ISBN: 978-3-95450-261-5 ISSN: 2673-7000 doi: 10.18429/JACoW-NAPAC2025-MOP094

TWO-STAGE CONSTRAINED BAYESIAN OPTIMIZATION FOR PARTICLE ACCELERATOR TUNING *

Fuhao Ji†, Zihan Zhu, Ryan Roussel, Lee Zucker-Murray[1], Mianzhen Mo, Yusong Liu, Patrick Kramer, Frederick (Erik) Cropp, Sara Miskovich, Indranil Nayak, Stephen Weathersby, Zhe Zhang, Sean Gasiorowski, Alexander Reid, Daniel Ratner, Auralee Edelen, Robert Joel England
SLAC National Accelerator Laboratory, Menlo Park, CA, USA
[1]also at University of California, Los Angeles, CA, USA

Abstract

Particle accelerators are highly complex, non-linear systems that require rapid tuning during operation to meet requirements on beam qualities for applications in different scientific disciplines. Multi-objective Bayesian Optimization (MOBO) has been recently demonstrated at SLAC MeV-UED facility for speeding up online electron beam tunings and obtaining Pareto Fronts giving trade-offs between key beam properties of interest. One challenge in algorithm-based tuning is the alignments of beam through the collimators, screens and timing diagnostics under different system working points. This usually requires trial-and-error based hand tunings and strongly limits the data taking efficiency. Here, we utilize two-stage constrained Bayesian optimizations (CBE-MOBO) for beam tunings at MeV-UED. Instead of directly optimizing objectives of interest, beam property constraints are first modeled in the input space using constrained Bayesian exploration. Based on the information learned, MOBO is then used to efficiently search the parameter space and resulted in dramatically improved valid data efficiency. Our results show potential of CBE-MOBO for autonomous tunings of particle accelerators under various constraints.

INTRODUCTION

Artificial intelligence and machine learning (AI/ML) have recently reached sufficient maturity for applications in real-world tasks, including natural language processing [1], autonomous driving, protein folding prediction [2] and fusion reactor control [3]. Speeding up and aiding online optimizations of complex particle accelerators is one of the key areas where AI/ML can make substantial contributions [5–13]. Bayesian Optimization (BO) with Gaussian processes (GPs) represents a class of ML algorithms in which a probabilistic model and its uncertainty at each measurement step are used to select the optimal next observation point. BO-based optimization methods have been widely applied by the accelerator physics community in recent years due to its advantages in addressing complex optimization tasks, especially for online accelerator tunings and computational expensive simulations [10]. At SLAC MeV-UED [4], Multi-objective Bayesian Optimization (MOBO) [14] has been recently demonstrated for speeding up online electron beam tunings and obtaining Pareto Fronts (PFs) giving trade-offs between key beam properties of interest [13]. Practically, one challenge in online accelerator tuning is the alignments of beam through the collimators, screens and timing diagnostics under different system working points. This often requires extensive hand tunings based on a trial-and-error approach and strongly limits the data taking efficiency of MOBO. To this end, we utilize two-stage constrained Bayesian optimizations for automating beam tunings at MeV-UED. Instead of directly optimizing objectives of interest, constrained Bayesian Exploration (CBE) [11] is first conducted for constructing Gaussian Processes (GPs) to model the beam property constraints in the input domain. Based on the information learned using CBE, constrained MOBO is then conducted to efficiently search the parameter space and optimize the PF. The CBE-MOBO approach results in dramatically improved valid data efficiency with minimal human inputs, and show potential of autonomous tunings of complex, correlated particle accelerators under various beam property constraints.

TWO-STAGE CONSTRAINED MULTI-OBJECTIVE BAYESIAN OPTIMIZATION

Practically, accelerator tunings often require meeting constraints such as transmission through collimators or apertures, as well as alignments through diagnostic elements. CBE treats each unknown constraint function $c_i(\mathbf{x})$ as an independent GP [15], and uses those models to bias exploration toward regions where each constraint exceeds a required threshold Q_i. Given a set of N observations:$D_N = \{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), ..., (\mathbf{x}_N, \mathbf{y}_N)\}$, The objective and constraint functions can be models as:

$$f(\mathbf{x}) \sim \mathcal{GP}\big(\mu_f(\mathbf{x}),\, k_f(\mathbf{x}, \mathbf{x}')\big) \tag{1}$$

$$c_i(\mathbf{x}) \sim \mathcal{GP}\big(\mu_{c_i}(\mathbf{x}),\, k_{c_i}(\mathbf{x}, \mathbf{x}')\big) \tag{2}$$

For a "$> Q_i$" constraint, the probability that c_i is satisfied at $\mathbf{x}$ is

$$P_i(\mathbf{x}) = P\big(c_i(\mathbf{x}) > Q_i\big) = \Phi\Big(\frac{\mu_{c_i}(\mathbf{x}) - Q_i}{\sigma_{c_i}(\mathbf{x})}\Big) \tag{3}$$

* Work supported by U.S. Department of Energy Office of Science, Basic Energy Sciences under Contract No. DE-AC02-76SF00515
† Corresponding author: fuhaoji@slac.stanford.edu

ISBN: 978-3-95450-261-5 ISSN: 2673-7000 doi: 10.18429/JACoW-NAPAC2025-MOP094

where $\sigma_{c_i}(\mathbf{x}) = \sqrt{k_{c_i}(\mathbf{x},\mathbf{x})}$ and Φ is the standard normal cumulative distribution function. Assuming m independent constraints, the joint feasibility is

$$P_{\text{feas}}(\mathbf{x}) = \prod_{i=1}^{m} \Phi\Big(\frac{\mu_{c_i}(\mathbf{x}) - Q_i}{\sigma_{c_i}(\mathbf{x})}\Big). \tag{4}$$

This feasibility term is then used to weight the unconstrained Bayesian Exploration acquisition function $\alpha_{\text{BE}}(\mathbf{x})$, yielding the constrained acquisition function

$$\alpha_{\text{CBE}}(\mathbf{x}) = \alpha_{\text{BE}}(\mathbf{x}) \times P_{\text{feas}}(\mathbf{x}). \tag{5}$$

By doing so, CBE naturally steers sampling toward points that both promise characterization in f and satisfy all $c_i(\mathbf{x}) > Q_i$ with high probability. Building on the feasible region that CBE learns, one can then employ constrained MOBO to learn the Pareto front giving optimal trade-offs between objectives of interest. The constrained EHVI acquisition function can be written as:

$$\alpha_{\text{cEHVI}}(\mathbf{x}) = \int_{\mathbb{R}^M} H_I(\mathcal{P}, \mathbf{y}, \mathbf{r})\, \mathcal{N}_{\boldsymbol{\mu}(\mathbf{x}),\boldsymbol{\Sigma}(\mathbf{x})}(\mathbf{y})\, d\mathbf{y} \times \prod_{i=1}^{m} \Phi\Big(\frac{\mu_{c_i}(\mathbf{x}) - Q_i}{\sigma_{c_i}(\mathbf{x})}\Big). \tag{6}$$

By separating feasible region exploration from Pareto-front learning, the CBE-MOBO approach significantly improves data efficiency and avoids the risk of the optimizer becoming trapped in local optima outside the true feasible Pareto front.

IMPLEMENTATION AT SLAC MEV-UED

We conduct experiments at SLAC MeV-UED facility to demonstrate feasibility of applying CBE-MOBO for online constrained exploration and optimizations of multiple electron beam properties in a multi-dimensional parameter space. The experimental setup and control workflow are illustrated in Refs. [12, 13]. Beam spot size (denoted as σ_x) at sample plane was measured using a YAG screen coupled to a Questar Microscope with spatial resolution better than 10 μm, Momentum space resolution (q-resolution, denoted as σ_q) was measured using a P43 phosphor screen coupled to Andor EMCCD. Pulse charge was obtained using integrated beam intensity with calibration based on Faraday cup measurement. CBE-MOBO was implemented using the Xopt package [16]. Trust-region BO (TurBO) [17] was used to maintain a trust region around the best feasible point and adjusts its size based on constraint feedback: it expands after successful evaluations to explore more broadly and contracts after violations to focus on valid areas. The dynamic sizing ensures efficient, locally informed optimization under constraints. Automated objective functions were constructed for evaluating objective and constraints by switching between different detectors and diagnostics so that measurements of beam properties at different beamline locations can be quickly performed in a serialized manner [12]. Time-stamped system parameters were archived simultaneously for GP modeling. 10^2 – 10^3 data points can be obtained during a several hours run.

RESULTS

CBE-MOBO was conducted at SLAC MeV-UED. Beamline parameters including gun phase, solenoid strengths and steering magnets, etc., were varied to produce electron beams with various pulse lengths, spot size, q-resolution and transmitted pulse charges. Table 1 illustrates the beamline parameters varied during CBE-MOBO optimizations. For each run, multiple objectives, e. g., spot size vs q-resolution, along with constraint of pulse charge transmitted through the collimator, were assigned. The optimizer first acquire 10 randomly sampled points in the parameter space and construct the GP surrogate model, then run CBE for 20 steps to explore the parameter space and learn the feasible region meeting pulse charge constraints. Based on the information learned using CBE, constrained MOBO is then conducted to efficiently search the parameter space and obtain the PF giving trade-offs between objectives.

Figure 1 illustrates the σ_x vs σ_q CBE-MOBO result at 2.31 MeV beam energy with initial pulse charge 10 fC and 100 µm collimator inserted. PF was obtained within 80 measurements. The Pareto Frontier data points (blue stars) show clear trade-off between σ_x at sample plane and σ_q on the diffraction detector, the smallest σ_x was down to 40 µm rms with σ_q up to 0.064 Å^{-1} (left most data point), while smallest σ_q of 0.0602 Å^{-1} can be achieved with σ_x up to 51 µm rms (lowest point on the PF).

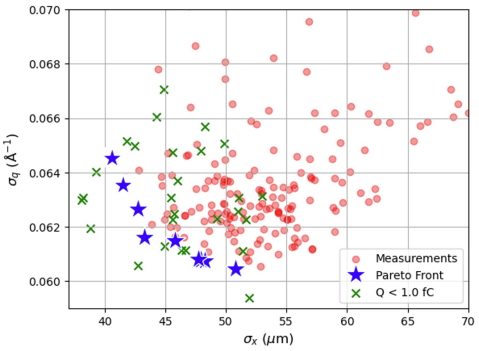

Figure 1: CBE-MOBO for σ_q vs σ_x optimizations, Pareto Frontier data points are highlighted with blue stars, data points violating the constraint are highlighted with green cross

Figure 2 illustrates data distributions in 2-dimensional subspaces of the parameter-objective-constraint joint domain. Each data point's color indicates its iteration in the evaluation sequence. As one can clearly see, data points end up converging to the optimal areas in the input parameter space. The Q vs. Sol1 plot shows a clear inverse correlation: as the gun solenoid strength increases, the transmitted charge

ISBN: 978-3-95450-261-5 ISSN: 2673-7000 doi: 10.18429/JACoW-NAPAC2025-MOP094

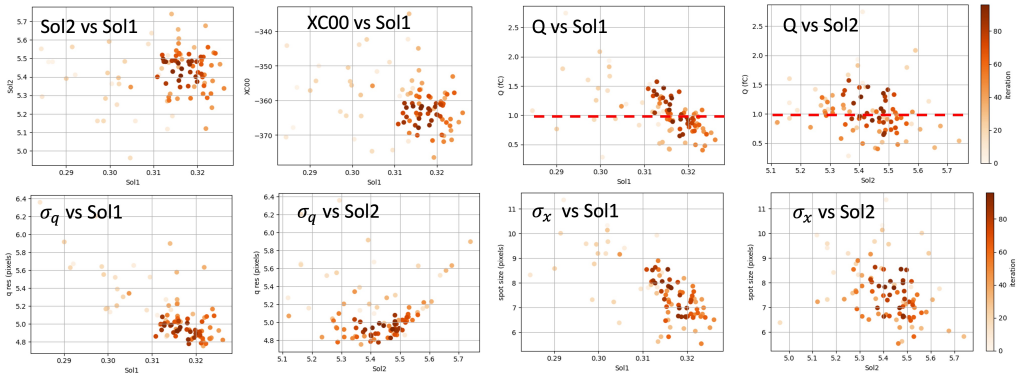

Figure 2: Data distributions in 2-dimensional subspaces of the parameter-objective-constraint space. Each data point's color indicates its iteration in the evaluation sequence. Dashed line represent the pulse charge constraint (Q > 1 fC) applied during CBE-MOBO process

Q decreases. The threshold $Q = 1.0$ fC occurs at Sol1 = 0.318. Meanwhile, both σ_x and σ_q decrease monotonically with Sol1, leading the constrained MOBO optimizer to converge at Sol1 = 0.318 to obtain optimal PF while satisfying the $Q > 1.0$ fC constraint. On the other hand, Sol2 shows a strong correlation only with σ_q, with a clear optimum at Sol2 = 5.4 where σ_q reaches its minimum. Figure 3 illustrates data points in the Sol1-XC00 and Sol1-YC00 subspaces, GP surrogate model of constraint Q was constructed using measured data points, color map represents the predicted mean of Q. To satisfy the $Q > 1.0$ fC constraint, XC00 must be co-optimized with Sol1, as the feasibility plot clearly delineates the valid region in the Sol1–XC00 space. In contrast, YC00 shows little correlation with Q, which is in consistent with prior experience. As evident, CBE-MOBO is capable of learning the feasible region in the multi-dimensional input space and can strategically propose the next optimal

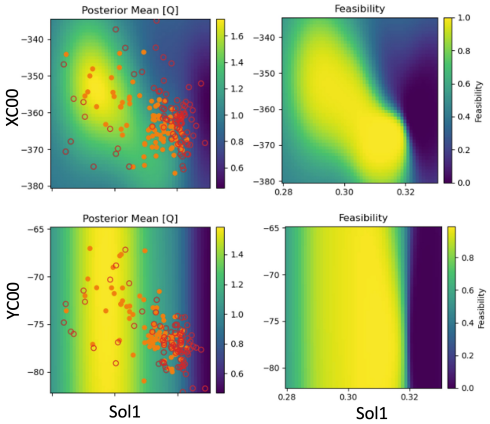

Figure 3: (left) data points in the Sol1-XC00 and Sol1-YC00 subspaces, GP surrogate model was constructed using measured data points, solid/hollow points represents valid/invalid data points, color map represents the predicted mean of transmitted pulse charge (Q). (right) corresponding feasibility map learned by CBE, the feasible region showing correlation between Sol1 and XC00 can be clearly seen.

Table 1: Beamline Parameters varied during CBE-MOBO

Parameter	**Range/Value**	**Unit***
Initial pulse charge	10	fC
Gun amplitude	66	MV/m
Gun phase	[20, 40]	degree
collimator diameter	100	μm
Gun solenoid strength	[0.24, 0.36]	-
2^{nd} solenoid strength	[4.4, 6.4]	-
Steering magnet XC00	[-420, -300]	-
Steering magnet YC00	[-93, -53]	-
Steering magnet XC01	[-4.2, 6.2]	-
Steering magnet YC01	[3.3, 5.3]	-
Steering magnet XC02	[11.0, 19.0]	-
Steering magnet YC02	[-3.7, 1.7]	-
Objective/Constraint	**Target**	
Spot size (σ_x)	minimize	
q-resolution (σ_q)	minimize	
Transmitted pulse charge (Q)	> 1.0 fC	

*a. b. unit if not specified.

observation point, thereby improving valid data efficiency and advancing the PF simultaneously. Table 2 illustrates the valid data efficiency $\rho = N_{valid}/N_{meas}$ at different stages of the CBE-MOBO process. As can be seen, after conducting CBE, the valid data efficiency of MOBO stage was dramatically improved to 81.6%, and shows clear advantage over random sampling (57.7%).

Table 2: Valid Data Efficiency at Different Stages

Stage	**Random**	**CBE**	**MOBO-post CBE**
ρ ($Q \geq 1.0$ fC)	57.7%	63.3%	81.6%

CONCLUSION

In this work, we demonstrate the two-stage CBE-MOBO algorithm for online electron beam tunings with beam prop-

ISBN: 978-3-95450-261-5 ISSN: 2673-7000 doi: 10.18429/JACoW-NAPAC2025-MOP094

erty constraints at SLAC MeV-UED facility. Instead of directly running MOBO, constrained Bayesian Exploration is first utilized for constructing GPs to model the beam property constraints in the multi-dimensional input parameter space. Based on the information learned, constrained MOBO is then conducted to efficiently search the parameter space and optimize the PF. Our results show that the CBE-MOBO approach is capable of learning the feasible region in the input parameter space, thereby dramatically improving valid data efficiency and advancing the PF simultaneously. The CBE-MOBO approach requires minimal human inputs, and show potential of autonomous tunings of complex, correlated particle accelerators under various beam property constraints.

ACKNOWLEDGEMENTS

This work was supported by the U.S. Department of Energy Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. M.M. acknowledges the support from the U.S. DOE Office of Science, Laboratory Directed Research and Development program at SLAC under contract DE-AC02-76SF00515, and Fusion Energy Sciences under FWP 101242

REFERENCES

[1] D. Guo *et al.*, "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning," arXiv:2501.12948, 2025.

[2] J. Jumper *et al.*, "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, pp. 583-589, 2021.

[3] J. Degrave *et al.*, "Magnetic control of tokamak plasmas through deep reinforcement learning," *Nature*, vol. 602, pp. 414–419, 2022.

[4] SLAC, MeV-UED (Megaelectronvolt Ultrafast Electron Diffraction), https://lcls.slac.stanford.edu/instruments/mev-ued

[5] A. Edelen, N. Neveu, M. Frey, Y. Huber, C. Mayes, and A. Adelmann, "Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems," *Phys. Rev. Accel. Beams*, vol. 23, p. 044601, 2020.

[6] A. Sheinker *et al.*, "Demonstration of model-independent control of the longitudinal phase space of electron beams in the Linac-Coherent Light Source with femtosecond resolution," *Phys. Rev. Lett.*, vol. 121, p. 044801, 2018,

[7] S. C. Leemann *et al.*, "Demonstration of machine learning-based model-independent stabilization of source properties in synchrotron light sources," *Phys. Rev. Lett.*, vol. 123, p. 194801, 2019.

[8] J. Duris *et al.*, "Bayesian Optimization of a Free-Electron Laser," *Phys. Rev. Lett.*, vol. 124, p. 124801, 2020.

[9] R. Roussel *et al.*, "Multiobjective Bayesian optimization for online accelerator tuning", *Phys. Rev. Accel. Beams*, vol. 24, p. 062801, 2021.

[10] R. Roussel *et al.*, "Bayesian optimization algorithms for accelerator physics," *Phys. Rev. Accel. Beams*, vol. 27, p. 084801, 2024.

[11] R. Roussel *et al.*, "Turn-key constrained parameter space exploration for particle accelerators using Bayesian active learning," *Nat. Commun.*, vol. 12, p. 5621, 2021.

[12] F. Ji *et al.*., "Multi-Objective Bayesian Optimization at SLAC MeV-UED", in *Proc. IPAC'22*, Bangkok, Thailand, Jun. 2022, pp. 995-998. doi:10.18429/JACoW-IPAC2022-TUPOST056

[13] F. Ji *et al.*, "Multi-objective Bayesian active learning for MeV-ultrafast electron diffraction,", *Nat. Communic.*, vol. 15, p. 4726, 2024.

[14] M. Emmerich, K. Yang, A. Deutz, H. Wang and C. M. Fonseca, "A multi-criteria generalization of Bayesian global optimization" in Advances in Stochastic and Deterministic Global Optimization, Springer Optimization and Its Applications, pp. 229-242.

[15] C. E. Rasmussen and K. I. Williams, "Gaussian Processes for Machine Learning" in Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, 2006

[16] R. Roussel., *et al.*, "Xopt: A simplified framework for optimization of accelerator problems using advanced algorithms," in *Proc. IPAC'23*, Venice, Italy, May 2023, pp. 4847–4850. doi::10.18429/JACoW-IPAC2023-THPL164

[17] D. Eriksson, *et al.*, "Scalable global optimization via local Bayesian optimization," Advances in Neural Information Processing Systems, 2019.

[18] S. Daulton, M. Balandat, and E. Bakshy, "Differentiable expected hypervolume improvement for parallel multi-objective Bayesian optimization" in *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020 (NeurIPS 2020)*, Dec,, 2020, virtual, Vancouver, BC, Canada, Dec. 2020, https://researchr.org/publication/DaultonBB20