12th Int. Particle Acc. Conf. IPAC2021, Campinas, SP, Brazil JACoW Publishing
ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-THPAB197

ENHANCING EFFICIENCY OF MULTI-OBJECTIVE NEURAL-NETWORK-ASSISTED NONLINEAR DYNAMICS LATTICE OPTIMIZATION VIA 1-D APERTURE OBJECTIVES & OBJECTIVE FOCUSING*

Y. Hidaka†, D. A. Hidas, F. Plassard, T. V. Shaftan, G. M. Wang
Brookhaven National Laboratory, Upton, NY, USA

Abstract

Mutli-objective optimizers such as multi-objective genetic algorithm (MOGA) have been quite popular in discovering desirable lattice solutions for accelerators. However, even these successful algorithms can become ineffective as the dimension and range of the search space increase due to exponential growth in the amount of exploration required to find global optima. This difficulty is even more exacerbated by the resource-intensive and time-consuming tendency for the evaluations of nonlinear beam dynamics. Lately the use of surrogate models based on neural network has been drawing attention to alleviate this problem. Following this trend, to further enhance the efficiency of nonlinear lattice optimization for storage rings, we propose to replace typically used objectives with those that are less time-consuming and to focus on a single objective constructed from multiple objectives, which can maximize utilization of the trained models through local optimization and objective gradient extraction. We demonstrate these enhancements using a NSLS-II upgrade lattice candidate as an example.

INTRODUCTION

Application of machine learning (ML) is gaining popularity among many scientific fields including accelerator physics. One category of such applications is to train surrogate models of complex nonlinear functions, traditionally performed with time-consuming but accurate simulations, with ML techniques to significantly speed up the calculations of the functions at the cost of some loss in accuracy. For example, the strengths of sextupoles and/or octupoles in a storage ring lattice need to be often optimized to increase its on-momentum dynamic aperture (DA) for adequate injection efficiency and its momentum aperture (MA) for longer Touschek beam lifetime. However, evaluations of these objectives via particle tracking are expensive in terms of computation cost and time.

To mitigate this bottleneck, a new optimization algorithm called NBMOGA [1], based on the well-known multi-objective genetic algorithm (MOGA) [2], has been recently proposed. This optimizer initially proceeds exactly the same way as MOGA, but past a certain number of generations (warm-up period), it starts to train a neural-network-based surrogate model on the multi-objective values evaluated with a physics-based simulation code to approximately predict the objectives. From that generation on, it uses the objective values estimated from the neural network (NN) model, instead of the physics model, for the purpose of individual selection. This leads to substantial speed-up in the overall optimization process, as the objective evaluation duration is reduced from hours to milliseconds.

In this paper, we propose two approaches to further enhance the speed of the nonlinear lattice optimization problem for storage rings utilizing NBMOGA.

OBJECTIVE REPLACEMENTS

The first approach is to replace the time-consuming objectives themselves to less time-consuming ones. NBMOGA enables much more exploration in the search space via fast NN model evaluations, but still requires many physics-based model evaluations to obtain training data and improve the NN model. Typically for the optimization problem we consider here, we evaluate full 2-D on-momentum DAs and local MAs (LMA), as they are the most direct indicators of the most critical machine performance metrics, namely, injection efficiency and Touschek beam lifetime. Particularly, LMA can take hours even for one lattice to calculate a reliable lifetime estimate.

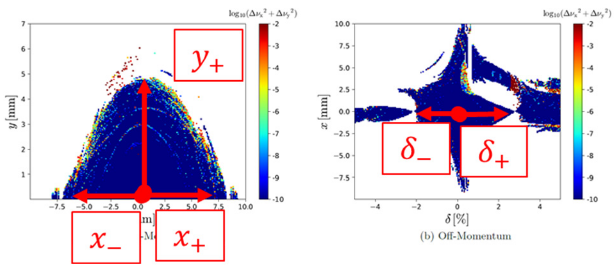
We propose to use the following five 1-D apertures (as shown in Fig. 1) to significantly reduce the time taken by physics-based model evaluations: Horizontal 1-D DAs (x_+, x_-), vertical 1-D DA (y_+), and 1-D MAs (RF & radiation off) at the injection point (δ_+, δ_-). Each aperture is determined by tracking a particle for ~ 100 turns for a full ring, gradually increasing the initial coordinate from zero, until a particle does not survive, or an integer/half-integer tune is crossed. These are chosen because every lattice that satisfies the 2-D DA and LMA requirements should necessarily have large values for these objectives. By no means, having large values for these new objectives guarantees acceptable 2-D DA and LMA objectives, as they are not sufficient conditions. But the advantage of reducing the evaluation time from 1-2 hours to 3-10 minutes for each lattice in our test case can justify the disadvantage of potentially converging to a lattice that does not quite meet the minimum criteria. At least with these objectives, we can quickly identify unpromising base linear lattices and shift optimization efforts to other unexamined lattices. In other words, this approach serves as a rapid screening tool for promising linear lattices.

This was tested on one of the upgrade candidate lattices for NSLS-II that has natural horizontal emittance of 40 pm and is a triple-complex-bend achromat (TCBA) [3, 4]. This optimization involved 12 chromatic and 8 harmonic octupoles as knobs, while the 1-D DA and MA scan ranges were limited to ±10 mm and ±4%, respectively. These limits were applied in the interest of further reducing

* Work supported by U.S. DOE under Contract No. DE-SC0012704.
† yhidaka@bnl.gov

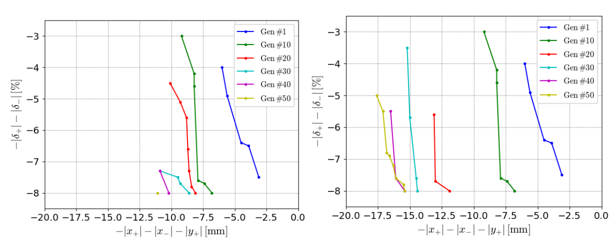
ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-THPAB197

evaluation time, as we expected these ranges give enough margin over our tentative soft requirements of ±5 mm horizontal DA and 1-2 hr lifetime. The results of optimization progress for 50 generations for MOGA and NBMOGA are shown in Fig. 2. The population size was 1000 for both. The warm-up period of NBMOGA was 10 generations. Hence the Pareto fronts are the same until that generation. A sudden acceleration in NBMOGA is clear after the warm-up period. MOGA quickly converged, but to a far inferior lattice. Python DEAP library [5] was used for genetic algorithm implementation in both MOGA and NBMOGA.



(a) (b)

Figure 1: Proposed five 1-D aperture definitions for faster physics-based model evaluations in an example of (a) on-momentum frequency map and (b) off-momentum one.



(a) (b)

Figure 2: Pareto fronts of (a) MOGA and (b) NBMOGA for a 40-pm NSLS-II upgrade candidate lattice.

The NBMOGA run took 0.5-2 hr per generation (i.e., 1000 lattice evaluations) with 200 cores (100 physical cores with hyperthreading; Intel Xeon CPU E5-2699 v4 2.20 GHz). If more computing resources are available, the run time should decrease almost linearly down to about 6-24 minutes per generation with 1000 cores, for example, as these evaluations are embarrassingly parallel calculations. Note that the optimization takes less time initially and gradually takes longer as it finds better lattice solutions for which DA and MA are larger and hence more tracking calculations are required.

As mentioned earlier, if the optimization results at this point are not promising at all, we can decide to move onto another linear lattice for which we have not yet optimized. However, sometimes it is worth adjusting the objectives and re-optimize instead. For instance, as shown in Fig. 3, a lattice with a minimum LMA of 1.8% (Fig. 3a) and lifetime of 0.86 hr (100% coupling) was found after an optimization. Upon inspecting its off-momentum frequency map (Fig. 3b), it appeared that the poor horizontal DA on the negative momentum side is ruining the good DA on the positive side due to synchrotron motions. If we could pull the off-momentum aperture envelope to the negative momentum side at the points denoted by the red circles ($x = \pm 2.5$ mm), we may be able to increase LMA. With this conjecture, we added 2 more objectives: $\delta_-(x = +2.5$ mm) and $\delta_-(x =$ -2.5 mm). After re-optimization, we found a new lattice with a minimum LMA of 2.4% (Fig. 4a), resulting in increased lifetime of 2.6 hr. Figure 4b shows that this optimization filled in the previously void regions in the frequency map as intended. This demonstrates usefulness of iterative customization of multi-objectives. Note that this iterative process is enabled because of the much-reduced optimization run time with the nominal 5 objectives.

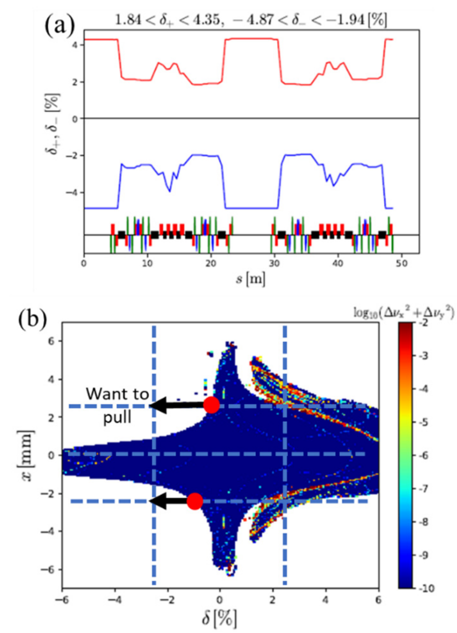


Figure 3: (a) LMA and (b) off-momentum frequency map for a lattice after optimizing with the 5 multi-objectives.

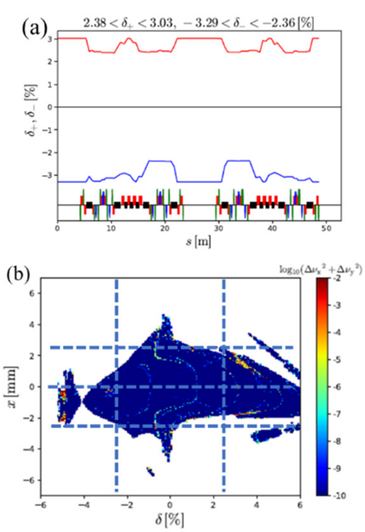


Figure 4: (a) LMA and (b) off-momentum frequency map for a lattice after optimizing with 2 additional MA objectives δ_- at $x = \pm 2.5$ mm.

ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-THPAB197

OBJECTIVE FOCUSING

Another approach to improve efficiency is to utilize single objectives with local optimizers. At the end of a multi-objective optimization run, we typically obtain a Pareto front with many lattices that have very good DA, but very poor lifetime, or vice versa. These do not satisfy minimum requirements and will never be selected as a viable solution. To avoid potentially wasteful computations for these extreme unpromising individuals in the population, we can sum the multi-objectives into a single objective in a way to impose minimum required objectives, then use it to guide the optimization direction. With multiple objectives focused into a single objective, we can also utilize many available local optimizers that are usually much faster and efficient than global multi-objective optimizers. Furthermore, since a NN surrogate model built with popular machine learning libraries such as tensorflow [6] and PyTorch [7] is equipped with automatic differentiation, we can easily extract gradients of the single objective when evaluating the function values from the NN model. This leads to significant speed boost of a gradient-based local optimizer.

We constructed the following single objective f to be minimized:

$$f = -\{|x_+| + |x_-| + |y_+|\}[\mathrm{mm}]$$
$$+\{\mathrm{selt}(\delta_+, +3\%, 0.1\%) + \mathrm{segt}(\delta_-, -3\%, 0.1\%)\},$$

where "selt" and "segt" are the soft-edge less/greater-than penalty functions defined in ELEGANT [8]:

$$\mathrm{selt}(x, x^*, \mathrm{tol}) = \begin{cases} 0 \text{ if } x \geq x^* \\ \left(\dfrac{x - x^*}{\mathrm{tol}}\right)^2 & \text{otherwise,} \end{cases}$$

and "segt" is similarly defined except that the inequality direction is opposite. This objective encourages optimization to reach $|\delta| > 3\%$ to have acceptable Touschek lifetime by adding a relatively large positive value in the second term. Once this minimum requirement on MAs is cleared, the second term becomes zero and stays neutral, while the optimizer continues to keep decreasing the first term (i.e., increasing the DA objectives).

In NBMOGA we implemented, in each generation, multi-objectives are predicted once from the NN model and use the non-dominated sorting genetic algorithm II (NSGA-II) [9] on these predicted values to select individuals to keep for next generation. The main difference in the objective-focused NBMOGA is that we run a local optimizer for each individual in an expanded candidate pool as in NBMOGA to minimize the single objective based on the multi-objective values predicted from the surrogate model repeatedly. From the optimized candidate individuals, a new population is selected according to the best predicted scores of the single objective and later evaluated with physics-based simulations. We used the L-BFGS-B algorithm [10] as our local optimizer, as it can impose bounds on input variables as well as utilize gradients.

The comparison between NBMOGA and objective-focused NBMOGA for the same test lattice used in the previous section is shown in Fig. 5. Once the warm-up period was over, the objective-focused case made an enormous jump. But the progress immediately slowed down and was eventually taken over by NBMOGA. Upon inspection of individual distributions, it appears that the excessive focusing by selecting new generations entirely based on the single objective resulted in a rapid loss of diversity. This likely made the population unable to create sufficiently diverse candidate individuals via genetic operations and thus getting stuck at a local minimum. A further algorithmic search is currently underway to balance between objective focusing and genetic diversity to boost the convergence speed while maintaining sufficient exploration capacity.

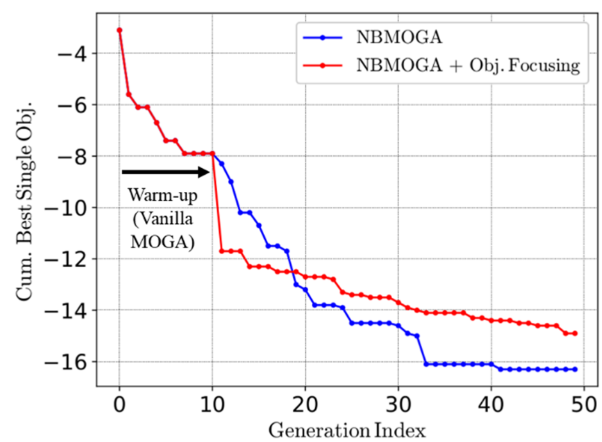


Figure 5: Cumulative best single objective history for NBMOGA with and without objective focusing.

Lastly, use of gradients for the local optimizer reduced the number of model evaluations required to converge by an order of magnitude, compared to the case when gradients were not utilized. Although the local optimization steps are not the dominant part of the overall optimization process in terms of run time, it is still worth avoiding unnecessary calculations wherever possible.

CONCLUSION

A speed boost by an order of magnitude for NBMOGA nonlinear beam dynamics optimization for storage rings was demonstrated by replacing typical time-consuming 2-D DA and lifetime evaluations with 1-D DA and MA evaluations. This allows a much quicker assessment on whether the base linear lattice is promising, although it comes at the potential cost of converging to sub-optimal results. When the results are not satisfactory, the multi-objectives can be also iteratively tweaked to guide optimization for improvement. Objective focusing can further accelerate NBMOGA, but needs more work to balance focusing against diversity to find better solutions.

ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-THPAB197

REFERENCES

[1] J. Wan, P. Chu, and Y. Jian, "Neural network-based multi-objective optimization algorithm for nonlinear beam dynamics", *Phys. Rev. Accel. Beams*, vol. 23, p. 081601, 2020. doi:10.1103/physrevaccelbeams.23.081601

[2] T. Murata and H. Ishibuchi, "MOGA: Multiobjective genetic algorithms", in *Proc. of IEEE International Conference on Evolutionary Computation*, Perth, Australia, Nov. 1995, pp. 289–294. doi:10.1109/ICEC.1995.489161

[3] G. Wang *et al.*, "Low Emittance Lattice Design in Synchrotron Light Sources by Using Complex Bends", presented at the 8th Low Emittance Rings Workshop, Frascati, Italy, Oct. 2020, unpublished.

[4] G. Wang *et al.*, "Complex bend II: A new optics solution", *Phys. Rev. Accel. Beams*, vol. 22, p. 110703, 2019. doi:10.1103/physrevaccelbeams.22.110703

[5] F. A. Fortin *et al.*, "DEAP: Evolutionary Algorithms Made Easy", *J. Mach. Learn. Res.*, vol. 13, pp. 2171-2175, 2012.

[6] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems", unpublished. arXiv:1603.04467

[7] A. Paszke *et al.*, "PyTorch: An Imperative Style, High-Performance Deep Learning Library", presented at the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada, Dec. 2019, unpublished. arXiv:1912.01703

[8] M. Borland, "elegant: A Flexible SDDS-Compliant Code for Accelerator Simulation", in *Proc. 6th International Computational Accelerator Physics Conf. (ICAP 2000)*, Darmstadt, Germany, Sep. 2000, paper LS-287. doi:10.2172/761286

[9] K. Deb, A. Pratap, S. Agrawal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II", *IEEE Trans. Evol. Comput.*, vol. 6, pp. 182-197, 2002. doi:10.1109/4235.996017

[10] R. H. Byrd, P. Lu, and J. Nocedal, "A Limited Memory Algorithm for Bound Constrained Optimization", *SIAM J. Sci. Comput.*, vol. 16, pp. 1190-1208, 1995. doi:10.1137/0916069