Multiobjective optimization of dynamic aperture

Lingyun Yang, Yongjun Li, Weiming Guo, and Samuel Krinsky

Photon Sciences Directorate, Brookhaven National Laboratory, Upton, New York 11973, USA
(Received 18 November 2010; published 2 May 2011)

Dynamic aperture (DA) is one of the key nonlinear properties for a storage ring. Although there have been both analytical and numerical methods to find the aperture, the reverse problem of how to optimize it is still a challenging problem. A general and flexible way of optimizing the DA is highly demanded in accelerator design and operation. In this paper, we discuss the use of multiobjective optimization for DA. First we consider using objective functions based only on numerical tracking results. Data mining of these results demonstrated a correlation between DA and low-order nonlinear driving terms. Next we considered using objective functions which included both numerical tracking results and analytical estimates of low-order nonlinear driving terms. This resulted in faster convergence. The National Synchrotron Light Source II (NSLS-II) lattice was taken as an example to illustrate this method. This multiobjective approach is not limited by particular linear or nonlinear lattice settings, and can also be applied for optimizing other properties of a storage ring.

PACS numbers: 41.85.−p, 29.20.db, 29.27.−a

I. INTRODUCTION

Dynamic aperture (DA) is one of the key nonlinear properties of a storage ring lattice in both design and operation stages. It strongly connects to beam lifetime and injection efficiency. Without any exception, the new design or upgrade of a storage ring has to consider this quantitatively [1–6]. The DA is limited by the nonlinear forces on circulating beams. The forces can be systematic or random nonlinear fields due to sextupoles used for adjusting the chromaticity, octupoles for stabilizing the particles collective motion, wigglers and undulators in synchrotron light sources, or beam-beam interaction in colliders. There have been both analytical methods which estimate the DA size [7,8] and direct tracking methods for precise DA calculation [9,10], but the reverse problem, i.e., how to optimize it, is still challenging for accelerator physicists.

For the optimization of DA, previous work has been focused on correcting the multipole errors and minimizing the nonlinear driving terms [11–13], where usually partial realistic random errors are included. Detailed particle tracking which has the effects from random misalignment errors, multipole errors, insertion devices, and small apertures is then applied afterward to validate the optimization. Tools like frequency map analysis (FMA) can be used to examine the tune diffusion and resonance structures [14].

In this paper, we present multiobjective approaches to DA optimization. First, we consider using objective functions based only on numerical tracking results. Data

mining of these results demonstrated a correlation between DA and low-order nonlinear driving terms. Next we considered using objective functions which included both numerical tracking results and analytical estimates of low-order nonlinear driving terms. This resulted in faster convergence. These approaches have been applied with success to the National Synchrotron Light Source II (NSLS-II) lattice [15] and similar methods and results are also recognized in works from the Advanced Photon Source [16]. The NSLS-II lattice was taken only as an example to illustrate this method and this paper does not present the final results of the NSLS-II DA optimization.

Section II is a short introduction to nonlinear driving terms. Section III introduces the conceptual basis for multiobjective optimizer, the objective functions, and constraints we are using for DA optimization. In Sec. IV A we discuss constraints that we impose in order to make the optimization robust. In Sec. IV B we outline two approaches that we have studied. The first considers only numerical tracking results and optimizes both the on-momentum and the off-momentum DA area. The second approach optimizes the on- and off-momentum DA and in addition analytic expressions for tune shift with amplitude. It is found that including the tune shifts in the optimizer improves convergence speed. Both approaches are illustrated using the NSLS-II lattice as an example. Results of the first approach are discussed in Sec. IV C. In Sec. IV D, we exhibit the correlation between DA and the low-order nonlinear driving terms, including the magnitude of tune shift with amplitude, by performing a data mining on the results of Sec. IV C. In Sec. IV E we illustrate the use of an optimizer using both numerical tracking data and tune shift with amplitude as objective functions. Conclusions are presented in Sec. V. The NSLS-II lattice properties are explained in more detail in Appendix A, and Appendix B

gives more algorithm details on generating new children in the optimization iterations.

II. NONLINEAR DRIVING TERMS

The single-particle dynamics of a circulating beam are governed by the one turn map. In the normal form analysis, the invariant or effective Hamiltonian of the whole ring is normalized in resonance bases [17]. The nonlinear driving terms are those that can excite certain resonances in the Hamiltonian. Minimization of the driving terms has been used for improving the DA [13,18]. In this paper we call this method the analytical approach.

For a ring with n elements, we can normalize the one turn map $\mathcal{M}_{1 \to n}$ as [13]

$$\mathcal{M}_{1 \to n} = \mathcal{A}_1^{-1} e^{:h:} \mathcal{R}_{1 \to n} \mathcal{A}_1, \qquad (1)$$

where $\mathcal{R}$ is a rotation, $e^{:h:}$ is the nonlinear Lie map, $\mathcal{A}_1$ is a normalizing map, and subscripts 1 and n are the indices of ring magnets. For linear lattice without coupling, we can apply perturbation theory and approximate the effects of the normalizing map $\mathcal{A}_i$ at the ith element as

$$\mathcal{A}_i x = \sqrt{\beta_{x,i}} x + \eta_{x,i} \delta,$$
$$\mathcal{A}_i p_x = \frac{-\alpha_{x,i} x + p_x}{\sqrt{\beta_{x,i}}} + \eta'_{x,i} \delta, \qquad (2)$$

where $\beta_{x,i}$ and $\eta_{x,i}$ are the horizontal beta function and dispersion at the ith magnet, and δ is the momentum deviation. Similar results also apply for the y plane. In the resonance basis $h_x^\pm \equiv \sqrt{2J_x} e^{\pm i\phi_x} = x \mp i p_x$, $\mathcal{R}_{i \to j}$ is merely a rotation of the phase advance between the ith and the jth element, (J, ϕ) are action-angle variables, and we have

$$\mathcal{R}_{i \to j} h_x^\pm = \mathcal{R}_{i \to j} \sqrt{2J_x} e^{\pm i\phi_x} = e^{\pm i\mu_{i \to j,x}} h_x^\pm, \qquad (3)$$

where $\mu_{i \to j,x}$ is the phase advance of $i \to j$. Therefore the potentials $V(x, y)$ of a multipole magnetic field, e.g. $V(x, y) = K_2(x^3 - 3xy^2)/6$ for a sextupole with strength K_2, can also be expanded in the resonance bases h_{abcde}:

$$h = \sum h_{abcde} h_x^{+a} h_x^{-b} h_y^{+c} h_y^{-d} \delta^e. \qquad (4)$$

The nonlinear driving terms h can be grouped by their order $n = a + b + c + d + e$ and each h_{abcde} is an explicit analytical function of magnet strengths, beta functions, and dispersions [13].

Each of the h_{abcde} drives a certain quantity or resonance, and some of the low-order driving terms are shown in Table I. Each of these terms has a closed analytical expression. By minimizing the driving terms h_{abcde} and tune shift with amplitude, the resonances and nonlinear effects get suppressed. Compared with the particle tracking for the DA, this analytical approach is indirect but requires less computing. Depending on the working point and magnet

TABLE I. Various low-order driving terms.

h_{abcde}	Driving effects
h_{11001}, h_{00111}	Linear chromaticity ζ_x, ζ_y
h_{21000}, h_{12000}	ν_x
h_{30000}, h_{03000}	$3\nu_x$
h_{10110}, h_{01110}	ν_x
h_{10020}, h_{01200}	$\nu_x - 2\nu_y$
h_{10200}, h_{01020}	$\nu_x + 2\nu_y$
h_{22000}, h_{00220}, h_{11110}	$d\nu_x/dJ_x$, $d\nu_y/dJ_y$, $d\nu_{x,y}/dJ_{y,x}$

type, different lattices may have different combinations of strong resonances; therefore tools like frequency map or tune footprint are needed to identify them and a minimization routine is then applied to suppress related driving terms [9,13,18]. In the following sections, we introduce the multiobjective optimization which combines driving terms and direct tracking to optimize the DA.

III. MULTIOBJECTIVE OPTIMIZATION

A very general optimization problem is to minimize a set of objective functions by tuning several variables under certain constraints:

$$\begin{aligned} \text{minimize} \quad & f_m(\mathbf{x}), \qquad m = 1, 2, \ldots, M; \\ \text{subject to} \quad & g_j(\mathbf{x}) \geq 0, \qquad j = 1, 2, \ldots, J; \\ \text{and} \quad & h_k(\mathbf{x}) = 0, \qquad k = 1, 2, \ldots, K; \\ \text{with} \quad & x_i^L \leq x_i \leq x_i^U \quad i = 1, 2, \ldots, N. \end{aligned} \qquad (5)$$

In the case of DA optimization, the free parameters $\mathbf{x}$ are magnet strengths, objective functions f_m are DA properties, and constraints can be DA shape, chromaticities, or nonlinear driving terms.

The solutions to Eq. (5) are not unique, but comprise a set, if the f_i conflict with each other. Increasing one f_i but decreasing the other will produce a new solution which is neither better nor worse than the original one. They are equally good but only with different trade-offs. We say that one solution dominates the other when it is better in at least one objective function f_i but not worse in any other objective function. A solution that meets the constraints g_j and h_k dominates any solution which does not. The solutions to Eq. (5) form a set called Pareto optimal set, in which any solution is not dominated by any possible solution in the whole space [19].

We use a population based evolution algorithm to find the Pareto optimum iteratively [19–21]. Each iteration is called one generation. At the first step, a fixed number of candidates are initialized as the first generation, and in our case they are uniformly distributed in $[x_1^L, x_1^U] \times [x_2^L, x_2^U] \times \cdots \times [x_N^L, x_N^U]$. Then one pair of them is randomly chosen as parents to generate two new children according to a certain probability density function (see Appendix B). This process is called crossover and is

repeated until the population is doubled. The third step is called mutation, where new children are perturbed slightly (see Appendix B). The objective functions $f_i(\mathbf{x})$, constraints $g_j(\mathbf{x})$, and $h_k(\mathbf{x})$ are evaluated for each of these new children. The whole population, including the parents, is then sorted or ranked according to their dominance relations. Candidates not dominated by anyone are in the first rank. The second rank candidates are only dominated by the rank-one candidates, and the third ranks are only dominated by the first and second ranks, etc. The final step is a population control process, where only half of the better candidates are kept. This is done by dropping candidates with larger rank. Within same rank, candidates in a high population density region have lower priority to be kept. Up to this point, the population is kept the same but the overall qualities in terms of objective functions and constraints are evolved not worse than the previous population before crossover. This is called one generation or one iteration. The population is evolved generation by generation until it converges or reaches the maximum number of iterations.

Figure 1 shows the last generation of objective functions, where we are maximizing both the on-momentum DA (horizontal axis) and the off-momentum DA (vertical axis). The trend of the optimization is that the whole population moves toward upper right, and converges to lower rank. By this generation, the Pareto optimal has only three candidates at the upper right corner. In a realistic optimization, the full convergence may take too long to achieve, while stopping at a set of reasonably good solutions which are sufficient for further applications is desirable. In Fig. 1, although the solutions are not fully converged, they have many candidates with DA >120 mm^2. It is a good starting point for detailed
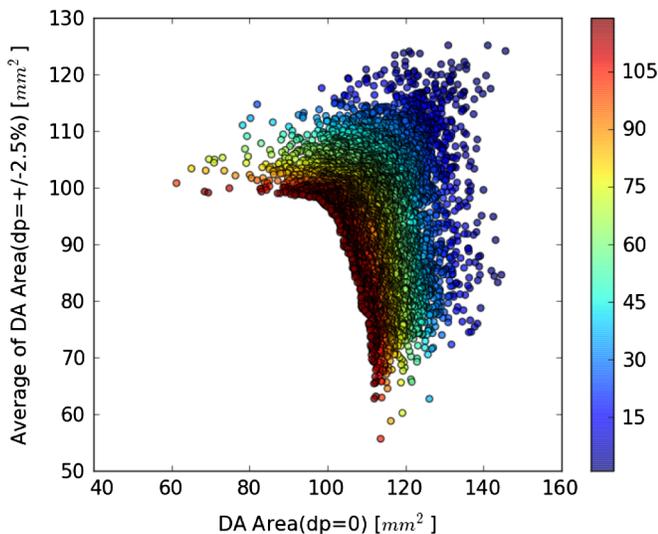
simulations that include various magnet errors and misalignment.

We are all familiar with the method of minimizing a weighted sum $f(\mathbf{x}) = \lambda_1 f_1^2[(x)] + \lambda_2 f_2^2[(x)] + \cdots + \lambda_M f_M^2[(x)]$, when dealing with multiobjective functions. In this case, solution $\mathbf{x}$ depends on the weight λ_i, $i = 1, 2, \ldots, M$, and the true solutions, i.e., the Pareto optimal set, need all combinations of weight λ_i. Unlike the weighted sum minimization, the multiobjective optimizer searches for an optimal set of solutions and leaves the choice of weight to the decision maker. The choice is made upon the global view of trade-off between optimal solutions instead of initial guesses of relative importance of each objective function. In this method, the sacrifice and gain are more clearly illustrated.

IV. DA OPTIMIZATION ON NSLS-II LATTICE

A. Robust dynamic aperture tracking and quantification

DA is usually defined as the maximum stable area in transverse plane at injection point. Particles with initial condition within this area will survive after a certain number of turns of tracking.

Obviously, the area of this 2D bounded region alone cannot represent the quality of DA. For different momentum deviation $\delta = (p - p_0)/p_0$, the stable area or DA may be different. Larger on-momentum DA may help the injection, while off-momentum DA helps Touschek lifetime.

One common way to calculate DA is particle tracking along several radial lines (see Fig. 2) with fourth-order or even higher symplectic integrator [22]. An ideal solution of DA would have an elliptical type of shape with no cut-in unstable area (as the two red points shown in Fig. 2). Because of betatron oscillation, particles are crossing the transverse plan at different (x, y) coordinate, and after long enough time, they will be lost at the cut-in unstable area. A good searching method should have good precision to detect these cut-in unstable areas while not requiring too much trial tracking near the DA boundary, since the tracking is expensive in computing time. Based on this, we

FIG. 1. The last generation of objective functions: DA of on-momentum (horizontal axis) and off-momentum (vertical axis) particles. Points are colored according to their rank.
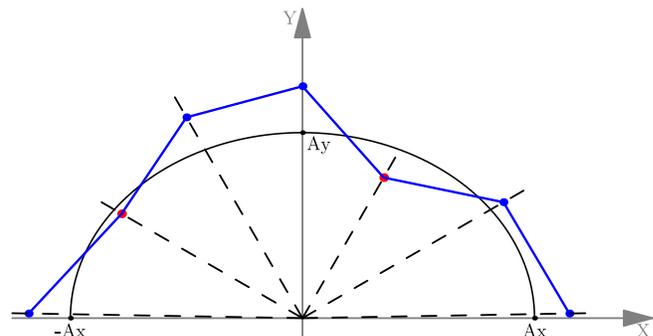
FIG. 2. Larger DA area may not necessarily provide a better solution, unless it covers an ellipse fully.

introduce a look-back feature in our tracking code. Along each searching line, if the current point failed to survive, we go inward one step, and continue the tracking. This robustness may not be necessary for a code running only once with human interpretation, but is critical for an automatic optimizer which is less tolerant.

In maximizing the DA area, we put a constraint for the DA boundary requiring that it should cover an ellipse defined by two half-axes lengths A_x and A_y. On each radial line, the boundary stable point (x, y) has $x^2/A_x^2 + y^2/A_y^2 \geq 1$, i.e., outside of the ellipse. This makes the maximization of DA area meaningful. The semiaxes A_x and A_y can be different for on-momentum and off-momentum DA. This flexibility does not increase any difficulty in our optimization method.

B. Optimization strategies

We shall discuss two different strategies to optimize DA. The first strategy uses the set of objective functions f_1 and f_2 characterizing the DA area of on-momentum and off-momentum particles:

$$f_1 = S(\delta = 0) \quad f_2 = S(\delta = -2.5\%) + S(\delta = 2.5\%), \quad (6)$$

where $S(\delta)$ is the DA area of momentum deviation δ. As shown in Fig. 3, the DA is tracked by seven radial lines for momentum deviation $\delta = -2.5\%$, 0, and 2.5%. A constraint ellipse with half axes $A_x = 18$ mm and $A_y = 2$ mm is set up to illustrate that the DA should cover this ellipse fully to prevent cut-in resonance structures.

The second strategy uses as objective functions f_1 the sum of the on- and off-momentum DA and f_2 the geometric sum of tune shifts with amplitude:

$$f_1 = \sum_{\delta} S(\delta, y = 1\mu m)$$

$$f_2 = \left(\frac{\partial \nu_x}{\partial J_x}\right)^2 + \left(\frac{\partial \nu_x}{\partial J_y}\right)^2 + \left(\frac{\partial \nu_y}{\partial J_y}\right)^2, \quad (7)$$

where $S(\delta, y = 1\mu m)$ is the DA area at certain momentum deviation δ and vertical offset $y = 1\mu m$. In Fig. 4, the top part shows the DA in the x-y plane for 13 δ equally spaced in $[-3\%, 3\%]$, and the bottom figure only shows the radial lines close to x axis but plotted in the δ-x plane with vertical offset $y \approx 1~\mu m$ and 1 mm. The DA in the δ-x plane decreases slowly with increased vertical offset. In this approach to DA optimization, an extra constraint shown as a green dashed line in Fig. 4 is applied. A preferred solution would have stable area lie outside the rectangle in the δ-x plane. However, in this particular solution, the constraint is not fully met near $\delta = -2.5\%$.

The DA is tracked by a fourth-order symplectic integrator. To increase accuracy, we use the exact solution for rectangular dipole [23]. This makes it unnecessary to slice a magnet into hundreds of pieces and has significant improvement on tracking speed. The damping wiggler is modeled by a kick-map with aperture size $[-30~\text{mm}, 30~\text{mm}] \times [-5~\text{mm}, 5~\text{mm}]$. As shown in Figs. 2 and 3, the searching is along several radial lines on the upper x-y plane. A small offset is introduced for the radial lines close to x axis, since zero y for the sextupole potential $V = k_2(x^3 - 3xy^2)/6$ will not have vertical kick (this is not necessary when the lattice has misalignment errors). Several points on each line are tracked simultaneously, and the most inward survival will be taken as the starting point for the next iteration. The look-back feature will pick a more inward point and track points between starting point and the first lost point along the same radial line. The searching stops when the grid is fine enough.

FIG. 3. DA of on/off-momentum particles on the x-y plane. The DA area S in Eq. (6) are based on seven radial lines tracking for each δ. The constraint ellipse has semiaxes $A_x = 18$ mm and $A_y = 2$ mm.
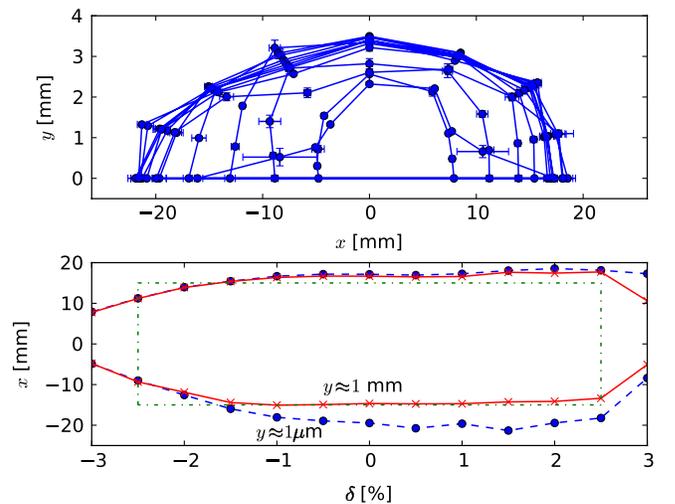
FIG. 4. DA in both the x-y plane (top) and the x-δ plane (bottom). The DA area f_1 in Eq. (7) is based on 13 different δ equally spaced in $[-3\%, 3\%]$ shown in the bottom plot.

In the optimization iteration, where speed is very important, we track only 64–128 turns. From our experience this is good enough for the overall size of DA, but not the detailed resonance structures. Frequency map analysis (FMA) comes as a filter to hand-pick the good candidates, where 2048 turns of tracking is done.

In the following sections, we will illustrate two different approaches by using the NSLS-II lattice (see Appendix A for the NSLS-II linear lattice) as an example. The first case is optimizing DA with the objective functions of Eq. (6) by varying six geometric sextupoles in straight sections. This linear lattice has tunes (33.43, 16.35) and chromaticities (1.5, 0.1) (see Sec. IV C). The second case uses the objective functions of Eq. (7) and has more independent variables (see Sec. IV E).

C. DA Optimization based solely on numerical tracking

In this example, we optimize the DA of a lattice with ($\nu_x = 33.43$, $\nu_y = 16.35$). The bare lattice has 15 superperiods with high-low-beta DBA cells (and long-short straight sections, see Appendix A, Fig. 13). The chromaticity is adjusted from natural value ($\xi_x = -103$, $\xi_y = -40$) to (1.5, 0.1) by three chromatic sextupoles between dipoles in each DBA cell. The other six geometric sextupoles labeled as SH1, SH3, SH4 in the high-beta section and SL1, SL2, SL3 in the low-beta section are used to optimize the DA. Three damping wigglers modeled by kickmap are included in the 2nd, 7th, and 12th long straight.

For the two objective functions of Eq. (6), i.e., DA of on-momentum and off-momentum particles, we use a population of 6000, and run for 300 generations. With 96 xeon 2.33 GHz CPUs in a Sun Grid Engine cluster , it takes less than a week to get a final population shown in Fig. 1. During the optimization, random errors such as multipole errors, 30 μm misalignment, and 5 μrad rotation errors of sextupoles are included. These errors are also introduced to all the quadrupoles at a smaller level to produce 20 μm–40 μm orbit distortion and x-y coupling.

In the optimization, when evaluating new children, the misalignment and rotation errors are changing from candidate to candidate, and each candidate does not use the same set of errors. Compared with optimization on a fixed set of random errors, this could reduce the speed of convergence; however, we benefit from the robustness of the solutions which are more independent of the choice of random seeds.

A postprocessing script is applied as a filter to pick the better solutions after more turns of tracking and averaging DA over more random seeds. As an example shown in Fig. 3, the script has applied a constraint ellipse with semiaxes $A_x = 18$ mm and $A_y = 2$ mm and about 50–70 are selected. As a final analysis, FMA and tune footprint are used. The properties of one of the better solutions are shown in Figs. 5–7 for FMA, tunes shift with amplitude and the tune footprint.
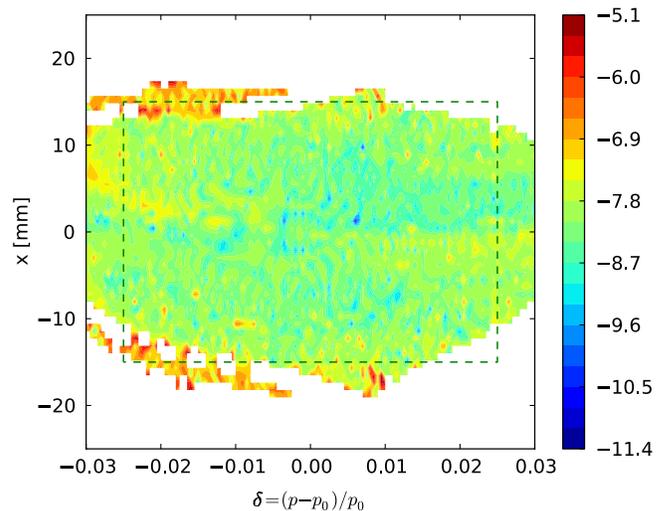
FIG. 5. Frequency map of one solution at working point (33.43, 16.35) with chromaticity (1.5, 0.1). 3 damping wiggler included. Multipole errors, misalignment, and rotation errors are also considered.
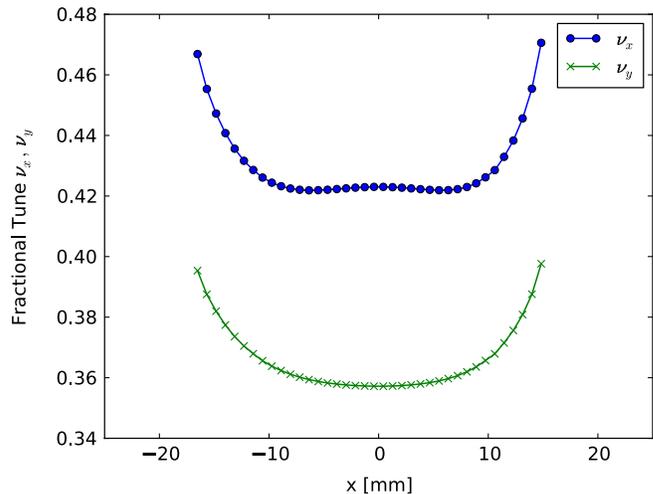
FIG. 6. Horizontal and vertical tune shifts with horizontal amplitude for working point (33.43, 16.35) and chromaticity (1.5, 0.1). All points are on-momentum particles with initial $y = 1$ μm.

The frequency map in Fig. 5 is for the x-δ plane with vertical offset $y = 1$ μm and uses the common definition of tunes drift $d\nu/dt = \log\sqrt{(d\nu_x/dt)^2 + (d\nu_y/dt)^2}$, where $d\nu_x/dt = (\nu_{x,1025\rightarrow2048} - \nu_{x,1\rightarrow1024})/1024$ and $d\nu_y/dt = (\nu_{y,1025\rightarrow2048} - \nu_{y,1\rightarrow1024})/1024$ are tunes drift per turn after 1024 turns. For high precision of frequency drift, e.g. 10^{-7}, Numerical Analysis of Fundamental Frequencies is used on 1024 turns of orbit [24].

The tune footprint is shown in Fig. 7. A subset with $x \in [-15 \text{ mm}, 15 \text{ mm}]$ and $\delta \in [-2.5\%, 2.5\%]$, i.e., the points inside the green dashed rectangle in Fig. 6, are plotted in bigger blue points.
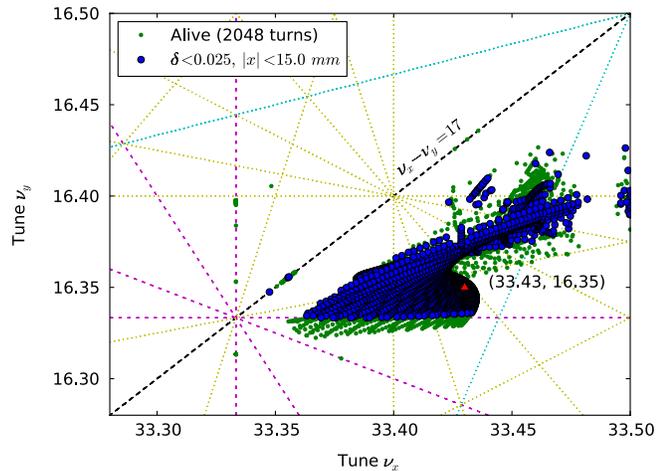
FIG. 7. Tune footprint for particles alive after 2048 turns of tracking. Its subset with initial $x \in [-15 \text{ mm}, 15 \text{ mm}]$, $y = 1 \ \mu\text{m}$, $\delta \in [-2.5\%, 2.5\%]$ is shown as bigger blue points. Resonance lines up to 5th order are plotted as dashed lines.

D. Correlation of DA and driving terms

Intuitively, we expect there is a correlation between DA and driving terms, i.e., the direct tracking and analytical approach. This is the reason that nonlinear driving terms can be used for DA optimization [13,18]. In this section, we investigate their correlation.

We performed a data mining on the results of Sec. IV C, where the optimization was done by direct particle tracking using constraints of Eq. (6) where no driving terms were calculated or used as a guide for the optimizer. For one of the early generations, we calculated the nonlinear driving terms up to the 2nd order, 33 terms in total and 24 terms after removing the degeneracy, for all candidates. To bring different terms to a same amplitude scale, a normalization is done within the whole generation for each driving term:

$$h^n_{abcde} = \frac{h_{abcde}}{\frac{1}{N_p} \sum_{\text{all candidates}} h_{abcde}}, \qquad (8)$$

where N_p is the number of individuals in all generations, h_{abcde} the nonlinear driving terms. Their sum $s(h) = \sqrt{\sum (h^n_{abcde})^2}$ is plotted against the DA area in Fig. 8.

As shown in Fig. 8, there is a strong correlation between DA area and the driving terms h_{abcde}. Similar correlations are observed also for $s(h)$ versus off-momentum DA area. We found that the optimal candidates always have small $s(h)$, while a small $s(h)$ does not always imply a good DA. This leads to a conclusion that, in the optimization process, small driving terms is a necessary but not sufficient condition for large DA. In Fig. 8, for many cases with DA area less than 40 mm^3, $s(h) \approx 5$ does not always give the largest DA area; while for solutions near 140 mm^2, all
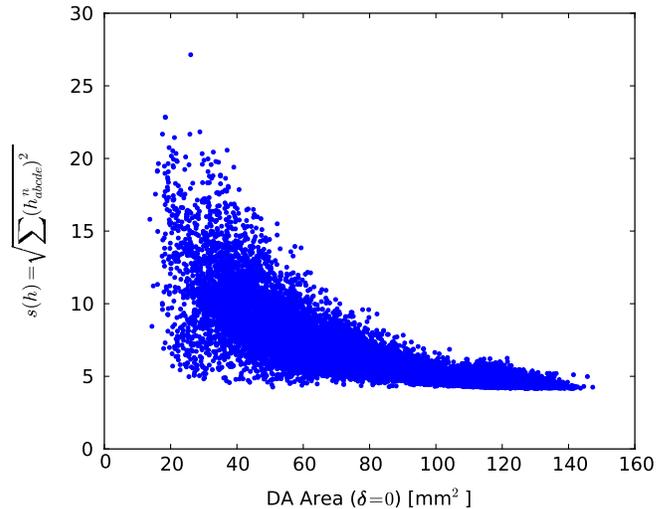
FIG. 8. Correlation between DA area and normalized driving terms, $s(h) = \sqrt{\sum (h^n_{abcde})^2}$, up to the second order. Small driving terms is a necessary but not sufficient condition for large DA area. Similar correlations are observed for off-momentum DA.

candidates have small $s(h) \approx 5$. This correlation can be used in the optimization as one of the objective functions or constraints. Depending on the lattice, contributions from each driving term varies. For example, the chromaticity related driving terms have no correlation to the DA, since this optimization is done with fixed chromaticity. In Sec. IV E, we set the constraints only on the most significant terms, i.e., tune shifts with amplitude, to be smaller than certain level. The exact value varies with lattice and we just estimate it from the first few generations. We have found that by using this correlation, the speed of convergence can be increased.

E. DA optimization based on numerical tracking and analytical tune shift with amplitude

Since many of the 3rd generation light source are operating at positive chromaticity to suppress the collective effects, we would like to optimize a nonlinear lattice of chromaticity $(+5, +5)$. The multiobjective evolution algorithm provides the flexibility to treat problems with many independent variables and constraints.

In this section we use the constraint functions of Eq. (7), i.e., including tune shift with amplitude, to optimize the DA in the x-δ plane with $y = 1 \ \mu\text{m}$. The particle tracking is done on a regular grid. All nine sextupoles are used as parameters, six for geometric optimization, two for fixing linear chromaticity at $(5, 5)$, and one for varying the nonlinear horizontal chromaticity. Two groups of quadrupoles, in long and short straight sections respectively, are also used as parameters to adjust tunes. Each group has three quadrupoles, two for boundary condition matching and one for varying tunes. In this case, 15 parameters are tuned and among them nine are free parameters.

With two free quadrupoles and four matching quadrupoles, the tunes are varied within a certain range in the optimization (see Fig. 9 for the footprint of tune variation). This method varies the tunes by adjusting the beta functions only in the straight sections (not including the three damping wiggler sections). Therefore, the dispersion and emittance are not varied in our optimization.

In each DBA cell, i.e., half of the mirror symmetric superperiod, there are three quadrupoles in high-beta straight section and low-beta straight section respectively (see Fig. 13). By varying QH1, QH2, and QH3, we make the beta functions β_x and β_y change only locally between QH3 and its mirror symmetric one on the left, i.e., within long straight sections. Two variables, QH1 and QH2, are used for the boundary conditions of $\alpha_x = \alpha_y = 0$ at the center of long straight section. The third quadrupole is a free variable that can vary the tunes. The short straight section can also do similar matching with QL1, QL2, and QL3 to have a second free variable for tunes. The matching can be done for each one of the 30 straight sections independently but the three long straight sections with damping wigglers are excluded in our optimization intentionally. As one matching example, the overall changes of beta functions are shown in Fig. 10 for 1/3 of the ring. It shows the relative lattice functions change before and after the matching. After the matching, the tunes vary by $(\delta\nu_x, \delta\nu_y) = (0.016, -0.042)$ from (33.15, 16.27). When matching long and short straight sections independently, the full range of tune variation is shown in Fig. 9. This plot compares the coverage of tune variation from the initial population, final population, and optimal candidates. Because of the random errors involved, the ranking of candidates are random-seed dependent. These dependencies could be reduced by a postprocessing analysis.

FIG. 10. Beta functions variation resulting from tune change from (33.15, 16.27) to (33.16, 16.23). Changes in short and long straight are independent, and the damping wiggler section is not changed. In each DBA cell, linear lattice between dipoles is not changed.

In the postprocessing, larger errors and longer tracking runs are applied while the rule for selecting best ones are same as the objective functions in the optimizer. After the postprocessing, we found the example shown in Fig. 11 where the working point is (33.24, 16.36). This new working point guarantees that at chromaticity (5, 5) the particles with $\delta = -0.03$ stay far enough from the integer tune $\nu_x \approx 33$ and survive in the tracking.

The FMA of this same candidate is shown in Fig. 11 and the surviving turns for particle tracking is shown in Fig. 12. In the survival plot, classical radiation damping and rf acceleration are included. The rectangle in the green dashed line in Fig. 11 is the constraint area that a candidate

FIG. 9. The tune variation range when matching long and short straight sections independently. The blue dots are initial tune coverage of the first generation. The green circles are from the final population, and red triangles are optimal candidates.
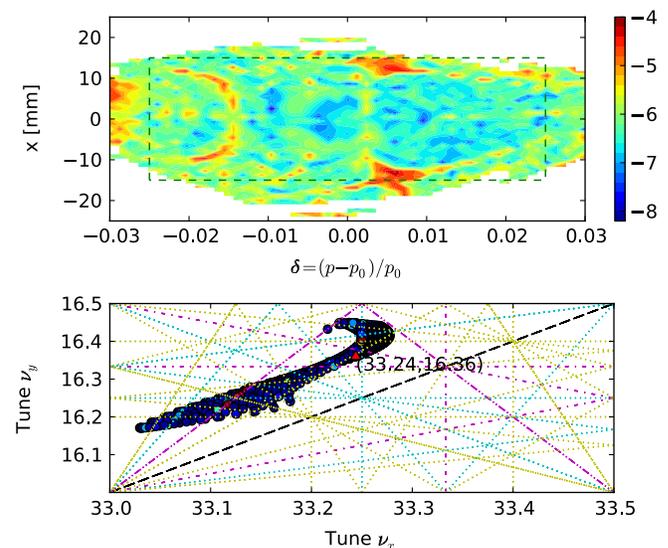
FIG. 11. Frequency map and tune footprint of a candidate lattice with tunes (33.24, 16.36) and fitted chromaticity (5, 5).
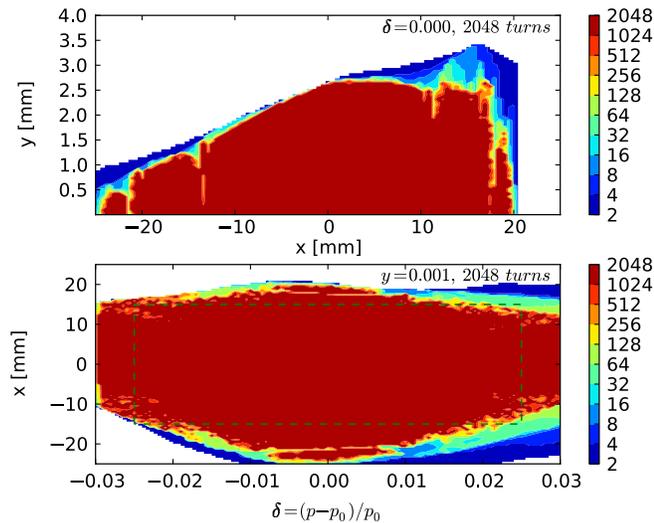
FIG. 12. The surviving turns in 2048 turns of particle tracking. In the x-y plane $\delta = 0$ and in the δ-x plane $y = 1$ um. Realistic misalignment errors, apertures, classical radiation, and rf cavity are included.

should cover (see Sec. IV B), however it is not fully met by this particular candidate.

By varying quadrupoles as well as sextupoles, we see that the multiobjective optimization provides an efficient method for exploring potential working points.

V. CONCLUSION

Taking the NSLS-II lattice as an example, we have shown that multiobjective optimization is an effective way of optimizing nonlinear properties, mainly DA area for both on-momentum and off-momentum particles. Two different examples are discussed for different tunes, chromaticities and adjustable magnets. The first case only uses geometric sextupoles at fixed tunes and chromaticities and objective functions of Eq. (6) based only on numerical tracking data. The data mining on this case shows strong correlations of DA area with normalized nonlinear driving terms. We observed that small low-order nonlinear driving terms are a necessary but not sufficient condition for a large DA area. The second case combines DA direct searching and nonlinear driving terms minimization. It varies both nonlinear chromaticities and tunes in the iterations of optimization. By checking with FMA, the results are proved to be satisfactory and the approaches we have been using are found to be effective and robust. Our conclusion is that multiobjective optimization provides a complementary approach which can be used either together with driving terms optimization or independently using only tracking data.

The approaches we have investigated have been illustrated using the NSLS-II lattice, but the applications are not limited to NSLS-II. Similar optimization techniques can be applied to other nonlinear properties in accelerator physics.

APPENDIX A: NSLS-II LATTICE PROPERTIES

National Synchrotron Light Source II (NSLS-II) is a state-of-the-art third-generation light source under construction at Brookhaven National Laboratory [1]. It has challenging requirements on brightness and stability. The optimizations on nonlinear lattice properties are therefore critical.

NSLS-II lattice has 30 double bend achromatic (DBA) cells. Two DBA cells with mirror symmetry inside have low- and high-beta functions at short and long straight sections (see Fig. 13). The whole bare lattice has 15-fold symmetry. Each DBA cell has nine sextupoles, three sitting in between the dipoles have nonzero dispersion, and are used for chromaticity adjustment. The other six, three on each side, are at the straight section where linear dispersion is zero. Three damping wigglers are included in the 2nd, 7th, and 12th long straight sections. The threefold linear lattice is matched to be approximately 15-fold symmetric.

APPENDIX B: CROSSOVER AND MUTATION

Crossover is an operation applied to two parent candidates to generate two children. These two parents are randomly chosen from the survivals of the last generation.
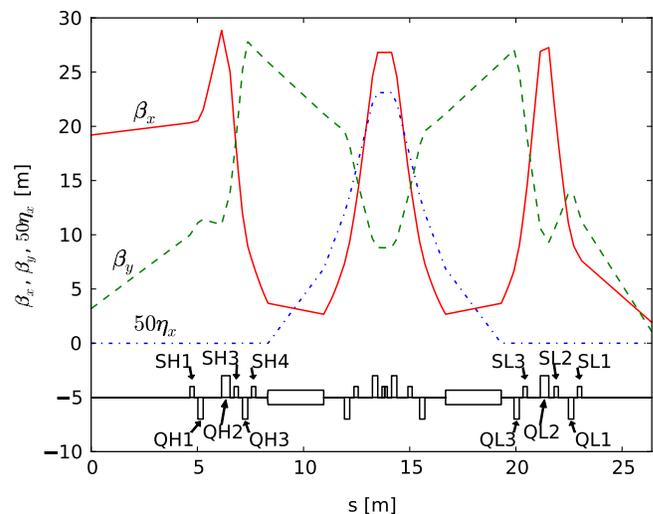
FIG. 13. Lattice functions and magnets layout of one DBA cell. It is half of the mirror symmetric superperiod. Each DBA cell has nine sextupoles in total, three for chromaticity adjustment and six for nonlinear lattice optimization.

In our case, we use simulated binary crossover method where the children are generated following a polynomial probability density function (PDF) and every candidate is chosen as a parent once and only once [19]. It is intuitive to have children generated with higher probability around the parents and lower when further away from the parents, i.e., the PDF peaks at parents. The polynomial form will reduce the complexity in simulation when we use inversion method to generate the random children based on a uniform distribution, since the polynomial form is invertible.

This PDF is controlled by a single parameter η_c which is the order of the polynomial PDF [see Fig. 14(a)]:

$$\beta_q = |(2x - (p1 + p2))/(x_U - x_L)| \qquad \text{(B1)}$$

$$f(\beta_q) = \begin{cases} \frac{1}{\alpha(x_U - x_L)}(\eta_c + 1)\beta_q^{\eta_c} & \text{if } \beta_q < 1.0 \\ \frac{1}{\alpha(x_U - x_L)}(\eta_c + 1)\beta_q^{-\eta_c - 2} & \text{otherwise,} \end{cases} \qquad \text{(B2)}$$

where x is the independent variable, x_U and x_L are the range of x, p_1 and p_1 two parents, and $\alpha = 2 - (1 + 2(p_1 - x_L)/(p_2 - p_1))^{-\eta_c - 1}$ the normalization factor. For simplicity we assume $x_L \le p_1 \le p_2 \le x_U$. The procedures of generating can be found in [25] [in Eq (A.4), we have correct $\delta = (x - x^l)/(x^u - x^l)$ for $u \le 0.5$ and $\delta = (x^u - x)/(x^u - x^l)$ otherwise].

The whole population is doubled after the crossover operation, but later the sorting process will find out the best half and keep them as the next generation.

After the crossover, mutation operations are followed to bring perturbations into the new children. Following similar method in crossover, we use polynomial PDF peaked at the new child. The perturbed value has lower probability

when they are further away from the current value [see Fig. 14(b), one of the children from crossover is now the parent in the mutation]. The shape of PDF can also be controlled by a parameter η_m.

Both the crossover and mutation can be done within some probability. When outside of this probability, in crossover we simply copy the parents to children and in mutation we do nothing.

FIG. 14.   Probability density function for crossover and mutation operations. In the crossover, two example parents are $x = 3.0(P1)$ and $x = 5.0(P2)$ and in the mutation the parent is $x = 3.5(P)$.

[1] "NSLS-II preliminary design report," http://www.bnl.gov/nsls2/project/PDR/.
[2] S. C. Leemann, A. Andersson, M. Eriksson, L.-J. Lindgren, E. Wallén, J. Bengtsson, and A. Streun, Phys. Rev. ST Accel. Beams **12**, 120701 (2009).
[3] "PETRA III technical design report," http://petra3.desy.de/general/tdr/.
[4] Y. Li, K. Balewski, and W. Decking, in *Proceedings of the 10th European Particle Accelerator Conference, Edinburgh, Scotland, 2006* (EPS-AG, Edinburgh, Scotland, 2006), http://accelconf.web.cern.ch/AccelConf/e06/PAPERS/THPLS021.PDF.
[5] "MAX IV conceptual design report," http://www.maxlab.lu.se/maxlab/max4/index.html.
[6] S. Y. Chen, H. J. Xu, and Z. T. Zhao, in *Proceedings of the 18th Particle Accelerator Conference, New York, 1999* (IEEE, New York, 1999), Vol. 1, pp. 209–211, http://accelconf.web.cern.ch/AccelConf/p99/PAPERS/TUDL1.PDF.
[7] J. Gao, Nucl. Instrum. Methods Phys. Res., Sect. A **451**, 545 (2000).
[8] J. Gao, Nucl. Instrum. Methods Phys. Res., Sect. A **516**, 243 (2004).
[9] M. Borland, Advanced Photon Source Technical Report No. LS-287, 2000.
[10] J. Bengtsson (unpublished).
[11] J. Shi, Nucl. Instrum. Methods Phys. Res., Sect. A **444**, 534 (2000).
[12] W. Wan and J. R. Cary, Phys. Rev. Lett. **81**, 3655 (1998).
[13] J. Bengtsson, SLS Technical Report No. 9/97, 1997.
[14] J. Laskar and D. Robin, Part. Accel. **54**, 183 (1996).
[15] L. Yang, in *48th ICFA Beam Dynamics Workshop on Future Light Sources* (SLAC National Accelerator Laboratory, Menlo Park, California, 2010), http://www-conf.slac.stanford.edu/icfa2010/.
[16] M. Borland, in 48th ICFA Beam Dynamics Workshop on Future Light Sources (Ref. [15]).
[17] E. Forest, *Beam Dynamics—A New Attitude and Framework* (Harwood Academic Publishers, Amsterdam, The Netherlands, 1998).
[18] W. Guo, S. Krinsky, and L. Yang, in Proceedings of the First International Particle Accelerator Conference, Kyoto, Japan (2010), http://accelconf.web.cern.ch/AccelConf/IPAC10/papers/wepea076.pdf.
[19] K. Deb and D. Kalyanmoy, *Multi-Objective Optimization Using Evolutionary Algorithms* (Wiley, New York, 2001), 1st ed.

[20] I. V. Bazarov and C. K. Sinclair, Phys. Rev. ST Accel. Beams **8**, 034202 (2005).

[21] L. Yang, D. Robin, F. Sannibale, C. Steier, and W. Wan, Nucl. Instrum. Methods Phys. Res., Sect. A **609**, 50 (2009).

[22] H. Yoshida, Phys. Lett. A **150**, 262 (1990).

[23] E. Forest, J. Phys. A **39**, 5321 (2006).

[24] J. Laskar, C. Froeschl, and A. Celletti, Physica D (Amsterdam) **56**, 253 (1992).

[25] K. Deb, Comput. Methods Appl. Mech. Eng. **186**, 311 (2000).