13th Int. Particle Acc. Conf. IPAC2022, Bangkok, Thailand JACoW Publishing
ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOST053

BEAM TUNING AT THE FRIB FRONT END USING MACHINE LEARNING*

Kilean Hwang†, Tomofumi Maruta, Alexander Plastun, Kei Fukushima, Tong Zhang, Qiang Zhao, Peter Ostroumov, Samuel Nash
Facility for Rare Isotope Beams, Michigan State University, East Lansing, Michigan, USA

Abstract

The Facility for Rare Isotope Beams (FRIB) at Michigan State University produced and identified the first rare isotopes demonstrating the key performance parameter and completion of the project. An important next step toward FRIB user operation includes fast tuning of the Front End (FE) decision parameters to maintain optimal beam optics. The FE consists of the ion source, charge selection system, LEBT, RFQ, and MEBT. The strong coupling of many ion source parameters, strong space-charge effects in multi-component ion beams, and a not well-known neutralization factor in the beamline from the ion source to the charge selection system make the FE modeling difficult. In this paper, we present our first effort toward the Machine Learning (ML) application for automatic control of the beam exiting the FE.

INTRODUCTION

FRIB houses a powerful heavy-ion linear accelerator to produce wide spectrum of rare isotope species of variety of charge states [1–3]. Such an unprecedented capability requires a frequent switch of primary beam ion species followed by optics tuning in FE. Therefore, fast FE beam optics tunning as well as maintaining the beam quality is crucial for the mission. We have been successfully utilizing the NelderMead [4] – a simplex method – for automated fast tuning of FE upon initialization of the beam commissioning. This optimization algorithm decides the next optimal decision parameters to be evaluated based only on a few data points that form a simplex over the decision parameter domain. If we can exploit all the data points we visited since the start of the optimization, we may get more sample-efficient[1] decision. Furthermore, if the all the historical (or archived) data from past operations is somewhat consistent[2], we may able to exploit it to enhance the sample efficiency even more. In this regard, we develop and test the prior-mean-assisted Bayesian optimization (pmBO) where the prior model is trained over the historical (or archived) data. We are also creating surrogate models of physics simulation of FE for fast and large data collection purposes so as to realize the application of pmBO for FE tuning.

PRIOR MEAN ASSISTED BAYESIAN OPTIMIZATION

The ML methods for the online beam tuning that are reported until today may be in largely two categories: reinforcement learning (RL) and surrogate model assisted optimization (SMAO) [5–7]. The Bayesian optimization (BO) belongs to the latter category. In general, compared to RL, SMAO is more sample efficient but less robust to the machine drift, and heavier in numerical complexity [8]. Furthermore, SMAO is, generally, not suitable for continuously tunning due to the assumption of static problems and the numerical complexity[3]. Nevertheless, we find SMAO, (especially BO) is a good fit for our purpose due to the ability to incorporate the historical (or archived) data naturally in terms of the prior model as well as the sample efficiency. And the numerical complexity problem can be relieved if optimization converges in a few steps thanks to the good prior model. Finally, once the optimization converged, we can use either a pre-trained (off-policy, offline) RL or traditional control algorithms like the Extremum-Seeking for continuous tunning for adaption to the machine drift [9].

In this section, we present how we model the data reflecting the effect of the machine drift, and exploit the data for optimization on two test problems: Rosenbrock and Rastrigin functions:

$$
\begin{aligned}
\mathfrak{Rosen}\,(x_1, x_2, \ldots, x_d) &= \sum_{i=1}^{d-1} \left(x_{i+1} - x_i^2\right)^2 + \frac{(1-x_i)^2}{100} \\
\mathfrak{Rastr}\,(x_1, x_2, \ldots, x_d) &= \sum_{i=1}^{d-1} \frac{x_i^2 - \cos(2\pi x_i)}{d} - 1 \qquad (1)
\end{aligned}
$$

These are commonly used objective functions for benchmarking optimization algorithms. Figure 1 shows them for the 2-dimensional case.

Historical Data Model

We assume that the system dynamics in terms of the objective $f(\cdot)$ of interest is fully describable by the decision parameters $\boldsymbol{x}_{decision}$, and known $\boldsymbol{x}_{known}$ and unknown $\boldsymbol{x}_{unknown}$ environmental parameters except for small noises $\epsilon \sim \mathcal{N}(0, 0.01)$:

$$y = f\left(\boldsymbol{x}_{decision}; \boldsymbol{x}_{known}, \boldsymbol{x}_{unknown}\right) + \epsilon \qquad (2)$$

* Work supported by the Director of the Office of Science of the US Department of Energy under Cooperative Agreement DE-SC0000661, the State of Michigan and Michigan State University

† hwang@frib.msu.edu

[1] By "sample efficiency", we mean a faster optimization in terms of the number of objective evaluations. This does not necessarily mean a faster optimization in terms of the wall clock.

[2] Here we mean the consistency of historical (or archived) data in a sense that the present machine status in terms of the relationship between the decision parameters and the objective of interest is not very far from most of the past machine statues (due to machine drift) when the data was collected.

[3] It involves with the surrogate model training and optimization over the surrogate model or an acquisition function that is a function of the surrogate model.

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOST053

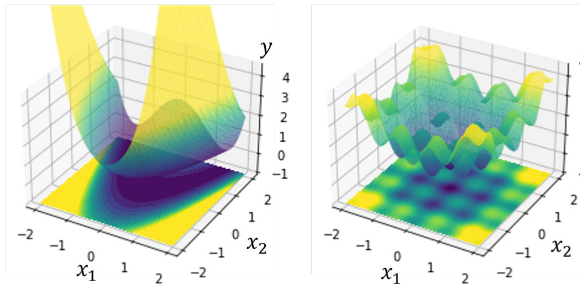

Figure 1: Rosenbrock 2D (left) and Rastrigin 2D (right).

where y is the observed objective. For example, $\boldsymbol{x}_{decision}$ may include beam steering magnets, $\boldsymbol{x}_{known}$ may include magnet settings that are not included in $\boldsymbol{x}_{decision}$, and $\boldsymbol{x}_{unknown}$ may include unknown injected beam condition, magnet misalignments, calibrations and etc. This way, we are modeling the machine drift over time in terms of the variation of the unknown environmental parameters.

We write the full historical data set by:

$$\mathcal{D}_{full} = \{y, \boldsymbol{x}_{decision}, \boldsymbol{x}_{known}, \boldsymbol{x}_{unknown}\} \tag{3}$$

and distinguish it from the data that can be recorded:

$$\mathcal{D} = \{y, \boldsymbol{x}_{decision}, \boldsymbol{x}_{known}\} \tag{4}$$

For test purposes, we generate a hypothetic historical data by uniform random sampling over the input domain $[-2, 2]^d$ of the Rosenbrock and Rastrigin functions where d is the dimension of the input domain. And assign decision parameters and known and unknown environmental parameters over the input domain.

Prior Mean Model

We train a neural-network (NN) F_θ, where θ is the NN parameters, to represent the prior mean model as a function of $\boldsymbol{x}_{decision}$ and $\boldsymbol{x}_{known}$ such that

$$\theta = \underset{\theta}{\operatorname{argmin}} \mathbb{E}_{\mathcal{D}} \left(F_\theta\left(\boldsymbol{x}_{decision}, \boldsymbol{x}_{known}\right) - y\right)^2 \tag{5}$$

To illustrate, we visualize a prior mean model for 2-dimensional Rosenbrock function assuming 1 decision parameter $x_1 = x_{decision}$ and 1 unknown environmental parameter $x_2 = x_{unknown}$ such that

$$f\left(x_{decision}; x_{unknown}\right) = \mathfrak{Rosen}\left(x_{decision}, x_{unknown}\right)$$

A hypothetic historical data $\mathcal{D}_{full}$ is shown on the left plot of Fig. 2. Since only x_1 is known, our prior mean model trained over $\mathcal{D}$ is 1-dimensional. Note that the prior mean model shown on the right of the Fig. 2 captures the overall shape of the projections (onto x_1) of the 2D Rosenbrock.

Although it may depend on the problem at hand, this example illustrates that the dynamics of the objective in terms of $x_{decision}$ can be approximately captured into the prior mean model even if the training data was taken from various machine status (that is modeled by change of $x_{unknown}$).

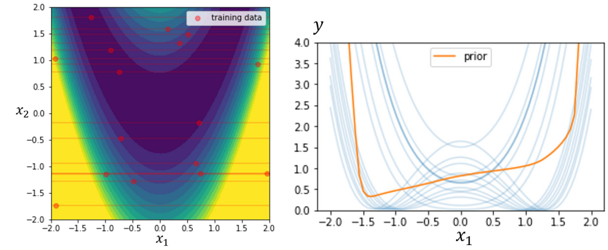

Figure 2: Left: hypothetic historical data $\mathcal{D}_{full}$ shown in red dots. Background color represents ground truth 2D Rosenbrock. Red lines correspond to the projections shown on the right plot. Right: Trained prior mean model (orange) over $\mathcal{D}$ and projections (blue) of the 2D Rosenbrock onto $x_{decision}$ at fixed $x_{unknown} \in \mathcal{D}_{full}$

This allows us to train a posterior model, conditioned to the prior mean model F_θ, that can adapt to the current machine status (i.e. current values of $x_{unknown}$) using fresh data that is collected on the fly during the optimization. Then SMAO, with the posterior as the surrogate model, can quickly guide us to the next decision parameters for evaluation.

Benchmark

For better confidence, we benchmark pmBO, BO, and NelderMead with 100 trials. For each trial, the initial decision parameters and environmental parameters are randomly chosen. The result for 2 dimensional Rosenbrock with 1 decision parameters, 1 unknown environmental parameters is shown in Fig. 3. The results for 14 dimensional Rosenbrock and Rastrigin functions with 10 decision parameters, 2 known and 2 unknown environmental parameters are shown in Fig. 4 and Fig. 5 respectively. In addition to these, in all other benchmarks we did with different combinations of parameters' dimensionality, pmBO [4] outperformed [5] vanilla BO that is not aware of the historical data.

SURROGATE MODEL OF SIMULATION

The reliability of the prior mean model depends on the training data distribution and size. For example, for a 10-dimensional problem, one may need about 10^6 data points (assuming 4 points in each dimension). If the experimental data collection rate is about 5 seconds, the 1 million data points require 2 months of operation. In addition, machine operation tune is often limited to near-optimal settings which prevent the generality of the historical data distribution. In this regard, we are creating surrogate models of physics simulations. In FE simulation, nearly half of the computation time is spent in RFQ. As a visual illustration, we created a longitudinal aperture surrogate model for RFQ as in Fig 6.

[4] We used the size of the training data (which represent hypothetic historical data) for the prior mean model to be $\min(4^d, 10^6)$ where d is the dimensionality of the problem so that the data size is exponential to the dimensionality while limited to 1 million.

[5] NelderMead performance strongly depends on the initial size of the simplex. Here, we used 0.05% of the bounds of each dimension to create points of the initial simplex around the initial decision parameters.

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOST053

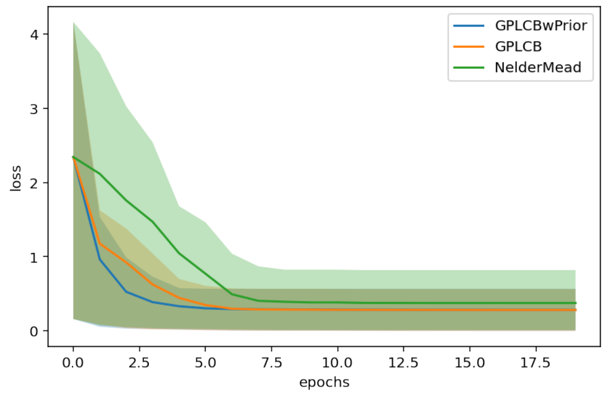

Figure 3: Benchmark of pmBO, BO, and NelderMead for 2D Rosenbrock problem assuming 1 decision and 1 unknown environmental parameter. 100 trials of optimization performed while $x_{unknown}$ are randomly chosen and fixed during each optimization. The thick line is the average over all the trials. The shade represents the trial population from 20% to 80%. The width of the shade is large because of different values of $x_{unknown}$. “GPLCB” represents BO with the Gaussian Process surrogate model and Lower Bound Limit acquisition function. “GPLCBwPrior” represents pmBO.

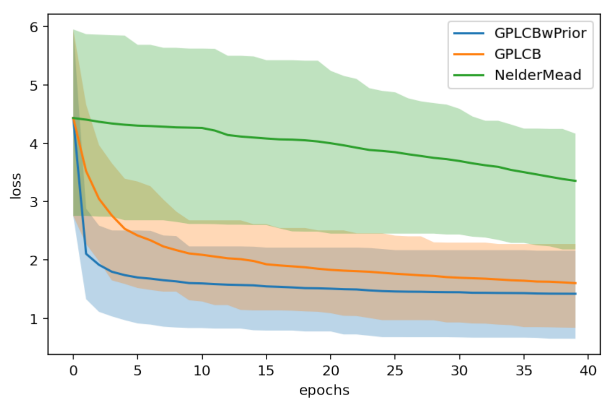

Figure 4: Benchmark: 14-dimensional Rosenbrock (10 decision, 2 known, 2unknown environmental parameters). 100 trials of optimization performed while x_{known}, and $x_{unknown}$ are randomly chosen and fixed during each optimization.

Note also that, if simulation time is about 1 minute, the 1 million data requires 1 year of computation with a single CPU. To be practical in terms of the number of simulations, we reduce the dimensionality by splitting the FE, section by section, so that each section contains only 7 or 8 tunable parameters. Such a simulational surrogate model can be used to generate low-fidelity (in the sense that the real machine may be different from the physics model) data. The high-fidelity data that is archived from real machine operation may be used to correct the prior mean model that is trained using low-fidelity data. Further details and results regarding the simulational surrogate model will be reported elsewhere.

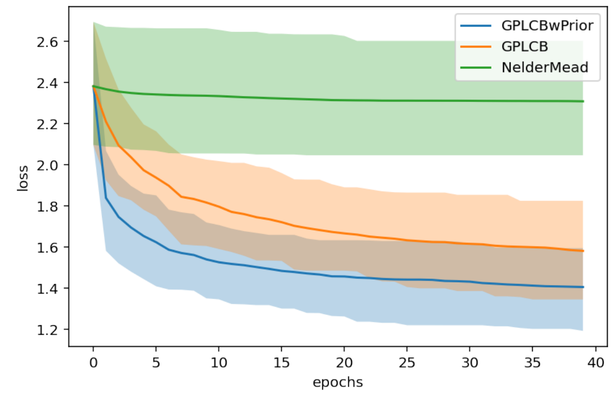

Figure 5: Benchmark: 14 dimensional Rastrigin (10 decision, 2 known, 2unknown environmental parameters).

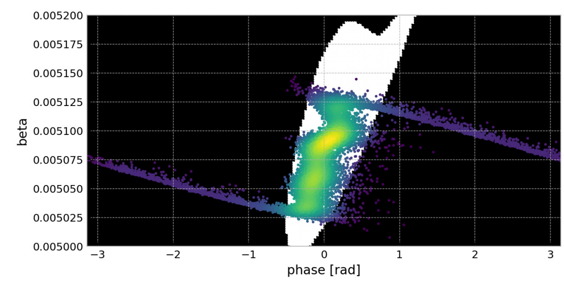

Figure 6: Longitudinal RFQ aperture surrogate model in black and white and the beam density.

SUMMARY AND CONCLUSION

In an effort toward fast tunning of FRIB FE, we developed and tested prior-mean-assisted Bayesian optimization (pmBO) on arbitrary dimensional test functions: Rosenbrock and Rastrigin functions. We modeled hypothetic historical data over the test functions in consideration of machine drift over time. Throughout our tests, we observed pmBO outperforming vanilla BO. Such success is conditional on enough data size for the prior mean model training. We are creating surrogate models of physics simulations for FRIB FE so as to generate large data for the high-dimensional prior mean model. Further details of the simulational surrogate modeling will be reported elsewhere.

REFERENCES

[1] J. Wei et al, “Advances of the FRIB project", International Journal of Modern Physics EVol. 28, No. 03, 1930003 (2019). https://www.worldscientific.com/doi/abs/10.1142/S0218301319300030

[2] P. N. Ostroumov, T. Maruta, S. Cogan, K. Fukushima, S. H. Kim, S. Lidia, F. Marti, A. S. Plastun, J. Wei, T. Yoshimoto, T. Zhang, and Q. Zhao, “Beam commissioning in the first superconducting segment of the Facility for Rare Isotope Beams", *Phys. Rev. ST Accel. Beams*, vol. 22, p. 080101, Aug. 2019. 10.1103/PhysRevAccelBeams.22.080101

[3] P. N. Ostroumov, K. Fukushima, T. Maruta, A. S. Plastun, J. Wei, T. Zhang, and Q. Zhao, “First Simultaneous Acceleration

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOST053

of Multiple Charge States of Heavy Ion Beams in a Large-Scale Superconducting Linear Accelerator", *Phys. Rev. Lett*, vol. 126, p. 114801, Mar. 2021.
10.1103/PhysRevLett.126.114801

[4] Virtanen, Pauli et. al., “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python", *Nature Methods*, vol. 17, p. 261–272, 2020.
https://rdcu.be/b08Wh

[5] Edelen, Auralee and Neveu, Nicole and Frey, Matthias and Huber, Yannick and Mayes, Christopher and Adelmann, Andreas, “Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems", *Phys. Rev. ST Accel. Beams*, vol. 23, p. 044601, Apr. 2020.
10.1103/PhysRevAccelBeams.23.044601

[6] Miskovich, S. A. et. al., “Online Bayesian optimization for a recoil mass separator", *Phys. Rev. ST Accel. Beams*, vol. 25, p. 044601, Apr. 2022.
10.1103/PhysRevAccelBeams.25.044601

[7] O’Shea, F. H. and Bruchon, N. and Gaio, G., “Policy gradient methods for free-electron laser and terahertz source optimization and stabilization at the FERMI free-electron laser at Elettra", *Phys. Rev. ST Accel. Beams*, vol. 23, p. 122802, Dec. 2020.
10.1103/PhysRevAccelBeams.23.122802

[8] Levine, Sergey and Kumar, Aviral and Tucker, George and Fu, Justin, “Offline Reinforcement Learning: Tutorial, Review, and Perspectives on Open Problems", *arXiv*, 2005.01643. 2020.
https://doi.org/10.48550/arxiv.2005.01643

[9] Alexander Scheinker, En-Chuan Huang and Charles Taylor, “Extremum Seeking-Based Control System for Particle Accelerator Beam Loss Minimization", in *IEEE*, Transactions on Control Systems Technology. Dec. 2021.
doi:10.1109/TCST.2021.3136133