20

Variational quantum algorithms

The authors are grateful to Marco Cerezo and Xiao Yuan for reviewing this chapter.

Rough overview (in words)

The so-called noisy intermediate-scale quantum (NISQ) era is a term used to describe the regime in which the best quantum processors have fifty to a few hundred noisy qubits [843]. In this regime, one does not have enough qubits or low enough error rates to carry out fault-tolerant quantum computation, and so one is constrained to run low-depth quantum circuits. Under these constraints, structured quantum algorithms with prescribed circuits and provable guarantees are unknown. In light of this, variational quantum algorithms (VQAs) have been proposed. We remark that, despite this original setting, it would also be possible to run VQAs on fault-tolerant devices. While many VQAs have been proposed for a wide range of applications, they all share the same core primitive, which we describe below.

The main idea is to encode the target problem into an optimization task of minimizing the expectation value of some parameterized quantum circuit, or a function thereof. In each optimization step, a quantum computer is used to evaluate expectation values at chosen parameter values, which are read by a classical optimizer that updates the parameters for the next step. The motivation for this framework is to offload some of the computational complexity onto the classical optimization algorithm, with an aim for the quantum subroutines to perform classically intractable calculations.

https://doi.org/10.1017/9781009639651.023

Rough overview (in math)

Given some parameterized unitary $U(\boldsymbol{\theta})$ with adjustable parameters $\boldsymbol{\theta}$, input state ρ, measurement operator O, and function $f(\cdot)$, one evaluates $C(\boldsymbol{\theta}) = f\Big(\mathrm{Tr}\Big[OU(\boldsymbol{\theta})\rho U^\dagger(\boldsymbol{\theta})\Big]\Big)$ on a quantum computer, which is known as a cost function. A classical optimizer is then tasked to solve the problem $\boldsymbol{\theta}_* = \mathrm{argmin}_{\boldsymbol{\theta}} f\Big(\mathrm{Tr}[OU(\boldsymbol{\theta})\rho U^\dagger(\boldsymbol{\theta})]\Big)$. By careful choice of $f(\cdot)$, ρ, and O, one can encode a problem of interest such that $U(\boldsymbol{\theta}_*)$ enables an (approximate) solution to the problem. For instance, the solution could correspond to the computational basis state with which the output state $U(\boldsymbol{\theta}_*)\rho U(\boldsymbol{\theta}_*)^\dagger$ has the largest overlap, or to the value of $f(\mathrm{Tr}[OU(\boldsymbol{\theta}_*)\rho U(\boldsymbol{\theta}_*)^\dagger])$ itself. In general, one can also construct a more elaborate cost function comprising a sum of observable-dependent functions with different input states and measurement operators.

The parameterized circuit $U(\boldsymbol{\theta})$ is commonly referred to as the "ansatz circuit." The choice of cost function and ansatz are key components in designing a VQA. Namely, they should ideally satisfy the following properties:

(i) Smaller values of the cost function should correspond to better quality of solution.
(ii) The ansatz should be sufficiently expressive to contain a unitary $U(\boldsymbol{\theta}_*)$, which yields an acceptable solution.
(iii) The ansatz should lead to a trainable cost landscape in parameter space, such that a sufficiently good solution can be found efficiently by the classical optimizer.
(iv) The cost function should be classically hard to simulate, given the choice of ansatz.

It should be noted that while one would expect any VQA to satisfy the first point by design, in general, it can be hard to satisfy all of the above requirements simultaneously via theoretical guarantees or even heuristically in practice. These caveats are discussed in more detail below.

Dominant resource cost (gates/qubits)

The gate complexity is wholly dependent on the choice of ansatz. Satisfying properties ((ii)) and ((iv)) may place lower bounds on the required circuit depth. In addition, the connectivity of the device may also significantly affect the depth of the circuit. For instance, compilation of a generic two-qubit gate acting on an n-qubit state on hardware with 1D nearest-neighbor connectivity incurs $\mathcal{O}(n)$ circuit depth.

Throughout the optimization, the cost function is evaluated at different parameter settings $\boldsymbol{\theta}$, chosen adaptively based on the outcome of prior evaluations. (In the case of gradient-based optimization, one can use the parameter shift rule [766, 918, 321, 1043] or finite difference methods.) Each evaluation of the cost function corresponds to approximating an expectation value to some additive error ε using finite measurement shots, where ε should be chosen to be sufficiently small for accurate optimization over the landscape. Specifically, it should be expected that ε is at most $\mathcal{O}(\sqrt{\mathbb{V}_{\boldsymbol{\theta}}[C(\boldsymbol{\theta})]})$ in order to accurately distinguish arbitrarily chosen points on the parameter landscape, where $\mathbb{V}_{\boldsymbol{\theta}}$ denotes the variance over uniformly distributed parameter settings.

Caveats

The optimization of certain parameterized quantum circuits is known to be subject to the detrimental phenomena of "barren plateaus," in which deviations between different cost values with high probability (or deterministically, depending on the setting) vanish exponentially with increasing number of qubits [663, 756, 240, 534, 742, 926, 662, 396, 850]. This is often characterized by observing that $\mathbb{V}_{\boldsymbol{\theta}}[C(\boldsymbol{\theta})] = \mathcal{O}(2^{-\beta n})$ for some $\beta > 0$ [60]. This mandates an exponential shot complexity for each evaluation of a cost value in order to reliably navigate the cost landscape (when probabilistic, this should be considered an average-case phenomenon). Note that this affects both gradient-based and gradient-free optimization strategies. Moreover, it has been found that many standard techniques to avoid vanishing gradients render the VQA classically simulable [243].

If VQAs are run on noisy devices, the effects of noise are known to severely restrict the scope for computation [12, 114, 1021, 400, 336]. This effect is amplified on devices with limited hardware connectivity, where one has to use additional circuit depth to compile generic gates [400, 1021].

Finally, in general, there is a lack of end-to-end theoretical guarantees for variational quantum algorithms. In order to show advantage over classical algorithms, at minimum one has to satisfy all of the properties laid out above. In particular, the classical parameter optimization is generally left as a heuristic subroutine. This optimization task is in general NP-hard, and can be burdened by many local minima of poor quality [154, 41]. This leads to a slow optimization process and many cost values may need to be evaluated.

Example use cases

- Quantum chemistry and condensed matter physics (ground state energy): The ground state and ground state energy of a given Hamiltonian H can be found by minimizing the cost $\langle\psi(\boldsymbol{\theta})|H|\psi(\boldsymbol{\theta})\rangle$, where $|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|\psi_0\rangle$ for some input state $|\psi_0\rangle$ [833]. This is known as the variational quantum eigensolver (VQE) algorithm. A widely used ansatz for fermionic Hamiltonians is the unitary coupled cluster (UCC) ansatz [982, 833, 197, 668, 786, 748, 629, 922]. Beyond direct variational optimization on a quantum computer, alternative hybrid classical-quantum algorithms have been proposed to use a variational circuit as trial states for quantum Monte Carlo [550, 539] or combining the use of VQE with density matrix embedding theory [557, 923].

- Combinatorial optimization: In the quantum approximate optimization algorithm (QAOA), combinatorial problems on bit-strings can be encoded in the Pauli-Z basis with Hamiltonian H_P [384]. By finding the state that minimizes $\langle\phi(\boldsymbol{\theta})|H_P|\phi(\boldsymbol{\theta})\rangle$, where $|\phi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|0\rangle$, the optimal bit-string can be extracted by sampling the optimized state in the computational basis. A widely studied ansatz for this problem is the quantum alternating operator ansatz (which bears the same acronym as the algorithm), inspired by Trotterized adiabatic evolution [482]. The ansatz takes the form $U(\boldsymbol{\gamma},\boldsymbol{\beta}) = \prod_{l=1}^{p} \mathrm{e}^{-\mathrm{i}\beta_l H_M}\mathrm{e}^{-\mathrm{i}\gamma_l H_P}$ where H_M is a specific "mixing" Hamiltonian. This ansatz is known to be computationally universal (when $p \to \infty$) for certain classes of Hamiltonians [706, 780]. Moreover, under reasonable complexity-theoretic assumptions, it is known that sampling from the output of the QAOA at $p = 1$ is classically hard [380]. On the other hand, there is evidence that shallow (small p) QAOA does not perform well [194, 507, 385, 386], leading to intuition that p may need to grow with problem size to produce better approximate solutions than what can be easily found classically. Alternatively, there is some evidence that an exponential number of samples from shallow QAOA circuits may yield polynomial speedups over classical methods for finding exactly optimal solutions [179, 928]; see Section 4.2 on beyond-quadratic speedups for combinatorial optimization.
- Linear system solvers: Given matrix A and vector b encoded in a quantum state $|b\rangle$, the goal is to variationally prepare a quantum state $|x\rangle$ with amplitudes proportional to elements of the vector $x = A^{-1}b$ [187, 1060, 541]. The strategy employed is to minimize the cost $\langle\tilde{x}(\boldsymbol{\theta})|H_L|\tilde{x}(\boldsymbol{\theta})\rangle$, where $|\tilde{x}(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|0\rangle$ and $H_L = A^\dagger(I - |b\rangle\langle b|)A$. These approaches require the assumption that A has a decomposition into a sum of a small number of efficiently implementable unitaries. Here the absolute value of the cost function bounds the approximation error. A numerical study up to 30 qubits showed favorable scaling in the time to solution with respect to the matrix size, condition number, and precision [187].
- Compiling: An interesting near-term application could be to approximate a given unitary V with native gate sequence $U(\boldsymbol{\theta})$. This can lead to a compressed approximate implementation of the unitary. One option is to construct a cost function via the Hilbert–Schmidt test circuit to evaluate $1 -$

$|\langle\Phi|V^* \otimes U(\boldsymbol{\theta})|\Phi\rangle|^2 = 1 - \left|2^{-n}\,\mathrm{tr}[V^\dagger U(\boldsymbol{\theta})]\right|^2$, where $|\Phi\rangle$ is the maximally entangled state [613].

- Quantum dynamics: VQAs can also be constructed to simulate real-time and imaginary-time evolution. We discuss two examples. First, by using the compiling primitives as outlined above, compressed gate sequences for short-time evolution can be found [119, 757, 967, 983, 159]. This in turn can also be used to fast forward and simulate evolution at long times [294, 305, 420]. Second, a line of work has considered simulating open and closed dynamics, as well as imaginary-time evolution, via McLachlan's variational principle [758], which gives a linear equation for parameter dynamics $\sum_j M_{i,j}\dot{\theta}_j = V_i$ describing the evolution of a parameterized state $|\psi(\boldsymbol{\theta})\rangle$ [684, 1074, 751]. Here, $M_{i,j}$ and V_i can both be found on a quantum computer with basic circuit primitives, leading to a hybrid classical-quantum algorithm. While a full review of variational quantum algorithms for dynamics is beyond the scope of this section, we refer the reader to more complete reviews in [1074, 241].
- Factoring: Variational methods for factoring have been proposed that exploit a mapping between the factoring problem and that of finding the ground state of an Ising Hamiltonian [39]. The authors use the QAOA ansatz and heuristically find that $p = \mathcal{O}(n)$ rounds of the ansatz can lead to a good solution overlap for small system sizes.
- Machine learning: Here one employs a parameterized quantum circuit to construct a hypothesis family. Variational methods have been proposed for both classical and quantum data for classification [919, 766, 914, 513, 306], generative models [1008, 117, 362], autoencoders [879, 1014, 1009], and beyond [878, 547]. Specific ansatzes have been proposed in these contexts, sometimes referred to as *quantum neural networks*, in analogy with their classical counterparts. "Classically inspired" quantum neural networks have been proposed, such as perceptron-based QNNs [22, 1014, 381, 112] and a quantum analog to the convolutional neural network [306], as well as approaches based on tensor networks [451, 548].

Further reading

- See [241, 148] for extensive reviews of VQAs, including a summary of different widely studied ansatzes, applications, and challenges.