Bayesian Optimization Priors for Efficient Variational Quantum Algorithms

Farshud Sorourifar,a,b,c Diana Chamaki,b,c Norm M. Tubman,b Joel Paulson,a David E. Bernal Neirab,c,d

a Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, OH, USA

b Quantum Artificial Intelligence Laboratory (QuAIL), NASA Ames Research Center, Moffett Field, CA, USA

c USRA Research Institute for Advanced Computer Science, Mountain View, CA, USA

d Davidson School of Chemical Engineering, Purdue University, West Lafayette, IN, USA

dbernaln@purdue.edu

Abstract

Quantum computing (QC) currently relies on hybrid quantum-classical methods known as Variational Quantum Algorithms (VQAs) to solve problems. Still, there are several challenges with VQAs on the classical computing side: they correspond to non-convex black-box optimization problems, the observations from quantum hardware are noisy, and quantum computing time is expensive. The first point is inherent to the problem structure, requiring the classical part of VQAs to be solved using global optimization strategies. However, there is a trade-off between cost and accuracy. QC returns a set of bitstrings, each referred to as a shot. The probabilistic nature of quantum computing demands many shots to measure its state accurately. Since QC time is charged per shot, reducing their number yields cheaper and less accurate observations. Recently, there has been an increasing interest in using Bayesian optimization (BO) methods to globally optimize quantum circuit parameters. This work proposes two modifications to the BO framework to provide a shot-efficient optimization strategy for VQAs. Specifically, we provide the means to place a prior on the periodicity of the rotation angles and a framework to set a topological prior using few-shot observations. We demonstrate the effectiveness of our proposed approach through an ablation study, showing that using both proposed features statistically outperforms a standard BO implementation within VQAs for computational chemistry simulations.

**Keywords**: Quantum computing, Bayesian optimization, Variational quantum algorithms

* 1. Introduction

Quantum computing (QC) has been the subject of growing interest in chemical engineering, owing to its potential to solve computationally challenging problems (Bernal et al., 2022). However, as a nascent technology, it has a limited number of processing units (qubits), and its quantum state rapidly decoheres. To circumvent current devices quick decoherence and a limited number of qubits, Variational Quantum Algorithms (VQAs) have been proposed, wherein a classical machine selects parameters for a quantum circuit representing a problem of interest, and a quantum machine evaluates it (Cerezo et al., 2020). This circuit encodes the evolution of prepared qubits through a series of parameterized operators or gates. The final qubit system state should follow a distribution representing the solution to a computational problem. Physically, the quantum gate parameters represent rotation angles, which modify the quantum state of the qubits system. The final states of the qubits are then measured by projecting them into a set of classical bits (bitstrings), with each bitstring measurement referred to as a shot. Thus, VQAs entail a feedback loop where a classical optimization algorithm selects the parameters for the quantum circuit based on a measure of the bitstrings.

Bayesian Optimization (BO) is a family of sample-efficient zeroth-order optimizers and has successfully solved various black-box problems, including VQAs. BO's sample efficiency results from using observations to construct a statistical surrogate model known as a Gaussian Process (GP), which generalizes a multivariate normal distribution to function space (i.e., a probability distribution over functions that fit the data). The GP's ability to quantify the model uncertainty allows us to systematically trade off exploring the parameter space and exploiting promising regions. BO has been gaining increased research interest for solving VQAs, spanning introductions to BO for VQAs (Tibaldi et al., 2023), benchmarking (Ciavarella and Chernyshev, 2022), and initialization strategies (Muller et al., 2022; Tamiya and Yamasaki, 2022). There is evidence that standard BO algorithms may benefit from lower-shot queries (Iannelli and Jansen, 2021).

This work provides background on BO algorithms and proposes two principled modifications to the vanilla BO algorithm (i.e., a standard GP built from a Matérn kernel with a zero-mean prior) to improve its efficiency in solving VQAs. Specifically, we propose encoding the parameter's 2 periodicity into the GP kernel function and a strategy for encoding a sample-based topological prior learned by fitting a second GP to low-shot measurements. Through an ablation study, we show that these modifications can significantly improve BO performance on VQAs.

* 1. Bayesian Optimization Preliminaries

First, let be the true quantum circuit value evaluated at a given vector of rotation angles . When we query the circuit, we generate a noisy observation,

|  |  |
| --- | --- |
|  | (1) |

where , is an independent and identically distributed Gaussian noise whose variance is given by the number of observations (shots) of the circuit. We let be the noise evaluated at the largest number of shots used to observe the circuit. BO starts with an initial data set consisting of initial observations, which can be used to build a statistical surrogate model. The model choice in this framework is general, only requiring that the model is statistical, as it can quantify epistemic uncertainty in terms of a covariance function. This function, coupled with the model's mean, is used to guide the sample points selection that balances exploration with exploitation through constructing and optimizing an acquisition function. Typically, this is done in-between observations and can be cheaply optimized using gradient methods. While many statistical models exist, the GPs are the most common choice due to their rigorous statistical quantification and non-parametric nature.

We assume that the circuit has a GP prior to the form where is the prior mean, and is the prior covariance function. There are many possible choices for the covariance function. Here, we define the Matérn kernel function,

|  |  |
| --- | --- |
|  | (2) |

where is a length-scale parameter, is the measurement noise variance, are the modified Bessel and gamma functions, and is a Euclidean distance function. The choice of is based on how smooth the function is believed to be, where larger values indicate a smoother function. Under the GP prior, the function evaluations are jointly Gaussian with mean , covariance , and . This implies the corresponding function value at any test point must be jointly Gaussian with . Due to the properties of jointly Gaussian random variables, we find that the posterior distribution of the objective given all available noisy observations , is Gaussian with the following mean and covariance,

|  |  |
| --- | --- |
| , | (3a)  (3b) |

With the mean and variance functions defined, we introduce the acquisition function. We can select a point that provides the most information by optimizing the acquisition function. For example, the lower confidence bound acquisition function,

|  |  |
| --- | --- |
|  | (4) |

balances exploration and exploitation by assigning an optimistic value to each candidate point. The exploitation term is represented by the mean, and the exploration term by the standard deviation scaled by a parameter . Small values of result in an exploitative strategy, and large values in a more exploratory strategy.

A graph of a graph

Description automatically generatedA graph of a graph

Description automatically generated

*Figure 1: Periodic (left) vs. non-periodic (right) BO kernel on a periodic function. The periodic kernel has noticeably higher accuracy for θ > 5 by recognizing the parameter space periodicity.*

* 1. Specialized Quantum Computing Priors

3.1 Periodic Parameter Prior

While BO requires box constraints on the parameters, the periodic boundary conditions are not typically known or enforced. We can, however, use periodic kernels to codify the periodicity of the circuit measurements and uncertainty in the GP model using a periodic kernel (MacKay, 1998). The periodic kernel can be defined as

|  |  |
| --- | --- |
|  | (5) |

where is the period, and is the length scale corresponding to the parameter. Typically, is an unknown parameter fitted to data as most problems have no known periodicity. Given that for VQAs, this parameter is fixed during the model fitting.

In Fig. 1, we provide an example of how knowledge of the periodic boundaries improves the surrogate model accuracy. The true function is represented as a black line, the data as red stars, and the mean and 95% CI as a blue line and clouds. Note the region , where uncertainty is substantially lower for the periodic kernel due to a measurement near . The non-periodic GP recommends the following sample near exploring the high uncertainty region, thus wasting samples.

3.2 Topological Prior

Although reducing the measured distribution of bitstrings to a single value, such as by taking the mean or conditional value at risk, allows traditional optimization strategies to be easily adapted to solve VQAs, they also result in a loss of information.

A graph of a graph showing a curve

Description automatically generated with medium confidenceA graph of a sine curve

Description automatically generated

*Figure 2: Effect of cheap and noisy (left) vs. expensive and precise data (right). Notice how samples with ×5 noise for ×1/5 cost may be more valuable than fewer, more accurate samples.*

Consider Fig. 2, where a larger volume (five times more data points) of noisier data (with five times the standard deviation of additive noise) may still result in a better surrogate model than fitting to a smaller, although more precise, dataset. This example can be likened to a quantum circuit, where the number of shots for each measurement is tunable.

To use the low-shot measurements as a topological prior, we propose using a low-shot residual (LSR) inspired by (Lu et al., 2021), which we adapt for VQAs by regulating the number of shots per query. For notation simplicity, we assume that , i.e., is sufficiently high such that noise can be ignored. Given that we aim to minimize a high-shot circuit output which uses shots per circuit call, we assume that a fraction of the total shot budget may be spent on observations of a low-shot circuit , which uses shots per circuit call. We assume that the high-shot model can be defined as

|  |  |
| --- | --- |
|  | (6) |

where function is a residual between the high and low-shot observations for a given . However, this assumption must be satisfied by construction. Let be the set of low-shot observations obtained from spending uniformly over , and be the mean and variance functions obtained from fitting a GP to . With the low-shot budget exhausted, becomes a deterministic function that approximates , the topological prior used to improve learning . With each measurement , we construct a data set to build a mean and variance function of the residual , while satisfying Eq. (6). Eq. (3a) becomes

|  |  |
| --- | --- |
|  | (7) |

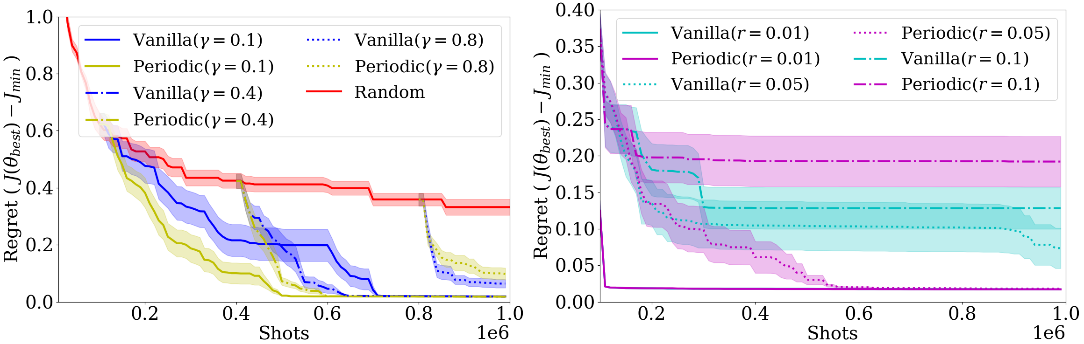
and can optimize the circuit by minimizing . Since the low-shot model can no longer reduce its variance, it serves no value in informing the exploration of the parameter space. Instead, we can formulate the Eq. (4) acquisition function as

|  |  |
| --- | --- |
|  | (8) |

where we use the mean function in Eq. (7) but only use the residual standard deviation for exploration. Note that the low-shot residual acquisition needs a larger exploration constant , given that the residual magnitudes are significantly smaller than the means.

* 1. Experimental Results

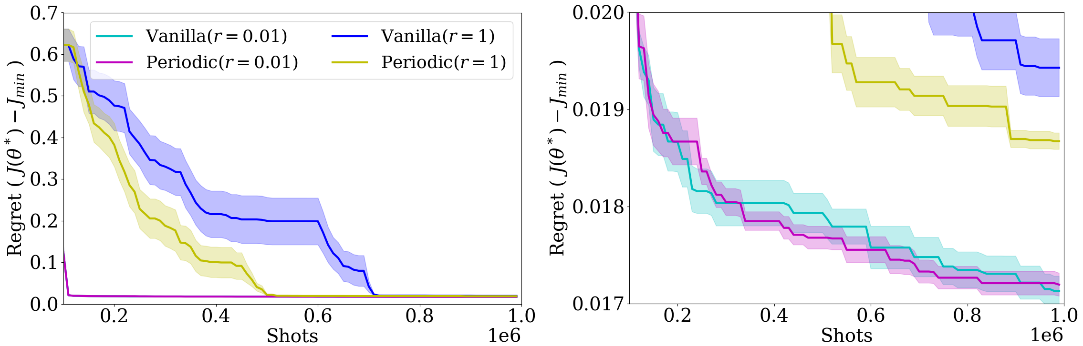
In Fig. 3.a, we compare the periodic kernel (yellow lines) versus a Matérn kernel (blue lines) using statistical convergence on a VQA simulation to find molecular hydrogen's ground state using as number of shots per evaluation , and as total budget. Additionally, we vary the fraction of the total shot budget for initialization , spent on random samples over the parameter space. Note that is a random sampling strategy (red line) with the worst performance, whereas smaller values of improve convergence for the periodic kernel. For the Matérn kernel, the trend isn't as apparent; convergence is faster with (blue-dashed line) than (blue-dotted line). Note that with a larger random sampling budget, the performance difference between the kernels diminishes since the highly explorative random sampling reduces the need for further exploration and, thus, the benefits derived from the periodic kernel. Finally, large allows the BO algorithms to focus on exploitation, evident by the faster convergence when switching from random to Bayesian sampling strategies.



*Figure 3: (left) Comparison of Matérn versus periodic kernel-based GPs in BO for multiple γ, (right) comparison of LSR-BO using a Matérn and periodic kernels for different ratios of r*

We present the results of the LSR-BO strategy in Fig. 3.b, where for the ratio of shots between the high and low shot circuits, we use , with . These results show a clear trend favoring residual models constructed from very few shots. In contrast, with small values, the algorithm finds near-optimal parameters on the high-shot circuit within the first few queries. However, larger values may diminish performance, as seen in the case with a vanilla kernel. While the results suggest that smaller could further improve performance, it would result in numerical issues. As shown in Eq. (3b), the GP fitting requires a matrix inversion, which results in cubic scaling and increased risk of singularity as the number of data points grows.

The ablation study in Fig. 4 compares the base strategy to one with a periodic kernel and two LSR strategies using Matérn and periodic kernels. The vanilla strategies can be considered a case of the LSR where . The vanilla periodic kernel (yellow) provides a clear advantage over the vanilla Matérn kernel (blue), and both LSR strategies provide an advantage over the vanilla strategies. Although the difference between the two kernels in the LSR is slight, the periodic kernel shows slightly lower regret and variance.



*Figure 4: Full ablation study results. The periodic kernel and the LSR independently improve the BO performance relative to a Vanilla strategy. (right) Zoomed-in regret performance.*

* 1. Conclusions

This work proposes two modifications to the standard Bayesian optimization implementation to improve shot-based efficiency when solving variational quantum algorithms. We show that a significant increase in performance can be achieved by encoding priors into the GP kernel function and surrogate model. The kernel prior endows the GP with knowledge of the parameter's 2 periodicity, which we find helpful in the limited circuit observation regime. At the same time, the topological prior provides a better starting model by utilizing large quantities of low-shot circuit measurements. These improvements highlight the possibility of using BO within VQAs.

**Acknowledgments**

We acknowledge the funding of the NASA ARMD Transformational Tools and Technology (TTT) Project. F.S. and D.C. participated in the NASA/USRA Feynman Quantum Academy internship program. D.B.N., F.S., and D.C. were supported by the NASA Academic Mission Services, Contract No. NNA16BD14C. The NSF Graduate Research Fellowship partially supported F.S.. J.P. acknowledges funding from the National Science Foundation, Award No. 2237616.

References

D. E. Bernal, A. Ajagekar, S. M. Harwood, S. T. Stober, D. Trenev, F. You, 2022. Perspectives of quantum computing for chemical engineering. AIChE Journal 68 (6), e17651.

M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, P. J. Coles, 2020. Variational quantum algorithms. Nature Reviews Physics 3, 625–644

A. N. Ciavarella, I. A. Chernyshev, Apr 2022. Preparation of the SU(3) lattice Yang-Mills vacuum with variational quantum methods. Phys. Rev. D 105, 074504.

G. Iannelli, K. Jansen, 2021. Noisy Bayesian optimization for variational quantum eigensolvers.

Q. Lu, L. D. Gonz ́alez, R. Kumar, V. M. Zavala, 2021. Bayesian optimization with reference models: A case study in MPC for HVAC central plants. Computers & Chemical Engineering 154, 107491.

D. J. C. MacKay, 1998. Introduction to Gaussian processes.

J. Muller, W. Lavrijsen, C. Iancu, W. de Jong, sep 2022. Accelerating Noisy VQE Optimization with Gaussian Processes. In: 2022 IEEE International Conference on Quantum Computing and Engineering (QCE). IEEE Computer Society, Los Alamitos, CA, USA, pp. 215–225

S. Tamiya, H. Yamasaki, Jul 2022. Stochastic gradient line Bayesian optimization for efficient noise-robust optimization of parameterized quantum circuits. npj Quantum Information 8 (1), 90

S. Tibaldi, D. Vodola, E. Tignone, E. Ercolessi, 2023. Bayesian Optimization for QAOA