Latent Bayesian Optimization for the Autonomous Alignment of Synchrotron Beamlines

T. W. Morris[a], Y. Du[a], M. Fedurin[a], A. C. Giles[a], P. Moeller[b], B. Nash[b], M. Rakitin[a], B. Romasky[a, c], A. L. Walter[a], N. Wilson[a, d], and A. Wojdyla[e]

[a]Brookhaven National Laboratory, Upton, NY 11973, USA
[b]RadiaSoft LLC, Boulder, CO 80301, USA
[c]Rutgers University, New Brunswick, NJ 08901, USA
[d]Cornell University, Ithaca, NY 14853, USA
[e]Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

ABSTRACT

The autonomous alignment of synchrotron beamlines is typically a high-dimensional, high-overhead optimization problem, requiring us to predict a fitness function in many dimensions using relatively few data points. A model that performs well under these conditions is a Gaussian process, upon which we can apply the framework of classical Bayesian optimization methods. We show that even with no prior data, a tailored Bayesian optimization algorithm is capable of autonomously aligning up to eight dimensions of a digital twin of the TES beamline at NSLS-II in only a few minutes. We implement this approach in a software package for automatic beamline alignment, which is available out-of-the-box for any facility that leverages the Bluesky environment for beamline manipulation and data acquisition.

Keywords: Bayesian optimization, machine learning, alignment, synchrotron beamlines, digital twins

1. INTRODUCTION

Manual alignment of synchrotron beamlines is typically a time-consuming process, with hours or days spent on diagnostics and not data collection. A system for automatically aligning beamlines would be an invaluable tool for beamline scientists and users alike, as it would allow for better and more rapid alignment leading to more efficient experimentation for users. However, there are a number of challenges to applying machine learning to beamline alignment, which is a noisy, high-dimensional optimization problem: inputs may be highly coupled, data acquisition is not robust, and sampling inputs is slow as it requires moving motors extremely precisely. Bayesian optimization excels in this class of optimization problem, as it performs well in high-dimensional spaces with relatively little data, and can robustly deal with unknown levels of noise; Bayesian algorithms are also general enough that we can tailor them to beamline optimization problems in particular. In this paper, we focus our attention on the Tender-Energy X-ray Absorption Spectroscopy (TES) beamline at the National Synchrotron Light Source-II (NSLS-II). We consider two components of the TES beamline: the toroidal mirror which focuses on the secondary source aperture (SSA) and the Kirkpatrick-Baez (KB) mirror pair which focuses the beam after the SSA. These two mirrors have orthogonal tasks of maximizing flux and spot size, respectively.

In Section 2, we give a general outline Bayesian optimization. In Section 3 we discuss specifics of applying it to beamline alignment, and Section 4 discusses the implementation of these specifics as a set of software tools. Section 5 tests the software on a real beamline's digital twin. Section 6 remarks on the next steps and goals for our beamline optimization tools.

2. BAYESIAN OPTIMIZATION IN GENERAL

Bayesian optimization iterates over the following steps in order to find the optimum for an expensive-to-sample black box function $f(x)$:

- Sample the data $y^* \sim f(x^*)$ at some inputs x^*.

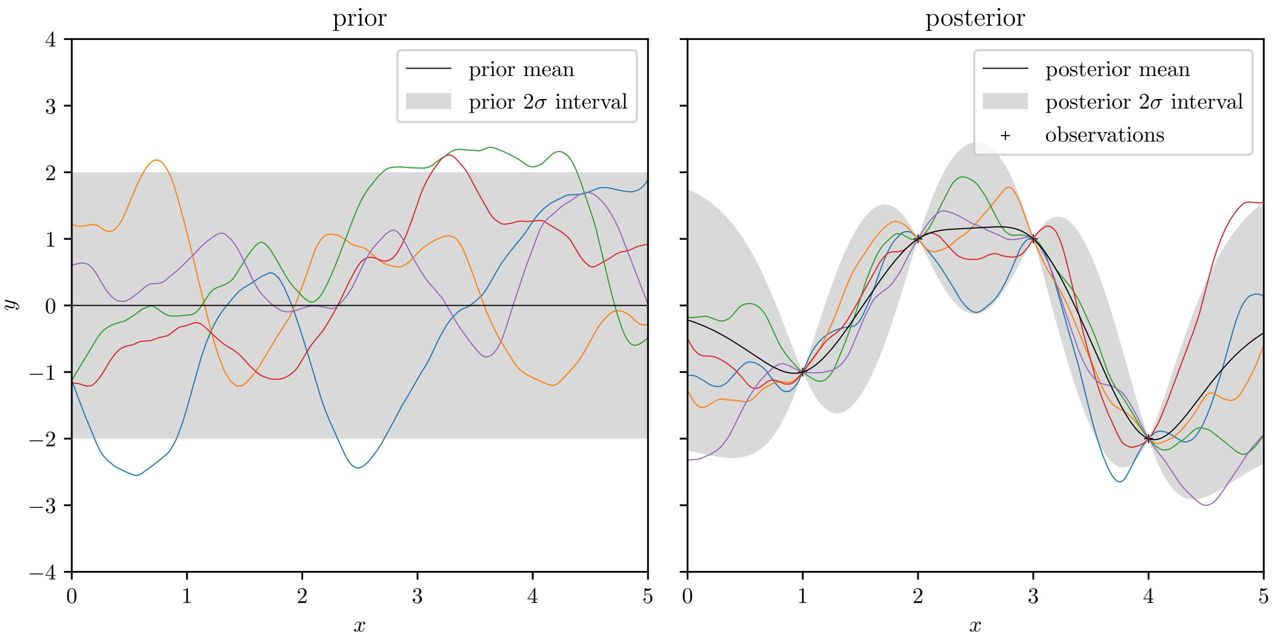

Figure 1: The prior and posterior distributions for a process.

- Assign a *prior* to the function $f(x)$ in general.
- Use Bayesian inference to compute a *posterior* estimate about the function at different inputs from the data and the prior.
- Use an acquisition function to predict the next most desirable point(s) to sample.
- Repeat until optimized.

2.1 Gaussian processes

Bayesian optimization necessitates a method of computing priors and posteriors over inputs; this is typically done with a Gaussian process. A Gaussian process is a stochastic process (random variables organized in space) where every collection of variables is a multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$. As an affine transform of a Gaussian process is also a Gaussian process, we can assume our processes have zero mean and unit variance for notational simplicity without loss of generality. In this case, the Gaussian process is defined entirely by the covariance. In practice, we do not know the covariance *a priori*, but given some samples we can find an approximation that fits the observed data. We can model it with a function called the kernel: for two points x_i and x_j, we model the covariance of the associated process variables as the output of a function $k(x_i, x_j, \theta)$, where θ represents a set of hyperparameters to be optimized.

2.2 Optimizing hyperparameters

Using our kernel function $\mathbf{K}(\theta)_{ij} = k(x_i, x_j, \theta)$, we fit the hyperparameters to the data $\mathbf{y}$ by maximizing the marginal likelihood of the process

$$p\big(\mathbf{y} \mid \mathbf{x}, \theta\big) = \frac{1}{(2\pi)^n \det \mathbf{K}(\theta)} \exp\Big(-\frac{1}{2}\mathbf{y}^\top \mathbf{K}(\theta)^{-1}\mathbf{y}\Big) \tag{1}$$

and adopting the optimal hyperparameters

$$\hat{\theta} = \arg\max_\theta p\big(\mathbf{y} \mid \mathbf{x}, \theta\big). \tag{2}$$

2.3 Computing posteriors

Given our optimal hyperparameters, constructing the respective posterior is straightfoward; the posterior of the value of the function at inputs $\mathbf{x}^*$ given measured values at $\mathbf{x}$ is then the MVN distribution $f(\mathbf{x}^*) \sim \mathcal{N}(\mathbf{A}f(\mathbf{x}), \mathbf{B})$, where

$$\mathbf{A} = k(\mathbf{x}^*, \mathbf{x}, \theta) k(\mathbf{x}, \mathbf{x}, \theta)^{-1} \tag{3}$$

$$\mathbf{B} = k(\mathbf{x}^*, \mathbf{x}^*, \theta) - \mathbf{A}k(\mathbf{x}, \mathbf{x}^*, \theta) \tag{4}$$

Figure 1 shows a set of functions randomly sampled from the prior distribution corresponding to a Gaussian process with kernel $k(x_i, x_j) = \frac{|x_i - x_j|^{5/2}}{2^{3/2}\Gamma(5/2)} K_{5/2}(\mid x_i - x_j \mid)$ (called the Matérn kernel; here $K_{5/2}$ is the modified Bessel function of the second kind of order 5/2).

2.4 Acquisition functions

Using our posterior, we can optimize an acquisition function over the inputs to figure out the most desirable next point to sample. One acquisition function is the *upper confidence bounds*

$$\mathrm{UCB}(x) = \mu(x) + z\sigma(x) \tag{5}$$

which values both a large expected value, as well as a large uncertainty. Another is *expected improvement*

$$\mathrm{EI}(x) = \mathbb{E}\big[\max(f(x) - f^\star, 0)\big] = \int_{f^*}^{\infty} y\phi(y)dy = \sigma(x)\big(\phi(z) + z\Phi(z)\big) \tag{6}$$

where $z = (\mu(x) - f^*)/\sigma(x)$, and $\phi(y)$ and $\Phi(y)$ are the PDF and CDF of the standard normal distribution.[1] We can also define the *probability of improvement*

$$\mathrm{PI}(x) = \mathbb{P}\big[f(x) > f^\star\big] = \int_{f^*}^{\infty} \phi(y)dy = \Phi(z) \tag{7}$$

Figure 2 shows these three acquisition function computed over the posterior in Figure 1. When optimized, different acquisition function will prefer different inputs, and may perform better or worse in the context of converging rapidly to an optimum.

3. BAYESIAN OPTIMIZATION FOR BEAMLINES

3.1 Latent dimensions

Beamlines have highly correlated inputs, which we may not know precisely. To model these correlations, we use the kernel

$$k(x_i, x_j, \theta) = a^2 f\big(\ ||\ \mathbf{D}\exp(\mathbf{S})\big(x_i - x_j\big)\ ||\ \big) \tag{8}$$

where $f(r)$ is a positive-definite function, $\mathbf{D}$ is a diagonal matrix with positive entries, $\exp(\cdot)$ is the matrix exponential, and $\mathbf{S}$ is a skew-symmetric matrix; the entries of these matrices and the total variance a^2 are determined by the hyperparameters θ.

Figure 3 shows an example of two-dimensional kernels corresponding to different latent dimensions. Figure 4 shows the latent dimensions for a two-dimensional parameter space at the TES beamline corresponding to the translation and rotation of a toroidal mirror.

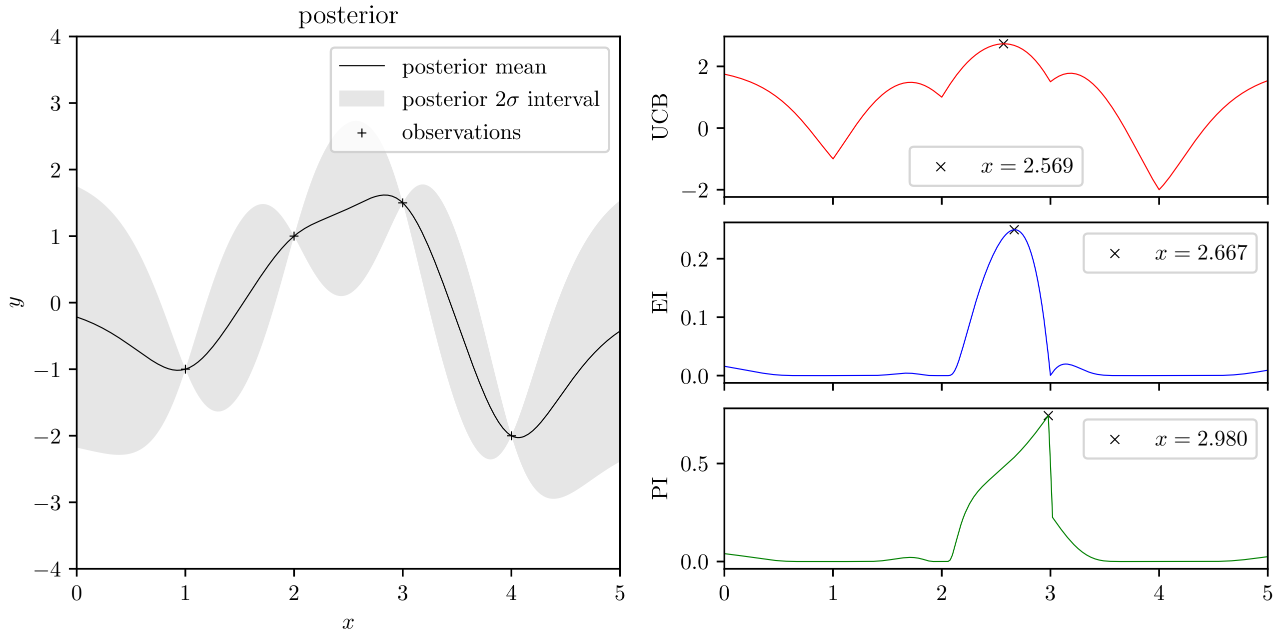

Figure 2: *Left*: a posterior condition on several observations. *Right*: Three different acquisition functions and their maxima: upper confidence bound (UCB), expected improvement (EI), probability of improvement (PI)

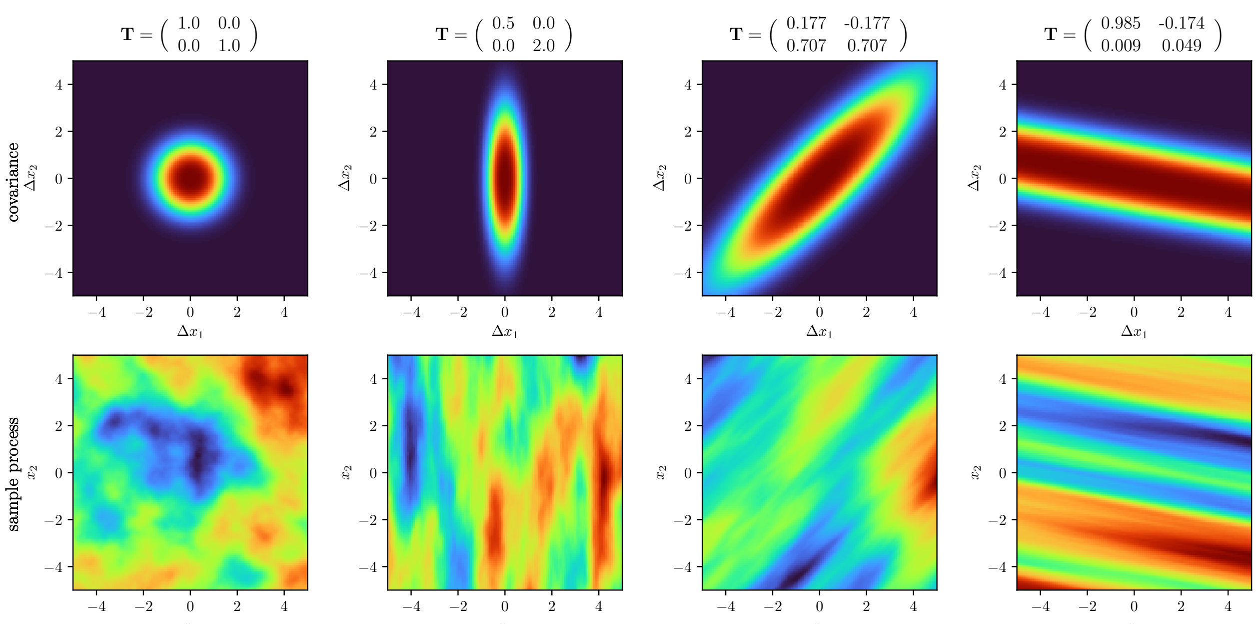

Figure 3: Different kernels defined by different latent dimensions; here $\mathbf{T} = \mathbf{D}\exp(\mathbf{S})$ is the matrix which transforms the inputs into the latent space.

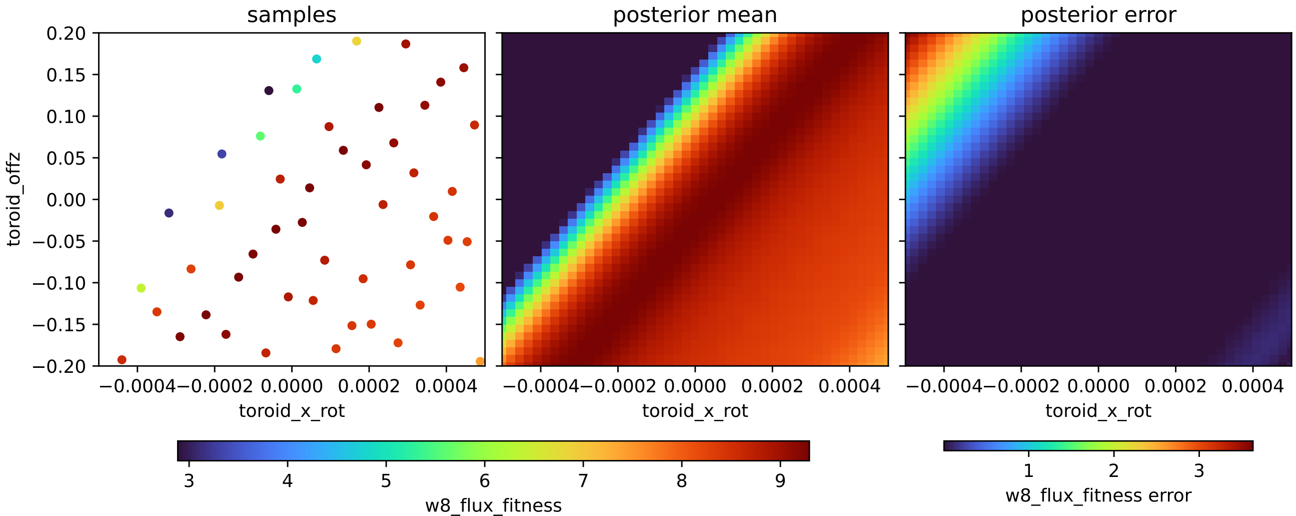

Figure 4: The flux through the SSA as a function of a rotation (x-axis) and translation (y-axis) of the toroidal mirror at the TES beamline; the two dimensions are highly coupled.

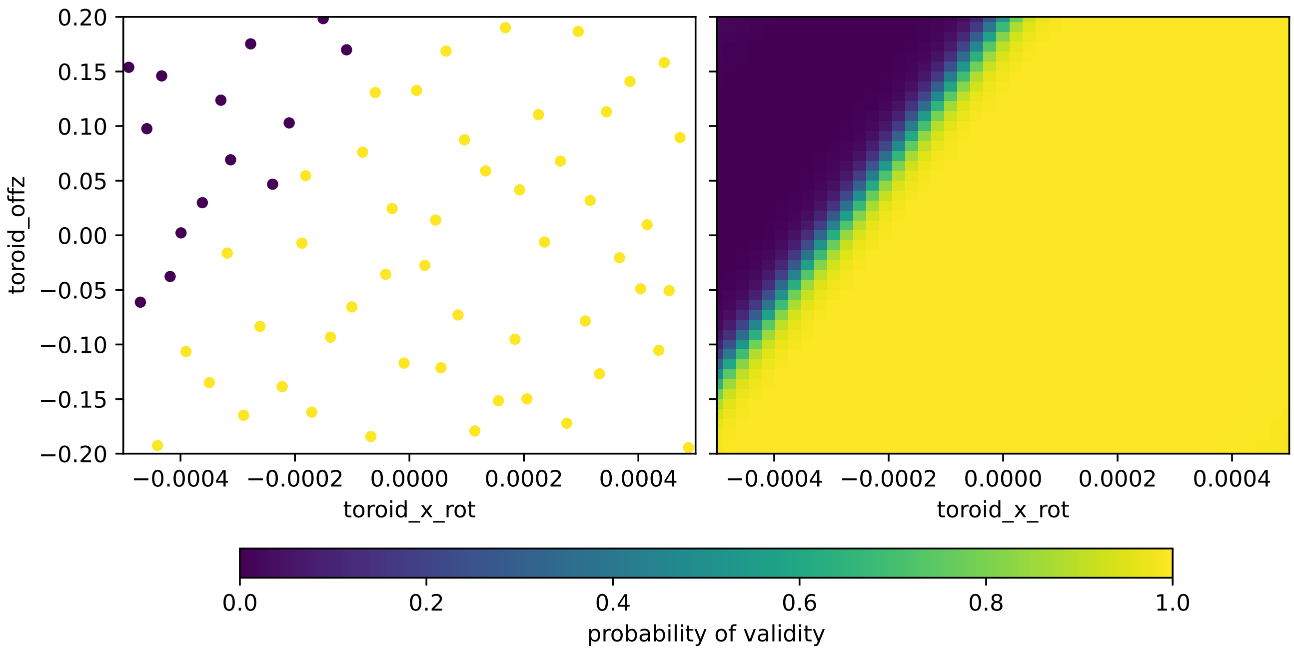

Figure 5: *Left*: Observations marked as either valid (yellow) or invalid (blue). *Right*: A Dirichlet-based probability model predicting the validity of future observations.

3.2 Validity modeling

In a practical sense, a sample can be invalid and thus disqualified for several reasons (it could be that there is not enough flux for a reliable estimate of the beam size, that the beam goes off the detector, there is a glitch in the detector, etc.). We construct a model to try to classify and predict observational validity as a function of inputs. In particular, we condition Gaussian processes to model the concentration parameters $\alpha_i(x)$ of a Dirichlet distribution.[2] The conjugate of the resulting distribution tells us the probability of a beam being valid. Figure 5 shows the validity model for a two-dimensional parameter space of the toroidal mirror, where some points are marked as invalid due to having insufficient flux.

3.3 Composite optimization

Bayesian optimization algorithms are capable of multi-tasking, which is to consider and optimize many aspects at the same time. While we consider only a scalar fitness in this paper, we still use separate Gaussian process models to predict the posterior of different aspects of the beamline: in the case of the TES beamline, this is the total flux, the vertical beam spread, and the horizontal beam spread. Even though we combine these into a single fitness, there is still a benefit in modeling them separately, as some aspects are independent to some motors leading to an optimization problem with a smaller effective dimensionality. This is typically referred to as *composite optimization.*

3.4 Passive degrees of freedom

Some parameters affect the beam, but cannot be reliably or directly manipulated (e.g. a drifting ring current affecting the flux, the changing temperature and pressure of a sample, a general drifting of beam dynamics over time, etc.). We fit to these parameters, but we don't allow the agent to change them when it computes the acquisition function.

4. IMPLEMENTATION

We implement our approach in the bloptools package, which establishes a user-friendly interface between the the Bluesky acquisition framework and customized BoTorch and GPyTorch models suited to modeling beamlines.[31]

The Sirepo-Bluesky[4] package also gives us the ability to develop code which is the same for beamlines and their digital twins, which allows for easier deployment to real beamlines.

5. BENCHMARKING

Several different methods can be used to simulate a beamline and define its digital twin, includes wavefront propagation,[5] ray-tracing,[6] reduced statistical models,[7] machine-learning based methods.[8] While all of these can be used for benchmarking, we find that the best method is ray-tracing, which for a microscopy beamline like TES can accurately and efficiently recreates the beam size and flux. In Figure 6, we show the results of benchmarks of four different optimization problems of 2, 4, 4, and 8 dimensions respectively, utilizing different combinations of degrees of freedom of the Kirkpatrick-Baez mirror pair and the pre-secondary source aperture toroidal mirror of the TES beamline. For the toroidal mirror alignment we optimize the total flux of the beam and for the KB mirrors, we try to optimize the flux density of the beam (total flux divided by area). During each run, the agent samples quasi-randomly to get a workable estimate of the hyperparameters; for higher-dimensional problems, we increase the number of quasi-random samples. For each of these optimization problems, the agent is able to consistently reach the global optimum even from a substantial misalignment.

6. DISCUSSION OF FURTHER DEVELOPMENTS

Further development will involve both adding new features (e.g. the ability to dynamically select acquisition functions during optimization and the ability to map the Pareto front) and improving the implementation of existing features (e.g. the efficiency and robustness of hyperparameter optimization). We will also apply the agent to more diverse alignment problems requiring more sophisticated feedback systems, such as optimizing the focus of nanoimaging beamlines (requiring knife-edge scans) or the resolution of a spectrometer (requiring energy scans.)

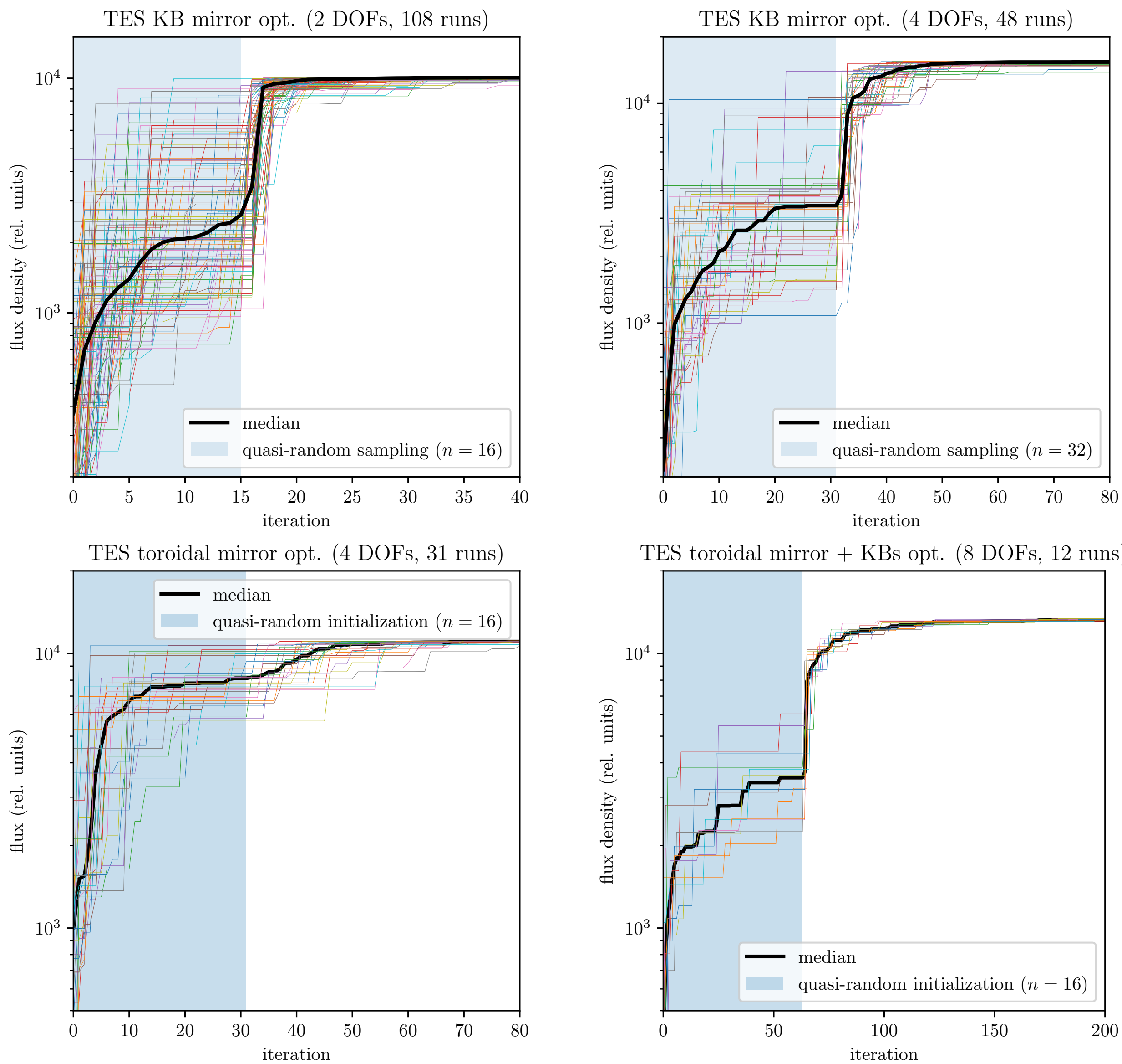

Figure 6: Four optimization problems for the TES beamline, simulating using a SHADOW (ray-tracing) backend. *Top left:* KB mirror rotation (two dimensions) *Top right:* KB mirror rotation + translation (four dimensions) *Bottom left:* pre-SSA toroidal mirror (four dimensions) *Bottom right:* KBs + toroid (eight dimensions)

ACKNOWLEDGMENTS

The work was supported in part by BNL's LDRD-22-031 project titled "Simulation-aided Instrument Optimization using Artificial Intelligence and Machine Learning Methods" and the DOE SBIR project (Award No. DE-SC00020593) titled "X-ray Beamline Control with an Online Model for Automated Tuning and Reconfiguration". Additionally, the project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI) during the summer of 2023.

REFERENCES

[1] Balandat, M., Karrer, B., Jiang, D., Daulton, S., Letham, B., Wilson, A. G., and Bakshy, E., "Botorch: A framework for efficient monte-carlo bayesian optimization," *Advances in neural information processing systems* **33**, 21524–21538 (2020).

[2] Milios, D., Camoriano, R., Michiardi, P., Rosasco, L., and Filippone, M., "Dirichlet-based gaussian processes for large-scale calibrated classification," *Advances in Neural Information Processing Systems* **31** (2018).

[3] Allan, D., Caswell, T., Campbell, S., and Rakitin, M., "Bluesky's ahead: A multi-facility collaboration for an a la carte software project for data acquisition and management," *Synchrotron Radiation News* **32**(3), 19–22 (2019).

[4] Rakitin, M. S., Giles, A., Swartz, K., Lynch, J., Moeller, P., Nagler, R., Allan, D. B., Caswell, T. A., Wiegart, L., Chubar, O., et al., "Introduction of the sirepo-bluesky interface and its application to the optimization problems," in [*Advances in Computational Methods for X-Ray Optics V*], **11493**, 209–226, SPIE (2020).

[5] Chubar, O., Fluerasu, A., Berman, L., Kaznatcheev, K., and Wiegart, L., "Wavefront propagation simulations for beamlines and experiments with" synchrotron radiation workshop"," in [*Journal of Physics: Conference Series*], **425**(16), 162001, IOP Publishing (2013).

[6] Sanchez del Rio, M., Canestrari, N., Jiang, F., and Cerrina, F., "Shadow3: a new version of the synchrotron x-ray optics modelling package," *Journal of Synchrotron Radiation* **18**(5), 708–716 (2011).

[7] Nash, B., Abell, D. T., Keilman, M., Du, Y., Giles, A., Lynch, J., Morris, T., Rakitin, M., Walter, A. L., and Goldring, N., "Online models for x-ray beamlines," *Bulletin of the American Physical Society* (2022).

[8] Nash, B., Abell, D., Nagler, R., Moeller, P., Keilman, M., Pogorelov, I., Goldring, N., Rakitin, M., Lynch, J., Giles, A., et al., "Combining diagnostics, modeling, and control systems for automated alignment of the tes beamline," in [*Journal of Physics: Conference Series*], **2380**(1), 012103, IOP Publishing (2022).