ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-TUPAB290

DEMONSTRATION OF MACHINE LEARNING FRONT-END OPTIMIZATION OF THE ADVANCED PHOTON SOURCE LINAC

A. Hanuka*, J. Duris, SLAC National Accelerator Laboratory, Menlo Park, CA, USA
H. Shang, Y. Sun, Advanced Photon Source, Argonne National Laboratory, Lemont, IL, USA

Abstract

The electron beam for the Advanced Photon Source (APS) at Argonne National Laboratory is generated from a thermionic RF gun and accelerated by a s-band linear accelerator – the APS linac. While the APS linac lattice is set up using a model developed with ELEGANT [1], the thermionic RF gun front-end beam dynamics has been difficult to model. One of the issues is that beam properties from thermionic gun can vary. As a result, linac front-end beam tuning is required to establish good matching and maximize the charge transported through the linac. A traditional Nelder-Mead simplex optimizer has been used to find the best settings for the sixteen magnets. However it takes a long time, and requires a higher signal to noise ratio than Gaussian Process to get started tuning. The Gaussian Process (GP) optimizer does not have the initial condition limitation and runs several times faster. In this paper, we report our data collection and analysis for the training of the GP hyperparameters, and discuss the application of GP optimizer on the APS linac front-end optimization for maximum bunch charge transportation efficiency through the linac.

INTRODUCTION

Front-end optimization of accelerators is a challenging task due to the large input parameter space and their nonlinear and correlated behaviour. Additional sources of complexity are the uncontrollable drift in the experimental conditions and measurement noise. In order to find the optimal machine configuration, design values and simplified models are often used to provide the initial settings. However, the optimal performance cannot always be reached this way. This is because some beam dynamics are complicated to model and the accelerator has small compounding errors in every element (for example, magnet calibration errors, alignment errors, power supply regulation errors, etc.). Simplifications and approximations adopted in the design model also contribute to the differences observed between the empirical machine behavior and the design model. Therefore, deterministic approaches are insufficient and adjustments to the machine setting are needed.

Traditional optimization methods that do not rely on a model, such as Nelder-Mead Simplex [2], have been used successfully for *online* optimization of accelerator systems. In recent years there has been significant progress in the development of online optimization algorithms for particle accelerators. These include local methods based on approximations of the gradient, such as robust conjugate direction search (RCDS) [3], and global methods, such as evolutionary genetic algorithms [4]. However, the methods described above usually take a long time to converge, or could be stuck in a local optimum. The ability to find the *global* optimum in a complex parameter space with high efficiency (in terms of finding the optimum with the minimum number of function evaluations) are two critical requirements for a suitable tuning algorithm. ML-based optimization methods may be able to improve both the speed of convergence and final solution quality obtained during tuning.

Model-guided Bayesian optimization is a gradient-free approach [5, 6], where a local surrogate model (for example Gaussian Processes) is updated iteratively to guide the selection of new points in the search for global optimum. Bayesian optimization using Gaussian Processes has been successfully demonstrated [7–11]. In this paper, we demonstrate this approach for electron bunch charge optimization in the Advanced Photon Source linear accelerator [12] (see Fig. 1). We first describe the optimization task at APS linac, along with the data collection and the choice of the optimizer hyperparameters. We then compare the optimization results of the Bayesian optimizer to other methods (RCDS and simplex).

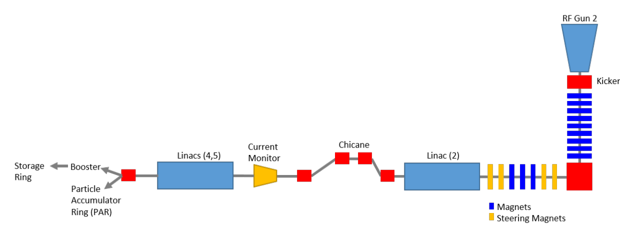


Figure 1: Schematics of the APS linac. The blue boxes show the magnet locations of the RG2 (RF gun 2) front-end used in the optimization. The current monitor (objective) is downstream in the linac (L3:CM1).

THE APS PHOTO-INJECTOR LINAC

The Advanced Photon Source (APS) at the U.S. Department of Energy's Argonne National Laboratory is a user facility that provides ultra-bright, high-energy storage ring-generated X-ray beams. The APS accelerator complex is composed of electron linear accelerator, Particle Accumulator Ring (PAR), booster and a 7 GeV storage ring. The APS electron linear accelerator is the first critical step in producing high-energy X-rays used for frontier research. Thus, optimizing the linac electron beam is a crucial task.

While the APS linac lattice is set up using a model developed with ELEGANT [1], the thermionic RF gun front-end

* adiha@slac.stanford.edu

ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-TUPAB290

beam dynamics have been difficult to model, especially for a very low energy beam just emitted from a thermionic cathode, where space-charge of the beam itself as well as electromagnetic field variations caused by small geometrical changes in thermionic cathode area can effect beam properties significantly. The beam properties from the thermionic gun can vary from run to run, or when the RF gun is swapped, resulting in variations of beam transport efficiency from the RF gun to the injector of the PAR. Therefore, linac front-end beam tuning is required to establish good matching and maximize the charge transported through the linac.

Currently, the semi manual tuning procedure can take several hours. A first step to the linac tuning procedure involves maximizing charge delivered through the chicane using a traditional simplex optimizer to set up the gun front-end magnets. This is followed by a three-screen emittance measurement downstream of the chicane to tune several quadrupoles in order to match beam parameters to the designed linac lattice. Then, the gradient and phase of the last two RF sectors of the linac are finely tuned for optimal beam energy and energy spread, and adjusting the Linac-To-PAR trajectory to finally achieve 100% injection efficiency through the PAR. In this work we optimize the gun front-end magnets for maximum charge using the recently developed Bayesian optimizer [7].

OPTIMIZATION OF THE LINAC ELECTRON BEAM

The goal is to optimize the RF gun front-end magnets to match the beam into the linac for maximal charge transportation efficiency. The S-band electron linac (375 - 500 MeV) has two thermionic RF guns (RG2 and RG1) and one Photo-Cathode RF gun, with thirteen 10-ft long accelerating structures. During normal operations, RG2 provides electron beams for the storage ring to generate X-rays for the APS users. We measure the bunch charge using a current monitor (L3:CM1) which is downstream in the linac as shown in Fig. 1.

The control variables include a total of 16 bipolar magnets installed between the RF gun 2 (RG2) and the first accelerating structure, among which there are 12 magnets (shown in blue) and 4 steering magnets (shown in orange). The 4 steering magnets are also used in the steering controller to adjust the linac trajectory. Downstream of the front-end magnets, this steering controller uses a total of 15 steering magnets at each plane to center the linac beam. If the steering controller is running during the optimization, it competes with the optimizer trying to adjust the 4 steering magnets. Therefore, we used the 12 magnets (without the 4 steering magnets) in the GP optimization tests.

In order to test the optimizer, we set a bad initial condition for the bunch charge (0.1 nC) as a starting point. We tested the L3:CM1 charge optimization using the 12 RG2 magnets with and without the steering controller running. All of the tested optimizers interfaced to the APS control system via the Ocelot optimization framework [13].

BAYESIAN OPTIMIZATION USING GAUSSIAN PROCESS

The Bayesian optimization framework uses an online probabilistic model (e.g. Gaussian process) of the objective function and an acquisition function to select the next point to sample as the space is searched. We used the upper confidence bound that incorporates the surrogate model's prediction and uncertainty compatible with observations up to the current observed point. The model is then updated with the new point and this routine continues until an optimum has been found.

The Gaussian Process (GP) models the bunch charge (objective) dependence on magnets currents (control variables) at tuning time. In this work, we deployed an online GP model [14]. The computation time per step is shorter than the acquisition time (about 5 seconds to set magnets and allow feedback systems to correct the trajectory and about 15 seconds to measure the bunch charge).

The GP model utilizes a kernel function that encodes the similarity between two points drawn from the objective function. Motivated by the observation that the charge response to the magnets looks bell-shaped when all other parameters are fixed, we chose to compare the radial basis function (RBF) kernel and the Matern52 kernel. We added a Gaussian noise kernel to capture the noise in the objective measurements.

The kernel's hyperparameters (length scales, amplitude, noise) are determined using a GP regression to historical data achieved by maximizing the GP marginal likelihood [15]; an approach known as ML-II. For the amplitude and noise parameters, we used archived data from the past two years which includes optimization data and typical operation data. For the magnets' lengthscales, we performed dedicated raster scans in order to capture the full behaviour of each magnet. Figure 2 shows raster scans of each control magnet without steering controller to maintain the linac trajectory during scanning. After scanning the 12 magnets, they were individually fit by a GP regressor using the scikit-learn package with 5 restarts.

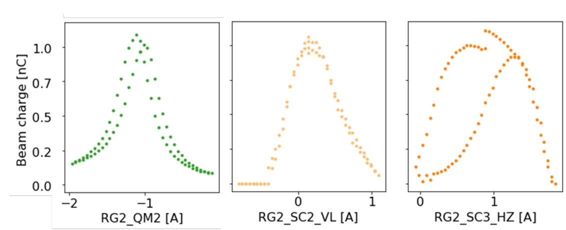
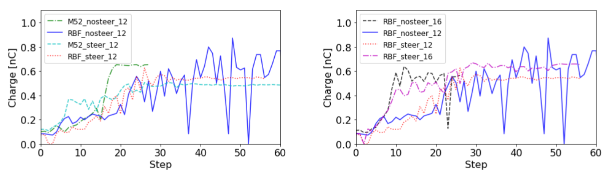


Figure 2: Raster scan examples of the 3 controlled magnets without the steering controller running.

It is evident that some of the magnets are affected by hysteresis. The Hysteresis might complicate the optimization process since the surrogate GP model of the beam charge as a function of the magnets depends now not only on the current magnet values but also on the historical sequence of the magnet values. The hysteresis behavior could be modeled and

ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-TUPAB290

incorporated into the GP in a way that the GP model would jointly learn the hysteresis and the beam charge response. However, in this work, we chose to limit the magnet's step size and constrain the acquisition bounds to restrain hysteresis errors. Such proximal exploration was studied in [8], where during optimization each input parameter is allowed to travel only a small distance.



(a) (b)

Figure 3: Comparison of the GP optimization performance with (a) either RBF or Matern52 kernels using 12 magnets and with/without the steering controller running, and (b) RBF kernel only with various hyperparameters.

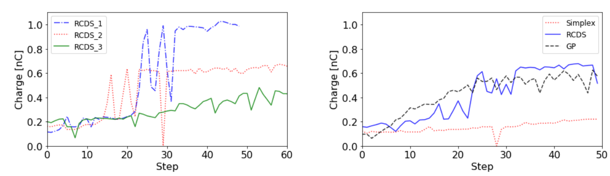
RESULTS & DISCUSSION

We compare the optimization performance of the GP optimizer with traditional, model-independent optimizers such as simplex and RCDS. Figure 3 shows the results of the GP optimizer with different kernels, various hyperparameters and with/without the steering controller running. In this figure, "steer" or "nosteer" represents optimization with or without the steering controller running correspondingly. First, it is evident that the optimization performance is without steering controller running was slightly better than the optimization with the steering controller. This means that the 12 magnets have some effect on the beam trajectory.

Figure 3a shows the optimization performance with RBF and Matern52 kernels using 12 magnets and with/without steering controller running. GP optimizer with RBF and "nosteer" was very unstable but reached a bunch charge of 0.87 nC with 48 steps, whereas the GP optimizer with Matern 52 reached a bunch charge of 0.65 nC after 26 steps. When the steering controller was running the GP optimizers with both kernels reached a lower charge (0.62 nC for the RBF and 0.5 nC for the Matern 52 kernels).

Figure 3b shows a comparison of the GP optimizer with RBF kernel and various hyperparameters; '12' or '16' means that the hyperparameters are obtained from raster scans performed without or with the steering controller running correspondingly. Fitting the GP regressor to obtain the lengthscales for the 12 magnets, using the raster scans data with the steering controller running, means that the 4 steering magnets, that were not used in optimization tests, were not fixed in the raster scan. This resulted in different lengthscale values for the 12 magnets used in the optimization. The optimization of the GP with hyperparameters obtained from raster scans with the steering controller running (labeled as RBF_steer_16 and RBF_nosteer_16) increased quickly in the first 10 steps. Overall, the performance of the GPs with the different hyperparameters is relatively comparable.

We then performed three RCDS scans without the steering controller running, shown in Fig. 4a. Although the best RCDS run reached a maximum charge of 1 nC, there is a great difference between the runs. On average, the RCDS runs reached 0.6 nC. Finally, in Fig. 4b we compare the averaged performance of GP, RCDS and Nelder-Mead simplex runs. Simplex didn't perform well since the initial starting point was very far from the optimal value. Usually, when used in normal operation, simplex is used for fine tuning from an initial point close to the optimal value. GP and RCDS performed similarly on average, both much better than simplex.



(a) (b)

Figure 4: (a) Three runs of RCDS; the spread between the runs is quite large. (b) Averaged performance of GP, RCDS and simplex optimizers.

CONCLUSION

Tuning and control are critical in accelerator commissioning and operation. While traditional tuning algorithms have limitations in the efficiency or the ability to search large parameter spaces, machine learning based algorithms should be able to reveal the global optimum in a parameter space with high efficiency. In this paper, we demonstrated online Bayesian optimization with hyperparameters estimated from archived historical data or dedicated raster scans. This optimizer can tune APS charge more efficiently than the routinely used Nelder-Mead simplex. The GP optimizer was more consistent in its performance on consecutive scans as compared with RCDS.

Various interesting applications for the APS linac are yet to be tested. For example, optimizing lattice (4 quads) and 10 steering magnets downstream of the magnetic chicane for maximum linac-to-ring injection efficiency. Moreover, further development of more expressive kernels [16, 17] or better approaches to obtain hyperparameters that are more representative of the machine [11] will benefit the commissioning and operation of existing and future accelerators and help automate accelerator operation in general.

ACKNOWLEDGEMENTS

This work was supported by the Department of Energy, Laboratory Directed Research and Development program at SLAC National Accelerator Laboratory, under contract DE-AC02-76SF00515. The authors would like to thanks Jane Shtalenkova (SLAC) for her support with Ocelot-Optimizer.

ISBN: 978-3-95450-214-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2021-TUPAB290

REFERENCES

[1] M. Borland, "ELEGANT: A flexible SDDS-compliant code for accelerator simulation", Argonne National Lab., IL, USA, Tech. Rep. LS-287 Advanced Photon Source, 2000. doi:10.2172/761286

[2] J. A. Nelder and R. Mead, "A simplex method for function minimization", *Comput. J.*, vol. 7, no. 4, pp. 308–313, 1965. doi:10.1093/comjnl/7.4.308

[3] X. Huang and J. Safranek, "Online optimization of storage ring nonlinear beam dynamics", *Phys. Rev. ST Accel. Beams*, vol. 18, p. 084 001, 2015. doi:10.1103/PhysRevSTAB.18.084001

[4] J. Wu *et al.*, "Multi-dimensional optimization of a terawatt seeded tapered free electron laser with a multi-objective genetic algorithm", *Nucl. Instrum. Methods. Phys. Res. A*, vol. 846, pp. 56–63, 2017. doi:10.1016/j.nima.2016.11.035

[5] J. Močkus, "On Bayesian methods for seeking the extremum", in *Proc. Optimization Techniques IFIP Tech. Conf.*, Novosibirsk, Russia, Jul. 1975, pp. 400–404. doi:10.1007/3-540-07165-2_55

[6] E. Brochu, V. M. Cora, and N. de Freitas, "A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning", 2010. arXiv: 1012.2599.

[7] J. Duris *et al.*, "Bayesian optimization of a free-electron laser", *Phys. Rev. Lett.*, vol. 124, p. 124 801, Mar. 2020. doi:10.1103/PhysRevLett.124.124801

[8] R. Roussel, A. Hanuka, and A. Edelen, "Multiobjective bayesian optimization for online accelerator tuning", *Phys. Rev. Accel. Beams*, vol. 24, p. 062 801, Jun. 2021. doi:10.1103/PhysRevAccelBeams.24.062801

[9] R. J. Shalloo *et al.*, "Automation and control of laser wakefield accelerators using bayesian optimization", *Nature Communications*, vol. 11, p. 6355, 2020. doi:10.1038/s41467-020-20245-6

[10] S. Jalas *et al.*, "Bayesian optimization of a laser-plasma accelerator", *Phys. Rev. Lett.*, vol. 126, p. 104 801, 2021. doi:10.1103/PhysRevLett.126.104801

[11] A. Hanuka *et al.*, "Physics model-informed gaussian process for online optimization of particle accelerators", *Phys. Rev. Accel. Beams*, vol. 24, p. 072 802, Jul. 2021. doi:10.1103/PhysRevAccelBeams.24.072802

[12] M. White, "Construction, commissioning and operational experience of the advanced photon source (APS) linear accelerator", in *Proc. 18th Int. Linear Accelerator Conf. (LINAC'96)*, Geneva, Switzerland, Aug. 1996, paper TU301, pp. 315–319.

[13] S. I. Tomin *et al.*, "Progress in automatic software-based optimization of accelerator performance", in *Proc. 7th Int. Particle Accelerator Conf. (IPAC'16)*, Busan, Korea, May 2016, pp. 3064–3066. doi:10.18429/JACoW-IPAC2016-WEPOY036

[14] M. McIntire, D. Ratner, and S. Ermon, "Sparse Gaussian processes for Bayesian optimization", in *Proc. 32nd Conf. Uncertainty in Artificial Intelligence*, Jersey City, NJ, USA, Jun. 2016, pp. 517–526.

[15] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[16] S. Sun, G. Zhang, C. Wang, W. Zeng, J. Li, and R. Grosse, "Differentiable compositional kernel learning for gaussian processes", in *Proc. 35th Int. Conf. on Machine Learning*, Stockholm, Sweden, Jul. 2018, pp. 4828–4837.

[17] A. G. Wilson, Z. Hu, R. R. Salakhutdinov, and E. P. Xing, "Stochastic variational deep kernel learning", in *Proc. Advances in Neural Information Processing Systems (NIPS'16)*, Barcelona, Spain, Dec. 2016, pp. 2586–2594.