ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-THAL02

BAYESIAN TECHNIQUES FOR ACCELERATOR CHARACTERIZATION AND CONTROL

R. Roussel*, A. Edelen, C. Mayes, SLAC National Accelerator Laboratory, 94025 Menlo Park, USA
J. P. Gonzalez-Aguilera, Y.K. Kim, University of Chicago, 60637 Chicago, USA

Abstract

Accelerators and other large experimental facilities are complex, noisy systems that are difficult to characterize and control efficiently. Bayesian statistical modeling techniques are well suited to this task, as they minimize the number of experimental measurements needed to create robust models, by incorporating prior, but not necessarily exact, information about the target system. Furthermore, these models inherently take into account noisy and/or uncertain measurements and can react to time-varying systems. Here we will describe several advanced methods for using these models in accelerator characterization and optimization. First, we describe a method for rapid, turn-key exploration of input parameter spaces using little-to-no prior information about the target system. Second, we highlight the use of Multi-Objective Bayesian optimization towards efficiently characterizing the experimental Pareto front of a system. Throughout, we describe how unknown constraints and parameter modification costs are incorporated into these algorithms.

INTRODUCTION

Tuning accelerators to meet operational goals of a given facility is a time-consuming task that limits valuable beam time resources for experimental users. This process represents a difficult to solve optimization problem, where there are many free parameters and limited, expensive to conduct measurements available to diagnose target objectives. Accelerator optimization problems also exist in tightly constrained parameter spaces where large regions of parameter space prevent even simple measurements of the beam. Finally, due to the complexity of accelerator systems, measurements are often noisy and/or have large uncertainties.

Model based optimization methods have been shown to speed up convergence of optimizing black box problems, where derivative information about the target function is not accessible, making routine operations faster and previously impossible to solve problems solvable in realistic settings. Of particular interest is the use of Bayesian optimization (BO) techniques for solving optimization problems [1, 2] including experimental optimization of accelerators [3, 4]. Bayesian statistical models aim to represent measurements as probability distributions, instead of scalar values. This naturally lends itself to characterizing experimental accelerator measurements, which have inherent noise and uncertainty. Bayesian optimization explicitly takes these uncertainties into account when performing optimization, resulting in an algorithm that is robust to noise (an issue faced by many other types of algorithms) [5]. This method is especially proficient at *efficient* global optimization, since the Bayesian surrogate model encodes high level information about the target function behavior (such as function smoothness), allowing it to make accurate predictions about the function with limited data sets, thus significantly improving optimization performance over other methods.

Bayesian optimization consists of two elements, a statistical surrogate model that makes predictions about a given target function and an acquisition function which uses those predictions to choose future points in input space to measure. The surrogate model is usually chosen to be a Gaussian process (GP) [6]. This model treats the value of the target function at each point in input space $\mathbf{x}$ as a random variable taken from a normal distribution $f \sim \mathcal{N}(\mu(\mathbf{x}), \sigma(\mathbf{x})^2)$ where $\mu(\mathbf{x})$ is the mean and $\sigma(\mathbf{x})$ is the standard deviation. Gaussian processes treat the joint probability distribution of the function values at each point in input space as a Multivariate normal distribution, specified by a mean function μ and a covariance matrix $\mathbf{\Sigma}$. We encode the expected behavior of the target function by specifying $\mathbf{\Sigma}$ via a kernel function $K(\mathbf{x}, \mathbf{x}')$ that describes how strongly function values at locations $\mathbf{x}$, $\mathbf{x}'$ are correlated with one another. A common class of kernel functions, known as "stationary kernels", are based solely on the distance between the two points, $||\mathbf{x}-\mathbf{x}'||$. We can then specify the expected smoothness of our function with a length-scale hyperparameter in specific kernels, such as the radial basis function (RBF) or Matern kernels [6]. Once experimental data is incorporated into the model, it can then be used to predict the function mean and corresponding uncertainty everywhere in the input domain.

Once a model is generated, we can then specify an acquisition function $\alpha(\mathbf{x})$ characterizes how valuable future observations are as a function of input parameters. For example, if we have high confidence in our model, we can choose to make measurements where the predicted function mean is at an extremum, thus heavily weighting "exploitation". On the other hand, if we wish to improve our understanding of the target function, we can place a high value on making observations in regions of high uncertainty, thus reducing the overall uncertainty of the model by heavily weighting "exploration". The most popular acquisition functions for single objective global optimization balances these two aspects, either implicitly using expected improvement over the best previously observed function value [7], or explicitly using an optimization hyperparameter [8].

In this work, we describe two acquisition functions that are specifically tailored to solve accelerator control problems. The first example describes "Bayesian exploration", an algorithm that enables automatic, efficient characterization of target functions, replacing the need for grid-like parameter

* rroussel@slac.stanford.edu

ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-THAL02

scans. The second is Multi-Objective Bayesian optimization (MOBO) which can be used to determine the entire Pareto front of a multi-objective optimization problem, as is often the case for control systems of large experimental facilities, using serialized measurements. We describe experimental demonstrations of each of these algorithms at two experimental facilities, the Linac Coherent Light Source (LCLS) and the Argonne Wakefield Accelerator (AWA). While our demonstrations are focused on the use of these algorithms in accelerator control systems, they can be easily used in a variety of other large experimental physics control systems.

BAYESIAN EXPLORATION

Due to the complex and time-consuming nature of accelerator diagnostics, characterization of beam response to input parameters is often limited to simple, uniformly spaced, grid-like parameter scans in one or two dimensions. This limitation results from the poor scaling of grid-like scans to higher dimensional spaces, where the number of samples needed grows exponentially with the number of input parameters. Furthermore, when attempting to characterize unknown systems it is often difficult to determine the ideal parameters of a grid scan that would result in an efficient and valid parameter scan.

The existence of tight constraints on which measurements are viable further complicates this process. For example, transverse beam size measurements on diagnostic screens are limited by the screen size, which in turn, imposes limits on the strength of upstream focusing magnet parameters. Simulation studies or extra measurements are needed beforehand to determine these limits. While these limits can be easily determined for a single parameter experimentally, it becomes infeasible to efficiently determine limits in higher dimensional input spaces, as they are often correlated with multiple parameters. Limitations such as these are shared among many types of control and optimization problems [9].

Finally, it is desirable to prevent rapid changes in accelerator input parameters during characterization. In some cases, it is temporally expensive to make changes in parameters, such as when mechanical actuators are used to change the phase of accelerating cavities. In other cases, fast feedback algorithms used in accelerator subsystems rely on adiabatic changes in external parameters to maintain system stability. Large jumps in parameter space can delay convergence of these feedback systems to stability or worse, cause them to fail entirely [10]. Practical experimental considerations such as these must be considered when automated sampling algorithms are used, which should strike a balance between costs associated with changing input parameters and the information gained.

To solve this problem, we developed an acquisition function to enable automated characterization of a target function in an efficient manner [11]. The acquisition function is given by

$$\alpha(\mathbf{x}, \mathbf{x}_0) = \sigma(\mathbf{x})\Psi(\mathbf{x}, \mathbf{x}_0) \prod_{i=1}^{N} P_i[g_i(\mathbf{x}) \geq h_i] \quad (1)$$

$$\Psi(\mathbf{x}, \mathbf{x}_0) = \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mathbf{x}_0)\right) \quad (2)$$

where $\sigma(\mathbf{x})$ is the standard deviation of the GP model, $\mathbf{x_0}$ is the most recently observed point, $\boldsymbol{\Sigma}$ is a diagonal, positive semi-definite matrix chosen by the experimenter and $P_i[g_i(\mathbf{x}) \geq h_i]$ represents the probability that the i'th operational constraint $g_i(\mathbf{x}) \geq h_i$ is satisfied. The terms of this acquisition function are as follows.

The first term $\sigma(\mathbf{x})$ is used to highly value points where model uncertainty is largest and has been shown to maximize information gain about the model [8]. When used with a GP kernel that has separate length scales for each free parameter in a process known as automatic relevance determination [6], this term will cause the optimizer to increase sampling frequency along the dimension with the shortest length scale, as seen in Fig. 1.

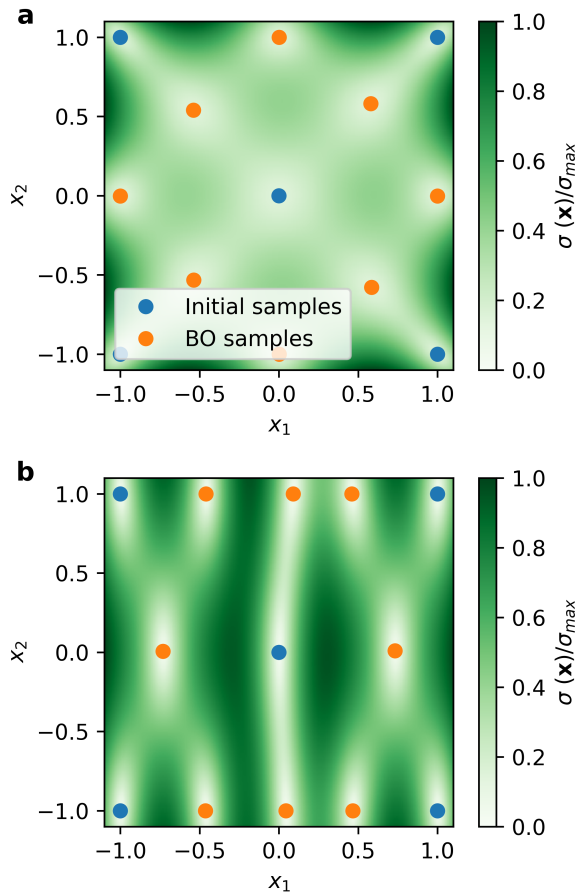

Figure 1: Plots showing Bayesian optimization sampling patterns depending on the kernel length scale using $\alpha(\mathbf{x}) = \sigma(\mathbf{x})$. Blue points are initial samples and orange points are determined via Bayesian optimization. (a) Length scales for both $[x_1, x_2]$ are set to 1. (b) Length scales for variables $[x_1, x_2]$ are set to $[0.25, 1]$. Reproduced from [11].

ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-THAL02

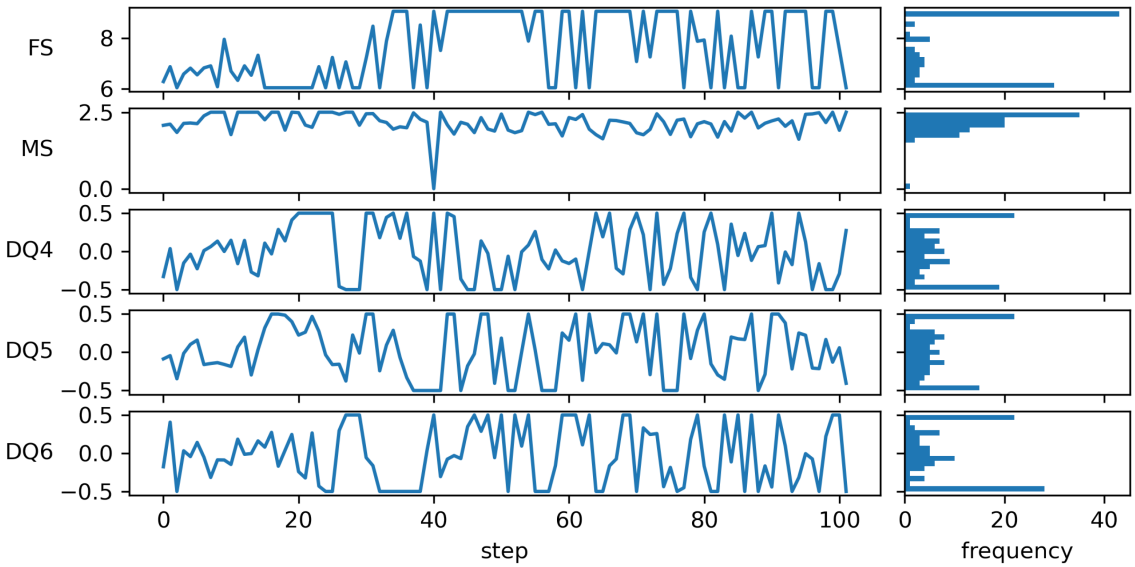

Figure 2: Parameter values (in arbitrary units) as a function of sample index during Bayesian exploration along with sample density histograms.

The second term $\Psi(\mathbf{x}, \mathbf{x_0})$ is a proximal weighting term, which biases the optimizer towards taking small steps in parameter steps. Another viable way to achieve this effect would be to place a hard limit on the maximum step size that the optimizer can take by setting the acquisition function to zero outside a given radius away from the last observed point. However, this negatively impacts the "exploration" advantages afforded by Bayesian optimization. By specifying a weighting that never reaches zero, we allow the optimizer to travel large distances in input space if it predicts a high potential value for the measurement.

Finally, the last term represents weighting due to constraining functions. For each constraint, we create a separate GP model that is used to calculate the probability that a given constraining function $g_i(\mathbf{x})$ satisfies $g_i(\mathbf{x}) \geq h_i$ where h_i is a constant. By weighting the acquisition function by these probabilities, it is less likely to propose points that violate any of the constraining functions, without prior knowledge of what those functions look like.

We conducted an experiment at the Argonne Wakefield Accelerator [12] to characterize the horizontal beam size at a diagnostic screen as a function of drive beam line parameters. These parameters included a solenoid around the photoinjector (FS), a solenoid just downstream of the photoinjector (MS) and three quadrupoles near the diagnostic screen (DQ4, DQ5, DQ6). The measurement was constrained by an upper bound on transverse beam size in both directions on the diagnostic screen and an upper bound on beam centroid distance from the screen center. We started the optimizer at a point in input space that satisfied the constraints and allowed it to explore for 100 iterations, taking 5 samples at each step and averaging the beam properties. At a repetition rate of 1 Hz, 100 iterations of the algorithm took less than 20 minutes, averaging less than 12 s per sample.

A trace of the sampled points in input space during exploration is seen in Fig. 2. Many of the parameters (FS, DQ4-6) did not have a strong effect on the bunch size over the given domain, which is inferred from the length scales of the fitted GP model for each parameter. As a result, most of the samples are on the domain boundaries and midpoints of each parameters' domain. Also note that these results imply that the next exploration run should happen over a larger domain if technologically feasible. On the other hand, we determined from its short length scale that the beam size was strongly correlated with the strength of the second solenoid (MS) relative to its input domain. Furthermore, we observe that only a narrow set of values for this solenoid satisfy the imposed constraints on the measurement. If we had attempted to characterize this input domain with a grid-like parameter scan, it is likely that a substantial number of measurements would have violated the constraints. Instead, our measurements using Bayesian exploration were valid about 80 percent of the time and sampling density was much higher for parameters that had strong correlations with our target measurement.

MULTI-OBJECTIVE BAYESIAN OPTIMIZATION

Many problems in accelerator physics optimization are multi-objective, where we must determine points in input space that optimally balance the trade-off between multiple, competing objectives, also known as the Pareto front. A popular set of techniques used to accomplish this are known as evolutionary algorithms. These algorithms, such as Non-dominated Sorting Genetic Algorithm II (NSGA-II) [13] or Multi-Objective Particle Swarm Optimization [14–16], are based on the generation of a diverse collection of candidate solutions, which are then observed via simulation or experiment, usually in a parallelized manner. The results from each observation are then sorted into non-dominated and dominated subsets. The non-dominated subset of can-

ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-THAL02

didate solutions is used to produce the next "generation" of candidate solutions using a stochastic heuristic, which are then re-evaluated. The process is repeated over several generations until the non-dominated set of observations converges to a stationary Pareto front or the hypervolume has converged to a maximum value. It has been shown that these methods are well suited for solving accelerator design optimization problems [17, 18].

However, these algorithms are poorly suited for online accelerator optimization. Evolutionary type algorithms often rely on parallelized evaluation of the objective functions. However, they become practically inefficient when restricted to serialized evaluations, as is the case for online accelerator optimization. This inefficiency rises from evolutionary algorithms use of the binary classification metric of Pareto dominance to generate the next generation of potential observation candidates. As a result, this metric does not guarantee optimal expansion of the Pareto front, as it does not consider the relative hypervolume improvement of individuals in the non-dominated subset of candidates.

The Multi-Objective Bayesian Optimization (MOBO) algorithm [19] on the other hand, maximizes optimization efficiency by using an explicit calculation of the hypervolume improvement as an acquisition function to select the best candidates to expand the Pareto frontier. The hypervolume improvement $\mathcal{H}_I$ is defined as the increase in Pareto front hypervolume by adding a new observation $\mathbf{y}$ (see Fig. 3). Each objective is modeled using a GP surrogate model which can be used to predict the hypervolume improvement as a function of input parameters. By maximizing the hypervolume improvement acquisition function, MOBO can determine a single point that maximally increases the Pareto front hypervolume at every step when evaluated in a serialized manner, making it ideal for online accelerator optimization.

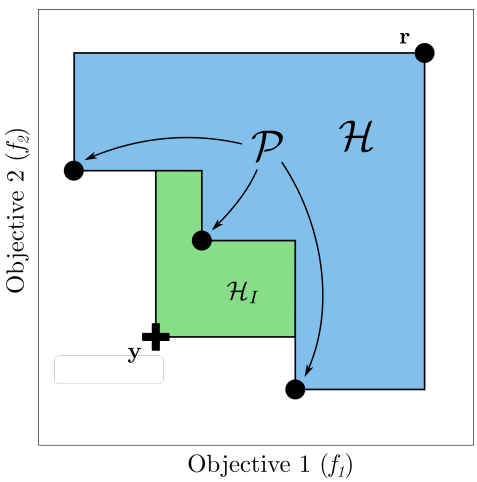

Figure 3: Cartoon of multi-objective optimization where each objective is to be minimized. Multi-objective optimization attempts to find a set of points known as the Pareto front $\mathcal{P}$ that dominate over a reference point $\mathbf{r}$ and any other observed points in objective space. The Pareto front hypervolume $\mathcal{H}$ (shown in blue) is the axis-aligned volume enclosed by the Pareto front and a reference point $\mathbf{r}$. Making a new observation $\mathbf{y}$, that dominates over points in the current Pareto front, leads to an increase in hypervolume (shown in green), referred to as the hypervolume improvement $\mathcal{H}_I$. Reproduced from [20].

The most common multi-objective acquisition function, expected hypervolume improvement (EHVI), is analogous to the popular single-objective expected improvement acquisition function [21]. This acquisition function calculates the average increase in hypervolume using the probability distribution of each objective function from the surrogate model. The EHVI acquisition function is formally defined as

$$\alpha_{EHVI}(\mu, \sigma, \mathcal{P}, \mathbf{r}) := \int_{\mathbb{R}^P} \mathcal{H}_I(\mathcal{P}, \mathbf{y}, \mathbf{r}) \cdot \xi_{\mu,\sigma}(\mathbf{y}) d\mathbf{y} \quad (3)$$

where $\mathcal{P}$ is the current set of Pareto optimal points, $\mathbf{r}$ is the reference point, $\mathcal{H}_I(\mathcal{P}, \mathbf{y}, \mathbf{r})$ is the hypervolume improvement from an observed point $\mathbf{y}$ in objective space, and $\xi_{\mu,\sigma}$ is the multivariate Gaussian probability distribution function with the GP predicted mean μ and standard deviation σ for each objective. A downside of this acquisition function is the calculation cost of determining the expected hypervolume improvement in high-dimensional objective spaces. A similar acquisition function, upper confidence bound hypervolume improvement (UCB-HVI) can be used as a cheaper to evaluate substitute with similar performance to EHVI [22].

We demonstrated the use of MOBO on optimizing the LCLS photoinjector [23]. Our goal was to determine the ideal trade-off between minimizing both the longitudinal bunch size σ_z and the vertical transverse beam size σ_y with respect to three tunable parameters, the gun solenoid, a focusing quadrupole magnet and a skew quadrupole magnet. The bounds of each free parameter were determined prior to experimentation from operator experience. The longitudinal bunch length was measured by a transverse deflecting cavity in the horizontal direction. Beam size measurements were conducted using optical transition radiation and statistics for each measurement were calculated from 5 samples each.

We began by randomly sampling 20 points in input space, producing the blue measurements in Fig. 4. We then used MOBO, with the UCB-HVI acquisition function to perform optimization for 20 iterations, initialized with data from the random search, shown in red in Fig. 4. Within these iterations MOBO was able to determine an approximate Pareto front (shown in the inset) that drastically improves over points found over random sampling. While this demonstration was a simple one, previous simulation results of optimizing the Argonne Wakefield Accelerator photoinjector with 7 objectives and 6 free input parameters show significant improvement over evolutionary algorithms commonly used by the accelerator community [20]. Furthermore, we also demonstrated that the same treatment of constraints and proximal biasing as was used in Bayesian exploration could also be incorporated into the MOBO acquisition function.

ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-THAL02

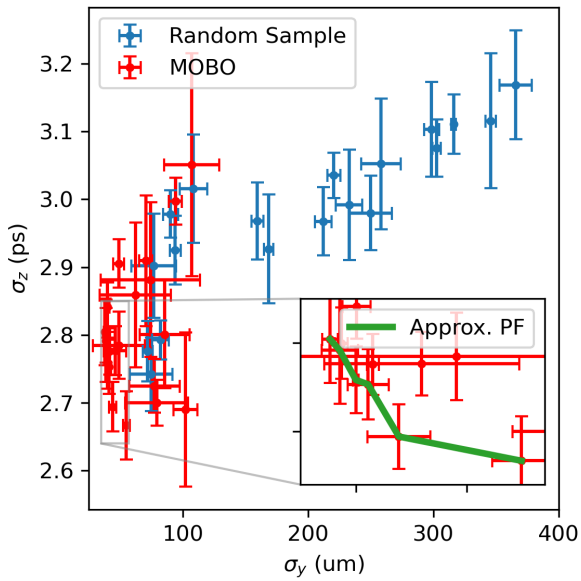

Figure 4: Results from running random search and then multi-objective Bayesian optimization to optimize the trade-off between electron bunch length σ_z and transverse beam size σ_y for the LCLS photoinjector. Free parameters included the gun solenoid, a focusing quadrupole and a skew quadrupole. Error bars denote 1σ error for each measurement due to jitter. Inset: Approximate Pareto front (PF) after 20 MOBO iterations.

Xopt: EASILY ACCESSIBLE IMPLEMENTATION OF ADVANCED ALGORITHMS

Using these algorithms on novel problems can take a significant amount of set-up and training to be successful. To reduce the upfront costs of using these and other complex algorithms for experimenters we have developed a flexible framework *Xopt* [24] for connecting advanced optimization algorithms to experimental control systems or simulations. Users need only to specify an algorithm by name and a python function that handles evaluation of target functions. Customization of optimization algorithms is also easily done, although their default configurations have been set to handle most problems off the shelf. *Xopt* has been used to both control the AWA accelerator and dispatch parallelized simulations on the Cori high performance computing cluster at NERSC [25]. The code is freely available at `https://github.com/ChristopherMayes/Xopt`.

CONCLUSION

In this work we have shown that Bayesian techniques can be used to efficiently characterize and optimize accelerators. They can navigate tightly constrained systems without prior knowledge regarding those systems. Furthermore, they can be modified by proximal biasing factors which promote adiabatic changes in controllable parameters, a requirement for efficient and stable optimization of large experimental systems.

ACKNOWLEDGEMENTS

We would like to thank the operations staff at both LCLS and AWA for their contributions to this work. This work was supported by the U.S. Department of Energy, under DOE Contracts No. DE-AC02-76SF00515, DE-AC02-06CH11357, the Office of Science, Office of Basic Energy Sciences as well as the Office of High Energy Physics, and U.S. National Science Foundation under Award No. PHY-1549132, the Center for Bright Beams.

REFERENCES

[1] B. Shahriari, K. Swersky, Z. Wang, R.P. Adams, and N. de Freitas, "Taking the Human Out of the Loop: A Review of Bayesian Optimization," in *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148-175, Jan. 2016. doi:10.1109/JPROC.2015.2494218

[2] S. Greenhill, S. Rana, S. Gupta, P. Vellanki, and S. Venkatesh, "Bayesian Optimization for Adaptive Experimental Design: A Review," in *IEEE Access*, vol. 8, pp. 13937-13948, 2020. doi:10.1109/ACCESS.2020.2966228

[3] 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, 2020. doi:10.1103/PhysRevLett.124.124801

[4] J. Kirschner, M. Mutný, N. Hiller, R. Ischebeck, and A. Krause, "Adaptive and Safe Bayesian Optimization in High Dimensions via One-Dimensional Subspaces," in *Proc. Int. Conf. on Machine Learning*, vol. 97, pp. 3429-3438, 2019. https://arxiv.org/abs/1902.03229v2

[5] J. Snoek, H. Larochelle, and R.P. Adams, "Practical Bayesian Optimization of Machine Learning Algorithms," in *Advances in Neural Information Processing Systems*, F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, Ed. New York, NY, USA: Curran Associates, Inc., 2012, pp. 2951–2959.

[6] C.E. Rasmussen and C.K.I. Williams. *Gaussian processes for machine learning*, Adaptive computation and machine learning. Cambridge, MA, USA: MIT Press 2006.

[7] J. Mockus, V. Tiesis, and A. Zilinskas, "The Application of Bayesian Methods for Seeking the Extremum," in *Towards Global Optimisation 2*, L.C.W.Dixon and G.P. Szego, Ed. North-Holland, pp. 117-129, 1978. https://www.researchgate.net/publication/248818761_The_application_of_Bayesian_methods_for_seeking_the_extremum

[8] N. Srinivas, A. Krause, S. Kakade, and M. Seeger, "Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design," in *Proceedings of the 27 th International Conference on Machine Learning (ICML'10)*, Haifa, Israel, 2010, pp. 1015–1022, Madison, WI, USA, June 2010. Omnipress. https://icml.cc/Conferences/2010/papers/422.pdf

[9] E.A. Baltz, E. Trask, M. Binderbauer, M. Dikovsky, H. Gota, R. Mendoza, J.C. Platt, and P.F. Riley, "Achievement of Sustained Net Plasma Heating in a Fusion Experiment with the

ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-THAL02

Optometrist Algorithm," *Scientific Reports*, vol. 7, p. 6425, July 2017. doi:10.1038/s41598-017-06645-7

[10] A.L. Edelen, S.G. Biedron, B.E. Chase, D. Edstrom, S.V. Milton, and P. Stabile, "Neural Networks for Modeling and Control of Particle Accelerators," *IEEE Transactions on Nuclear Science*, vol. 63, pp. 878–897, 2016. doi:10.1109/TNS.2016.2543203

[11] R. Roussel, J.P. Gonzalez-Aguilera, Y.-K. Kim, E. Wisniewski, W. Liu, P. Piot, J. Power, A. Hanuka, and A. Edelen, "Turn-key constrained parameter space exploration for particle accelerators using Bayesian active learning," *Nature Communications*, vol. 12, p. 5612, 2021. doi:10.1038/s41467-021-25757-3

[12] M.E. Conde, S.P. Antipov, D.S. Doran, W. Gai, Q. Gao, G. Ha, *et al.*, "Research Program and Recent Results at the Argonne Wakefield Accelerator Facility (AWA)," in *Proc. IPAC'17*, Copenhagen, Denmark, May 2017, pp. 2885–2887. doi:10.18429/JACoW-IPAC2017-WEPAB132

[13] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, pp. 182–197, 2002. doi:10.1109/4235.996017

[14] J. Kennedy and R. Eberhart, "Particle Swarm Optimization," in *Proc. of the IEEE Int. Conf. on Neural Networks (ICNN'95)*, vol. 4, pp. 1942-1948, 1995. doi:10.1109/ICNN.1995.488968

[15] X.B. Huang and J. Safranek, "Nonlinear dynamics optimization with particle swarm and genetic algorithms for SPEAR3 emittance upgrade," *Nucl. Instrum. Methods Phys. Res., Sect. A*, vol. 757, pp. 48–53, September 2014. doi:10.1016/j.nima.2014.04.078

[16] X. Pang and L.J. Rybarcyk, "Multi-objective particle swarm and genetic algorithm for the optimization of the LANSCE linac operation," *Nucl. Instrum. Methods Phys. Res., Sect. A*, vol. 741, pp. 124–129, March 2014. doi:10.1016/j.nima.2013.12.042

[17] Y.J. Li, W.X. Cheng, L. H. Yu, and R. Rainer, "Genetic algorithm enhanced by machine learning in dynamic aperture optimization," *Phys. Rev. Accel. Beams*, vol. 21, p. 054601, May 2018. doi:10.1103/PhysRevAccelBeams.21.054601

[18] N. Neveu, L. Spentzouris, A. Adelmann, Y. Ineichen, A. Kolano, C. Metzger-Kraus, C. Bekas, A. Curioni, and P. Arbenz, "Parallel general purpose multiobjective optimization framework with application to electron beam dynamics," *Phys. Rev. Accel. Beams*, vol. 22, p. 054602, May 2019. doi:10.1103/PhysRevAccelBeams.22.054602

[19] M. Emmerich, K. Yang, A. Deutz, H. Wang, and C.M. Fonseca, "A Multicriteria Generalization of Bayesian Global Optimization," in *Advances in Stochastic and Deterministic Global Optimization*, P. Pardalos, A. Zhigljavsky, J. Žilinskas, Eds. Springer Optimization and Its Applications, vol. 107, pp. 229–242. Springer International Publishing, Cham. doi:10.1007/978-3-319-29975-4_12

[20] R. Roussel, A. Hanuka, and A. Edelen, "Multiobjective Bayesian optimization for online accelerator tuning," *Phys. Rev. Accel. Beams*, vol. 24, p. 062801, June 2021. doi:10.1103/PhysRevAccelBeams.24.062801

[21] M.T.M. Emmerich, A.H. Deutz, and J.W. Klinkenberg, "Hypervolume-based expected improvement: Monotonicity properties and exact computation," *IEEE Congress of Evolutionary Computation (CEC)*, pp. 2147-2154, 2011. doi:10.1109/CEC.2011.5949880

[22] M.T.M. Emmerich, K. Yang, and A.H. Deutz, "Infill Criteria for Multiobjective Bayesian Optimization," in *High-Performance Simulation-Based Optimization. Studies in Computational Intelligence*, T. Bartz-Beielstein, B. Filipič, P. Korošec, E.-G. Talbi, Eds. Cham, Springer International Publishing, vol. 833, pp. 3–16, 2020. doi:10.1007/978-3-030-18764-4_1

[23] R. Akre, D. Dowell, P. Emma, J. Frisch, S. Gilevich, G. Hays, Ph. Hering, R. Iverson, C. Limborg-Deprey, H. Loos, A. Miahnahri, J. Schmerge, J. Turner, J. Welch, W. White, and J. Wu, "Commissioning the Linac Coherent Light Source injector," *Phys. Rev. ST Accel. Beams*, vol. 11, p. 030703, March 2008. doi:10.1103/PhysRevSTAB.11.030703

[24] C. Mayes, R. Roussel, and H. Slepicka, *ChristopherMayes/Xopt:* Xopt v0.5.0, October 2021. doi:10.5281/zenodo.5559141

[25] K. Antypas, N. Wright, N.P. Cardo, A. Andrews, and M. Cordery, "Cori: A Cray XC Pre-Exascale System for NERSC," in *Cray User Group Proceedings (CUG2014)*, Cray, 2014. https://cug.org/proceedings/cug2014_proceedings/includes/files/pap211.pdf