Physics makes the difference: Bayesian optimization and active learning via augmented Gaussian process

IOP Publishing Mach. Learn.: Sci. Technol. 3 (2022) 015003 https://doi.org/10.1088/2632-2153/ac4baa

MACHINE LEARNING
Science and Technology

OPEN ACCESS

RECEIVED
29 October 2021

REVISED
12 January 2022

ACCEPTED FOR PUBLICATION
14 January 2022

PUBLISHED
7 February 2022

PAPER

Physics makes the difference: Bayesian optimization and active learning via augmented Gaussian process

Maxim A Ziatdinov[1,2,*], **Ayana Ghosh**[1,2] **and Sergei V Kalinin**[1]

[1] Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America
[2] Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America
* Author to whom any correspondence should be addressed.

E-mail: ziatdinovma@ornl.gov

Keywords: Bayesian optimization, active learning, physics-informed, Gaussian process

Supplementary material for this article is available online

Abstract

Both experimental and computational methods for the exploration of structure, functionality, and properties of materials often necessitate the search across broad parameter spaces to discover optimal experimental conditions and regions of interest in the image space or parameter space of computational models. The direct grid search of the parameter space tends to be extremely time-consuming, leading to the development of strategies balancing exploration of unknown parameter spaces and exploitation towards required performance metrics. However, classical Bayesian optimization (BO) strategies based on the Gaussian process (GP) do not readily allow for the incorporation of the known physical behaviors or past knowledge. Here we explore a hybrid optimization/exploration algorithm created by augmenting the standard GP with a structured probabilistic model of the expected system's behavior. This approach balances the flexibility of the non-parametric GP approach with a rigid structure of physical knowledge encoded into the parametric model. The fully Bayesian treatment of the latter allows additional control over the optimization via the selection of priors for the model parameters. The method is demonstrated for a noisy version of a standard univariate test function used to evaluate optimization algorithms and further extended to physical lattice models. This methodology is expected to be universally suitable for injecting prior knowledge in the form of physical models and past data in the BO framework.

Modern scientific research is based on the synergy of experimental and theoretical methods for the exploration and prediction of states of matter. Computational methods [1–6] ranging from lattice models and classical molecular dynamics to advanced density functional and quantum Monte Carlo techniques provide wealth of information on thermodynamic, dynamic, and electronic properties of materials. Electron and scanning probe microscopies [7, 8] and neutron and x-ray scattering methods [9, 10] provide huge volumes of structural and functional data. Finally, combinatorial, and automated synthesis and high-throughput characterization [11–13] now allow rapid exploration of multidimensional compositional spaces for complex functional materials.

Common for the experimental and computational methods alike is the need for search across broad parameter spaces. In the theory domain, materials prediction necessitates exploration of large chemical spaces of molecules encoded via SMILES [14], SELFIES [15], or other descriptors, or compositional spaces of complex materials. Very similar challenges exist in the experimental domains ranging from automated synthesis to imaging and spectroscopy. For example, tuning of a physical instrument and searching for regions of interest are both searches in parameter space, to find optimal experimental conditions and physical behavior of interest within a material, respectively. However, the direct grid search of the parameter space tends to be extremely time-consuming, leading to the development of strategies balancing exploration of unknown parameter spaces and exploitation towards required performance metrics. The classical approach for this is the Gaussian process (GP) based Bayesian optimization (BO) [16–18]. This method

balances the learning of the correlations in the parameter space with the exploration-exploitation balancing the uncertainty and maximization of a certain target function combined in a single acquisition function.

However, classical GP-based BO strategies do not readily allow for the incorporation of the known physical behaviors or past knowledge. In other words, the GP-based BO methods create a fully non-parametric model of the system based on the prior observations, with the 'learned' physics of the system reflected in the kernel function. As such, the latter is roughly equivalent to the parameterized via a certain functional form correlation function. Hence, this approach is limited if behaviors of interest change differently in different regions of parameter space, e.g. have sharp boundaries as in the case of phase transitions in physical systems. While a large number of GP/BO models allowing for the greater flexibility of kernel function, warping of the parameter space, and consideration of input-dependent noise are continuously being developed [19–25], the overall limitations persist.

Comparatively, physics applications are typically associated with a large volume of domain-specific knowledge. For example, in exploring the lattice models in statistical physics, the asymptotic behaviors of the relevant parameters in the vicinity of phase transition and in the large/small temperature limits are generally known, but not the transition temperature or universality classes. Similarly, in materials discovery, the evolution of properties of interest such as photoluminescent peak position or stability are strongly correlated with the composition-dependent band gap or thermodynamic stability. In ferroelectric materials, the thickness dependence of the coercive field is defined by the phenomenological Kay–Dunn laws. Virtually in all areas of physical sciences, past knowledge is encoded in the form of symbolic models. In some cases, the models can be derived from the underpinning microscopic models or follow from the conservation laws; in many others, they represent the phenomenological behavior of the system. Nonetheless, this knowledge is generally not incorporated in the BO framework.

Here we explore a hybrid optimization/exploration algorithm created by augmenting the standard GP with a structured probabilistic model of the expected system's behavior. This approach balances the flexibility of the non-parametric GP approach with a rigid structure of physical knowledge encoded into the parametric model. The method is demonstrated for a noisy version of a standard univariate test function used to evaluate optimization algorithms and further extended to physical lattice models and is expected to be universally suitable for injecting prior knowledge in the form of physical models and past data into the BO framework.

1. Gaussian process (GP)

Given the dataset $D = \{x_i, y_i\}_{i=1,\ldots,N}$, such that x_i are input features and $f(x_i) = y_i$ are output targets (referred to as 'training' data in the machine learning community), the GP model can be defined as:

$$\mathbf{f} \sim \text{MVNormal}\left(\mathbf{m}, \mathbf{K}\right) \tag{1a}$$

$$K_{ij} = \sigma \exp(0.5 \text{ dist}\left(x_i, x_j\right)/l^2) \tag{1b}$$

$$\sigma \sim \text{LogNormal}\left(0, s_1\right) \tag{1c}$$

$$l \sim \text{LogNormal}\left(0, s_2\right) \tag{1d}$$

where MVNormal stands for multivariate normal distribution, $\mathbf{m}$ is a mean function which is usually chosen to be constant, and $\mathbf{K}$ is a covariance function (kernel), for which we chose a standard radial basis function with output scale σ and length scale l. We also assume there is normally distributed observation noise, $\varepsilon_n \sim \text{Normal}\left(0, \sigma_n^2\right)$, such that $y_n = f(x_n) + \varepsilon_n$. Implementation wise, this noise is absorbed into the computation of covariance function K. To get posterior samples for the GP model parameters, we use a Hamiltonian Monte Carlo (HMC) [26] algorithm. After inferring the parameters of GP model, we can use it to obtain the expected function values and associated uncertainty on new, previously not seen by the model, data (the so-called 'test' data). Specifically, we sample from the multivariate normal posterior over the model outputs at the provided points X_*:

$$\mathbf{f}_*^i \sim \text{MV Normal}\left(\mu_{\boldsymbol{\theta}^i}^{\text{post}}, \Sigma_{\boldsymbol{\theta}^i}^{\text{post}}\right) \tag{2a}$$

$$\mu_{\boldsymbol{\theta}^i}^{\text{post}} = \mathbf{m}\left(X_*\right) + \mathbf{K}(X_*, X|\boldsymbol{\theta}^i)\mathbf{K}\left(X, X|\boldsymbol{\theta}^i\right)^{-1}\left(\mathbf{y} - \mathbf{m}\left(X\right)\right) \tag{2b}$$

$$\Sigma^{\text{post}}_{\boldsymbol{\theta}^i} = \mathbf{K}\left(X_*, X_*|\boldsymbol{\theta}^i\right) - \mathbf{K}(X_*, X|\boldsymbol{\theta}^i)\mathbf{K}\left(X, X|\boldsymbol{\theta}^i\right)^{-1}\mathbf{K}(X, X_*|\boldsymbol{\theta}^i) \tag{2c}$$

where $\boldsymbol{\theta}^i = \left[\sigma^i, l^i\right]$ is a single HMC posterior sample containing kernel hyperparameters.

2. Bayesian optimization (BO)

The GP posterior in equation (2) can be used for selecting the next measurement/evaluation point in the active learning or BO setting. This is achieved by minimizing (or maximizing, depending on a particular problem) the so-called acquisition function. Perhaps the simplest (but nevertheless rather effective) version of the acquisition function is a Thompson sampler [27], which represents a single draw, $\mathbf{f}^i_*$, from posterior samples. Another acquisition function, known as the upper confidence bound (UCB), is derived from a linear combination of the predicted mean function value, $\overline{\mathbf{f}_*}$, and associated variance ('uncertainty'), $\mathbb{V}[\mathbf{f}_*]$, across the sampled predictions. Finally, in a pure exploratory regime [28], the selection of the next points is guided by the minimization of uncertainty $\mathbb{V}[\mathbf{f}_*]$.

Hence, a single BO step consists of (a) obtaining/updating the GP posterior over the model outputs at the provided points X_* given the sparse measurements D, (b) deriving the acquisition function, (c) selecting the next measurement point based on the minimum (or maximum) value of the acquisition function, (d) performing measurement in the 'suggested' point and adding the measured value to the dataset D. The goal of BO is usually to quickly identify regions where a particular behavior (a 'black-box function') is maximized or minimized. As such, it has been actively used in many domains, ranging from organic synthesis [29] to hyperparameter optimization in deep learning [30].

The limitation of the standard GP-based BO is that it does not readily allow for the incorporation of the known physical behaviors or past knowledge. This means that no domain-specific information or theoretical insights are factored in the selection of the next query point(s), potentially resulting in a sub-optimal sampling of the parameter spaces of physical systems. In other words, whereas GP-BO is an optimal exploration strategy in the absence of prior knowledge, this is not the case in most domain applications, including materials science, physics, and chemistry, where prior domain expertise and first-principles simulations are often the key factors guiding the exploration.

3. Augmented GP-BO and toy model

To overcome the limitations of the classical GP-BO we propose augmenting GP with a structured probabilistic model of the expected system's (physical) behavior. Specifically, we substitute a constant mean function $\mathbf{m}$ in equations (1) and (2) with a probabilistic model whose parameters are inferred together with the kernel parameters via the HMC. This probabilistic model reflects our prior knowledge about the system, but it does not have to be precise. In other words, the model can have a different functional form, if it captures the general trend in the data. In the language of machine learning, this can be interpreted as adding the so-called inductive bias [31] to our GP model. The equation (2) then becomes:

$$\mu^{\text{post}}_{\boldsymbol{\theta}^i\boldsymbol{\phi}^i} = \mathbf{m}\left(X_*|\boldsymbol{\phi}^i\right) + \mathbf{K}(X_*, X|\boldsymbol{\theta}^i)\mathbf{K}\left(X, X|\boldsymbol{\theta}^i\right)^{-1}(\mathbf{y} - \mathbf{m}(X|\boldsymbol{\phi}^i)) \tag{3}$$

where $\boldsymbol{\phi}^i$ is a single HMC posterior sample with the learned model parameters. Hence, the GP prediction now depends on kernel hyperparameters $\boldsymbol{\theta}$ as well as on the parameters $\boldsymbol{\phi}$ of a probabilistic model. This leads to the semi-parametric BO algorithm that makes 'physics-informed' decisions about which points in the parameter space to evaluate next. We note that our approach is different from studies that were using a deterministic GP mean function (either in analytical form [32] or as a neural network [33]) as it allows specifying a fully Bayesian probabilistic model by placing suitable priors on its parameters so that the uncertainty from the corresponding posterior samples is automatically taken into account when performing BO or active learning. At the same time, the flexibility of GP kernel allows us to use an imprecise and sometimes even 'incorrect' functional form, which would not be possible for a standalone probabilistic model (see supplementary materials available online at stacks.iop.org/MLST/3/015003/mmedia). We demonstrate our approach for the toy model based on the modified Forrester function [34] and for the 1D and 2D Ising models, but it is expected to be universally suitable for injecting prior knowledge in the form of physical/domain knowledge in the BO framework.

We start by illustrating our idea using a modified Forrester function [34], $f(x) = (5x-2)^2\sin(12x-4)$, which is commonly used to evaluate the optimization algorithms. We note that the end-goal is to have a robust optimization algorithm capable of working with noisy experimental or simulated data. Hence, we deliberately corrupted the observations with Gaussian noise. The noisy observations and the true function are shown in figure 1(a). The predictive means of the standard GP model trained on the entire data in

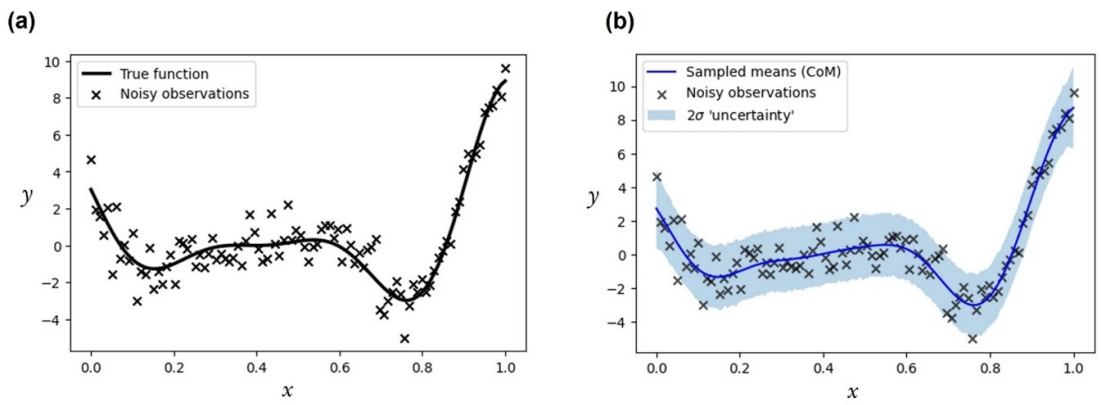

Figure 1. Forrester objective function. (a) Noisy observations of the Forrester objective function and the true function. (b) The posterior predictive distribution of the standard GP model trained on all the data points. The blue curve shows the center of mass (CoM) of GP predictive means (see equation (2)) for different HMC samples and the shaded area shows the 2σ 'uncertainty' in the sampled GP predictions.

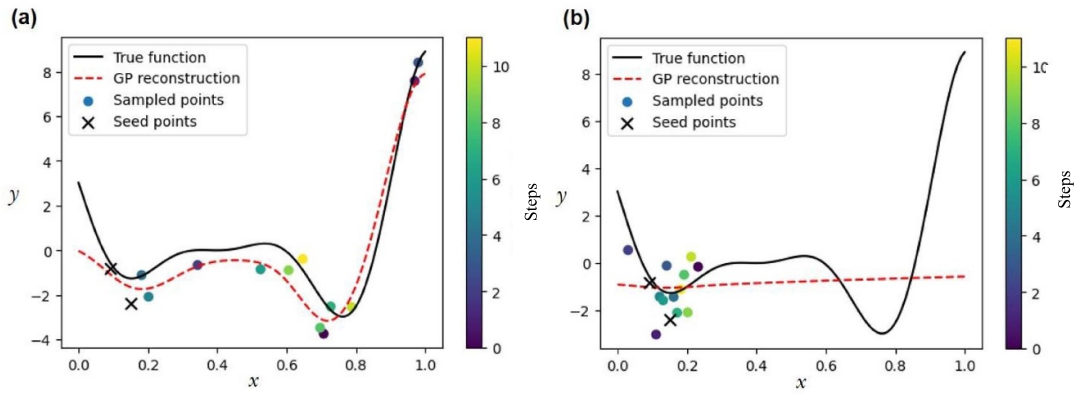

Figure 2. Results of the standard GP-BO for noisy observations of Forrester function. (a) Thompson sampling, (b) UCB. The 'GP reconstruction' shows the center of mass of the sampled predictive means from equation (2*b*).

figure 1(a) along with the associated uncertainty (variance across the sampled predictions) are shown in figure 1(b).

We now proceed to the BO whose goal, in this case, is to converge to the global minimum region (we cannot find the exact minimizer of the function since observations contain noise) using a minimal number of steps. First, we perform BO using the standard GP. We start with just two initial (seed) observations and run BO for 12 iterations using Thompson and UCB acquisition functions. The results are shown in figure 2. Clearly, the BO with Thompson sampler (figure 2(a)) did not achieve high fidelity in the approximation of the global minimum region whereas in the case of UCB (figure 2(b)) the optimization simply got stuck in a local minimum. The reasons for this somewhat disappointing performance are bad initialization (the seed points are near the local minimum) and relatively large observational noise. We note that for the clean data, the standard GP-BO identifies the global minimum with relative ease as shown in supplementary figure 1. However, the real-world measurements are always noisy. Furthermore, one usually does not have the luxury of restarting the experimental measurements (or even sometimes simulations) multiple times using different initializations.

Next, we show how injecting prior knowledge about our objective function/data can help mitigate the aforementioned issues. Specifically, we are going to augment GP with two different structured probabilistic models. Although both models do not describe the actual objective function, they do describe certain general trends in the data. Our first probabilistic model is defined as:

$$m = y_0 - \sum_{n=1}^{N} L_n \ (N = 2) \tag{4a}$$

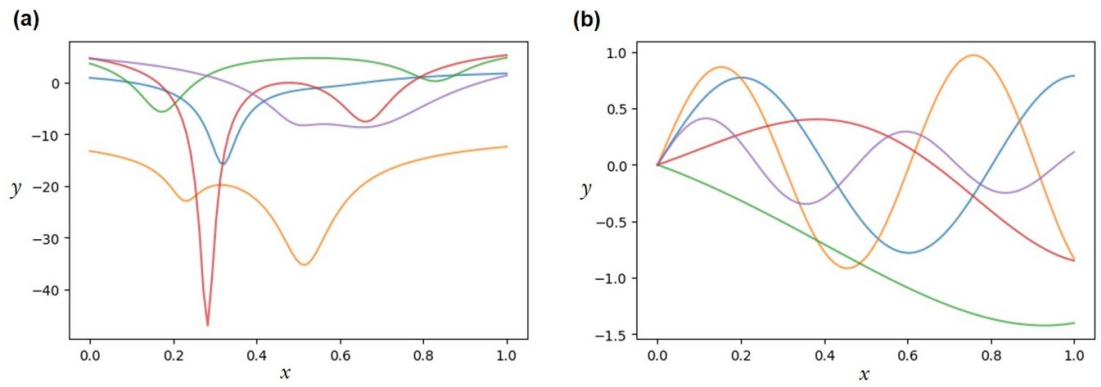

Figure 3. Prior predictive distributions for the two structured probabilistic models described in (a) equation (4) and (b) equation (5).

$$y_0 \sim \text{Uniform}\,(-10, 10) \tag{4b}$$

$$L_n \sim \frac{A_n}{\sqrt{(x - x_n^0)^2 + w_n^2}} \tag{4c}$$

$$A_n \sim \text{LogNormal}\,(0, 1) \tag{4d}$$

$$w_n \sim \text{HalfNormal}\,(.1) \tag{4e}$$

$$x_n^0 \sim \text{Uniform}\,(0, 1)\,. \tag{4f}$$

This model simply tells us that there are two minima in our data but does not assume to have any prior knowledge about their relative depth (i.e. which one of them is global) and width, nor it contains information about how far apart they are. The prior draws from the model are shown in figure 3(a). Next, we substitute the constant mean function in GP with our probabilistic model and run the BO the same way as we did earlier. The HMC is used at each step to obtain posterior samples for the model parameters ϕ and GP kernel hyperparameters $\boldsymbol{\theta}$ (see equation (3)).

The results for the Thomson sampler and UCB are shown in figures 4(a) and (b), respectively. Compared to the standard GP-BO (constant prior mean function), the GP-BO augmented with structured probabilistic model—hereafter referred to as *s*GP-BO—was able to converge to the global minima region within a relatively small number of steps. In addition, even though this was not a goal here, the *s*GP reconstruction of the objective function based on just $(12 + 2)$ points closely matches the shape of the true objective function. This shows that augmenting GP with a (imprecise) model of systems behavior that captures general trend in data is enough to significantly improve the efficiency of the optimization with noisy observations. We can also inspect the statistics of posterior samples for the mean GP function after the 12 steps of *s*GP-BO. The inferred center (x_0) of the deeper minimum is at 0.77 ± 0.03, which is close to the true minimum of the objective function (0.757). We note that the optimization efficiency can be further improved by using more informative priors in equations (4*b*)–(4*f*) with appropriately chosen acquisition functions. For example, the more informative priors about the minima locations can improve a search efficiency when using the acquisition functions that prioritize exploitation over exploration (see supplementary figure 2).

For our second probabilistic model, we have chosen a model that represents a 'wrong' function but still *partially* captures trends in the data such as the presence of more than one minimum. The second model is defined as:

$$m = Ae^{ax}\sin\,(bx) \tag{5a}$$

$$A \sim \text{LogNormal}\,(0, 1) \tag{5b}$$

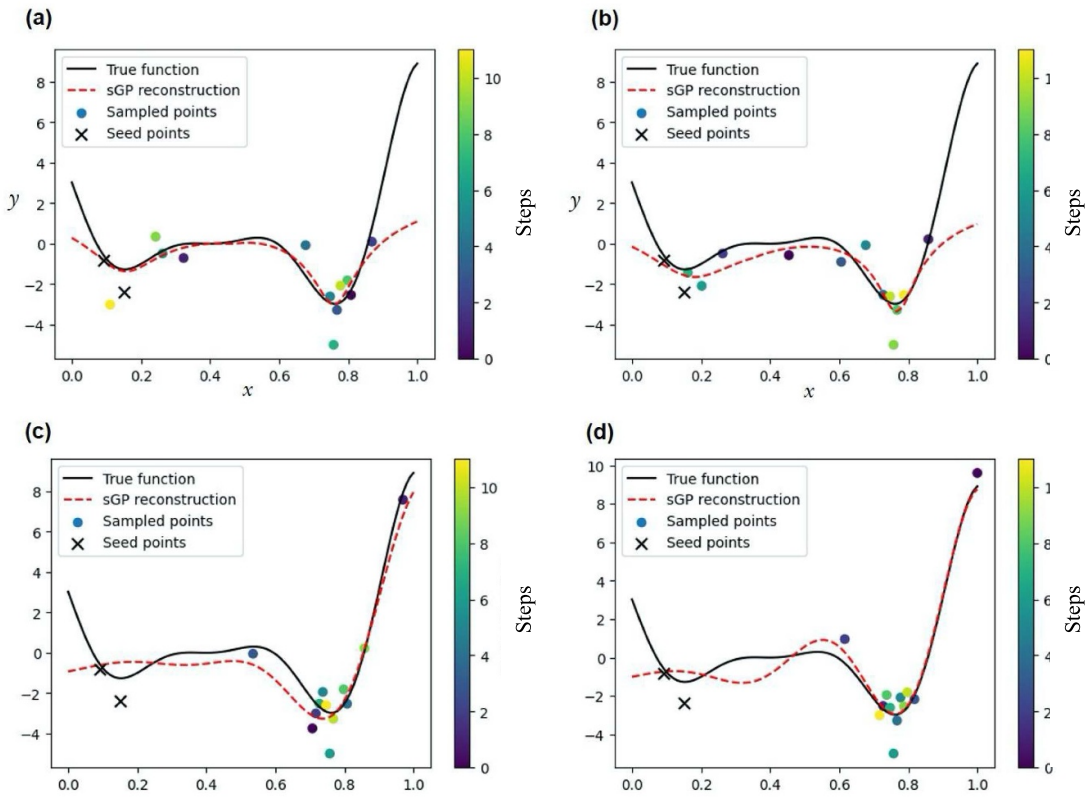

Figure 4. Results of the structured GP-BO for the noisy observations of Forrester function with a probabilistic model from equation (4) (a) and (b) and equation (5) (c) and (d) as GP's mean function. Figures (a) and (c) show results for Thompson sampler and (b) and (d) show results for UCB. The '*s*GP reconstruction' shows the center of mass of the sampled predictive means from equation (3).

$$a \sim \text{Normal}\,(1, 2) \tag{5c}$$

$$b \sim \text{Normal}\,(10, 5)\,. \tag{5d}$$

The prior draws for the second model are shown in figure 3(b) and the *s*GP-BO results for the Thomson sampler and UCB are shown in figures 4(c) and (d). One can see that while the quality of the overall reconstruction has suffered, we were able nevertheless to converge on the global minimum region. We note that this is only possible because we use a hybrid of the GP and probabilistic model and would not be possible if we use only the latter (see supplementary figure 3).

To better understand the difference between the standard GP-BO and the GP-BO augmented with a structured probabilistic model, we compared the acquisition functions for both algorithms at different steps of the optimization. In figures 5(a) and (b), we show the Thompson acquisition function at optimization steps 4 and 8 for the standard GP-BO (figure 5(a)) and the GP-BO augmented by the probabilistic model in equation (4) (figure 5(b)). For the standard GP-BO, the acquisition function is simply too noisy to guide the optimization algorithm. On the other hand, in the case of the *s*GP-BO, the acquisition function starts reflecting the general trends in the data already at the early steps of the optimization. In the case of the UCB acquisition function (figures 5(c) and (d)), it remains largely unchanged for the standard GP-BO (figure 5(c)) as the algorithm 'assumes' that it already found a global minimum. At the same time, for the *s*GP-BO (figure 5(d)), the injection of prior knowledge (that there is more than one peak) helps the algorithm to climb out of the local minimum and converge on the location of the global minimum.

To ensure the reproducibility of our results, we performed a systematic comparison of two GP-BO approaches for 20 different initializations of initial observations (using different pseudo-random number generator seeds). We introduced the efficiency (e) parameter that defines the number of points discovered by the algorithm in the ± 0.05 vicinity of the true minimum after ten steps divided by the total number of points lying in the same region for a given discretization of the X space. The ideal value is $e = 1$. The results shown

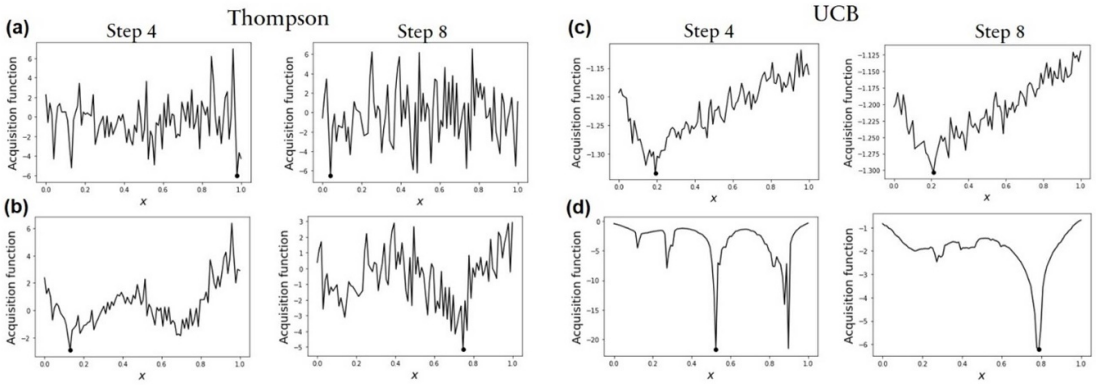

Figure 5. Acquisition functions at different optimization steps. (a) and (b) Thompson acquisition function for the (a) GP-BO and (b) *s*GP-BO. (c) and (d) UCB acquisition function for the (c) GP-BO and (d) *s*GP-BO.

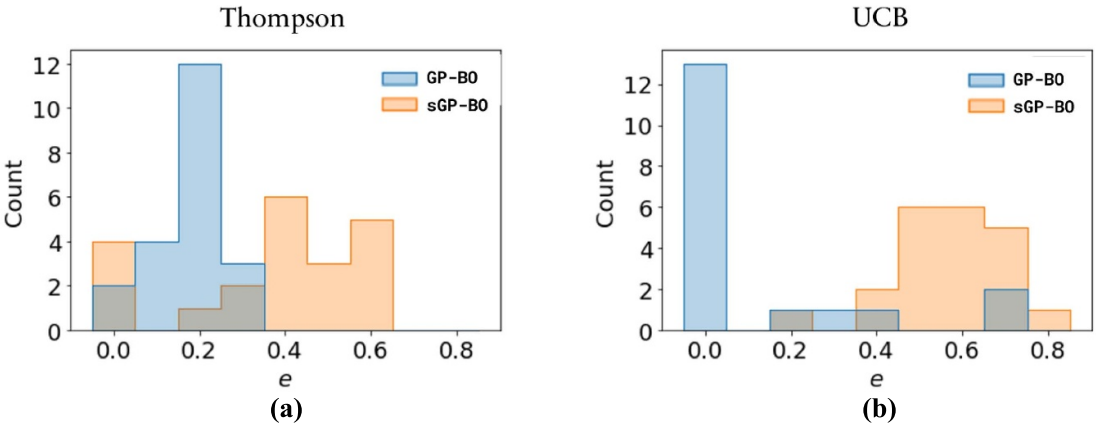

Figure 6. Efficiency of the two different BO approaches for (a) Thomson sampler and (b) UCB acquisition function. The efficiency (e) is calculated as a ratio of the points sampled by an optimization algorithm within ± 0.05 of the true minimum to the total number of points lying within the same range for a given discretization of the X space ($\Delta x = 0.01$). For the ten optimization steps used here, the highest possible efficiency is 1.

in figure 6 show that our approach clearly outperforms the standard GP-BO for both Thompson and UCB acquisition functions.

4. Physical systems

Having verified that our algorithm works on synthetic data, we next apply it to the 1D and 2D Ising models. The Hamiltonian of the Ising model with nearest-neighbor interactions can be written as $\mathcal{H} = -\sum_{\langle ij \rangle} J_{ij}\, \mathbf{S}_i \cdot \mathbf{S}_j$ where J represents the spin–spin interaction term and $\mathbf{S}_i$ are the individual spins on each of the lattice sites. The parameters of the Ising model simulations are the same as used by Kalinin *et al* [35]. Here, we are going to use the classical thermodynamic properties such as susceptibility and magnetization for our objective functions to compare the two GP-BO approaches.

We start by using BO to identify values of the J parameter that maximize the susceptibility in the 1D Ising model. For the *s*GP-BO, we use a simple Gaussian peak model of the form $Ae^{\left(-(J-J_0)^2/w^2\right)}$ with (log-)normal priors on its parameters as our GP mean function. The results for classical GP-BO and *s*GP-BO are shown in figures 7(a) and (b). One can see that while results are somewhat close (due to the relatively low noise and presence of only one peak) the incorporation of prior knowledge allows avoiding unphysical behaviors, such as at $J \approx 1.2$ in figure 7(a), when only a limited number of observations is available.

For the magnetization, which does not have a global minimum/maximum, we used a pure uncertainty-based exploration (kriging). Here, the goal is to explore and reconstruct its behavior in two

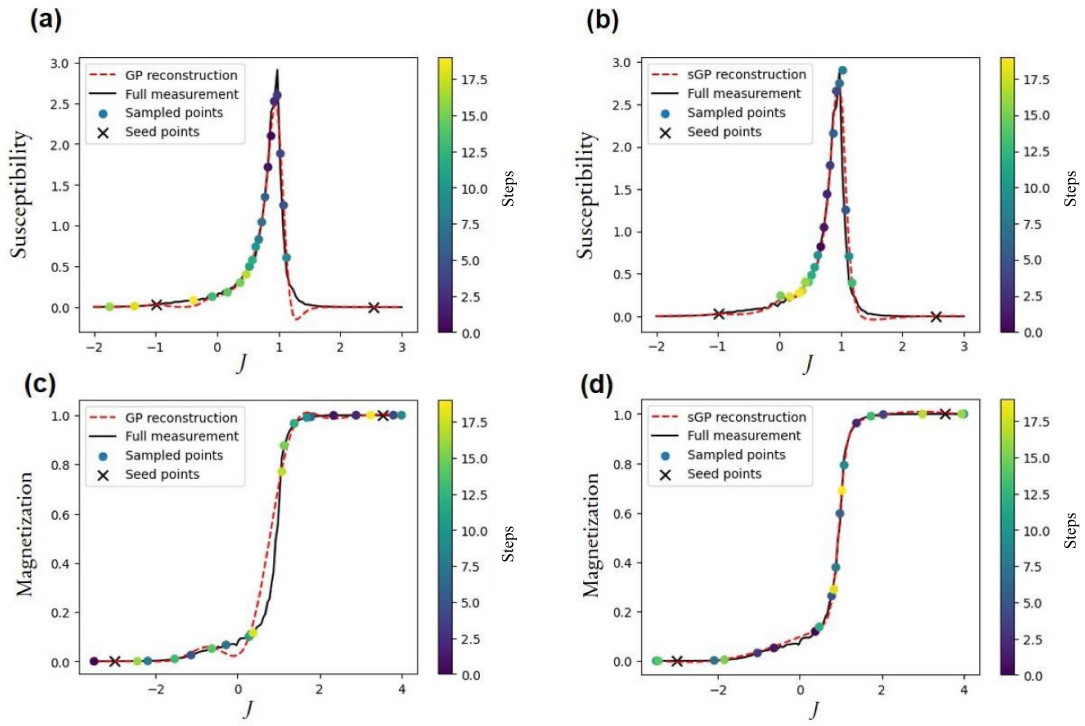

Figure 7. BO and active learning for 1D Ising model. (a) and (b) Optimization of the J parameter to maximize the susceptibility using UCB acquisition function with (a) standard GP-BO and (b) *s*GP-BO. (c) and (d) Kriging-based exploration of the magnetization across the phase transition with (c) standard GP kriging and (d) *s*GP kriging.

different phases and, most importantly, at the phase boundary. The results for the 1D case are shown in figures 7(c) and (d). For the *s*GP, we used a logistic function of the form $A/\tanh\left(\frac{J-J_0}{w}\right)$, with a uniform prior on J_0 and log-normal priors on A and w, as our structured probabilistic model. Clearly, the incorporation of prior knowledge about the system's behavior allowed for exploration of the transition region and better overall reconstruction (figure 7(d)). For the 2D case, we used a probabilistic model of the form $A/\tanh\left(\frac{f(J_1)+f(J_2)}{w}\right)$ where $f(J)$ is a third-degree polynomial with normal priors on its parameters. Note that while this is not the actual functional form of the phase transition in the system, it allows (with a proper choice of priors) to incorporate our knowledge that this phase transition can be in general described by a sigmoid-like structure but the actual phase boundary is a curve with a potentially complex shape. This should allow for a sufficient flexibility needed to map and reconstruct the phase boundary using a small number of measurements. The results for the standard GP kriging is shown in figure 8(a). Evidently, it could not localize the phase boundary and provided an overall poor reconstruction (figure 8(c)). In comparison, the *s*GP kriging was able to map the phase boundary (figure 8(b)) and provide a reconstruction of a sufficient quality (figure 8(d)) within the same number of exploration steps.

In summary, we have demonstrated how the augmentation of GP with a (fully Bayesian) probabilistic model of expected system's physical behavior allows for a more efficient optimization and active learning of system's properties. This was demonstrated for noisy observations of the standard objective function used to evaluate optimization algorithms and for the 1D and 2D Ising model in physics. The current approach can be extended to more complex physical systems where in the absence of any prior knowledge one can start with a standard GP and after sufficient number of observations (allowing to come up with a possible model of system's behavior) adds a structured probabilistic model, although a more sophisticated interplay between GP and *s*GP is possible. The systems with discontinuous phase transitions may in particular benefit from the current approach as they may not be amenable to a standard GP. Finally, in addition to incorporating our knowledge about possible functional forms, one can also incorporate a partial knowledge of causal links in a system potentially allowing for an even more efficient optimization and active learning.

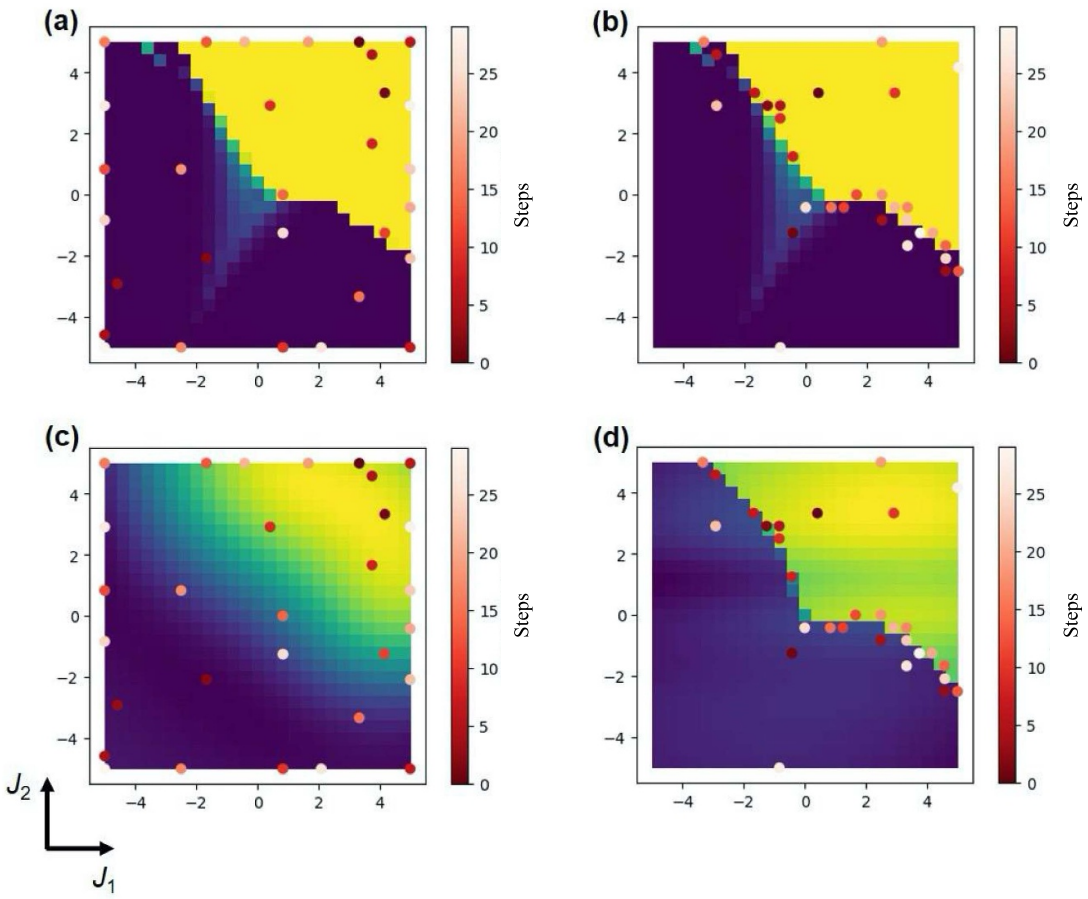

Figure 8. Active learning of magnetization behavior in 2D Ising model. (a) and (b) Points sampled by (a) standard GP kriging and (b) *s*GP kriging overlaid with the fully measured system ('ground truth'). (c) and (d) Data reconstructed from the points sampled using the (c) standard GP and (d) *s*GP.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

This work was supported (M Z) and performed at Oak Ridge National Laboratory's Center for Nanophase Materials Sciences (CNMS), a US Department of Energy, Office of Science User Facility, and supported by the Energy Frontier Research Centers program: CSSAS—The Center for the Science of Synthesis Across Scales—under Award No. DE-SC0019288, located at the University of Washington (S V K, A G).

Methods

The GP-BO and sGP-BO routines were implemented in JAX [36] using the iterative No-U-turn-sampler [37, 38] for HMC. The UCB acquisition function was defined as $\overline{\mathbf{f}_*} + k\sqrt{\mathbb{V}[\mathbf{f}_*]}$, with $k = -0.5$ for minimizing the Forrester objective function and $k = 0.5$ for maximizing the susceptibility in Ising model. The code is available at https://github.com/ziatdinovmax/AugmentedGaussianProcess.

ORCID iDs

Maxim A Ziatdinov https://orcid.org/0000-0003-2570-4592
Ayana Ghosh https://orcid.org/0000-0002-0432-3689
Sergei V Kalinin https://orcid.org/0000-0001-5354-6152

References

[1] Hafner J 2000 *Acta Mater.* **48** 71–92
[2] Ohno K, Esfarjani K and Kawazoe Y 2018 *Computational Materials Science: From Ab Initio to Monte Carlo Methods* (Berlin: Springer)
[3] Curtin W A and Miller R E 2003 *Modelling Simul. Mater. Sci. Eng.* **11** R33–68
[4] de Borst R 2008 *Comput. Mater. Sci.* **43** 1–15
[5] Hafner J, Wolverton C and Ceder G 2006 *MRS Bull.* **31** 659–68
[6] Hergert W, Ernst A and Dän M 2004 *Computational Materials Science: From Basic Principles to Material Properties* (Berlin: Springer Science & Business Media)
[7] Gerber C and Lang H P 2006 *Nat. Nanotechnol.* **1** 3–5
[8] Pennycook S J 2017 *Ultramicroscopy* **180** 22–33
[9] Keen D A and Goodwin A L 2015 *Nature* **521** 303–9
[10] Tokura Y 2006 *Rep. Prog. Phys.* **69** 797–851
[11] Epps R W, Bowen M S, Volk A A, Abdel-Latif K, Han S Y, Reyes K G, Amassian A and Abolhasani M 2020 *Adv. Mater.* **32** 2001626
[12] MacLeod B P *et al* 2020 *Sci. Adv.* **6** 8
[13] Higgins K, Valleti S M, Ziatdinov M, Kalinin S V and Ahmadi M 2020 *ACS Energy Lett.* **5** 3426–36
[14] Weininger D 1988 *J. Chem. Inf. Comput. Sci.* **28** 31–36
[15] Krenn M, Häse F, Nigam A, Friederich P and Aspuru-Guzik A 2020 *Mach. Learn.: Sci. Technol.* **1**
[16] Shahriari B, Swersky K, Wang Z, Adams R P and Freitas N D 2016 *Proc. IEEE* **104** 148–75
[17] Kushner H J 1962 *J. Math. Anal. Appl.* **5** 150–67
[18] Kushner H J 1964 *J. Basic Eng.* **86** 97–106
[19] Wilson A G and Nickisch H 2015 *Proc. 32nd Int. Conf. on Int. Conf. on Machine Learning* vol 37 *(Lille, France: JMLR.org)* pp 1775–84
[20] Wilson A G, Hu Z, Salakhutdinov R and Xing E P 2016 *Proc. 19th Int. Conf. Artificial Intelligence and Statistics, PMLR* **51** 370–8 (http://proceedings.mlr.press/v51/wilson16.html)
[21] Wilson A and Adams R 2013 *Proc. 30th Int. Conf. Machine Learning, PMLR* **28** 1067–75 (http://proceedings.mlr.press/v28/wilson13.html)
[22] Alexander W L *et al* 2020 (arXiv:2012.03826)
[23] Jasper Snoek K S, Zemel R S and Adams R P 2014 (arXiv:1402.0929)
[24] Griffiths R, Aldrick A, Garcia-Ortegon M and Lalchand V 2021 *Mach. Learn.: Sci. Technol.* **3**
[25] Makarova A, Usmanova I, Bogunovic I and Krause A 2021 *Adv. Neural Inf. Process. Syst.* **34**
[26] Duane S, Kennedy A D, Pendleton B J and Roweth D 1987 *Phys. Lett.* B **195** 216–22
[27] Thompson W R 1933 *Biometrika* **25** 285–94
[28] Williams C K I 1998 *Learning in Graphical Models* ed M I Jordan (Berlin: Springer) pp 599–621
[29] Korovina K, Xu S, Kandasamy K, Neiswanger W, Poczos B, Schneider J and Xing E 2020 *Proc. 23rd Int. Conf. Artificial Intelligence and Statistics, PMLR* **108** 3393–403 (http://proceedings.mlr.press/v108/korovina20a.html)
[30] Snoek J, Larochelle H and Adams R P 2012 *Proc. 25th Int. Conf. on Neural Information Processing Systems* vol 2 *(Lake Tahoe, NV: Curran Associates Inc.)* pp 2951–9
[31] Battaglia P W, Hamrick J B, Bapst V, Sanchez-Gonzalez A, Zambaldi V, Malinowski M, Tacchetti A, Raposo D, Santoro A and Faulkner R 2018 (arXiv:1806.01261)
[32] Noack M M *et al* 2021 *Nat. Rev. Phys.* **3** 685–97
[33] Fortuin V and Rätsch G 2019 (arXiv:1901.08098)
[34] Forrester A I J and Keane A J 2009 *Prog. Aerosp. Sci.* **45** 50–79
[35] Kalinin S V, Valleti M, Vasudevan R K and Ziatdinov M 2020 *J. Appl. Phys.* **128** 164304
[36] Bradbury J, Frostig R, Hawkins P, Johnson M, Leary C, Maclaurin D and Wanderman-Milne S 2021 GitHub (available at: http://github.com/google/jax) (Accessed 23 August 2021)
[37] Phan D, Pradhan N and Jankowiak M 2019 (arXiv:1912.11554)
[38] Homan M D and Gelman A 2014 *J. Mach. Learn. Res.* **15** 1593–623