\partial \boldsymbol{p}}(\boldsymbol{p}_0) \cdot \Delta\boldsymbol{p} + \frac{1}{2}\Delta\boldsymbol{p}' \cdot \mathrm{H}(\boldsymbol{p}_0) \cdot \Delta\boldsymbol{p}$$

$$H_{ij} = \frac{\partial^2 f}{\partial p_i \partial p_j}$$

At the minimum $\frac{\partial f}{\partial \boldsymbol{p}}(\boldsymbol{p}_m) = \frac{\partial f}{\partial \boldsymbol{p}}(\boldsymbol{p}_0)+\mathrm{H}(\boldsymbol{p}_0) \cdot \Delta\boldsymbol{p}=0$, $\Delta\boldsymbol{p} = \boldsymbol{p}_m - \boldsymbol{p}_0$, which leads to

$$\mathrm{H}(\boldsymbol{p}_0) \cdot \Delta\boldsymbol{p} = -\frac{\partial f}{\partial \boldsymbol{p}}(\boldsymbol{p}_0)$$

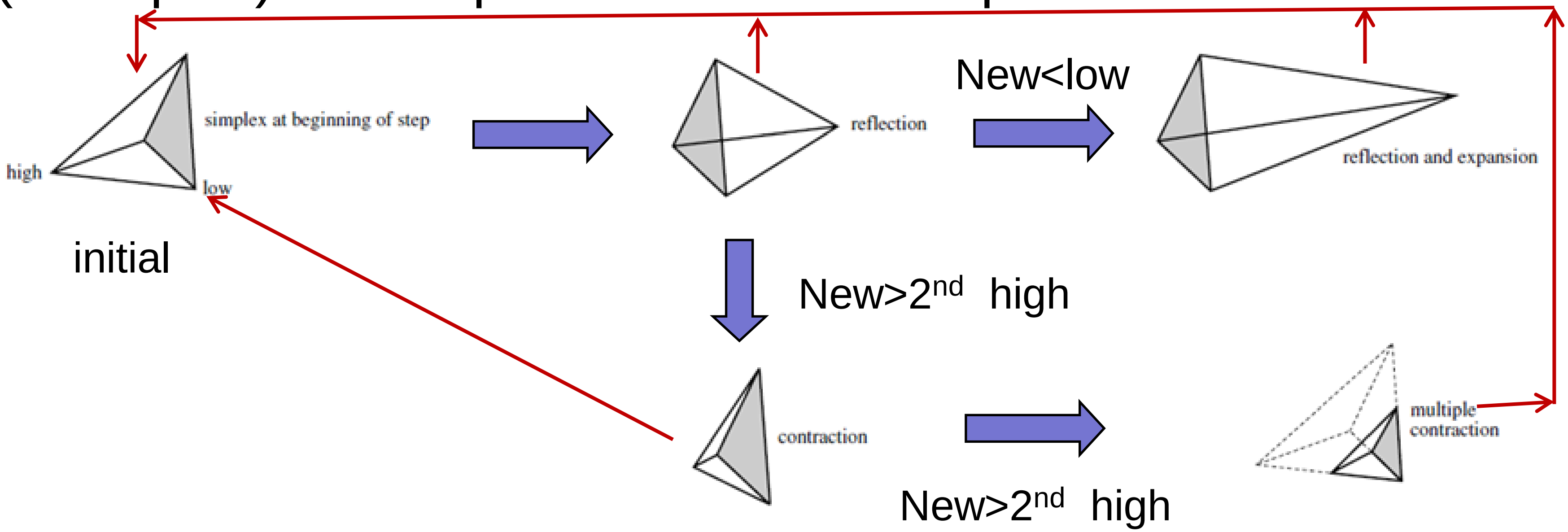
At each solution, solve for $\Delta\boldsymbol{p}$ to determine the next solution $\boldsymbol{p}_0 + \Delta\boldsymbol{p}$ (Gauss-Newton method)

Advantages: fast convergence.
Disadvantages:
(1) Requires the evaluation of the first and second order derivatives.
(2) Noise in the function introduces large errors to finite difference calculation.
(3) Solution $\Delta\boldsymbol{p}$ can be unrealistic if the approximation of f is too far off (can be remedied with trust region approach).

We don't consider gradient based methods for online optimization because of noise in online function evaluation.

Downhill Simplex Algorithm

- The downhill simplex algorithm is known for being a robust optimization method.
- For an N-variable problem, an $N+1$ vertex geometric body (a simplex) is manipulated with a few operations.



Advantage: efficient, simple to use.
Disadvantages: doesn’t work when noise in function alters comparison results of simplex vertices.

J.A. Nelder, R. Mead, Computer Journal vol 7, pp 308
W. H. Press, et al, Numerical Recipe, 3rd ed, (2007)

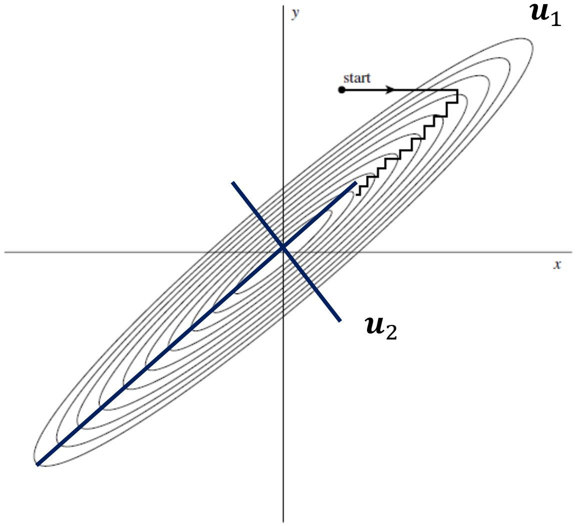
Powell's conjugate direction set method

- Powell's method has two components:
  - A procedure to update the direction set to make it a conjugate set
  - A line optimizer that looks for the minimum along each direction

Conjugate direction set: $\{\boldsymbol{u}_i\}, i = 1,2,\cdots,N_p$
in which for any $i \neq j$, it satisfied that

$$\boldsymbol{u}_i \cdot \mathrm{H} \cdot \boldsymbol{u}_j = 0$$

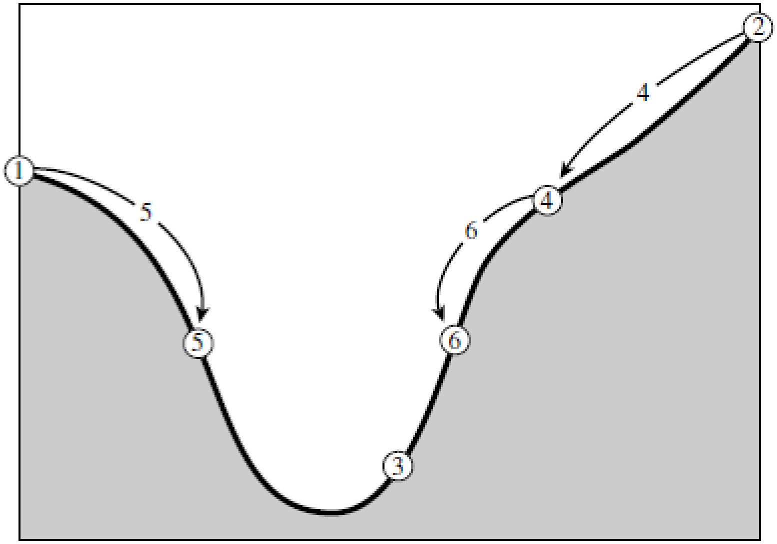
Search of optimum along one direction does not alter the previous search results along the conjugate directions (hence efficient).



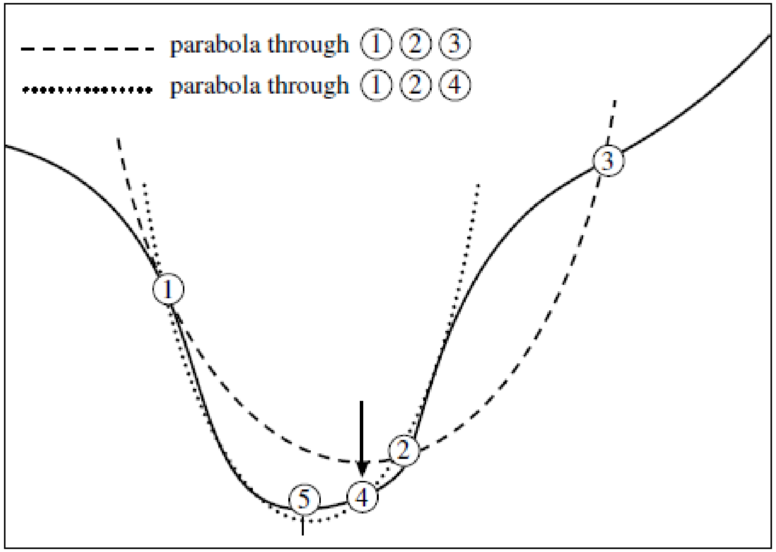
*M.J.D. Powell, Computer Journal 7 (2) 1965 155

Line optimizer

Line optimizer: start from the present solution $\boldsymbol{p}_0$, along direction $\boldsymbol{u}$, minimize function $g(\lambda) = f(\boldsymbol{p}_0 + \lambda \boldsymbol{u})$.



Golden section search



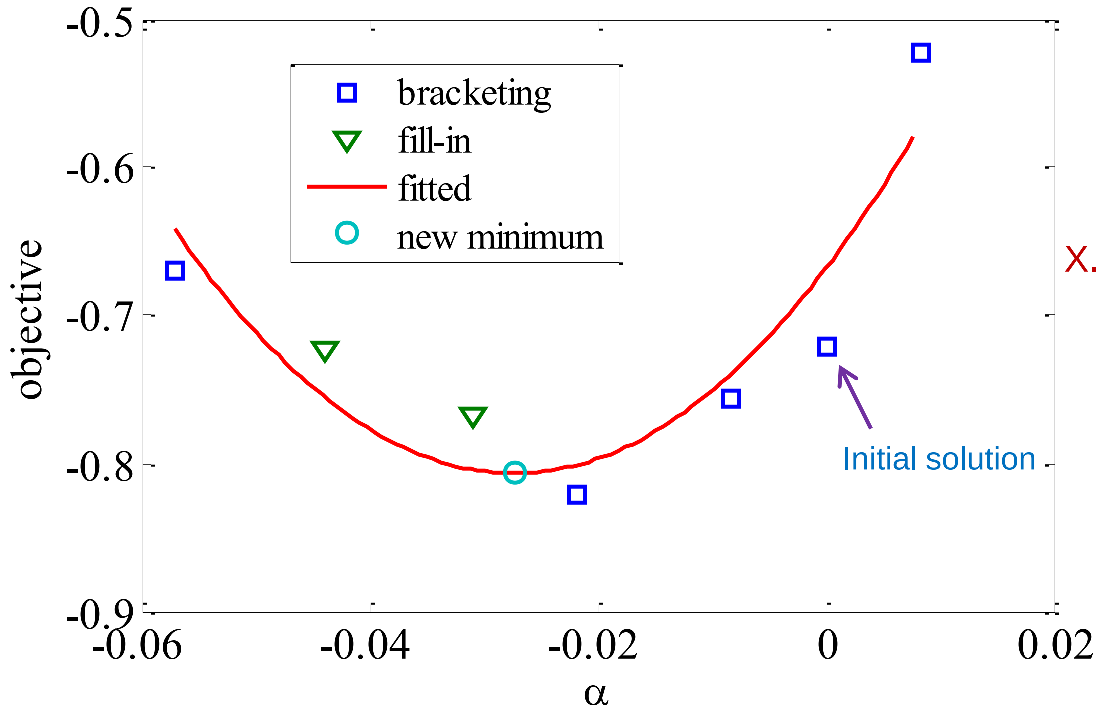
Parabolic interpolation.

First bracket the minimum (with a center point below two outside ones). Iteratively reduce the bracket size with additional intermediate point.

Both algorithms suffer from noise as it alters comparison results. .

W. H. Press, et al, Numerical Recipe, 3rd ed, (2007)

The robust line optimizer of RCDS



X. Huang, et al, NIMA 726 (2013)

Step 1: bracketing the minimum with noise considered.
Step 2: Fill in empty space in the bracket with solutions and perform quadratic fitting. Remove any outlier and fit again. Find the minimum from the fitted curve.

Global sampling within the bracket helps reducing the noise effect.

RCDS is Powell's conjugate method* + the new robust line optimizer.

*however, since the online run time is usually short, it is important to provide good an initial conjugate direction set which may be calculated with a model.

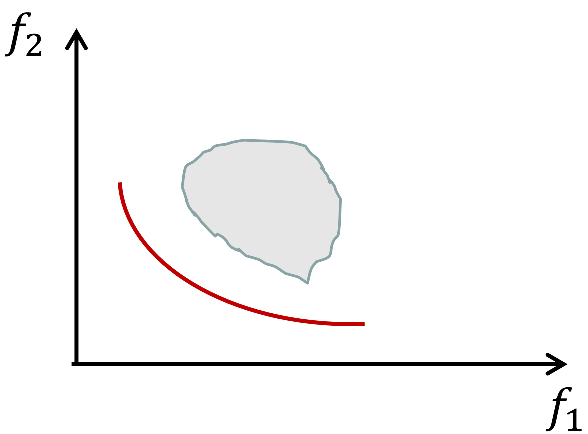
multi-objective optimization: comparison of solutions

Problem: minimize $f_m(\boldsymbol{x})$, $m = 1,2,\ldots,M$ with parameter ranges $x_i \in [x_i^L, x_i^U]$

Comparison of two solutions (definition of domination): Solution x_a dominates solution x_b if for all $m = 1,2,\ldots,M$, we have $f_m(x_a) \leq f_m(x_b)$ and for at least one objective m', $f_m(x_a) < f_m(x_b)$.

Pareto front: the set of all solutions in the search space that are non-dominated by any solutions.

Goal of multi-objective optimization is to obtain the Pareto front for further analysis.



A multi-objective genetic algorithm (MOGA): NSGA-II

- NSGA (non-dominated sorting genetic algorithm)–II

Selection (of parents)
Crossover
Mutation

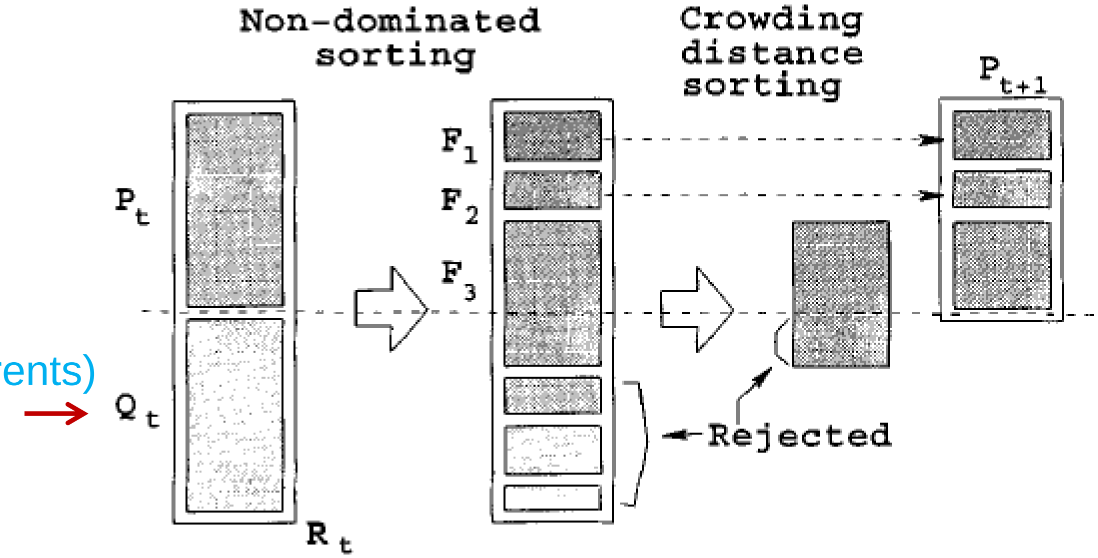
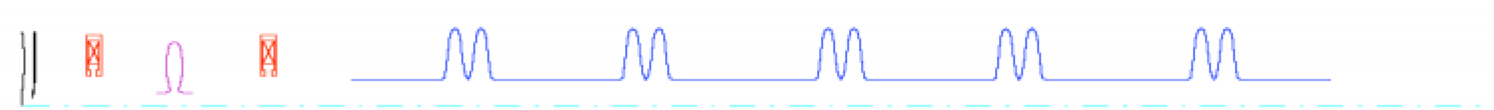


Fig. 2. NSGA-II procedure.

K. Deb, IEEE Transtions On Evolutionary Computation Vol 6, No 2,April 2002

Application of NSGA-II: injector optimization (Cornell)



Decision variables:

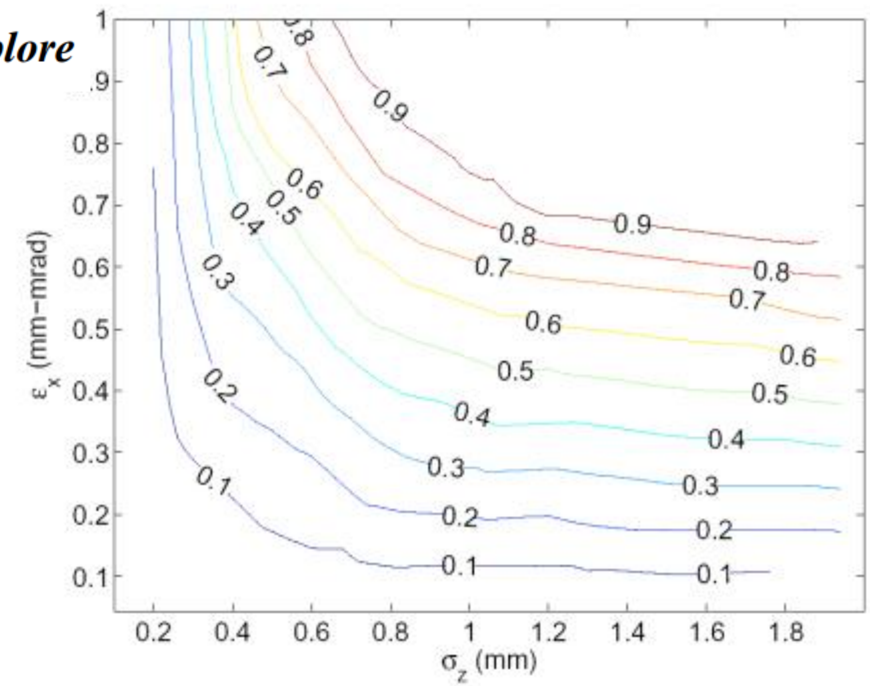


FIG. 10: Transverse emittance vs. bunch length for various charges in the injector (nC).

Total: 22-24 dimensional parameter space to explore

Fields:

DC Gun Voltage (300-900 kV)
2 Solenoids
Buncher
SRF Cavities Gradient (5-13 MV/m)
SRF Cavities Phase

Bunch & Photocathode:

$E_{thermal}$
Charge

Positions:

2 Solenoids
Buncher
Cryomodule

Laser Distribution:

Spot size
Pulse duration (10-30 ps rms)
{tail, dip, ellipticity} · 2

I. Bazarov, PRSTAB 8, 034202, (2005)

Multi-objective particle swarm optimization (MOPSO)

MOPSO also manipulates a population of solutions over many iterations with random operations.

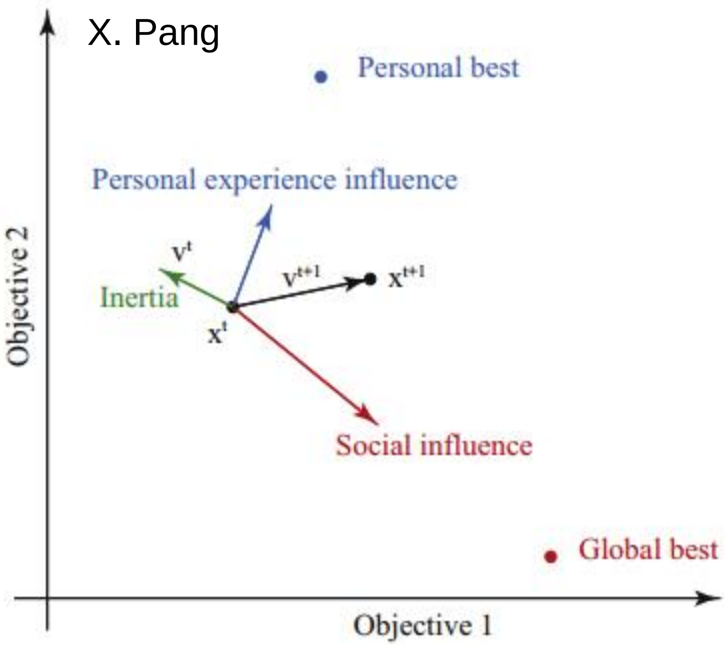


Fig. 1. Velocity and position updates in PSO.

Updating particle population in an iteration

$$\boldsymbol{x}_i^{t+1} = \boldsymbol{x}_i^t + \boldsymbol{v}_i^{t+1}$$
$$\boldsymbol{v}_i^{t+1} = w\boldsymbol{v}_i^t + c_1 r_1\left(\boldsymbol{p}_i^t - \boldsymbol{x}_i^t\right) + c_2 r_2\left(\boldsymbol{g}^t - \boldsymbol{x}_i^t\right)$$

$\boldsymbol{x}_i^t$ position (parameter vector) of particle i at iteration t.
$\boldsymbol{v}_i^t$ velocity (increment) of particle i at iteration t.
Control parameters: $w = 0.4$, $c_1 = c_2 = 1$.
r_1,r_2 are random within $[0, 1]$ or fixed values.

MOPSO can also include mutation operation.

X. Pang, L.J. Rybarcyk, NIMA 741 (2014)

Comparison of MOGA and MOPSO in a study

Optimizing transmission and matching to the DTL of the LANSCE linac at LANL:
Objectives: beam loss and mismatch factors
Knobs: four quadrupoles in the transport line prior to DTL.

X. Pang, L.J. Rybarcyk, NIMA 741 (2014)

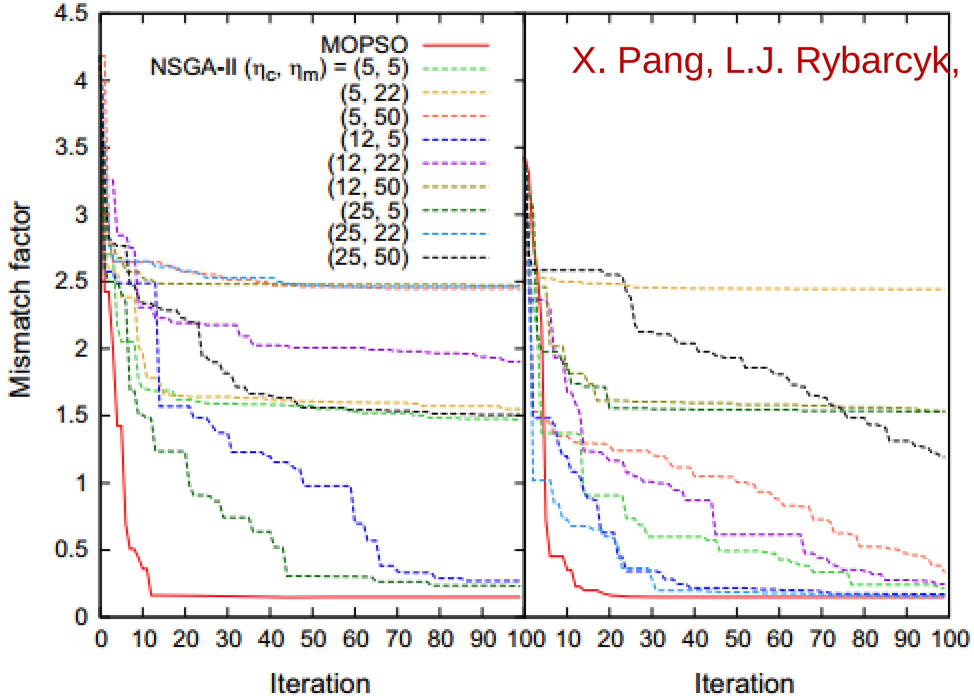
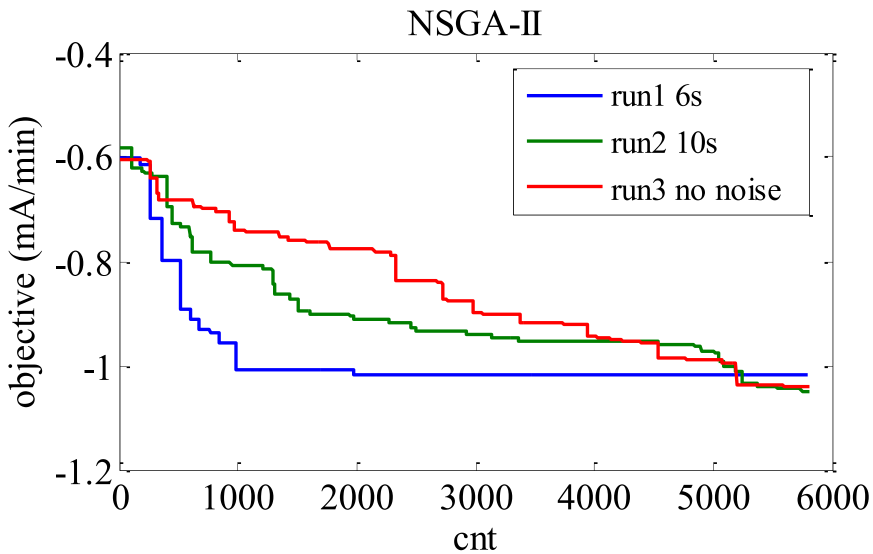


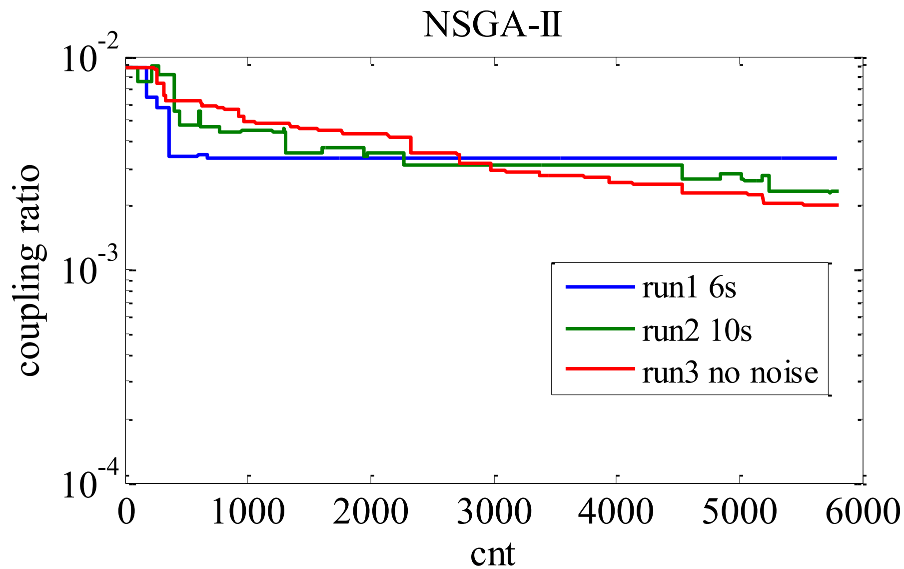
Fig. 2. The two plots depict simulations started with two different initial random populations. Both NSGA-II and MOPSO converged to similar final solution, but MOPSO did it in fewer iterations. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

A simulation study: coupling correction for SPEAR3

- Goal: minimize linear coupling (measured by coupling ratio $\frac{\epsilon_y}{\epsilon_x}$) in the storage ring.
- Setup
  - Using calculated beam loss rate as the objective function (for Toushcek lifetime dominated beams, loss rate is inversely proportional to coupling).
  - Noise is generated in the objective function by adding random noise to beam current values (used for loss rate calculation).
  - There are 13 coupling correction skew quads in SPEAR3.
  - Initial conjugate direction set is from SVD of the Jacobian matrix of orbit response matrix w.r.t. skew quads
  - With
    - 500 mA beam current with 1% random variation. On top of that a DCCT noise with sigma = 0.003 mA. The beam loss rate noise evaluated from 6-s duration is 0.06 mA/min.
    - 40 hour gas lifetime; 10 hour Touschek lifetime with 0.2% coupling.
    - The coupling ratio with all 13 skew quads off is 0.9% (with simulated error), corresponding to a loss rate of 0.6 mA/min.

MOGA (NSGA-II)

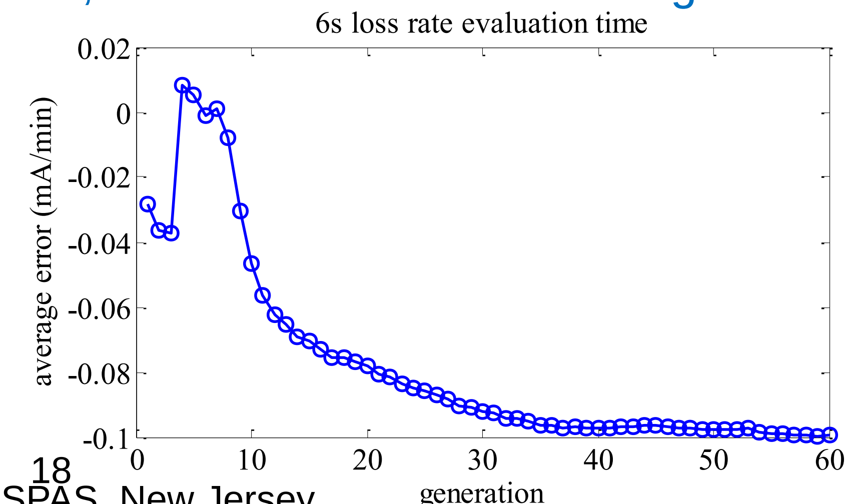




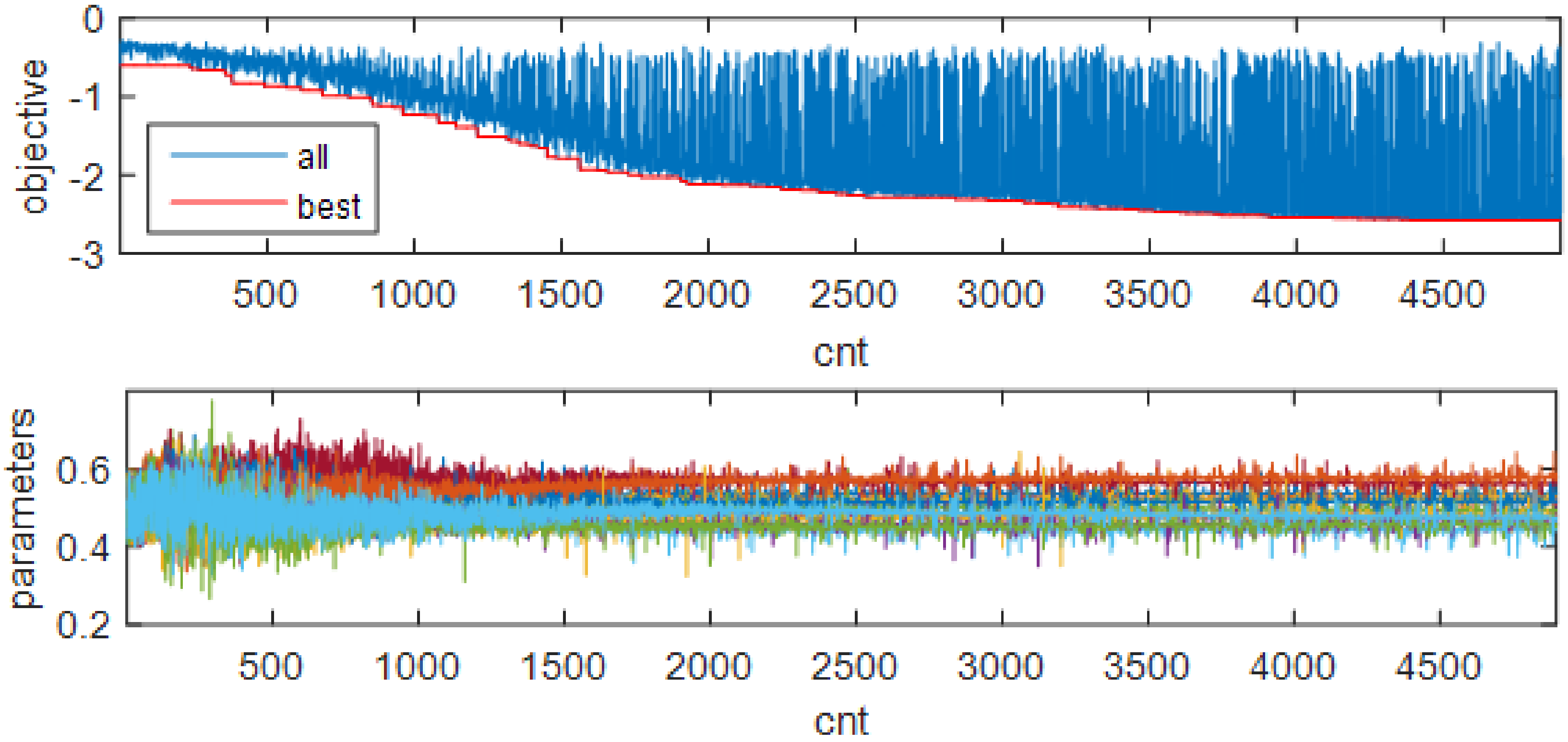
Population: 100; Ran 60 generations; 10% mutation, 90% crossover.

- MOGA is inefficient, even for cases without noise. In addition,
- Noise puts a bias on the selection operator, which limits it from reaching real optimum.

Solutions luckier than the average by nearly two sigma dominate the population. The selection operation is biased.

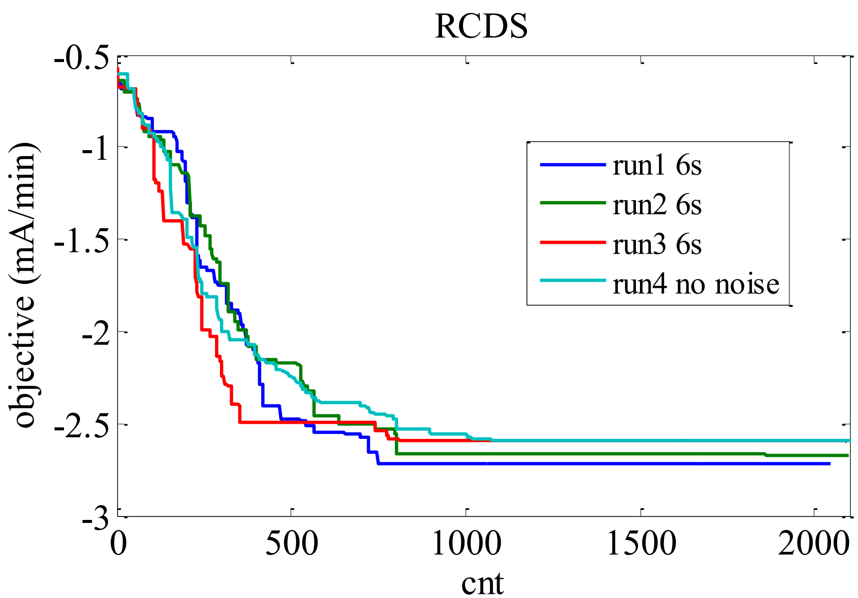


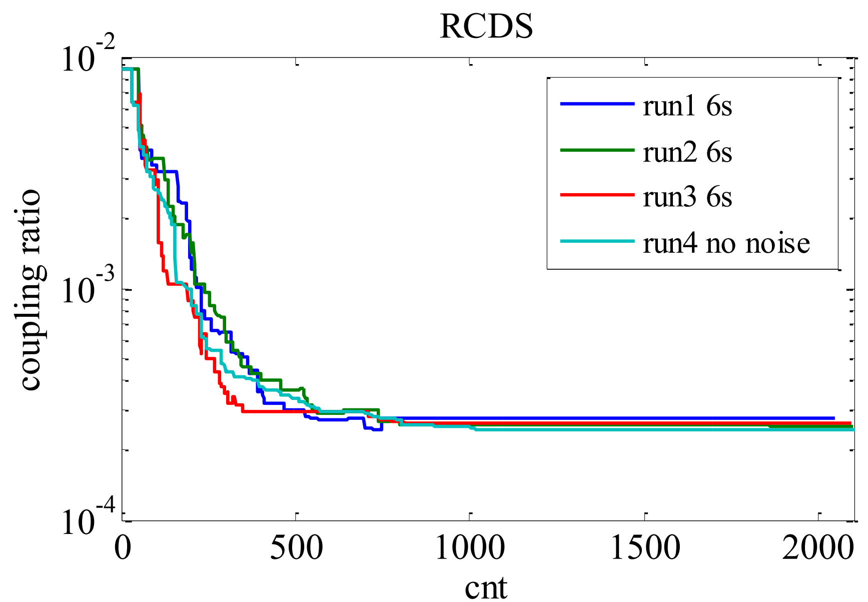
Particle swarm optimization (PSO)



Noise level corresponding to 10s loss.
PSO results are not biased by noise.
The convergence is faster, compared to MOGA.

RCDS





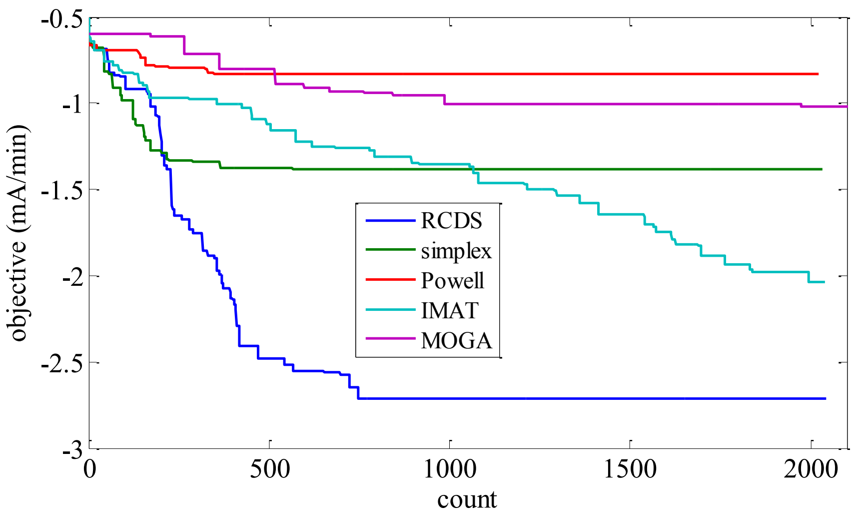
The RCDS method has nearly the same performance under noise (at a high level) as the case without noise.
Calculation of the conjugate direction set with model:
SVD of the Jacobian of the orbit response matrix w.r.t. the skew quadrupoles.

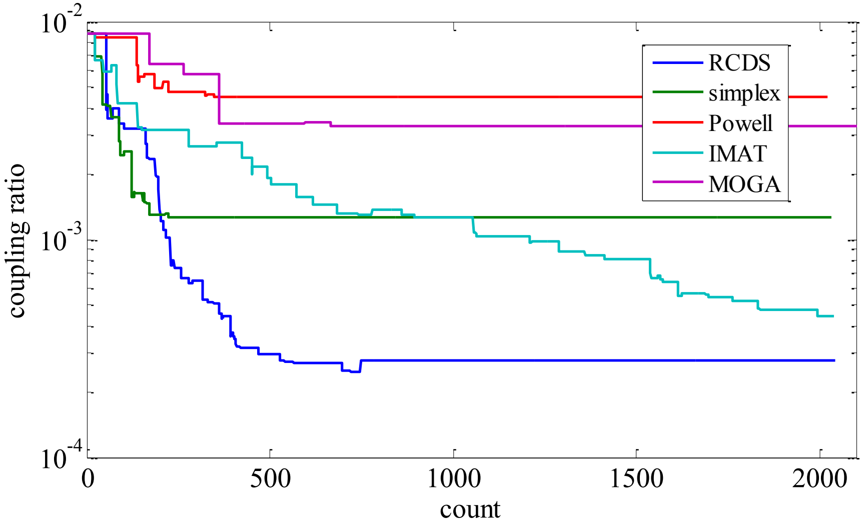
Jacobian $\mathbf{J}$, with $J_{ij} = \frac{\partial r_i}{\partial p_j}$, and $\mathbf{r}$ the residual vector ($\chi^2 = \mathbf{r}^{\mathbf{T}} \cdot \mathbf{r}$), $\mathbf{p}$ the parameter vector.

$$\mathbf{J} = \mathbf{USV}^T$$

Use columns in $\mathbf{V}$ as conjugate direction set.

Comparison of performance (best of 6-s cases)





The IMAT method uses the same RCDS line optimizer, but keep the direction set of unit vectors (not conjugate).

X. Huang, et al, NIMA 726 (2013)

Clearly,

(1) The line optimizer is robust against noise.

(2) Searching with a conjugate direction set is much more efficient.

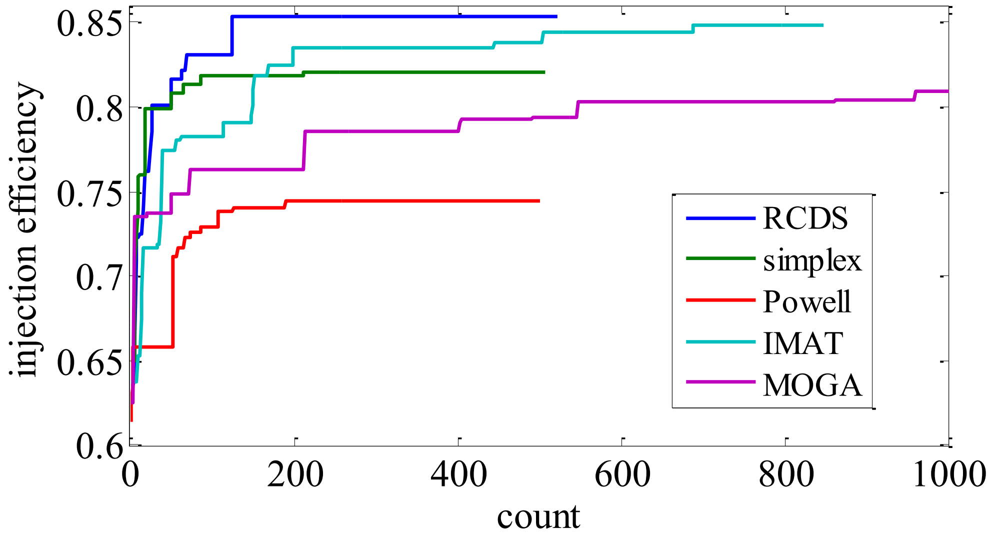
(3) Original Powell's method, downhill simplex and MOGA are not effective for noisy problems.

Note that the direction set has been updated only about 8 times after 500 evaluations (out of 13 directions). So the high efficiency of RCDS is mostly from the original direction set.

2nd simulation study: transport line optics

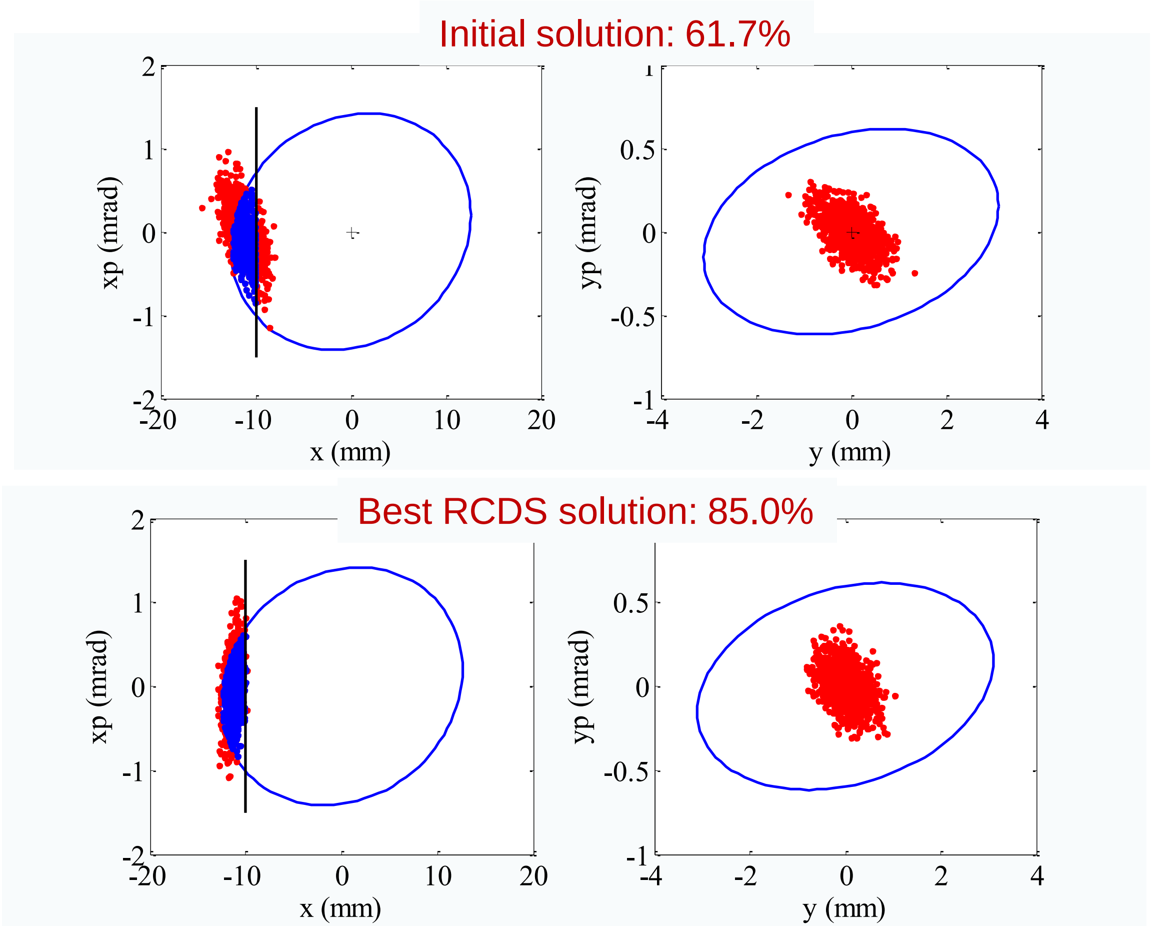
BTS transport line optics matching:

Varying 6 quads in BTS for optics matching between transport line and storage ring.
Noise is generated from the finiteness of the number of particles.
With 1000 particles in a distribution, the noise sigma of injection efficiency is 1.6%.
Initial conjugate direction set is from SVD of beam moments (elements of the σ-matrix) w.r.t. quads. Dynamic aperture of the ring is intentionally shrunk to 12.5 mm in the simulation.

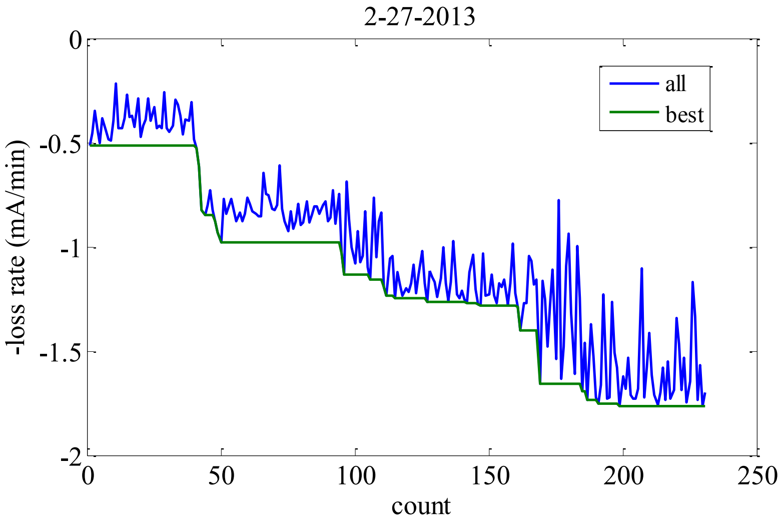


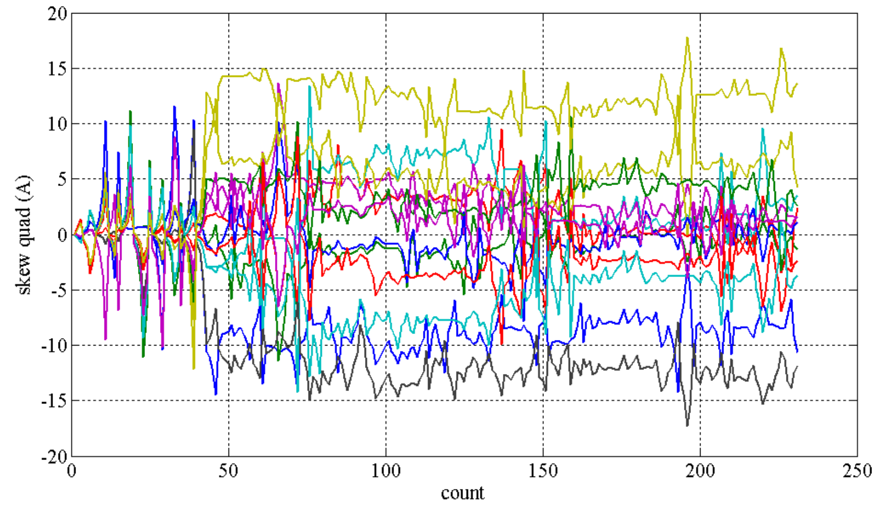
Performance similar to the coupling study is observed.

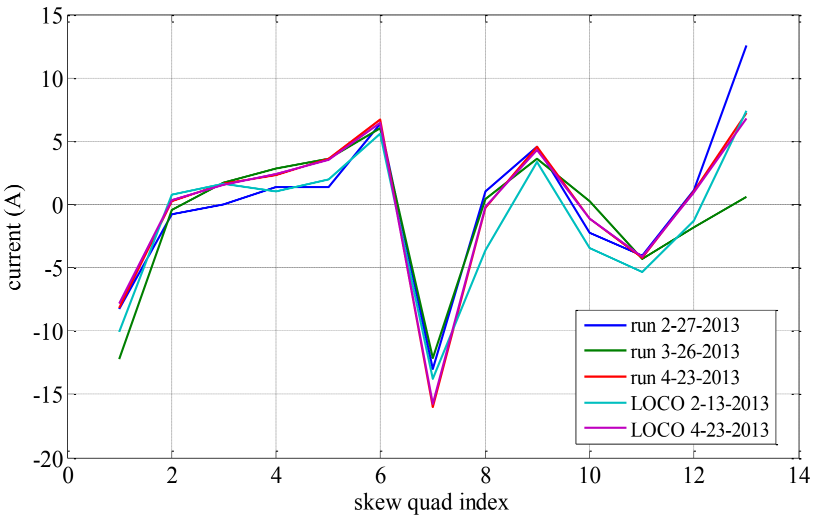
Best solution



Experiment: coupling correction with loss rate







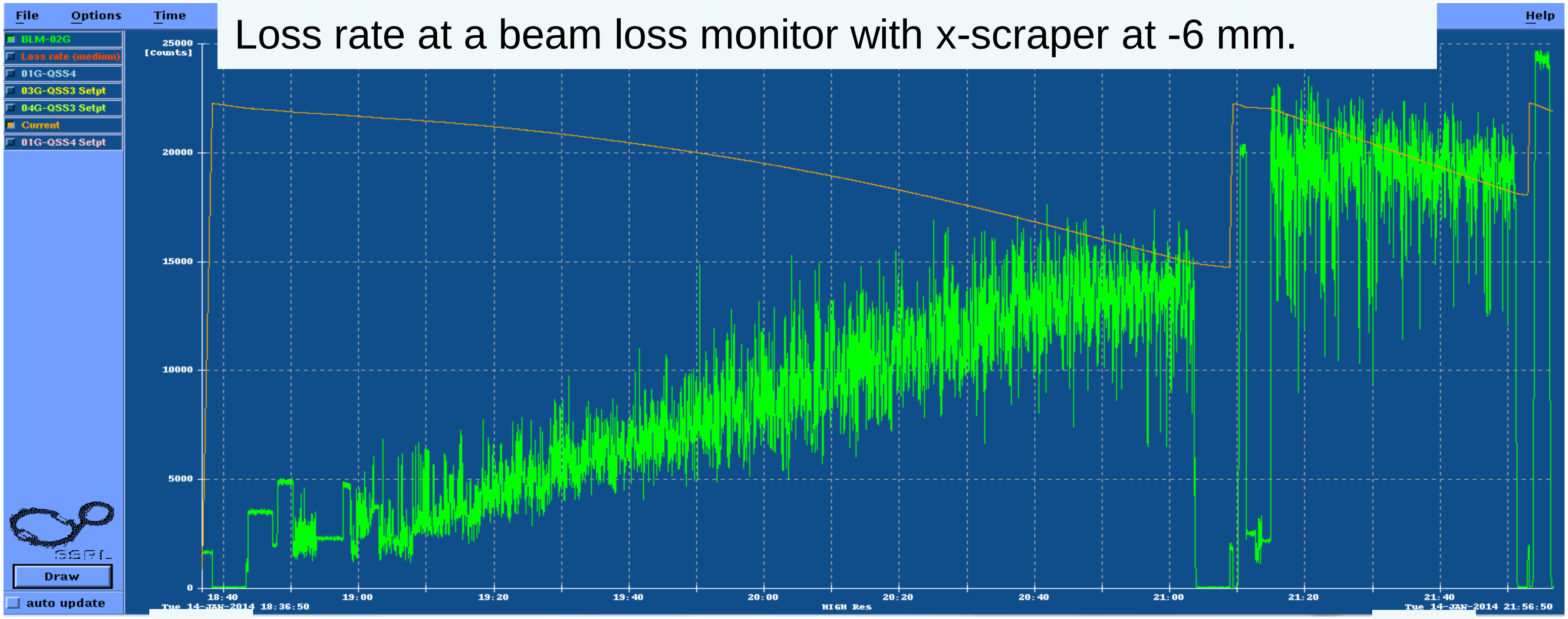
Beam loss rate is measured by monitoring the beam current change on 6-second interval (no fitting). Noise sigma 0.04 mA/min.
Data were taken at 500 mA with 5-min top-off.

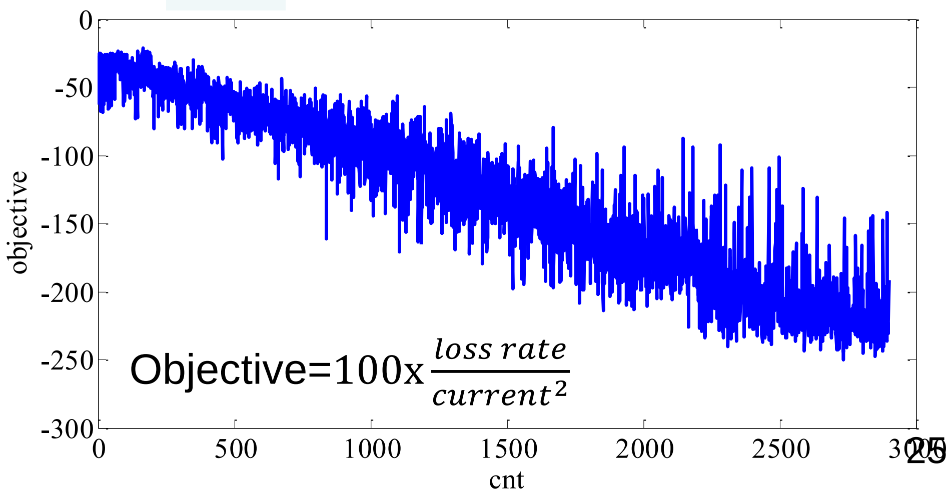
Initially setting all 13 skew quads off. Loss rate at about 0.4 mA/min.
Final loss rate at about 1.75 mA/min.
At 500 mA, the best solution had a lifetime of 4.6 hrs.
This was better than the LOCO correction (5.2 hrs)

Best result with RCDS is loss rate >2.0 mA/min and 500 mA lifetime 4.2 hrs.

Coupling correction with MOPSO – experiment

Loss rate at a beam loss monitor with x-scraper at -6 mm.



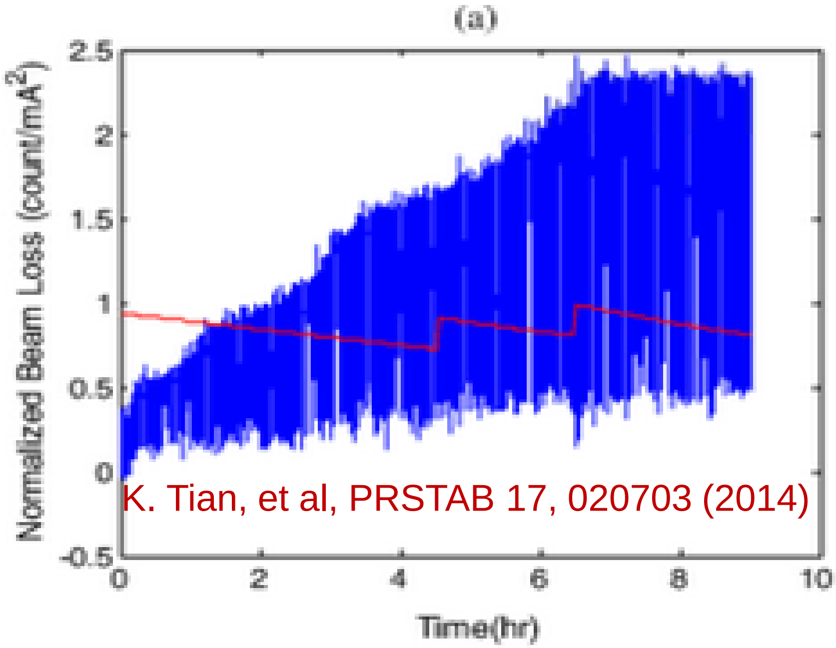


$$\text{Objective}=100\text{x}\frac{loss\ rate}{current^2}$$

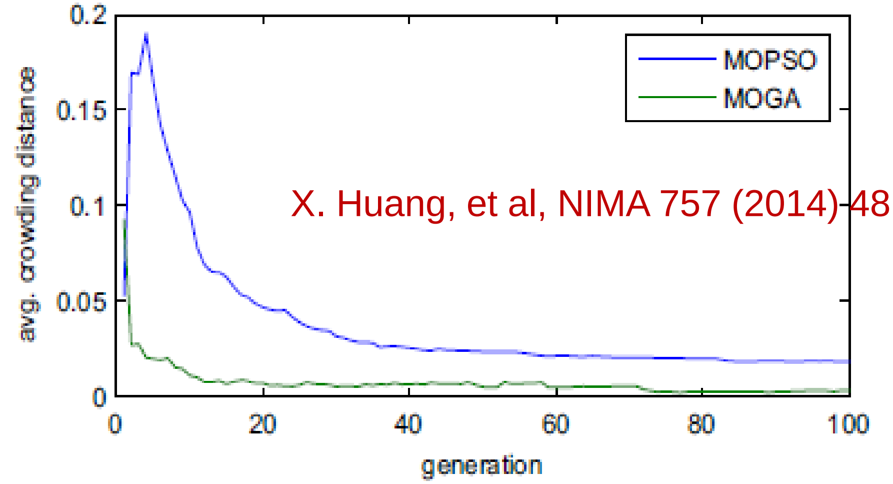
After optimization, for the best solution, at 500 mA, the lifetime is 3.78 hrs.
LOCO data showed that coupling ratio is 0.029%, lowest on SPEAR3.

The experiment took less than 3000 evaluations.

Coupling correction with MOGA



K. Tian, et al, PRSTAB 17, 020703 (2014)



X. Huang, et al, NIMA 757 (2014) 48

Diversity of PSO solutions is significantly better than GA solutions.

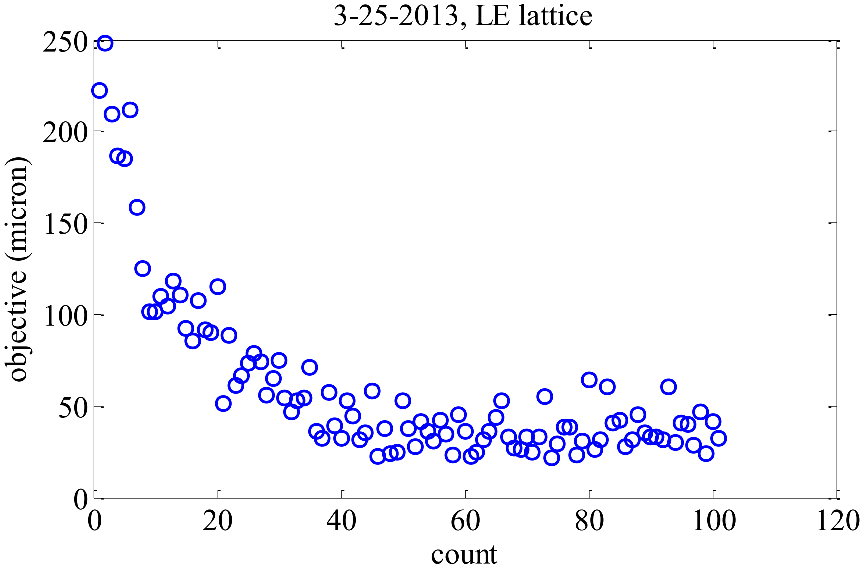
Algorithm used was NSGA-II.
Beam loss monitor data was used as objective function.
Horizontal scraper was moved in to concentrate beam loss in the loss monitor area.

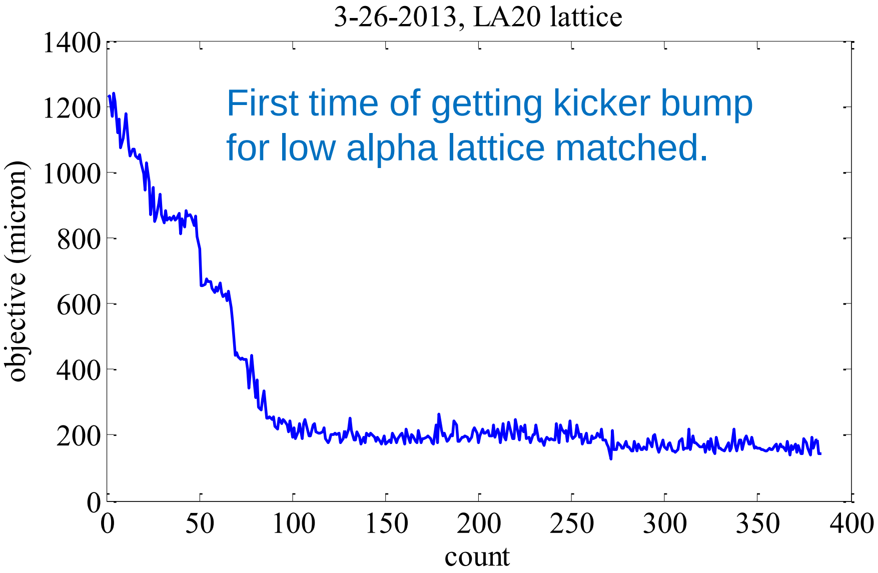
It took ~20000 evaluations

Experiment: kicker bump matching

Parameters: Adjusting pulse amplitude, pulse width and timing delay of K1 and K3 (with K2 fixed) and two skew quads for vertical plane (This is the setup by James for kicker bump matching), 8 parameters total.

Objective: sum of rms(x) and rms(y) of turn-by-turn orbit (for 30~300 turns).

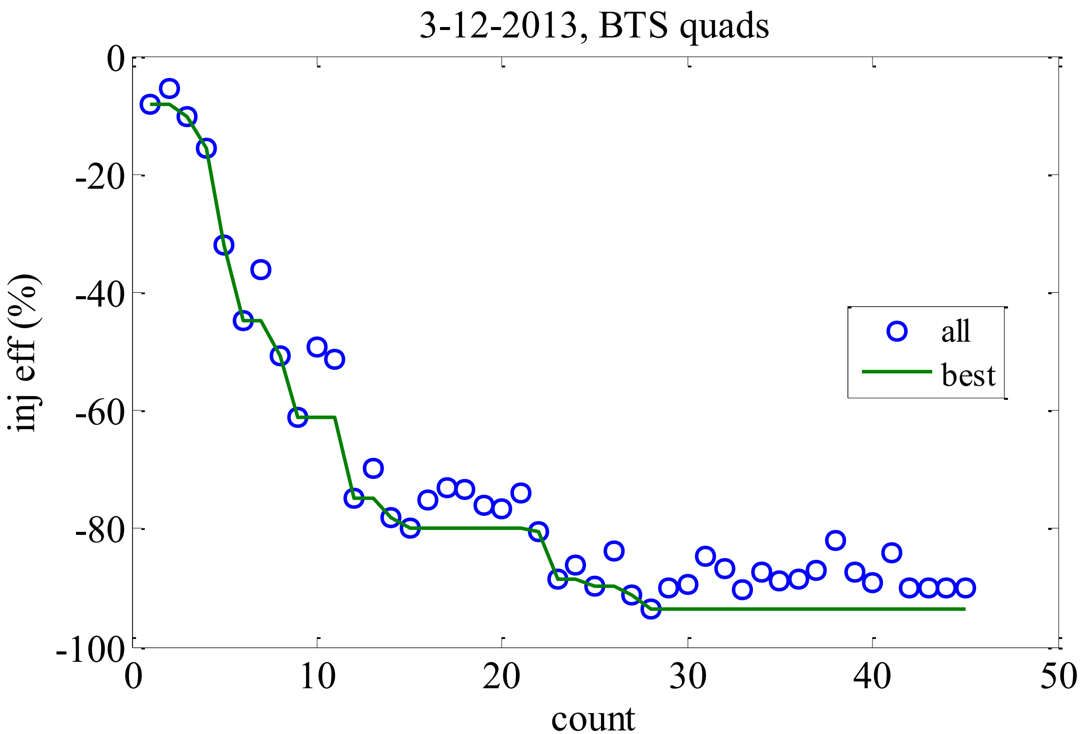




Experiment: BTS optics

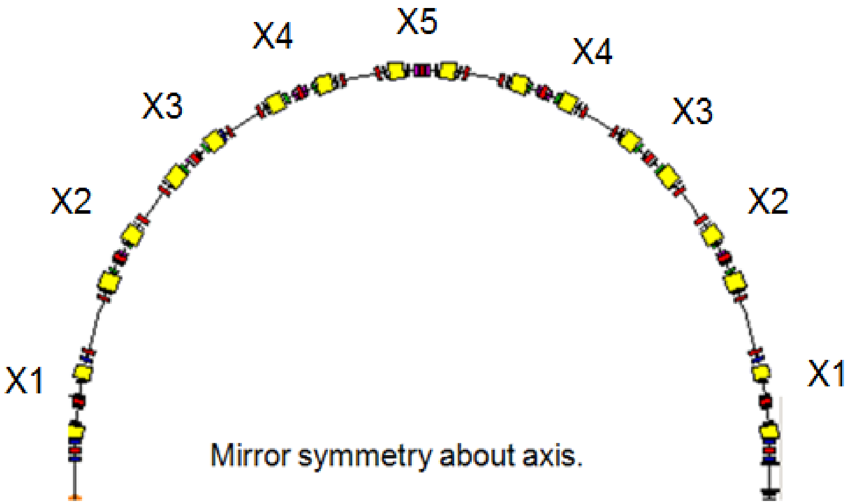
For injection optics matching, ideally we need to decouple the steering effect of quadrupole changes (when beam trajectory is off-center). But we haven't done that yet.
To test the algorithm, we changed the 9 BTS quads setpoints randomly to mess up injection and use the code to bring injection back.

Knobs: 9 BTS quads.
Objective: injection efficiency.

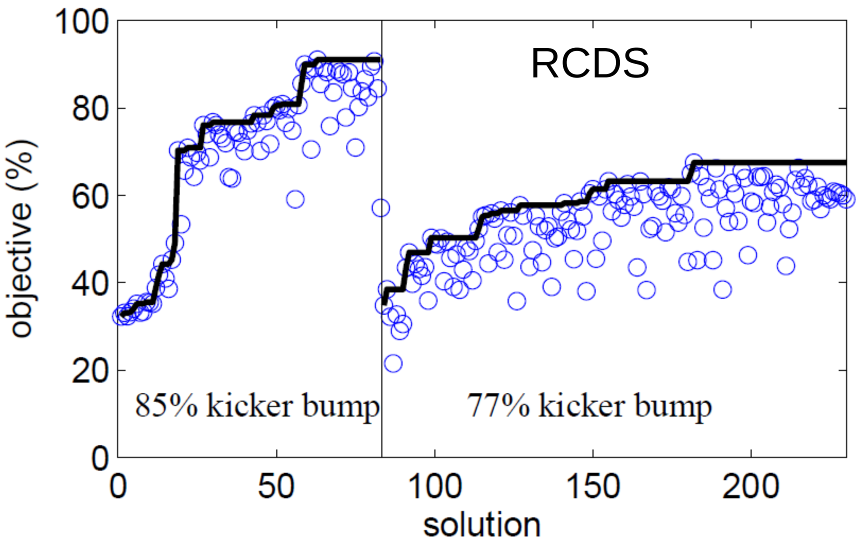


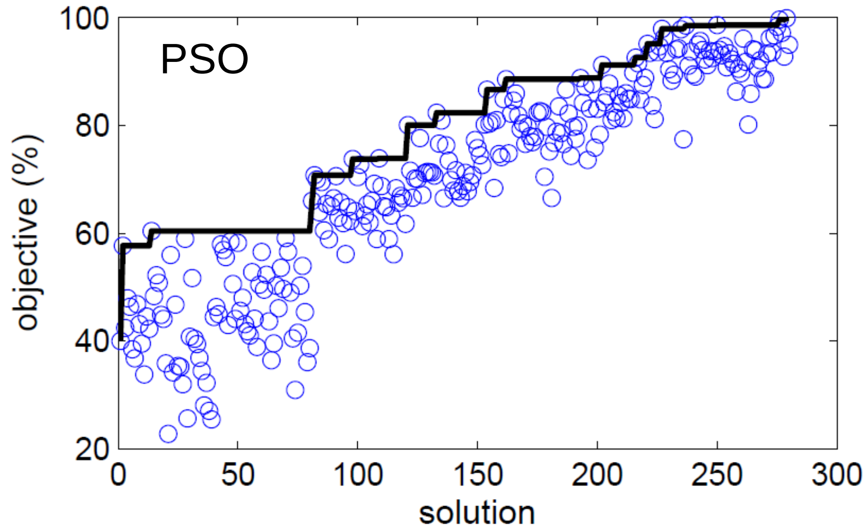
Optimization of nonlinear dynamics

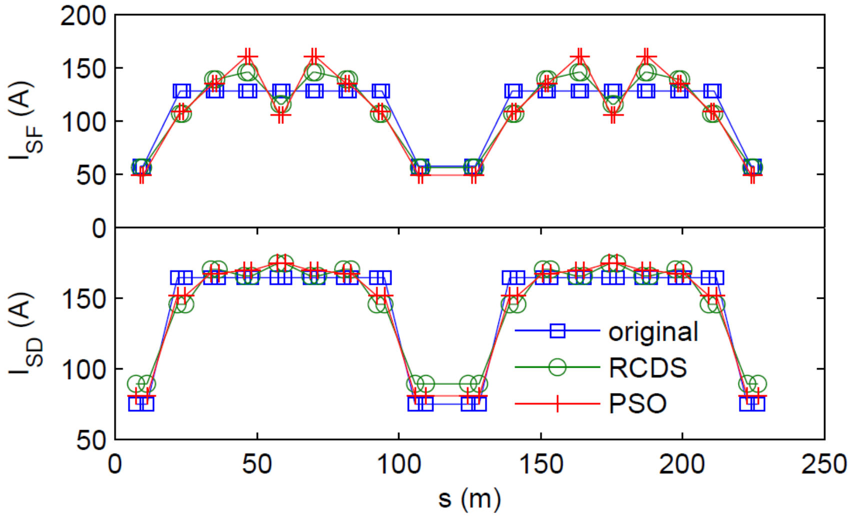
- 10 independently powered sextupole groups
- reduced the kicker bump by 34% to have a low initial injection efficiency
- Injector is detuned to lower total beam loss during experiment.
- Knobs: 8 variables in the subspace of the 10-dim parameter space that keep chromaticities fixed (basis of the null space of the chromaticity response matrix).
- Objective: Injection efficiency calculated as beam current change over 10 seconds normalized by average Booster beam current within the period.



Results with RCDS and PSO algorithms



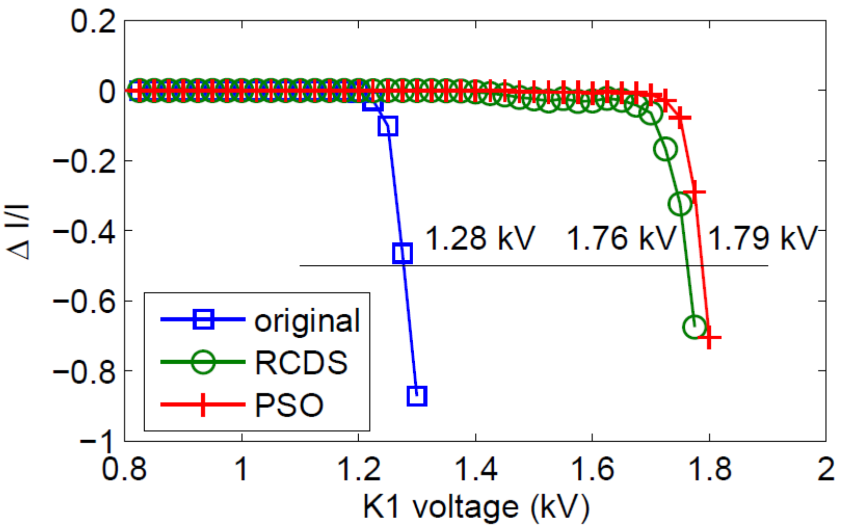


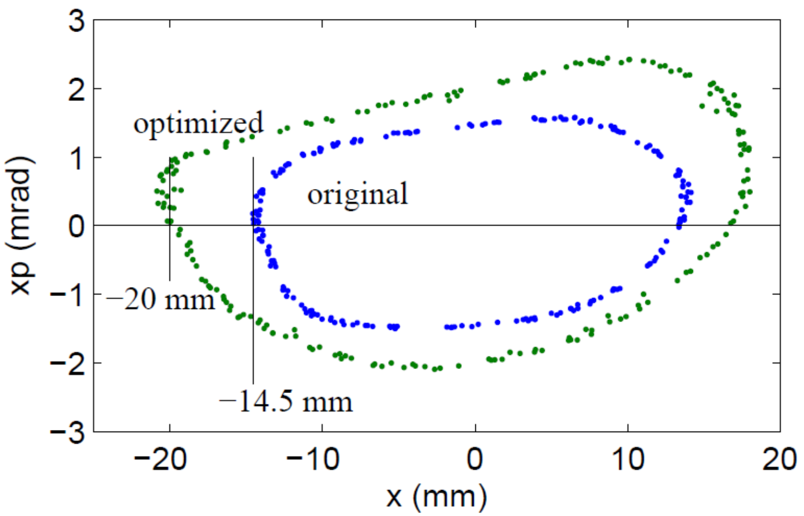


Application of RCDS and PSO algorithms resulted in significant improvement of dynamic aperture.

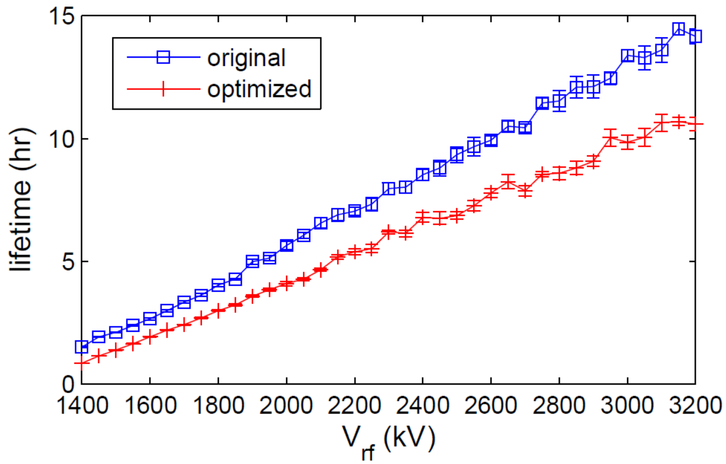
X. Huang, J. Safranek, http://arxiv.org/abs/1502.07799

Verification of best solution





Dynamic aperture measurement indicates an increase of > 5 mm.



Momentum aperture is not negatively affected.

Summary

- There is a need of online optimization algorithm in the era of computerized control. Biggest challenge to conventional algorithms is noise in function evaluation.
- The RCDS method is demonstrated to be robust against noise and efficient when conjugate direction set is supplied.
- The RCDS method has been successfully applied to many accelerator optimization problems.
- When stochastic algorithms are desired, the particle swarm method (PSO) is preferred.