



STIMULATE
European Joint Doctorates

BAYESIAN OPTIMIZATION OF VARIATIONAL QUANTUM EIGENSOLVERS



Giovanni Iannelli, Karl Jansen

Lattice 2021



Introduction

Testing in
one
dimension

Transverse
Ising model

Conclusions
and Outlooks

Appendix

INTRODUCTION



Variational Quantum Eigensolver (VQE) (I)



University
of Cyprus



- ▶ Hybrid quantum-classical algorithm to find minimum eigenvalue(s) and eigenvector(s) of a Hermitian operator with the variational method.
- ▶ Physics application: find the ground state E^{\min} of a Hamiltonian H performing the minimization:

$$\min_{\theta_{\alpha}} E(\theta_{\alpha}) \equiv \min_{\theta_{\alpha}} \langle \psi_{\theta_{\alpha}} | H | \psi_{\theta_{\alpha}} \rangle \geq \langle \psi_{\min} | H | \psi_{\min} \rangle = E_{\min}$$

- ▶ Given a Hamiltonian in the following form:

$$H = \sum_{i\gamma} h_{\gamma}^i \sigma_{\gamma}^i + \sum_{ij\gamma\delta} h_{\gamma\delta}^{ij} \sigma_{\gamma}^i \sigma_{\delta}^j + \dots \quad (i, j = 1, \dots, n \vee \gamma, \delta = x, y, z)$$

Introduction

Testing in
one
dimension

Transverse
Ising model

Conclusions
and Outlooks

Appendix



Variational Quantum Eigensolver (VQE) (II)



- ▶ A QPU can compute $E(\theta_\alpha) \equiv \langle \psi_{\theta_\alpha} | H | \psi_{\theta_\alpha} \rangle$ with exponential speedup when state dimensionality grows.
- ▶ The minimization $\min_{\theta_\alpha} E(\theta_\alpha)$ is performed with a classical optimizer (subject of this work).
- ▶ VQE decomposes the computation into many small quantum programs \Rightarrow good for nowadays noisy QPU.
- ▶ VQE can be extended to find the first k quantum states instead of just the ground state.

Introduction

Testing in
one
dimension

Transverse
Ising model

Conclusions
and Outlooks

Appendix



Difficulties of the minimization



- ▶ Evaluating $E(\theta_\alpha)$ requires many probabilistic quantum measurements (shots) \Rightarrow uncertainties (noisy optimization of $E(\theta_\alpha)$).
- ▶ Quantum devices have decoherence noise \Rightarrow systematic errors on $E(\theta_\alpha)$.
- ▶ Quantum measurements are expensive in time and budget (for now) \Rightarrow optimization should be measurements-efficient.
- ▶ $E(\theta_\alpha)$ is non-convex with many possible local minima \Rightarrow global optimization.

Introduction

Testing in
one
dimension

Transverse
Ising model

Conclusions
and Outlooks

Appendix



Why Bayesian Optimization (BO)?



- ▶ Is very measurements-efficient compared to other methods: more computational effort on the classical side in order to alleviate the quantum side.
- ▶ Very good for noisy optimization: the $E(\theta_\alpha)$ can be treated as a stochastic process and the measured energy variances take part to the inference procedure using Gaussian Process Regression (GPR).
- ▶ There are some properties of $E(\theta_\alpha)$ that can be exploited for the optimization. If θ_α parametrize qubit rotations:

$$\begin{cases} E(\theta_\alpha) \in C^\infty \\ E(\theta_\alpha + 2\pi\delta_{\alpha\beta}) = E(\theta_\alpha) \quad \forall \beta. \end{cases}$$

- ▶ Defect of BO: difficult (but possible) to extend to high dimensional spaces. Many works on this topic.



Bayesian Optimization (BO)



Steps of the algorithm:

1. Generate n (quasi-)random d -dimensional points $\theta_{1\alpha}, \dots, \theta_{n\alpha}$ and evaluate the corresponding energies E_1, \dots, E_n (with errors $\Delta E_1, \dots, \Delta E_n$).
 2. Use E_1, \dots, E_n to construct a surrogate model using GPR that predicts the energy in all domain.
 3. Use the surrogate model to choose the next point $\theta_{n+1,\alpha}$ and measure there E_{n+1} (exploration vs exploitation).
 4. Iterate from point (2) with $n \leftarrow n + 1$.
- We implemented the algorithm using/customizing the libraries Ax, BoTorch and GPyTorch.

Introduction

Testing in
one
dimension

Transverse
Ising model

Conclusions
and Outlooks

Appendix



Introduction

Testing in
one
dimension

Transverse
Ising model

Conclusions
and Outlooks

Appendix

TESTING IN ONE DIMENSION



Test in 1d

- ▶ The behaviour of the algorithm is here shown with a 1d example:

$$E_{1d}(\theta) = \sin(\theta) + \frac{1}{2} \sin(4\theta)$$

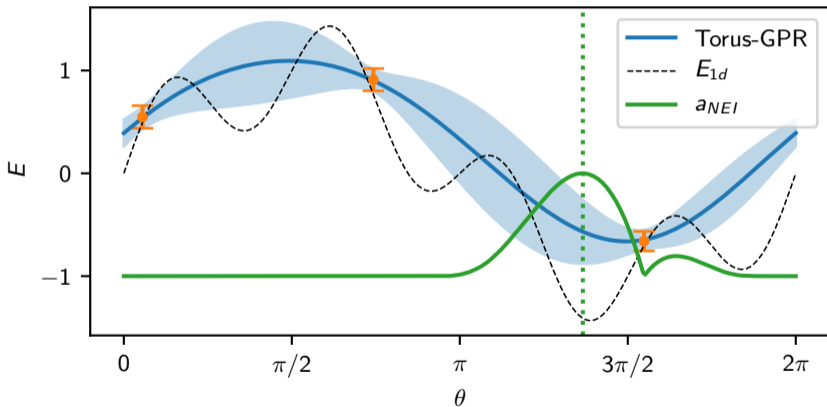
- ▶ Measurements are smeared with a Gaussian error with $\sigma = 0.5$.
- ▶ Error measurements are smeared with a Gaussian error with $\sigma = 0.05$.
- ▶ 3 local minima and 1 global minimum.



Torus RBF in 1d (I)



Bayesian optimization in 1d with Torus RBF-GPR and NEI ($n = 3$)

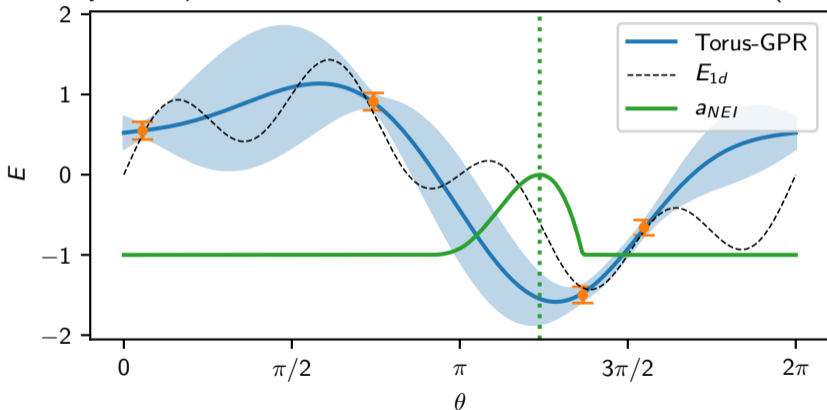




Torus RBF in 1d (II)



Bayesian optimization in 1d with Torus RBF-GPR and NEI ($n = 4$)





- Introduction
- Testing in one dimension
- Transverse Ising model**
- Conclusions and Outlooks
- Appendix

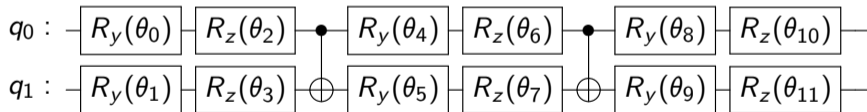
TRANSVERSE ISING MODEL

Testing on VQE

- ▶ 2 spin/qubit transverse-field Ising model:

$$H = -\sigma_x^0 \otimes \sigma_x^1 - \sigma_z^0 - \sigma_z^1$$

- ▶ $2^8 = 256$ shots per measurement. Noisy simulator of IBMQ Santiago.
- ▶ Ansatz used (qiskit's default EfficientSU2 circuit):





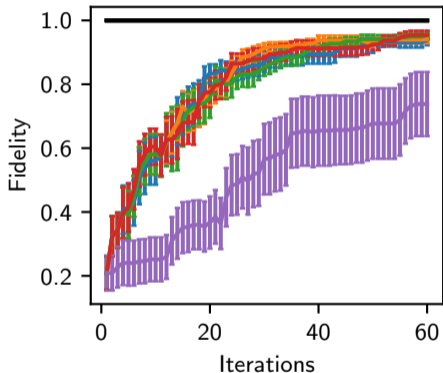
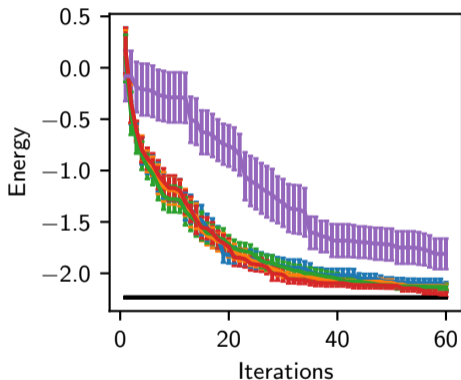
Results of Bayesian VQE test



University of Cyprus



Introduction
Testing in one dimension
Transverse Ising model
Conclusions and Outlooks
Appendix





Introduction

Testing in
one
dimension

Transverse
Ising model

**Conclusions
and Outlooks**

Appendix

CONCLUSIONS AND OUTLOOKS



Conclusions and Outlooks



Conclusions:

- ▶ BO outperforms SPSA in the tests (best optimizer in qiskit for noisy measurements).
- ▶ Implementing different symmetries in kernels is not showing significant advantages so far.

Outlooks:

- ▶ Design and test new kernels tailored for this problem.
- ▶ Improve scalability, e.g. optimizing sequences of domain subspaces (sequential minimal optimization as in [Nakanishi et al., 2020]).



References

- ▶ VQE: Peruzzo et al. 2014.
- ▶ Gaussian Process Regression: Williams and Rasmussen 2006.
- ▶ Bayesian Optimization tutorial: Frazier 2018.
- ▶ Sequential minimal optimization: Nakanishi et al. 2020.
- ▶ SPSA: Kandala, Mezzacapo et al. 2017.
- ▶ Noisy Expected Improvement: Letham, Karrer, Ottoni, Bakshy 2019.
- ▶ Ax platform: Facebook Inc. <https://ax.dev/>.
- ▶ Qiskit: IBM Inc. <https://qiskit.org/>.



STIMULATE
European Joint Doctorates

BAYESIAN OPTIMIZATION OF VARIATIONAL QUANTUM EIGENSOLVERS



Giovanni Iannelli, Karl Jansen

Lattice 2021



Introduction

Testing in
one
dimension

Transverse
Ising model

Conclusions
and Outlooks

Appendix

APPENDIX



Gaussian Process (GP)



- ▶ A GP f maps n d -dimensional domain points $\theta_{1\alpha}, \dots, \theta_{n\alpha}$ to n stochastic variables f_1, \dots, f_n distributed as:

$$p(f_i) = \det(2\pi K)^{-\frac{1}{2}} e^{-\frac{1}{2} \sum_{ij} (f_i - \mu_i)(K^{-1})_{ij}(f_j - \mu_j)}$$

- ▶ where:

$$\mu_i \equiv \mu(\theta_{i\alpha}) \quad K_{ij} \equiv k(\theta_{i\alpha}, \theta_{j\alpha})$$

- ▶ $\mu(\theta_{i\alpha})$ is the mean function: $\mathbb{E}[f(\theta_\alpha)] = \mu(\theta_{i\alpha})$.
- ▶ $k(\theta_{i\alpha}, \theta_{j\alpha})$ is the covariance function: $\text{Cov}[f(\theta_\alpha), f(\theta'_\alpha)]$.
- ▶ μ and k are sufficient to univocally define a GP.
- ▶ Functions can be sampled from a GP (same as for variables from distributions).



- ▶ We want to model the energy function with a GP probabilistic model (samples are possible energy functions) using the information given by measured data E_1, \dots, E_n and their uncertainties $\Delta E_1, \dots, \Delta E_n$.
- ▶ Bayesian inference:
 1. A prior GP is chosen before measuring data $\rightarrow \mu(\theta_\alpha)$ and $k(\theta_\alpha, \theta'_\alpha)$.
 2. The prior GP is updated using Bayes theorem given the measured E_i in order to find the posterior GP defined by $\mu(\theta_\alpha | E_i)$ and $k(\theta_\alpha, \theta'_\alpha | E_i)$.
- ▶ The prior choice determines properties of the energy model.
- ▶ $\mu(\theta_\alpha) \in C^k$ and $k(\theta_\alpha, \theta'_\alpha) \in C^{2k} \implies$ sampled functions are $\in C^k$.
- ▶ We know that the energy function is $\in C^\infty$ and 2π -periodic (n-torus). Although decoherence noise could break $\in C^\infty$.



Choice of the prior (I)

- ▶ Constant mean function (C^∞ and 2π -periodic):

$$\mu_{\text{const}}(\theta_\alpha) \equiv \mu$$

- ▶ Usually the cov. function is chosen to be stationary (unless it is possible to model inhomogeneities):

$$k(\theta_\alpha, \theta'_\alpha) = k(d(\theta_\alpha, \theta'_\alpha))$$

where $d(\theta_\alpha, \theta'_\alpha)$ is a distance function. Possibilities are:

$$d^E(\theta_\alpha, \theta'_\alpha) = \sqrt{\sum_\alpha \left(\frac{\theta_\alpha - \theta'_\alpha}{l_\alpha} \right)^2} \quad (\text{anisotropic Euclidean})$$

$$d^P(\theta_\alpha, \theta'_\alpha) = \sqrt{\sum_\alpha \left(\frac{P}{\pi l_\alpha} \sin \left(\frac{\pi}{P} (\theta_\alpha - \theta'_\alpha) \right) \right)^2} \quad (P \text{ periodic } n\text{-torus})$$



Choice of the prior (II)

- ▶ Using these two distances, different choices for k imply different differentiability classes:
- ▶ RBF covariance function (C^∞):

$$k^{\text{RBF}}(\theta_\alpha, \theta'_\alpha) \equiv \sigma^2 e^{-\frac{d(\theta_\alpha, \theta'_\alpha)^2}{2}}$$

- ▶ Matérn covariance function ($C^{2\lceil\nu-1\rceil}$):

$$k_\nu^{\text{Matérn}}(\theta_\alpha, \theta'_\alpha) \equiv \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} d(\theta_\alpha, \theta'_\alpha) \right)^\nu K_\nu \left(\sqrt{2\nu} d(\theta_\alpha, \theta'_\alpha) \right)$$

where $K_\nu(x)$ are the modified Bessel functions of second kind.



Choice of the prior (III)

- ▶ $k_\nu^{\text{Matérn}}$ can be used if we expect weaker level of smoothness than C^∞ . Typical in machine learning.
- ▶ In general, the result of the regression will be $\in C^{\lceil \nu - 1 \rceil}$. Typical choices are $\nu = \frac{5}{2}, \frac{3}{2}$. (K_ν in closed form for half-integer ν).
- ▶ For $\nu \rightarrow \infty$, $k_\nu^{\text{Matérn}} \rightarrow k^{\text{RBF}}$.
- ▶ μ , σ^2 and ℓ_α^2 are called hyperparameters and need to be fixed (later).
- ▶ ℓ_α are the typical lengths in different directions. σ^2 is the prediction variance. μ is the prior mean energy.
- ▶ If $\ell_\alpha = \ell \forall \alpha$, the prior is isotropic \Rightarrow variations of $E(\theta_\alpha)$ expected to have the same typical length in every direction.
- ▶ If ℓ_α are let to be different, the inference procedure will determine the most relevant directions, aka Automatic Relevance Determination (ARD).



Gaussian Process Regression (GPR) (I)

- ▶ The posterior GP values are distributed as:

$$p(f(\theta_\alpha|E_i)) = \int d^n f_i p(f(\theta_\alpha|f_i))p(f_i|E_i)$$

where f_i are possible (noiseless) energy values over which we can integrate.

- ▶ The integral is computable analytically, and is Gaussian:
 - ▶ $p(f(\theta_\alpha|f_i))$ is available in closed form (conditional Gaussian).
 - ▶ $p(f_i|E_i)$ (Gaussian) available with Bayes theorem:

$$p(f_i|E_i) = \frac{p(E_i|f_i)p(f_i)}{\int d^n f_i p(E_i|f_i)p(f_i)}$$



Gaussian Process Regression (GPR) (II)



- Analytic Posterior GP $f(\theta_\alpha|E_i)$ has mean and covariance:

$$\begin{cases} \mu(\theta_\alpha|E_i) = \mu(\theta_\alpha) + \sum_{ij} k(\theta_\alpha, \theta_{i\alpha})(\tilde{K}^{-1})_{ij}(E_j - \mu(\theta_{j\alpha})) \\ k(\theta_\alpha, \theta'_\alpha|E_i) = k(\theta_\alpha, \theta'_\alpha) - k(\theta_\alpha, \theta_{i\alpha})(\tilde{K}^{-1})_{ij}k(\theta_{j\alpha}, \theta_\alpha) \end{cases}$$

- $\tilde{K}_{ij} \equiv k(\theta_{i\alpha}, \theta_{j\alpha}) + \Delta E_i \delta_{ij}$. Gaussian errors taken into account.
- The posterior GP maintains smoothness and torus bounds of the prior.
- Most expensive part is the inversion (through Cholesky factorization) $\sim N^3/3$.



Choosing the hyperparameters

- ▶ The hyperparameters $\xi = (\mu, \sigma^2, \ell_\alpha^2)$ need to be fixed.
- ▶ The close form of the marginal likelihood is:

$$p(E_i) = \int d^n f_i p(E_i | f_i) p(f_i) = \det(2\pi \tilde{K})^{-\frac{1}{2}} e^{-\frac{1}{2}(E_i - \mu_i)(\tilde{K}^{-1})_{ij}(E_j - \mu_j)}$$

- ▶ \tilde{K} and μ depend on $\xi \Rightarrow p(E_i | \xi)$.
- ▶ ξ is selected optimizing $p(E_i | \xi) \Rightarrow$ model selection (simplest have priority).
- ▶ Gradient of $p(E_i | \xi)$ is available in closed form (messy). Evaluable with automatic differentiation.
- ▶ This is Maximum Likelihood Estimation of type II.
- ▶ Further regularizations are possible placing priors on ξ (default in BoTorch).



Expected Improvement (EI) (I)

- ▶ Point of next measurement is the maximum of an acquisition function (exploration vs exploitation).
- ▶ EI acq. function is defined through a utility function.
- ▶ Given a hypothetical energy outcome E at θ_α :

$$u_{EI}(\theta_\alpha | E, E_i) \equiv \begin{cases} 0 & \text{if } E \geq E^{\min} \\ E^{\min} - E & \text{if } E < E^{\min} \end{cases}$$

- ▶ Integrating u_{EI} over the GP predictions \rightarrow EI Acquisition function:

$$a_{EI}(\theta_\alpha | E_i) \equiv \int df(\theta_\alpha | E_i) u_{EI}(\theta_\alpha | f(\theta_\alpha | E_i), E_i) p(f(\theta_\alpha | E_i))$$

Expected Improvement (EI) (II)

- ▶ Analytical integration of EI:

$$a_{EI}(\theta_\alpha | E_i) = \sigma(\varphi(u) + u\Phi(u))$$

- ▶ where:

$$\sigma^2 \equiv k(\theta_\alpha, \theta_\alpha | E_i) \quad u \equiv (E^{\min} - \mu(\theta_\alpha | E_i)) / \sigma$$

$$\varphi(u) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \quad \Phi(u) \equiv \int_{-\infty}^u dt \varphi(t)$$

- ▶ Pros:

1. Both exploration and exploitation (a bit more exploitation).
2. Analytic formula also for the gradient \rightarrow gradient-based optimization.

- ▶ Cons: Bad for noisy measurements: Can be $a_{EI}(\theta_\alpha) > 0$ for already measured θ_α .



Noisy Expected Improvement (NEI)

- ▶ NEI is given integrating over all possible noiseless previous measurements f_1, \dots, f_n according to the posterior GP:

$$a_{NEI}(\theta_\alpha | E_i) = \int d^n f_i a_{EI}(\theta_\alpha | f_i) p(f_i | E_i)$$

- ▶ Not available analytically, but approximable with (quasi-)Monte Carlo:

$$a_{NEI}(\theta_\alpha | E_i) \approx \frac{1}{K} \sum_k a_{EI}(\theta_\alpha | f_i^k)$$

- ▶ a_{NEI} can be optimized using the analytical average of the K a_{EIS} .
- ▶ $a_{NEI}(\theta_\alpha) = 0$ if $E(\theta_\alpha)$ is already measured \rightarrow noisy issue eliminated.



STIMULATE
European Joint Doctorates

BAYESIAN OPTIMIZATION OF VARIATIONAL QUANTUM EIGENSOLVERS



Giovanni Iannelli, Karl Jansen

Lattice 2021