ISBN: 978-3-95450-236-3 ISSN: 2673-7027 doi:10.18429/JACoW-eeFACT2022-WEZAS0106

METHODS AND EXPERIENCES OF AUTOMATED TUNING OF ACCELERATORS*

Xiaobiao Huang†, SLAC National Accelerator Laboratory, Menlo Park, CA, USA

Abstract

Automated tuning, or beam-based optimization, is a general approach to improve accelerator performances. The approach is different from the other common approach of beam-based correction. The differences between these two approaches and the advantages of the optimization approach are discussed. Two online optimization methods, the robust conjugate direction search (RCDS) and the multi-generation Gaussian process optimizer (MG-GPO), are described. Experiences of apply the methods to storage ring nonlinear dynamics optimization at SPEAR3 and APS storage rings, as well as application to other machines, are presented.

INTRODUCTION

An accelerator typically has many error sources that cause its behavior to differ from the ideal design. The performance of the machine can be substantially degraded due to the errors. The machine also has many control parameters (i.e., knobs) that can be used to change its behavior, which could compensate the effects of the errors and restore the machine performance. Accelerator physicists use beam-based measurements to determine the desired knob adjustments. The methods employed to find the accelerator setting based on beam-based measurements could be classified into two categories: beam-based correction and beam-based optimization [1].

In this paper, we will first discuss the characteristics of these two approaches. This is followed by discussions on the methods and application of beam-based optimization. The methods to be focused on are the robust conjugate direction search (RCDS) method [2] and the multi-generation Gaussian process optimizer (MG-GPO) [3]. Considerations on application of the methods to real-life accelerator tuning problems are discussed. Some important applications, such as minimization of the vertical emittance in storage rings, tuning of linac front end, and optimization of nonlinear beam dynamics of storage rings, are described.

BEAM-BASED CORRECTION AND OPTIMIZATION

The performance of an accelerator can be characterized by various metrics, such as beam intensity, beam size, beam lifetime, beam loss, transmission efficiency, injection efficiency, and beam stability. These metrics could be constantly monitored, or in some cases, are measured on demand. Depending on the purpose of the machine, each accelerator may have a different set of performance metrics of importance. In many cases, a set of knobs can target one performance metric without affecting the others. However, in some cases, the same set of knobs that are used to tune one metric can simultaneously impact the other metrics.

The diagnostic system of the accelerator measure and monitor many signals that represent the state of the machine or the beam. For example, the orbit of the beam throughout the accelerator is typically monitored with beam position monitors (BPMs). The transverse beam profile and in turn the transverse beam size can be measured at some locations. In circular accelerators, the betatron tunes can be constantly monitored. Some machine state variables can be derived from the monitor signals. In some cases, the beam or the machine are intentionally perturbed in order to perform an observation of the machine state. For example, the betatron phase advances can be measured from turn-by-turn BPM data when the beam is kicked. The closed orbit response, measured by making a small change to an orbit corrector, is another example.

The machine state as characterized by the diagnostic system could be directly correlated with the performance metrics, such that restoring the machine state automatically also restores the performance. In other cases, the correlation is not as strong; yet, it is still generally preferred to operate under certain machine states. In those cases, a “golden” machine state can be defined as the target configuration. For example, a golden beam orbit is usually defined for a storage ring. Desired values of betatron tunes and chromaticities are also specified. In a linac or transport line, the desired orbit and beam distribution is often specified at some strategically important locations, for example, at the end of the transport line for injection to another accelerator or at the entrance of the undulators in a free electron laser.

Often times, a known set of knobs can be used to change a certain aspect of the machine state. If there are enough effective knobs, it may be possible to move the machine into any reasonable state with those knobs. Because usually each knob has a definitive and predictable effect to the machine state, given the current machine state, the current knob setting, and the target machine state, one could work out the required adjustment to the knobs in a deterministic fashion. As not everything is perfectly known, it may take several iterations to reach the target state. The process of of bringing the machine state as measured by the beam diagnostic system to a target state with control knobs via a deterministic procedure is called beam-based correction.

Beam-based correction requires beam diagnostics that can sufficiently characterize the machine states, a known target machine state, knobs that can effectively change the machine state, and a deterministic procedure to determine the required knob changes toward the target. Reaching the

* Work supported by DOE Contract No. DE-AC02-76SF00515
† xiahuang@slac.stanford.edu

ISBN: 978-3-95450-236-3 ISSN: 2673-7027 doi:10.18429/JACoW-eeFACT2022-WEZAS0106

target state does not necessarily leads to the highest machine performance, as the correlation between the machine state and the performance metrics is not always strong and the optimal target state could change with other controlled or uncontrolled machine conditions (e.g., drift w/ ambient temperature).

Beam-based optimization is another category of beam-based methods to tune knobs for improving accelerator performances. This approach is also referred to as automated tuning. Manual tuning is common in accelerator control. As accelerators adopted computerized control early on, there have been various attempts to automate the tuning process [2, 4].

Beam-based optimization aims at improving the performance metrics directly by changing the tuning knobs, using the measured performance metrics as the guide. It is the same as mathematical optimization - the performance metrics are the objective functions and the tuning knobs are the optimization variables. The objective functions are evaluated by performing a measurement on the machine, after the tuning knobs are dialed in. The accelerator can be considered a black box; the main requirements for the machine are that the knobs are effective in changing the performance and that it can reproduce the performance for the same knob setting. No measurement of the machine state is necessary, unless certain features of the machine state are part of the performance metrics.

The optimization approach has some advantages over the correction approach. It does not have high requirements for diagnostics, as measurements for the performance metrics are usually available. It does not need a target state. This would be important in the commissioning phase of an accelerator as the target state, needed by the correction approach, may be still undetermined. Nor does it require enough a prior knowledge about the system to relate the target to the knobs. Therefore, it is relatively easy to set up and perform beam-based optimization (see Fig. 1).

The key to beam-based optimization is robust, efficient optimization algorithms. The requirements for online optimization algorithms may differ from that for usual mathematical optimization. In the next section we will discuss the considerations and requirements for online optimization algorithms, as well as discuss some specific options.

BEAM-BASED OPTIMIZATION ALGORITHMS

Considerations

Mathematical optimization is a well-researched area. There are numerous optimization algorithms. However, online optimization has special requirements and not all algorithms are suitable for online application [2].

One important difference is that the objective function in online optimization is impacted by measurement errors. For the same machine configuration, corresponding to the same set of optimization variables, the objective function evaluated on the machine will have slightly different value every time. The difference comes from the measurement errors of the parameters that come into the definition of the objective function. These errors could be due to the diagnostic system involved in the measurements. They could also be due to fluctuations of the machine condition that cause the parameter to actually change. Errors in the objective function can severely impact the performance of some algorithms, sometimes causing them to fail completely. For example, in the Nelder-Mead simplex method [5], the operations of the algorithm depend on the comparison results of the function values. When the comparison results are altered by measurement errors, the algorithm takes wrong paths, which could slow down or even prevent the convergence to the optimum.

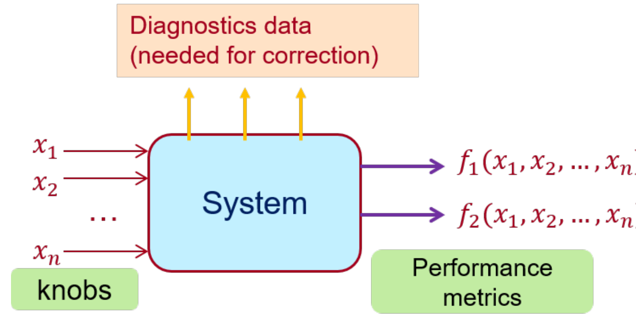

Figure 1: An illustration of the accelerator system for beam-based correction or optimization.

The measurement errors also cause difficulties to gradient-based optimization algorithms. These algorithms require the first-order or second-order derivatives of the objective function. Ordinarily, the derivatives can be approximated with numerical differences. However, when there are errors in function evaluations, the derivatives will have large errors as usually the step size used in numerical differential is small. The errors to second order derivatives are even larger. The alternative of using an accelerator model to compute the derivatives may not work as the model is inaccurate - otherwise online tuning would be unnecessary. Therefore, gradient-based algorithms, such as Newton's method or pseudo-Newton methods, are typically not used in online applications.

High efficiency is especially important for online optimization. This is because evaluation of the objective function takes time, and the overall time available for machine study is usually limited. The evaluation time include the time needed to change the machine setting until it settles in the new state and the time to measure the performance metrics. The time for a magnet to change to a new setpoint may be up to a few seconds, depending on the type of magnet and power supply. The measurement of performance metrics could vary from nearly instantaneous, seconds to tens of seconds, or even longer. An optimization session on the machine is usually up to a few hours in duration. Therefore, the number of function evaluations in one session can be

ISBN: 978-3-95450-236-3 ISSN: 2673-7027 doi:10.18429/JACoW-eeFACT2022-WEZAS0106

between tens to a few hundreds, with which the algorithm has to locate the optimum.

Special consideration is also needed on ensuring the safety of the machine. Safety caution could be implemented in the control system, for example, by setting software limit to the knob ranges. It can also be implemented in the objective function, in which more complex conditions or measures can be programmed. For example, a corner in the parameter space can be excluded; a "not an number" (NaN) value could be returned for an invalid beam condition; or the optimization can be paused if a certain beam condition is detected. The implementation of the algorithm should be aware of the scenarios that could occur during the evaluation of the objective function on the machine. By properly handling the scenarios, the optimization can be made safer, more efficient, and more reliable.

Algorithms

We discuss a few useful algorithms for online optimization, including Nelder-Mead simplex, robust conjugate direction search (RCDS), and multi-generation Gaussian process optimizer (MG-GPO).

The simplex algorithm is an efficient, gradient-free method. It converges to a minimum by morphing a simplex, a geometric body in the N-dimensional parameter space defined by $N + 1$ vertices, through a number of operations, including reflection, expansion, and contraction. In online application, the biggest challenge is that comparison of function values on the vertices can be altered by measurement noise, as the simplex size is reduced. The robust simplex (RSimplex) method can alleviate the issue by using extra sampling to reduce noise when necessary [6]. However, an accurate noise model is needed for it to be the most efficient. Using a large initial simplex size could also help reduce the impact the noise.

The RCDS method combines the power of conjugate direction searches and a robust, noise-aware 1-dimensional optimizer and is ideal for locating the optimum from a point in its vicinity. Search along one conjugate direction is independent of the search along another, which gives the method high efficiency. The conjugate direction set could be derived from a model by calculating the Hessian matrix of the objective function. The key of RCDS for its effectiveness comes from its ability to optimize under noise. During the step of bracketing the minimum, instead of merely comparing the function values between two points, the robust 1-D optimizer requires the end points to be higher than the lowest point (for a minimization problem) inside the bracket by 2 or 3 sigma. It also uses parabolic fitting to improve the accuracy in determining the minimum. The RCDS method has been successfully applied to many real-life accelerator optimization problems.

Many optimization algorithms, including simplex and RCDS, are inclined to converge to a local minimum. In many accelerator optimization problems, the true challenge is to find the global optimum in a high dimensional parameter space. Stochastic algorithms, such as random search, simulated annealing, genetic algorithms (GA) [7], and particle swarm optimization (PSO) [8, 9], are often used for global optimization. By using random operations to generate new candidate solutions, or a random decision process, these algorithms can break out from the attraction of local minima. However, these algorithms are typically not very efficient.

The MG-GPO method is a stochastic optimization algorithm with relatively high efficiency, which is enabled by machine learning. Similar to GA and PSO, it is population based and generates new solutions with random optimizations. However, it makes better use of the information contained in the solutions previously evaluated. A Gaussian process (GP) regression model is constructed for each objective function, using the existing solutions and a prior model characterized by the kernel matrix. The models can predict the performance of a trial solution. Instead of evaluating all trial solutions, MG-GPO uses the GP models to predict the performance of a large set of trial solutions and select only the ones expected to have good performance for evaluation. The algorithm has been benchmarked against many advanced stochastic algorithms and it was demonstrated it has superior convergence speed. It has also been tested in simulation and online problems. The MG-GPO algorithm is suitable for global optimization of complex, large parameter spaces.

Bayesian optimization (BO), also based on Gaussian process regression, has been adopted for accelerator tuning [10–12]. Bayesian optimization can be very efficient. Compared to BO, MG-GPO may be more robust and less dependent on the starting point and fine tuning of hyperparameters in the algorithm, for example, as experienced in the linac front-end tuning at APS [13].

APPLICATION EXAMPLES

There have been many successful applications of online optimization of accelerator performances. We will only discuss a few selected examples.

Storage Ring Vertical Emittance Minimization

In electron storage rings, the vertical emittance can come from the horizontal-vertical linear coupling and the vertical dispersion. Both effects are primarily due to random errors in the real machine, for example, misalignment of quadrupole (rolls) and sextupole magnets. Skew quadrupoles are effective knobs for controlling the vertical emittance as they can compensate both linear coupling and vertical dispersion. In storage ring light sources, a small vertical emittance corresponds to photon beam high brightness. In some cases the vertical emittance is set to a relative high level to achieve a reasonable Touschek lifetime. Even for these cases, it is preferable to first minimize the vertical emittance and then adjust the dispersion wave knob to increase it to the desired level.

At the SPEAR3 storage ring, the vertical emittance minimization problem has been used to test optimization algo-

ISBN: 978-3-95450-236-3 ISSN: 2673-7027 doi:10.18429/JACoW-eeFACT2022-WEZAS0106

rithms [2,12,14,15]. The ring has 13 free skew quadrupoles, which are used as tuning knobs. The beam loss rate, in a Touschek loss dominated parameter regime, normalized by the single bunch current, can be used as the objective function. A small vertical emittance corresponds to high beam loss rate. The beam loss rate can be measured by observing the beam current change over a fixed duration, or with beam loss monitors. In the latter case, it may be desirable to concentrate beam loss at where the loss monitor is located or use loss monitors distributed around the ring.

The RCDS method has been used to minimize the SPEAR3 vertical emittance [2]. The conjugate direction set is obtained with the Jacobian matrix of the off-diagonal blocks of the orbit response matrix with respect to the skew quadrupole knobs - it corresponds to the singular value decomposition (SVD) of the Jacobian matrix. In the experiments, all of the skew quadrupoles are initially set to zero strength. With about 200 function evaluations, the beam loss rate is increased to the maximum level. The skew quadrupole setting for the maximum beam loss rate is similar to the setting found with LOCO (correction with orbit response matrix) [16,17], while the maximum loss rate was higher than that of LOCO. The MG-GPO method has also been successfully applied to the SPEAR3 vertical emittance minimization problem [15].

Linac Transmission

Online optimization has been successfully used to tune the machine for optimal beam transmission in the linacs of both SPEAR3 and APS.

Some recent results for APS linac are reported in [13]. In the APS experiments, the goal is improve the transmission from the gun, around the alpha magnet, and in the first section of the linac. The objective function is the the charge measured in the L3 section. The tuning knobs are steering magnets and quadrupole magnets before and immediately after the alpha magnet. There are 12 tuning knobs. Several algorithms have been tested, including the simplex method, RCDS, PSO, and MG-GPO. Simplex works in many cases and converges fast, although it can also fail to make any improvement. For RCDS, it is important to correctly set the noise sigma parameter. For the MG-GPO algorithm, different population size of 8, 12, 20, and 30 was tried and was found to be robust. It converges faster than the PSO method and can find better solutions. The online tuning has been very helpful when a new gun was installed and it has been routinely used for linac front-end tuning.

Storage Ring Nonlinear Dynamics Optimization

Storage ring nonlinear dynamics tuning is extremely important for the commissioning of low emittance storage rings since these rings tend to have small dynamic aperture and momentum aperture, while there is no other reliable methods to compensate the inevitable errors in the real machine [18].

Online optimization of dynamic aperture with the RCDS method has been successively applied at several storage rings, including SPEAR3, MAX-IV [19], and NSLS-II [20]. Typically, the injection efficiency can be used as the objective function. If the initial injection efficiency is high, a reduced kicker bump or kicker bump mismatch may be used to decrease the injection efficiency and thus allow room for improvement. The sextupole and octupole (if available) magnets are used as tuning knobs. In the SPEAR3 case, a substantial improvement of more than 30% was achieved for DA; similarly large improvement was seen on MAX-IV and NSLS-II. PSO and MG-GPO have also been successfully used for DA optimization at SPEAR3.

At ESRF, sextupole knobs were used to maximize the Touschek lifetime [21]. The objective is lifetime normalized with beam current and the measured vertical beam size. In an experiment, the lifetime was improved from 11 h to 17 h.

In a recent study [22], simultaneous optimization of the dynamic aperture and Touschek lifetime was demonstrated on the APS storage ring, using MG-GPO, a multi-objective optimization method. The same 5 sextupole knobs are used for both objectives. These knobs are constructed from the 280 sextupole magnets, each with individual power supply, with symmetry considerations. A population size of 15 was used. To evaluate the objective functions without frequently dumping beam, for each generation, the injection efficiency was first measured for all solutions, which is followed by the lifetime measurements. Substantial improvements to both the dynamic aperture and the Touschek lifetime objectives were made.

SUMMARY

We discussed characteristics of the two beam-based approaches for improving accelerator performance: correction and optimization, in particular, the need for beam-based optimization and the special considerations for its implementation. Several online optimization algorithms, such as simplex, RCDS, and MG-GPO, are discussed. Their application to a few important real-life accelerator problems are described.

REFERENCES

[1] X. Huang, “Development and application of online optimization algorithms,” in *Proceedings of NAPAC2016*, Chicago, IL, 2016, pp. 1287–1291.

[2] X. Huang, J. Corbett, J. Safranek, and J. Wu, “An algorithm for online optimization of accelerators,” *Nucl. Instrum. Methods Phys. Res, Section A*, vol. 726, pp. 77 – 83, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900213006347

[3] X. Huang, M. Song, and Z. Zhang, “Multi-objective multi-generation gaussian process optimizer for design optimization,” *CoRR*, vol. abs/1907.00250, 2019. [Online]. Available: http://arxiv.org/abs/1907.00250

[4] L. Emery, M. Borland, and H. Shang, “Use of a general-purpose optimization module in accelerator control,” in *Proceedings of the PAC’03*, 2003, pp. 2330–2332.

[5] J. A. Nelder and R. Mead, “A simplex method for function minimization,” *The Computer Journal*, vol. 7,

ISBN: 978-3-95450-236-3 ISSN: 2673-7027 doi:10.18429/JACoW-eeFACT2022-WEZAS0106

no. 4, pp. 308–313, 01 1965. [Online]. Available: https://doi.org/10.1093/comjnl/7.4.308

[6] X. Huang, "Robust simplex algorithm for online optimization," *Phys. Rev. Accel. Beams*, vol. 21, p. 104601, Oct 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevAccelBeams.21.104601

[7] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *Trans. Evol. Comp*, vol. 6, no. 2, pp. 182–197, Apr. 2002. [Online]. Available: http://dx.doi.org/10.1109/4235.996017

[8] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, vol. 4, Nov 1995, pp. 1942–1948 vol.4.

[9] X. Pang and L. Rybarcyk, "Multi-objective particle swarm and genetic algorithm for the optimization of the LANSCE linac operation," *Nucl. Instrum. Methods Phys. Res, Section A*, vol. 741, pp. 124 – 129, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900213017464

[10] M. McIntire, T. Cope, and D. Ratner, "Bayesian optimization of FEL performance at LCLS," in *Proceedings of IPAC2016*, Busan, Korea, 2016, pp. 2972–2975.

[11] J. Duris, D. Kennedy, A. Hanuka, J. Shtalenkova, A. Edelen, P. Baxevanis, A. Egger, T. Cope, M. McIntire, S. Ermon, and D. Ratner, "Bayesian optimization of a free-electron laser," *Phys. Rev. Lett.*, vol. 124, p. 124801, Mar 2020. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.124.124801

[12] A. Hanuka, X. Huang, J. Shtalenkova, D. Kennedy, A. Edelen, Z. Zhang, V. R. Lalchand, D. Ratner, and J. Duris, "Physics model-informed gaussian process for online optimization of particle accelerators," *Phys. Rev. Accel. Beams*, vol. 24, p. 072802, Jul 2021. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevAccelBeams.24.072802

[13] H. Shang, Y. Sun, X. Huang, M. Song, and Z. Zhang, "Experience with on-line optimizers for APS linac front end optimization," in *Proceedings of the IPAC'21*, 2021, pp. 2151–2154.

[14] K. Tian, J. Safranek, and Y. Yan, "Machine based optimization using genetic algorithms in a storage ring," *Phys. Rev. ST Accel. Beams*, vol. 17, p. 020703, Feb 2014. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevSTAB.17.020703

[15] Z. Zhang, M. Song, and X. Huang, "Online accelerator optimization with a machine learning-based stochastic algorithm," *Machine Learning: Science and Technology*, vol. 2, no. 1, p. 015014, Dec 2020. [Online]. Available: https://dx.doi.org/10.1088/2632-2153/abc81e

[16] J. Safranek, "Experimental determination of storage ring optics using orbit response measurements," *Nucl. Instrum. Methods Phys. Res, Section A*, vol. 388, no. 1, pp. 27 – 36, 1997. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0168900297003094

[17] J. Safranek, G. Portmann, and A. Terebilo, "Matlab-based LOCO," in *Proceedings of EPAC'02*, 2002, pp. 1184–1186.

[18] X. Huang and J. Safranek, "Online optimization of storage ring nonlinear beam dynamics," *Phys. Rev. ST Accel. Beams*, vol. 18, p. 084001, Aug 2015. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevSTAB.18.084001

[19] D. K. Olsson, "Online optimisation of the MAX-IV 3 Gev ring dynamic aperture," in *Proceedings of IPAC2018*, Vancouver, BC, Canada, 2018, pp. 2281–2283.

[20] X. Yang *et al.*, "Online optimization of NSLS-II dynamic aperture and injection transient," in *Proceedings of IPAC2022*, Bangkok, Thailand, 2022, pp. 1159–1162.

[21] S. Liuzzo, N. Carmignani, L. Farvacque, B. Nash, T. Perron, P. Raimondi, R. Versteegen, and S. M. White, "RCDS optimizations for the ESRF storage ring," in *Proceedings of IPAC2016*, Busan, Korea, 2016, pp. 3420–3422.

[22] L. Emery, H. Shang, Y. Sun, and X. Huang, "Application of a machine learning based algorithm to online optimization of the nonlinear beam dynamics of the argonne advanced photon source," *Phys. Rev. Accel. Beams*, vol. 24, p. 082802, Aug 2021. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevAccelBeams.24.082802