

# Structure optimization for parameterized quantum circuits

Mateusz Ostaszewski<sup>1,2</sup>, Edward Grant<sup>2,3</sup>, and Marcello Benedetti<sup>2,4</sup>

<sup>1</sup>Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, Bałtycka 5, 44-100 Gliwice, Poland

<sup>2</sup>Department of Computer Science, University College London, WC1E 6BT London, United Kingdom

<sup>3</sup>Rahko Limited, N4 3JP London, United Kingdom

<sup>4</sup>Cambridge Quantum Computing Limited, CB2 1UB Cambridge, United Kingdom

We propose an efficient method for simultaneously optimizing both the structure and parameter values of quantum circuits with only a small computational overhead. Shallow circuits that use structure optimization perform significantly better than circuits that use parameter updates alone, making this method particularly suitable for noisy intermediate-scale quantum computers. We demonstrate the method for optimizing a variational quantum eigensolver for finding the ground states of Lithium Hydride and the Heisenberg model in simulation, and for finding the ground state of Hydrogen gas on the IBM Melbourne quantum computer.

## 1 Introduction

Methods for tuning the parameters of a quantum circuit to perform a specific task have several important applications. For example, these have been demonstrated in chemical simulation, combinatorial optimization, generative modeling and classification [1–8]. These tasks are usually performed by selecting a fixed circuit structure, parameterizing it using rotation gates, then iteratively updating the parameters to minimize an objective function estimated from measurements. Circuits of this type are known as parameterized quantum circuits. These methods are particularly promising for use with noisy intermediate-scale quantum (NISQ) computers because of their relative tolerance to noise compared to many other quantum algorithms.

In recent years a lot of progress has been made in improving the performance of parameterized quantum circuits including methods for calculating parameter gradients, hardware efficient ansätze, reducing the number of measurements required, and resolving problems with vanishing gradients [9–13]. Despite progress, many challenges remain. One of the most critical challenges is addressing the effects of noise. Generally, the effects of noise increase with the depth of the quantum circuit. For this reason it is highly desirable for a parameterized quantum circuit to be as shallow as possible whilst being expressive enough to perform the task at hand.

Few approaches have been proposed for optimizing the structure of a quantum circuit. In Ref. [14] the authors propose using a genetic algorithm for optimizing the circuit structure by selecting candidate gates from a set of allowed gates that are not parameterized. In Ref. [15] the authors propose growing the circuit by iteratively adding parameterized gates and re-optimizing the circuit using gradient descent.

In this work we propose a method for efficiently optimizing the structure of quantum circuits while optimizing an objective function, for example the energy given by a Hamiltonian operator. Gates are defined so that both the direction and angle of rotation are degrees of freedom. Whilst the angle is parameterized in a continuous manner, the direction is chosen from a set of generators, namely tensor products of Pauli matrices. Optimization is performed by selecting the first gate

Mateusz Ostaszewski: [mm.ostaszewski@gmail.com](mailto:mm.ostaszewski@gmail.com)

Edward Grant: [edward.grant@rahko.ai](mailto:edward.grant@rahko.ai)

Marcello Benedetti: [marcello.benedetti@cambridgequantum.com](mailto:marcello.benedetti@cambridgequantum.com)

and finding the direction and angle of rotation that yield minimum energy. This is performed iteratively for all the parameterized gates in the circuit. Once the last gate has been optimized the cycle is repeated until convergence.

In the Methods section we describe the approach in generality and then provide algorithms for the case of parameterized single-qubit gates and fixed two-qubit gates. This special case is interesting because it can be executed on all existing NISQ hardware and can provide a significant advantage as we show in the Results section. For single-qubit rotations about a fixed direction, the optimal angle can be found using three energy estimations. Independently from our work this property has been used as part of proposed methods for optimizing angles of rotation in quantum circuits [16, 17]<sup>‡</sup>. Our method extends this and finds also the optimal direction within a set of canonical ones. This comes with very little overhead, since only seven energy estimations are required. Although we demonstrate the method on parameterized single-qubit gates and fixed entangling gates, the method can be applied to higher order gates. In the Discussion section we present possible extensions and ways to reduce the number of energy estimations required for our algorithms. We leave all the details to the Appendix.

## 2 Methods

Let us consider an optimization problem where the objective function is encoded in a Hermitian operator  $M$  and the candidate solution is encoded in a parameterized quantum circuit  $U = U_D \cdots U_1$  acting on an  $n$ -qubit initial state  $\rho$ . Each gate is either fixed, e.g., a controlled-Z, or parameterized of the form  $U_d = \exp(-i\frac{\theta_d}{2}H_d)$ . Here,  $\theta_d \in (-\pi, \pi]$  are angles of rotation and  $H_d$  are Hermitian and unitary matrices, e.g., tensor products of Pauli matrices. We collect these parameters into a real vector  $\boldsymbol{\theta}$  and a vector of matrices  $\mathbf{H}$ , respectively.

Without loss of generality we consider the task of minimizing the objective function which we simply refer to as *the energy*. Using subscripts to indicate arguments of functions, we can state the problem as  $(\boldsymbol{\theta}^*, \mathbf{H}^*) = \arg \min_{\boldsymbol{\theta}, \mathbf{H}} \langle M \rangle_{\boldsymbol{\theta}, \mathbf{H}}$ , where  $\langle M \rangle_{\boldsymbol{\theta}, \mathbf{H}} = \text{tr}(MU\rho U^\dagger)$ .

To solve the problem we present two methods. The first fixes the circuit structure and optimizes only the angles; the second optimizes structure and angles simultaneously. Both methods rely on the fact that the expectation value as a function of an angle of rotation has sinusoidal form. That is, if we fix  $\mathbf{H}$  as well as all the degrees of freedom in vector  $\boldsymbol{\theta}$  except one, say  $\theta_d$ , we can express the expectation as  $\langle M \rangle_{\theta_d} = A \sin(\theta_d + B) + C$  where  $A, B$  and  $C$  are unknown coefficients. A detailed derivation is provided in Appendix A. Clearly, if we were able to estimate these coefficients, we could also characterize the sinusoidal form. This can be exploited to design optimization strategies.

Our first method is a coordinate optimization [18–20] algorithm applied to the angles of rotation. It finds the optimal angle for one gate while fixing all others to their current values, and sequentially cycles through all gates. This is rather simple to perform since at each step the energy has sinusoidal form with period  $2\pi$ . For gate  $U_d$ , the optimal angle has a closed form expression

$$\begin{aligned} \theta_d^* &= \arg \min_{\theta_d} \langle M \rangle_{\theta_d} \\ &= \phi - \frac{\pi}{2} - \arctan 2 \left( 2 \langle M \rangle_\phi - \langle M \rangle_{\phi + \frac{\pi}{2}} - \langle M \rangle_{\phi - \frac{\pi}{2}}, \langle M \rangle_{\phi + \frac{\pi}{2}} - \langle M \rangle_{\phi - \frac{\pi}{2}} \right) + 2\pi k, \end{aligned} \quad (1)$$

for any real  $\phi$  and any integer  $k$ . In practice we select  $k$  such that  $\theta_d^* \in (-\pi, \pi]$ .

The optimal angle can be found for all  $d = 1, \dots, D$  in order to complete a cycle. Once all angles have been updated, a new cycle is initialized unless a stopping criterion is met. A number of potential stopping criteria could be used here. For example, one could stop after a fixed budget of  $K_1$  cycles with the caveat that there may still be room for improvement. As another example, one could stop when the energy has not been significantly lowered for  $K_2$  consecutive cycles. We call this algorithm **Rotosolve** and summarize it in Algorithm 1.

The choice of circuit structure is often based on prior knowledge about the problem as well as hardware constraints, e.g., qubit-to-qubit connectivity and gate set. It is reasonable to expect the chosen circuit structure to be suboptimal in the majority of cases. We believe there is large room

---

<sup>‡</sup>To the best of our knowledge this idea appeared in Nakanishi et al. (March 2019), Parrish et al. (April 2019) and this work (May 2019). Our manuscript was originally titled ‘Quantum circuit structure learning’.

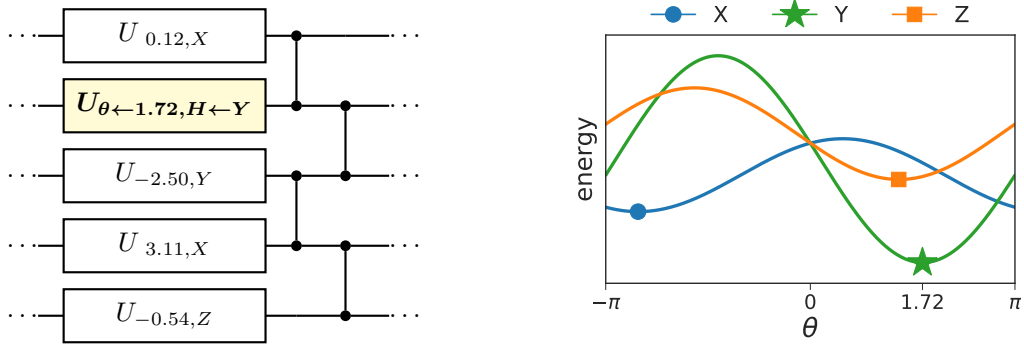


Figure 1: The gradient-free algorithm `Rotoselect` sequentially adjusts angles of rotation and circuit structure in order to efficiently minimize the energy. For each generator the energy is a sinusoidal function of  $\theta$  with period  $2\pi$ . In this example, the single-qubit gate coloured in yellow has been assigned generator  $Y$  and angle of rotation 1.72. This is the configuration attaining minimal energy, as shown in the right panel.

for improvement here. Our second method relaxes the constraint of the fixed circuit structure and optimizes it along with the angles. Similarly to the first method, we opt for a greedy approach optimizing one gate at a time.

Recall that in the parameterization considered here,  $n$ -qubit gates are generated by Hermitian and unitary matrices such as tensor products of Pauli matrices  $H_d \in \{I, X, Y, Z\}^{\otimes n}$ . Structure optimization aims at finding the optimal set of generators, which is clearly a daunting combinatorial problem. The general approach consists of using the expression in Eq. (1) to find minimizers  $\theta_d^*(P) = \arg \min_{\theta_d} \langle M \rangle_{\theta_d, P}$  for all generators  $P \in \{I, X, Y, Z\}^{\otimes n}$ . Here the second subscript indicates that  $P$  is used as the generator for the  $d$ -th gate. Then,  $H_d$  is set to the generator giving lowest energy, and  $\theta_d$  is set to the corresponding minimizer. This is repeated for  $d = 1, \dots, D$  to complete a cycle and the algorithm iterates until a stop criterion is triggered.

It is worth noting that to select the generator we need to know the energy attained by the corresponding optimal angle. In other words, given a generator  $P$  and its optimal angle  $\theta_d^*(P)$ , we need to calculate  $\langle M \rangle_{\theta_d^*(P), P} = -A + C$ . This can be extrapolated at no additional cost using the expressions for  $A$  and  $C$  provided in Appendix A.

The above is very general and indeed suffers from combinatorial explosion due to the  $4^n$  possible choices for the generator of each gate. However, in practice we shall still comply with the constraints of the underlying NISQ hardware. Since we are not considering compilation of logical gates to physical ones, only a very small subset of generators is available, namely the native gate set of the hardware. For simplicity, we now consider optimizing the structure of single-qubit gates while employing a fixed layout of two-qubit entangling gates similar to the one in Ref. [21]. Figure 1 illustrates the circuit layer and shows an example of how the algorithm updates both the generator and the angle.

When structure optimization is limited to single-qubit gates, the generators can be selected such that  $H_d \in \{X, Y, Z\}$ . The identity generator is not required because it can be trivially obtained from any other generator by setting the angle to zero. In fact, we can exploit this to reduce the number of circuit evaluations per optimization step. According to Eq. (1), each of the three generators requires three circuit evaluations, for a total of 9 evaluations per optimization step. Recalling that  $\phi$  can be chosen at will, we can set  $\phi \leftarrow 0$  to obtain the identity gate for all three generators. Since  $\langle M \rangle_{0, X} = \langle M \rangle_{0, Y} = \langle M \rangle_{0, Z}$ , we can estimate this quantity once, effectively reducing the number of evaluations to 7 per optimization step. We call this algorithm `Rotoselect` and summarize it in Algorithm 2. We emphasize that `Rotosolve` is a classical optimization algorithm for circuit parameters. `Rotoselect` is an extension to `Rotosolve` that also allows for some optimization of the structure of the ansatz itself. Neither algorithm is tied to a specific ansatz.

---

**Algorithm 1** Rotosolve

---

**Input:** Hermitian measurement operator  $M$  encoding the objective function, parameterized quantum circuit  $U$  with fixed structure, stopping criterion

- 1: Initialize  $\theta_d \in (-\pi, \pi]$  for  $d = 1, \dots, D$  heuristically or at random
  - 2: **repeat**
  - 3:   **for**  $d = 1, \dots, D$  **do**
  - 4:     Select a value  $\phi \in \mathbb{R}$  heuristically or at random
  - 5:     Fix all angles except the  $d$ -th one
  - 6:     Estimate  $\langle M \rangle_\phi$ ,  $\langle M \rangle_{\phi+\frac{\pi}{2}}$  and  $\langle M \rangle_{\phi-\frac{\pi}{2}}$  from samples
  - 7:      $\theta_d \leftarrow \phi - \frac{\pi}{2} - \arctan 2(2\langle M \rangle_\phi - \langle M \rangle_{\phi+\frac{\pi}{2}} - \langle M \rangle_{\phi-\frac{\pi}{2}}, \langle M \rangle_{\phi+\frac{\pi}{2}} - \langle M \rangle_{\phi-\frac{\pi}{2}})$
  - 8: **until** stopping criterion is met
- 

---

**Algorithm 2** Rotoselect

---

**Input:** Hermitian measurement operator  $M$  encoding the objective function, parameterized quantum circuit  $U$ , stopping criterion

- 1: Initialize  $\theta_d \in (-\pi, \pi]$  and  $H_d \in \{X, Y, Z\}$  for  $d = 1, \dots, D$  heuristically or at random
  - 2: **repeat**
  - 3:   **for**  $d = 1, \dots, D$  **do**
  - 4:     Fix all angles and generators except for the  $d$ -th gate
  - 5:     **for**  $P \in \{X, Y, Z\}$  **do**
  - 6:       Compute  $\theta_d^*(P) = \arg \min_{\theta_d} \langle M \rangle_{\theta_d, P}$  using Eq. (1) with  $\phi \leftarrow 0$
  - 7:       Extrapolate  $\langle M \rangle_{\theta_d^*(P), P}$  using the expressions in Appendix A
  - 8:      $H_d \leftarrow \arg \min_P \langle M \rangle_{\theta_d^*(P), P}$
  - 9:      $\theta_d \leftarrow \theta_d^*(H_d)$
  - 10: **until** stopping criterion is met
- 

We conclude this section with a note about convergence. On a quantum computer the objective function is stochastic because  $\langle M \rangle_{\theta, \mathbf{H}}$  is estimated from samples. For implementation in NISQ computers the operator is typically described as the weighted sum of a polynomial number of terms  $M = \sum_i w_i M_i$ . In this case the expectation is given by  $\langle M \rangle_{\theta, \mathbf{H}} = \sum_i w_i \langle M_i \rangle_{\theta, \mathbf{H}}$ . Assuming an infinite number of measurements the objective function is deterministic and easier to analyze. Bertsekas [18] provides convergence results for coordinate minimization algorithms under rather general conditions and including non-convex objective functions. An analysis along these lines could be attempted for **Rotosolve**. In **Rotoselect** the analysis is complicated by the fact that multiple generators could lead to equally good minima. That is, at each given optimization step the solution may not be unique.

### 3 Results

In this section we first study the impact of structure optimization on the variational quantum eigensolver (VQE) for the Heisenberg model and for Lithium Hydride. **Rotoselect** differs from **Rotosolve** only because it employs structure optimization. Any difference in performance between these two algorithms can therefore be attributed to structure optimization. Second, we demonstrate **Rotoselect** on the IBM Melbourne quantum computer by performing VQE for Hydrogen. Thirdly, we show that an unfavorable initial choice of circuit structure can be boosted using our method. Finally, we compare our algorithms with other state-of-the-art optimizers and find significant improvements in terms of performance and scaling.

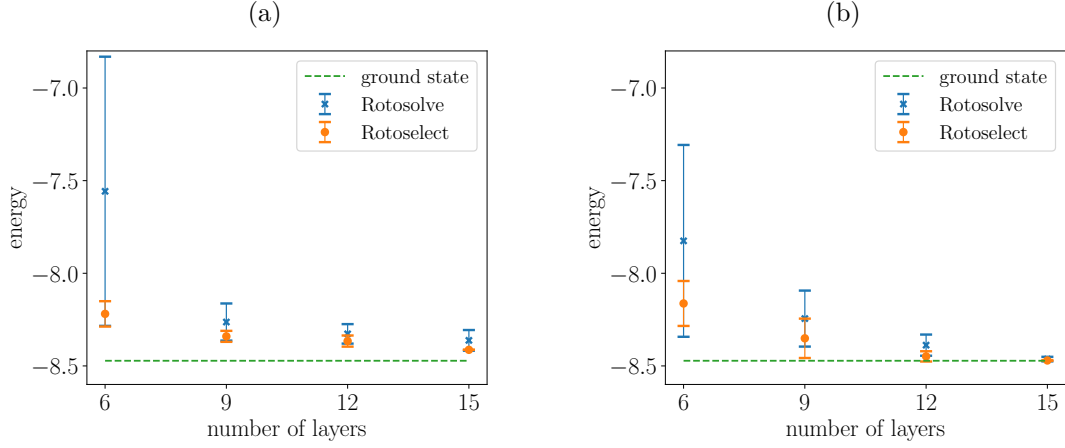


Figure 2: Mean and standard deviation of energy across trials as a function of number of circuit layers comparing Rotosolve and Rotoselect for optimizing a VQE to minimize the energy of the cyclic spin chain Heisenberg model on 5 qubits. In (a) 1000 measurements for each Hamiltonian term were used to approximate the energy. In (b) the exact energy was calculated.

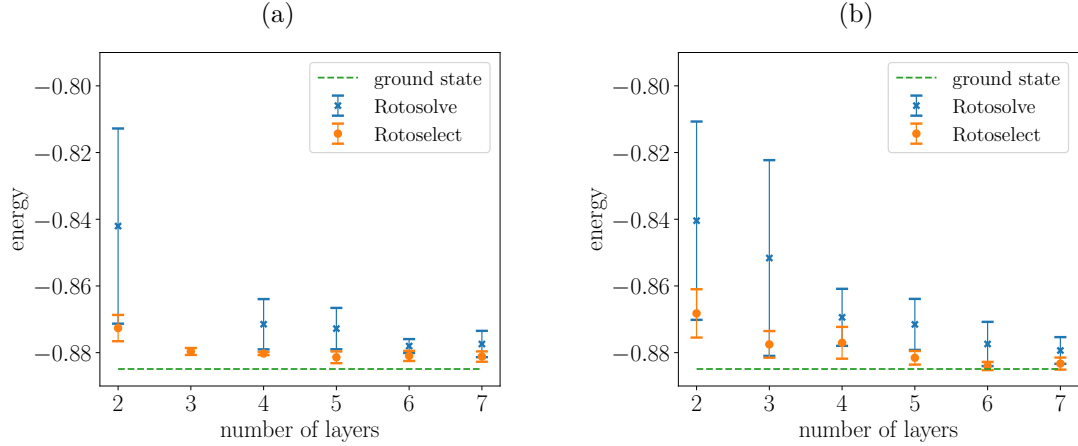


Figure 3: Mean and standard deviation of energy across trials as a function of number of circuit layers comparing Rotosolve and Rotoselect for optimizing a VQE to minimize the molecular Hamiltonian for LiH. In (a) 1000 measurements for each Hamiltonian term were used to approximate the energy. We do not include the results for Rotosolve on 3 layers because the mean energy lied above the plotted range due to the presence of an outlier. In (b) the exact energy was calculated.

### 3.1 Performance on the variational quantum eigensolver

We considered the problem of finding the ground state energy of the 5-qubit Heisenberg model on a 1D lattice with periodic boundary conditions and in the presence of an external magnetic field. The corresponding Hamiltonian reads

$$M = J \sum_{(i,j) \in \mathcal{E}} (X_i X_j + Y_i Y_j + Z_i Z_j) + h \sum_{i \in \mathcal{V}} Z_i, \quad (2)$$

where  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  is the undirected graph of the lattice with 5 nodes,  $J$  is the strength of the spin-spin interactions, and  $h$  is to the strength of the magnetic field in the  $Z$ -direction. For  $J/h = 1$  the ground state is known to be highly entangled (see Ref. [10] for VQE simulations on the Heisenberg model). We chose  $J = h = 1$ .

Circuits were initialized using the strategy described in Ref. [13]. The circuit form consisted of layers of parameterized single-qubit rotations followed by a ladder of controlled-Z gates [21]. An example of this type of layer is shown in Fig. 1. Optimization was performed for 1000 cycles.

Figure 2 reports mean and standard deviation of the lowest energy encountered during optimization across 10 trials for circuits with 6, 9, 12 and 15 layers. Panel (a) shows results when 1000 measurements were used to estimate the expectation of each term in the sum in Eq. (2), while panel (b) shows oracular results in the limit of infinite measurements. For all number of layers, `Rotoselect` achieved better mean, standard deviation, and absolute lowest energy than `Rotosolve`. In particular for the lower depth circuits `Rotoselect` achieves a much smaller standard deviation than `Rotosolve`. This is likely due to the fact that the generators are initialized at random and `Rotosolve` is unable to change an initial bad condition. Given the limited capacity of low-depth circuits, a good choice for the generators appears to be particularly important. As the number of layers increases, both algorithms find better approximations to the ground state. Both algorithms exhibit robustness to the finite number of measurement employed here.

The experiment was repeated on the 4-qubit chemical Hamiltonian for Lithium Hydride (LiH) at bond distance. To construct the Hamiltonian we used the coefficients and the Pauli terms given in Ref. [10], Supplementary Material. Figure 3 shows mean and standard deviation of the lowest energy encountered across 10 trials for circuits up to 7 layers. Panel (a) shows results when expectations are estimated from 1000 measurements, while panel (b) shows results in the limit of infinite measurements. The results are consistent with the previous experiment.

### 3.2 Demonstration on the IBM Melbourne computer

We compared the performance of `Rotoselect` on the 14-qubit IBM Melbourne quantum computer against a quantum simulator. For this test we chose to find the ground state energy of the 2-qubit Hydrogen Hamiltonian, using the coefficients and the Pauli terms given in Ref. [10], Supplementary Material. The circuit had 2 layers and a total of 4 adjustable angles. In contrast to previous experiments where we used controlled- $Z$  entangling gates, here we used CNOTs as they belong to the native gate alphabet of this device.

We also characterized the measurement noise using an off-the-shelf method provided by Qiskit Ignis [22], namely the tensored measurement calibration. A general measurement error calibration prepares each of the  $2^n$  basis states and immediately measures them. The statistics are then used to calculate a calibration matrix which is applied in the classical post-processing of subsequent runs. Tensored measurement calibration assumes that errors are local to subsets of qubits, hence reducing the requirements for calculating the calibration matrix. In our experiments, we assumed that measurement errors are local to each qubit. In the following analysis, we do not include the overhead from the calibration since it was performed only once at the beginning of each trial.

We ran 5 trials with randomly initialized circuits and taken 1000 measurements for each Hamiltonian term when estimating the energy. Figure 4 shows the mean and one standard deviation of the energy as a function of the number of evaluations for two full cycles of `Rotoselect`. Despite error mitigation, results on the quantum device (blue dots) are in average slightly worse than those from the simulator (orange crosses). Yet we were able to get close to the ground state within the 56 evaluations required to complete two cycles.

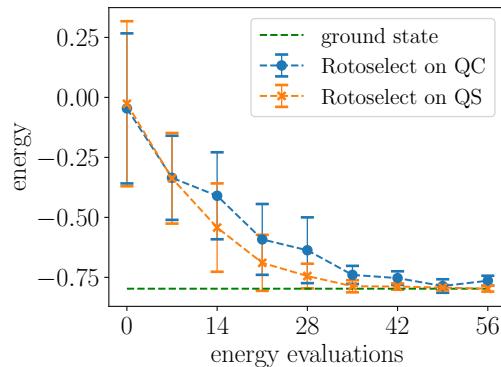


Figure 4: Mean and standard deviation of the energy as a function of the number of evaluations for the Hydrogen Hamiltonian. Blue dots correspond to experiments on the IBM Melbourne quantum computer (QC), and orange crosses corresponds to experiments on the quantum simulator (QS).

### 3.3 Optimization of circuits with limited expressibility

In Ref. [23] the authors define “expressibility” as the circuit’s ability to generate pure states that are well representative of the Hilbert space. To estimate expressibility, they repeatedly sample the adjustable angles of the circuit at random in order to generate a distribution of quantum states; then, they use a suitable divergence to compare this distribution to the uniform distribution of quantum states. Small divergence means high expressibility.

As a function of the number of layers, expressibility improves at different rates depending on the circuit structure. For some choices of structure, additional layers do not yield improvements and expressibility saturates rather quickly. The authors suggest that information about expressibility saturation could be valuable when selecting the circuit structure.

In this section, we use expressibility to select an unfavorable circuit and show that structure optimization can turn it into a useful one. In particular, we required a circuit structure for which expressibility is poor and saturates quickly. We chose circuit #15 from the pool of circuits studied in Ref. [23]. Figure 5 (a) shows the corresponding circuit diagram for 4 qubits. Our task consisted of maximizing the overlap of the state generated by the circuit with a target state sampled uniformly at random. To avoid potential misunderstanding, we stress that our task is not that of maximizing the expressibility, although the two may be related. In practice, we minimized the energy for operator  $M = -|\phi\rangle\langle\phi|$  where  $|\phi\rangle$  is the random target state.

For this simulation, we used the exact energy and we varied the number of layers in  $\{1, \dots, 7\}$ . We stopped the algorithms after 50 cycles. Figure 5 (b) shows the average trace distance and standard deviation as a function of layers across 10 random target states for circuit #15. **Rotosolve** is not able to take the trace distance below a certain value, even when adding more layers. On the other hand, **Rotoselect** achieves lower trace distance for all choices of the number of layers and approaches zero for 7 layers. This indicates that an unfavorable initial choice of circuit structure can be boosted using our method.

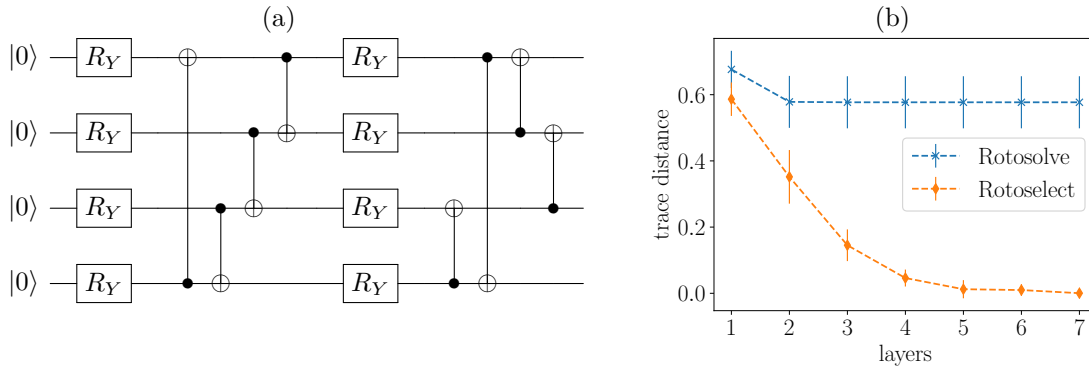


Figure 5: (a) Diagram for circuit #15 from Ref. [23]. (b) Mean and standard deviation of the trace distance to uniformly random states as a function of number of circuit layers for **Rotosolve** and **Rotoselect** on circuit #15.

### 3.4 Comparison of optimization algorithms and scaling

In this experiment we compare the optimization time and optimization time scaling in the system size for **Rotosolve**, **Rotoselect**, **SPSA** [24] and gradient descent with **Adam** [25].

Figure 6 (a) shows the energy and standard deviation as a function of the number of energy evaluations across 5 trials for the 5-qubit Heisenberg cyclic spin chain described in the Results section. The depth of the circuit was 30 layers. The exact energy was used to perform updates. The learning rate for **Adam** was set to 0.05. The hyperparameter for **SPSA** were the same as in Ref. [10]. Both **Rotosolve** and **Rotoselect** converged with significantly fewer energy evaluations than **Adam** or **SPSA**.

Figure 6 (b) shows the mean and standard deviation for the number of evaluations needed to find the ground state of the Heisenberg cyclic spin chain as a function of the size of the system. The ground state was considered found if the candidate solution had energy within 2% from the actual ground state energy. More precisely, this threshold is in relation to the normalized distance



between the candidate energy and the minimum eigenvalue expressed by  $\frac{\langle M \rangle - E_{\min}}{E_{\max} - E_{\min}}$ . It should be noted, that in the experiments there was an additional stop condition, which stopped algorithm after 100,000 cycles. Five trials were performed to estimate each mean and standard deviation. 1000 measurements were performed for each Hamiltonian term in order to approximate the energy for each evaluation. Circuit depth was  $3n^2/2 + 2n$  for an even number of qubits and  $3(n^2 - 1)/2 + 2n$  for odd where  $n$  was the number of qubits.

Both **Rotosolve** and **Rotoselect** converged faster than **Adam** or **SPSA** for the number of qubits tested, indicating a favorable scaling.

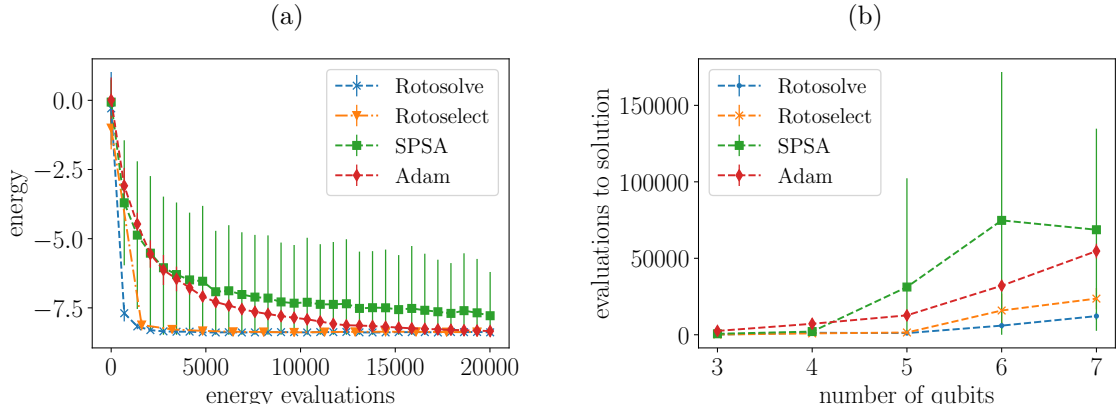


Figure 6: A comparison of **Rotosolve**, **Rotoselect**, **SPSA** and **Adam**. (a) Energy as a function of number of energy evaluations for the 5-qubit Heisenberg cyclic spin chain. (b) Number of energy evaluations to solution as a function of the number of qubits for the Heisenberg cyclic spin chain.

## 4 Discussion

The proposed algorithms can be extended in a number of ways. In the context of circuit optimization, **Rotosolve** can be generalized to find minimizers of  $K$  angles of rotation at the same time and at the cost of evaluating  $3^K$  circuits. While this is an exponential cost, for small  $K$  this approach is computationally feasible and may provide an advantage. In Ref. [17] the authors explore this idea for  $K \in \{1, 2, 3\}$  where subsets of angles are chosen at random. This approach belongs to the class of algorithms called coordinate block minimization [20].

In the context of circuit structure optimization, **Rotoselect** can be generalized to incrementally grow circuits from scratch rather than starting from random initial structures. This is similar in spirit to **Adapt-VQE** [15], but with the advantage that each new gate is optimized efficiently in closed form rather than using gradient-based optimization. Furthermore, we could efficiently remove gates by assessing whether they are contributing to the solution (e.g., a redundant gate would generate sinusoidal forms that appear to be flat).

Another interesting extension is to use **Rotoselect** to learn the optimal connectivity layout for the entangling gates. As an example consider trapped ion computers that can implement fully connected layers of Mølmer-Sørensen gates [26]. This choice provides high expressive power in low-depth circuits [4], but must be balanced against the potentially slow clock speed of these entangling gates [27] and potential gate errors. Our algorithm could find this sweet spot. Since with  $n$  qubits we must evaluate  $n(n-1)/2$  choices for non-directional two-qubit gates in a layer, this approach is also efficient. We discuss other generalizations in Appendix B.

Heuristics can be used to speedup our methods. The most striking example consists of reusing information to reduce the number of energy evaluations. By appropriately choosing the offset  $\phi$  in Eq. (1), one obtains an angle for which the energy is already known from the previous update. Thus one only needs to evaluate the remaining two energies to perform the current update. By applying this trick systematically one reduces the number of evaluations from 3 to 2 for **Rotosolve**, and from 7 to 6 for **Rotoselect**. While the result is exact in the limit of infinite number of measurement, in realistic scenarios this approximation introduces noise and its effects should be investigated in future work.



In our simulations we also found that `Rotoselect` tends to change the structure only in the early stage of optimization. One could detect this and switch from `Rotoselect` to `Rotosolve` further reducing the number of circuit evaluations.

Quantum circuit structure optimization provides an efficient means for improving the expressivity of low-depth quantum circuits with only a small computational overhead. This characteristic makes the described methods particularly suitable for deployment on near-term quantum computers where circuit depth and optimization time are significant bottlenecks.

## Acknowledgements

M.B. is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) and by Cambridge Quantum Computing Limited (CQC). E.G. is supported by EPSRC [EP/P510270/1]. M.O. acknowledges support from Polish National Science Center scholarship 2018/28/T/ST6/00429. We thank Leonard Wossnig for helpful discussions. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research. This research was supported in part by PLGrid Infrastructure.

## References

- [1] Alberto Peruzzo et al. “A variational eigenvalue solver on a photonic quantum processor”. In: *Nature Communications* 5.1 (2014), p. 4213. DOI: [10.1038/ncomms5213](https://doi.org/10.1038/ncomms5213).
- [2] E. Farhi et al. *Quantum Algorithms for Fixed Qubit Architectures*. 2017. arXiv: [1703.06199](https://arxiv.org/abs/1703.06199) [quant-ph].
- [3] Edward Farhi and Hartmut Neven. *Classification with Quantum Neural Networks on Near Term Processors*. 2018. arXiv: [1802.06002](https://arxiv.org/abs/1802.06002) [quant-ph].
- [4] Marcello Benedetti et al. “A generative modeling approach for benchmarking and training shallow quantum circuits”. In: *npj Quantum Information* 5.1 (May 2019). DOI: [10.1038/s41534-019-0157-8](https://doi.org/10.1038/s41534-019-0157-8).
- [5] Marcello Benedetti et al. “Adversarial quantum circuit learning for pure state approximation”. In: *New Journal of Physics* 21.4 (Apr. 2019), p. 043023. DOI: [10.1088/1367-2630/ab14b5](https://doi.org/10.1088/1367-2630/ab14b5).
- [6] Hongxiang Chen et al. *Universal discriminative quantum neural networks*. 2018. arXiv: [1805.08654](https://arxiv.org/abs/1805.08654) [quant-ph].
- [7] Edward Grant et al. “Hierarchical quantum classifiers”. In: *npj Quantum Information* 4.1 (2018), pp. 1–8. DOI: [10.1038/s41534-018-0116-9](https://doi.org/10.1038/s41534-018-0116-9).
- [8] Marcello Benedetti et al. “Parameterized quantum circuits as machine learning models”. In: *Quantum Science and Technology* 4.4 (Nov. 2019), p. 043001. DOI: [10.1088/2058-9565/ab4eb5](https://doi.org/10.1088/2058-9565/ab4eb5).
- [9] K. Mitarai et al. “Quantum circuit learning”. In: *Phys. Rev. A* 98 (3 Sept. 2018), p. 032309. DOI: [10.1103/PhysRevA.98.032309](https://doi.org/10.1103/PhysRevA.98.032309).
- [10] Abhinav Kandala et al. “Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets”. In: *Nature* 549.7671 (Sept. 2017), 242–246. DOI: [10.1038/nature23879](https://doi.org/10.1038/nature23879).
- [11] Kosuke Mitarai, Tennin Yan, and Keisuke Fujii. “Generalization of the Output of a Variational Quantum Eigensolver by Parameter Interpolation with a Low-depth Ansatz”. In: *Phys. Rev. Applied* 11 (4 Apr. 2019), p. 044087. DOI: [10.1103/PhysRevApplied.11.044087](https://doi.org/10.1103/PhysRevApplied.11.044087).
- [12] Artur F. Izmaylov, Tzu-Ching Yen, and Ilya G. Ryabinkin. “Revising the measurement process in the variational quantum eigensolver: is it possible to reduce the number of separately measured operators?” In: *Chem. Sci.* 10 (13 2019), pp. 3746–3755. DOI: [10.1039/C8SC05592K](https://doi.org/10.1039/C8SC05592K).
- [13] Edward Grant et al. “An initialization strategy for addressing barren plateaus in parametrized quantum circuits”. In: *Quantum* 3 (Dec. 2019), p. 214. DOI: [10.22331/q-2019-12-09-214](https://doi.org/10.22331/q-2019-12-09-214).
- [14] Rui Li et al. “Approximate Quantum Adders with Genetic Algorithms: An IBM Quantum Experience”. In: *Quantum Measurements and Quantum Metrology* 4.1 (26 Jul. 2017), pp. 1–7. DOI: <https://doi.org/10.1515/qmetro-2017-0001>.

- [15] Harper R. Grimsley et al. “An adaptive variational algorithm for exact molecular simulations on a quantum computer”. In: *Nature Communications* 10.1 (July 2019). DOI: [10.1038/s41467-019-10988-2](https://doi.org/10.1038/s41467-019-10988-2).
- [16] Ken M. Nakanishi, Keisuke Fujii, and Syngae Todo. “Sequential minimal optimization for quantum-classical hybrid algorithms”. In: *Phys. Rev. Research* 2 (4 Oct. 2020), p. 043158. DOI: [10.1103/PhysRevResearch.2.043158](https://doi.org/10.1103/PhysRevResearch.2.043158).
- [17] Robert M. Parrish et al. *A Jacobi Diagonalization and Anderson Acceleration Algorithm For Variational Quantum Algorithm Parameter Optimization*. 2019. arXiv: [1904.03206](https://arxiv.org/abs/1904.03206) [quant-ph].
- [18] D.P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999.
- [19] Ankan Saha and Ambuj Tewari. *On the Finite Time Convergence of Cyclic Coordinate Descent Methods*. 2010. arXiv: [1005.2146](https://arxiv.org/abs/1005.2146) [cs.LG].
- [20] Stephen J Wright. “Coordinate descent algorithms”. In: *Mathematical Programming* 151.1 (2015), pp. 3–34. DOI: [10.1007/s10107-015-0892-3](https://doi.org/10.1007/s10107-015-0892-3).
- [21] Jarrod R. McClean et al. “Barren plateaus in quantum neural network training landscapes”. In: *Nature Communications* 9.1 (Nov. 2018). DOI: [10.1038/s41467-018-07090-4](https://doi.org/10.1038/s41467-018-07090-4).
- [22] Héctor Abraham et al. *Qiskit: An Open-source Framework for Quantum Computing*. 2019. DOI: [10.5281/zenodo.2562110](https://doi.org/10.5281/zenodo.2562110).
- [23] Sukin Sim, Peter D. Johnson, and Alán Aspuru-Guzik. “Expressibility and Entangling Capability of Parameterized Quantum Circuits for Hybrid Quantum-Classical Algorithms”. In: *Advanced Quantum Technologies* 2.12 (2019), p. 1900070. DOI: <https://doi.org/10.1002/qute.201900070>.
- [24] J. C. Spall. “Multivariate stochastic approximation using a simultaneous perturbation gradient approximation”. In: *IEEE Transactions on Automatic Control* 37.3 (1992), pp. 332–341. DOI: [10.1109/9.119632](https://doi.org/10.1109/9.119632).
- [25] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2014. arXiv: [1412.6980](https://arxiv.org/abs/1412.6980) [cs.LG].
- [26] Anders Sørensen and Klaus Mølmer. “Quantum Computation with Ions in Thermal Motion”. In: *Phys. Rev. Lett.* 82 (9 Mar. 1999), pp. 1971–1974. DOI: [10.1103/PhysRevLett.82.1971](https://doi.org/10.1103/PhysRevLett.82.1971).
- [27] Norbert M. Linke et al. “Experimental comparison of two quantum computing architectures”. In: *Proceedings of the National Academy of Sciences* 114.13 (2017), pp. 3305–3310. DOI: [10.1073/pnas.1618020114](https://doi.org/10.1073/pnas.1618020114).

## A Sinusoidal form of expectation values

Recall that we consider circuits  $U = U_D \cdots U_1$  where each gate is either fixed, e.g., the controlled-Z, or parameterized. We consider parameterized gates of the kind  $U_d = \exp(-i\frac{\theta_d}{2}H_d)$ , where  $\theta_d \in (-\pi, \pi]$  and the generator  $H_d$  is a Hermitian and unitary matrix such that  $H_d^2 = I$ . Using the definition of matrix exponential we obtain

$$\begin{aligned}
 U_d &= \sum_{k=0}^{\infty} \frac{(-i)^k (\frac{\theta_d}{2})^k H_d^k}{k!} \\
 &= \sum_{k=0}^{\infty} \frac{(-i)^{2k} (\frac{\theta_d}{2})^{2k} H_d^{2k}}{(2k)!} + \sum_{k=0}^{\infty} \frac{(-i)^{2k+1} (\frac{\theta_d}{2})^{2k+1} H_d^{2k+1}}{(2k+1)!} \\
 &= \sum_{k=0}^{\infty} \frac{(-1)^k (\frac{\theta_d}{2})^{2k}}{(2k)!} I - i \sum_{k=0}^{\infty} \frac{(-1)^k (\frac{\theta_d}{2})^{2k+1}}{(2k+1)!} H_d \\
 &= \cos(\frac{\theta_d}{2}) I - i \sin(\frac{\theta_d}{2}) H_d.
 \end{aligned} \tag{3}$$

As an example of generators with this property, we could use tensor products of Pauli matrices  $H_d \in \{I, X, Y, Z\}^{\otimes n}$ , where  $n$  is the number of qubits.

We apply the circuit to a fiducial state  $\bar{\rho}$  and then measure a Hermitian operator  $\bar{M}$  encoding the objective function. From measurement outputs we can estimate the expectation

$$\langle \bar{M} \rangle = \text{tr} \left( \bar{M} U_D \cdots U_d \cdots U_1 \bar{\rho} U_1^\dagger \cdots U_d^\dagger \cdots U_D^\dagger \right). \quad (4)$$

We want to analyze this expectation as a function of a single parameter  $\theta_d$ . To simplify the notation we absorb all gates before  $U_d$  in the density operator, that is  $\rho = U_{d-1} \cdots U_1 \bar{\rho} U_1^\dagger \cdots U_{d-1}^\dagger$ . Similarly, we absorb all gates after  $U_d$  in the measurement operator  $M = U_{d+1}^\dagger \cdots U_D^\dagger \bar{M} U_D \cdots U_{d+1}$ . This can be done because unitary transformations preserve the Hermiticity of both density and measurement operators. Using this notation Eq. (4) can be written as  $\langle M \rangle = \text{tr} \left( M U_d \rho U_d^\dagger \right)$ .

In the following discussion we will need to evaluate the expectation at different parameter values for gate  $U_d$ . A further change in notation helps. From now on, we drop index  $d$  and use subscripts to indicate parameter value. Using the new notation we write

$$\begin{aligned} \langle M \rangle_\theta &= \text{tr} \left( M U_\theta \rho U_\theta^\dagger \right) \\ &= \text{tr} \left( M \left( \cos \left( \frac{\theta}{2} \right) I - i \sin \left( \frac{\theta}{2} \right) H \right) \rho \left( \cos \left( \frac{\theta}{2} \right) I + i \sin \left( \frac{\theta}{2} \right) H \right) \right) \\ &= \cos^2 \left( \frac{\theta}{2} \right) \text{tr} (M \rho) + i \sin \left( \frac{\theta}{2} \right) \cos \left( \frac{\theta}{2} \right) \text{tr} (M [\rho, H]) + \sin^2 \left( \frac{\theta}{2} \right) \text{tr} (M H \rho H). \end{aligned} \quad (5)$$

Let us inspect the third line of this equation. First, we note that  $\text{tr}(M \rho)$  is simply the expectation when the circuit is evaluated at  $\theta = 0$ . That is,  $\langle M \rangle_0 = \text{tr}(M \rho)$ . Second, the term  $i \text{tr}(M [\rho, H])$  can be written as the difference of two expectations. Indeed, using two independent circuits  $U_{\pm \frac{\pi}{2}} = \exp(\mp i \frac{\pi}{4} H) = \frac{1}{\sqrt{2}} (I \mp i H)$  we have

$$\begin{aligned} \langle M \rangle_{\frac{\pi}{2}} - \langle M \rangle_{-\frac{\pi}{2}} &= \text{tr} \left( M U_{\frac{\pi}{2}} \rho U_{\frac{\pi}{2}}^\dagger \right) - \text{tr} \left( M U_{-\frac{\pi}{2}} \rho U_{-\frac{\pi}{2}}^\dagger \right) \\ &= \frac{1}{2} \left( \text{tr} (M (I - i H) \rho (I + i H)) - \text{tr} (M (I + i H) \rho (I - i H)) \right) \\ &= -i \text{tr} (M H \rho) + i \text{tr} (M \rho H) \\ &= i \text{tr} (M [\rho, H]). \end{aligned} \quad (6)$$

Third, the last term  $\text{tr}(M H \rho H)$  is the expectation obtained evaluating the circuit at  $\theta = \pi$ . That is, using circuit  $U_\pi = \exp(-i \frac{\pi}{2} H) = -i H$  we have  $\langle M \rangle_\pi = \text{tr}(M H \rho H)$ .

Putting these three pieces back in Eq. (4) and using well known trigonometric identities, we obtain

$$\begin{aligned} \langle M \rangle_\theta &= \cos^2 \left( \frac{\theta}{2} \right) \langle M \rangle_0 + \sin \left( \frac{\theta}{2} \right) \cos \left( \frac{\theta}{2} \right) \left( \langle M \rangle_{\frac{\pi}{2}} - \langle M \rangle_{-\frac{\pi}{2}} \right) + \sin^2 \left( \frac{\theta}{2} \right) \langle M \rangle_\pi \\ &= \frac{1 + \cos(\theta)}{2} \langle M \rangle_0 + \frac{\sin(\theta)}{2} \left( \langle M \rangle_{\frac{\pi}{2}} - \langle M \rangle_{-\frac{\pi}{2}} \right) + \frac{1 - \cos(\theta)}{2} \langle M \rangle_\pi \\ &= \frac{\cos(\theta)}{2} (\langle M \rangle_0 - \langle M \rangle_\pi) + \frac{\sin(\theta)}{2} \left( \langle M \rangle_{\frac{\pi}{2}} - \langle M \rangle_{-\frac{\pi}{2}} \right) + \frac{1}{2} (\langle M \rangle_0 + \langle M \rangle_\pi). \end{aligned} \quad (7)$$

We now use the identity  $a \cos(x) + b \sin(x) = \sqrt{a^2 + b^2} \sin(x + \arctan(\frac{a}{b}))$  to obtain a compact expression. The expectation value then reads

$$\begin{aligned} \langle M \rangle_\theta &= A \sin(\theta + B) + C, \\ A &= \frac{1}{2} \sqrt{(\langle M \rangle_0 - \langle M \rangle_\pi)^2 + (\langle M \rangle_{\frac{\pi}{2}} - \langle M \rangle_{-\frac{\pi}{2}})^2}, \\ B &= \arctan2 \left( \langle M \rangle_0 - \langle M \rangle_\pi, \langle M \rangle_{\frac{\pi}{2}} - \langle M \rangle_{-\frac{\pi}{2}} \right), \\ C &= \frac{1}{2} (\langle M \rangle_0 + \langle M \rangle_\pi). \end{aligned} \quad (8)$$

That is, the expectation of  $M$  as a function of a single parameter has sinusoidal form with amplitude  $A$ , phase  $B$ , intercept  $C$ , and period  $2\pi$ . An example is shown in Figure 7.

Note that we must use the  $\arctan2$  function in order to correctly handle the sign of numerator and denominator, as well as the case where the denominator is zero. To see why, assume  $H$  commutes either with  $M$  or  $\rho$ . Then, Eq. (6) evaluates to zero and triggers a division by zero in the standard  $\arctan$ .

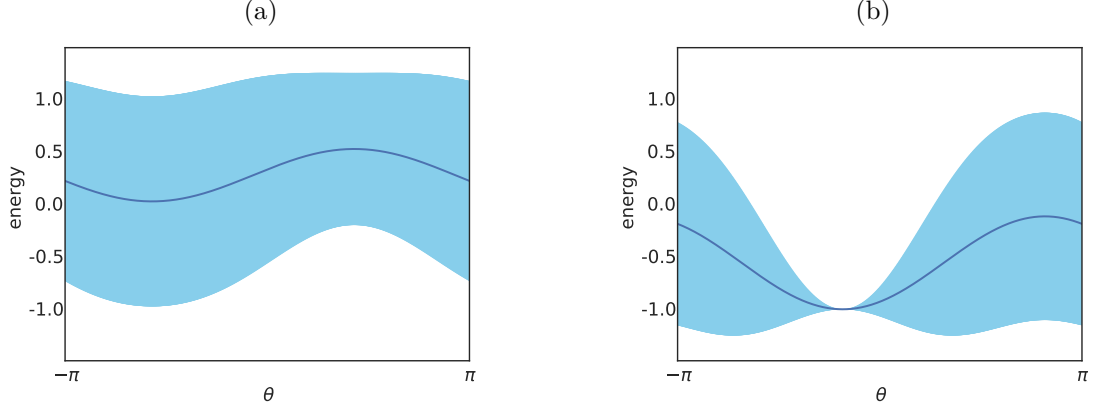


Figure 7: Expectation and variance of Hermitian observable  $ZXZX$  as a function of two different angles of rotation in a random circuit. The expectation value is always of sinusoidal form with period  $2\pi$ . (a) The variance is large because this angle cannot yield an eigenstate of  $ZXZX$ . (b) Both expectation and variance are minimized by setting this second angle to  $\theta^* \approx -0.5$ .

Now, from the graph of the sine function, it is easy to locate the minima at  $\theta^* = -\frac{\pi}{2} - B + 2\pi k$  for all  $k \in \mathbb{Z}$ . The estimator for  $B$  in Eq. (8) seems to require four distinct circuit evaluations. Using simple trigonometry we generalize the estimator and show that only three evaluations are required. Let us write

$$\begin{aligned} \langle M \rangle_{\phi + \frac{\pi}{2}} &= A \sin\left(\phi + \frac{\pi}{2} + B\right) + C \\ &= -A \sin\left(\phi - \frac{\pi}{2} + B\right) + C \\ &= -\langle M \rangle_{\phi - \frac{\pi}{2}} + 2C. \end{aligned} \quad (9)$$

From this we obtain the general estimator for  $C$

$$C = \frac{1}{2} \left( \langle M \rangle_{\phi + \frac{\pi}{2}} + \langle M \rangle_{\phi - \frac{\pi}{2}} \right). \quad (10)$$

We can write

$$\tan(\phi + B) = \frac{\sin(\phi + B)}{\sin(\phi + B - \frac{\pi}{2})} = \frac{\langle M \rangle_{\phi} - C}{\langle M \rangle_{\phi - \frac{\pi}{2}} - C} = \frac{2\langle M \rangle_{\phi} - \langle M \rangle_{\phi + \frac{\pi}{2}} - \langle M \rangle_{\phi - \frac{\pi}{2}}}{\langle M \rangle_{\phi + \frac{\pi}{2}} - \langle M \rangle_{\phi - \frac{\pi}{2}}}. \quad (11)$$

Taking the inverse tangent we obtain the general estimator for  $B$  requiring only three circuit evaluations

$$B = \arctan2\left(2\langle M \rangle_{\phi} - \langle M \rangle_{\phi + \frac{\pi}{2}} - \langle M \rangle_{\phi - \frac{\pi}{2}}, \langle M \rangle_{\phi + \frac{\pi}{2}} - \langle M \rangle_{\phi - \frac{\pi}{2}}\right) - \phi. \quad (12)$$

Finally, we express the minimizer in closed form

$$\begin{aligned} \theta^* &= \arg \min_{\theta} \langle M \rangle_{\theta} \\ &= -\frac{\pi}{2} - B + 2\pi k \\ &= \phi - \frac{\pi}{2} - \arctan2\left(2\langle M \rangle_{\phi} - \langle M \rangle_{\phi + \frac{\pi}{2}} - \langle M \rangle_{\phi - \frac{\pi}{2}}, \langle M \rangle_{\phi + \frac{\pi}{2}} - \langle M \rangle_{\phi - \frac{\pi}{2}}\right) + 2\pi k, \end{aligned} \quad (13)$$

where  $\phi \in \mathbb{R}$  and  $k \in \mathbb{Z}$ . Note that in practice we chose  $k$  such that  $\theta^* \in (-\pi, \pi]$ . At this point we may also want a general estimator for  $A$ . It is easy to verify that

$$A = \frac{1}{2} \sqrt{\left(2\langle M \rangle_{\phi} - \langle M \rangle_{\phi + \frac{\pi}{2}} - \langle M \rangle_{\phi - \frac{\pi}{2}}\right)^2 + \left(\langle M \rangle_{\phi + \frac{\pi}{2}} - \langle M \rangle_{\phi - \frac{\pi}{2}}\right)^2}. \quad (14)$$

This is useful when we need to extrapolate the energy attained by the minimizer,  $\langle M \rangle_{\theta^*} = -A + C$ , and can be done at no additional cost. In summary, Eqs. (10),(12) and (14) completely characterize the sinusoidal form of the energy and allow us to estimate both minimizer and minimum in closed form.

## B A few generalizations

We now discuss some generators  $H_d$  for which `Rotosolve` applies. Our derivation above is rather general, but in practice Algorithms 1 and 2 rely on the canonical axes of rotation, i.e, tensor products of Pauli matrices.

Our first generalization applies to single-qubits gates. Recall that in order to obtain a sinusoidal form of expectation values we need  $H_d^2 = I$ . For single-qubit gates this condition is met by any  $H_d = c_x X + c_y Y + c_z Z$  such that  $(c_x, c_y, c_z) \in \mathbb{R}^3$  is a unit vector. To see why,

$$\begin{aligned}
 H_d^2 &= (c_x X + c_y Y + c_z Z)(c_x X + c_y Y + c_z Z) \\
 &= c_x^2 X^2 + c_x c_y XY + c_x c_z XZ + c_y c_x YX + c_y^2 Y^2 + c_y c_z YZ + c_z c_x ZX + c_z c_y ZY + c_z^2 Z^2 \\
 &= c_x^2 I + i c_x c_y Z - i c_x c_z Y - i c_y c_x Z + c_y^2 I + i c_y c_z X + i c_z c_x Y - i c_z c_y X + c_z^2 I \\
 &= (c_x^2 + c_y^2 + c_z^2) I = I,
 \end{aligned} \tag{15}$$

where in the third line we used well-known identities for the Pauli matrices. Therefore,  $\langle M \rangle_{\theta_d}$  has sinusoidal form for single-qubit gates of the kind  $U_d = \exp(-i \frac{\theta_d}{2} (c_x X + c_y Y + c_z Z))$ .

This suggests a version of `Rotosolve` where a unit vector of coefficients is sampled at random, and where  $\theta_d$  is found using Algorithm 1. This version has the potential of finding interesting axes of rotation. Note that a NISQ implementation may require compilation of  $U_d$  to lower-level gates.

Our second generalization applies to gates acting on an arbitrary number of qubits. Starting from tensor products of  $n$  Pauli matrices, i.e.,  $H_d \in \{I, X, Y, Z\}^{\otimes n}$ , any unitary transformation  $V$  provides an alternative valid generator  $H'_d = V H_d V^\dagger$ . To see why,

$$(H'_d)^2 = V H_d V^\dagger V H_d V^\dagger = V H_d^2 V^\dagger = V V^\dagger = I. \tag{16}$$

The result is an  $n$ -qubit gate of the form  $U_d = \exp(-i \frac{\theta_d}{2} V H_d V^\dagger)$ , where  $\theta_d$  is again found using Algorithm 1. We leave the question of how to exploit gates of this kind for future work.