Variational Quantum Semi-Definite Programming

Mark M. Wilde

School of Electrical and Computer Engineering

Cornell University

Joint work with Hanna Westerheim, Jingxuan Chen, Zoë Holmes, Ivy Luo, Theshani Nuradha, Dhrumil Patel, Soorya Rethinasamy, and Kathie Wang

Available as arXiv:2312.03083 and arXiv:2312.03830

Research Team

Ivy Hanna Jenny Mark Kathie Theshani

Dhrumil

Soorya

Zoë

Motivation and Goal

- Optimization is a key task for which quantum computers could possibly provide a speedup over classical algorithms
- Semi-definite programs (SDPs) have wide applicability in computer science, engineering, mathematics, and physics
- This talk:
  1) brief overview of SDP theory
  2) QSlack method for estimating their optimal values

Background on
Semi-Definite Programming

Background on Semi-Definite Programming

- An SDP involves optimizing a linear objective function over the cone of positive semi-definite matrices, with affine constraints
- Every SDP (called primal SDP) has a corresponding dual SDP
- Dual SDP involves opposite kind of optimization and provides a bound on primal SDP

Primal and Dual SDPs

- SDP inputs are Hermitian matrices A and B and a Hermiticity-preserving superoperator Φ
- Primal SDP: $\alpha := \sup_{X \geq 0} \{\mathrm{Tr}[AX] : \Phi(X) \leq B\}$
- Dual SDP: $\beta := \inf_{Y \geq 0} \left\{\mathrm{Tr}[BY] : \Phi^\dagger(Y) \geq A\right\}$

Adjoint of a Superoperator

- Recall Hilbert-Schmidt inner product: $\langle C, D\rangle := \mathrm{Tr}[C^\dagger D]$
- Superoperator $\Phi^\dagger$ is adjoint of Φ
- It is the unique superoperator satisfying $\langle Y, \Phi(X)\rangle = \left\langle \Phi^\dagger(Y), X\right\rangle$ for all X, Y

SDP Duality Theory

- Weak duality $\alpha \leq \beta$ always holds

$$\alpha := \sup_{X \geq 0} \{\mathrm{Tr}[AX] : \Phi(X) \leq B\}$$

$$\beta := \inf_{Y \geq 0} \left\{\mathrm{Tr}[BY] : \Phi^\dagger(Y) \geq A\right\}$$

- Indeed, for all feasible X and Y,

$$\mathrm{Tr}[AX] \leq \mathrm{Tr}[\Phi^\dagger(Y)X] = \mathrm{Tr}[Y\Phi(X)] \leq \mathrm{Tr}[YB]$$

- Slater's condition $\Rightarrow$ strong duality $\alpha = \beta$

- Strong duality holds in practice for many SDPs

Application of SDP Duality

- SDP duality theory allows for bounding the optimal value α from both above and below
- Key feature that we make use of in QSlack, for providing certified bounds on the optimal value

Quantum Mechanics and Semi-Definite Constraints

- Postulates of quantum mechanics: all its constituents are described by constrained positive semi-definite matrices
- Quantum state $\rho \in \mathscr{D}$ satisfies $\rho \geq 0$ and $\mathrm{Tr}[\rho] = 1$
- Quantum measurement $(\Lambda_x)_{x \in \mathscr{X}}$ satisfies $\Lambda_x \geq 0 \ \ \forall x \in \mathscr{X}$ and $\displaystyle\sum_{x \in \mathscr{X}} \Lambda_x = I$
- Choi matrix of a channel $\mathscr{N}$ satisfies $\Gamma^{\mathscr{N}}_{RB} \geq 0$ and $\mathrm{Tr}_B[\Gamma^{\mathscr{N}}_{RB}] = I_R$
- Thus natural for SDPs to play a role in q. information and vice versa

Ground-State Energy as an SDP

- Finding the ground-state energy of a Hamiltonian H is an SDP:

$$\lambda_{\min}(H) = \inf_{\rho \in \mathcal{D}} \mathrm{Tr}[H\rho]$$

- Dual SDP is

$$\lambda_{\min}(H) = \sup_{\eta \in \mathbb{R}} \left\{ \eta : \eta I \leq H \right\}$$

- Key example of interest in physics

Normalized Trace Distance as an SDP

- Key distinguishability measure of quantum states ρ and σ is normalized trace distance

- Can be written as primal and dual SDPs:

$$\frac{1}{2}\|\rho-\sigma\|_1 = \sup_{\Lambda\geq 0}\left\{\mathrm{Tr}[\Lambda(\rho-\sigma)] : \Lambda \leq I\right\} = \inf_{Z\geq 0}\left\{\mathrm{Tr}[Z] : Z \geq \rho-\sigma\right\}$$

Four Steps of QSlack

(focusing on primal SDP)

Step 1 of QSlack: Introduce Slack Variables

- Introduce slack variable $W \geq 0$ to replace inequality constraint $\Phi(X) \leq B$ with equality constraint $B - \Phi(X) = W$:

$$
\begin{aligned}
\alpha &= \sup_{X \geq 0} \{\mathrm{Tr}[AX] : \Phi(X) \leq B\} \\
&= \sup_{X, W \geq 0} \{\mathrm{Tr}[AX] : B - \Phi(X) = W\}
\end{aligned}
$$

- Follows because $\Phi(X) \leq B$ is equivalent to existence of $W \geq 0$ such that $B - \Phi(X) = W$

Step 2 of QSlack: Introduce Penalty Terms

- Use penalty term to move equality constraint into objective function:

$$\alpha(c) := \sup_{X,W\geq 0} \left\{ \mathrm{Tr}[AX] - c \left\| B - \Phi(X) - W \right\|_2^2 \right\}$$

where $c > 0$ is a penalty constant

- From optimization theory, we know that $\lim_{c\to\infty} \alpha(c) = \alpha$

- Use finite but large value of c to approximate $\lim_{c\to\infty} \alpha(c) = \alpha$

Step 3 of QSlack: Introduce Density Matrices

- Replace optimization over PSD matrices with scaled density matrices:

$$\alpha(c) = \sup_{X,W \geq 0} \left\{ \mathrm{Tr}[AX] - c \left\| B - \Phi(X) - W \right\|_2^2 \right\}$$

$$= \sup_{\substack{\lambda, \mu \geq 0, \\ \rho, \sigma \in \mathcal{D}}} \left\{ \lambda \, \mathrm{Tr}[A\rho] - c \left\| B - \lambda\Phi(\rho) - \mu\sigma \right\|_2^2 \right\}$$

- As before, $\lim_{c \to \infty} \alpha(c) = \alpha$

Step 4 of QSlack: Parameterize States

- Final step: Replace optimization over all states with optimization over parameterized states:

$$\tilde{\alpha}(c) := \sup_{\lambda,\mu\geq 0,\theta_1,\theta_2\in\Theta} \left\{ \lambda \operatorname{Tr}[A\rho(\theta_1)] - c \left\| B - \lambda\Phi(\rho(\theta_1)) - \mu\sigma(\theta_2) \right\|_2^2 \right\},$$

- Then $\alpha(c) \geq \tilde{\alpha}(c)$ holds because set of parameterized states is contained in set of all states

Steps of QSlack for Dual SDP

- Steps are similar and shown concisely here

$$\beta := \inf_{Y \geq 0} \left\{ \mathrm{Tr}[BY] : \Phi^\dagger(Y) \geq A \right\}$$

$$= \inf_{Y, Z \geq 0} \left\{ \mathrm{Tr}[BY] : \Phi^\dagger(Y) - A = Z \right\}$$

$$\beta(c) := \inf_{Y, Z \geq 0} \left\{ \mathrm{Tr}[BY] + c \left\| \Phi^\dagger(Y) - A - Z \right\|_2^2 \right\}$$

$$\lim_{c \to \infty} \beta(c) = \beta$$

$$= \inf_{\substack{\kappa, \nu \geq 0, \\ \tau, \omega \in \mathscr{D}}} \left\{ \kappa \, \mathrm{Tr}[B\tau] + c \left\| \kappa \Phi^\dagger(\tau) - A - \nu\omega \right\|_2^2 \right\}$$

$$\geq \tilde{\beta}(c) := \inf_{\kappa, \nu \geq 0, \theta_3, \theta_4 \in \Theta} \left\{ \kappa \, \mathrm{Tr}[B\tau(\theta_3)] + c \left\| \kappa \Phi^\dagger(\tau(\theta_3)) - A - \nu\omega(\theta_4) \right\|_2^2 \right\},$$

QSlack

Primal		Dual
$\alpha = \sup_{X \geq 0} \{\mathrm{Tr}[AX] : \Phi(X) \leq B\}$	Constrained forms Weak duality: $\alpha \leq \beta$	$\beta = \inf_{Y \geq 0} \{\mathrm{Tr}[BY] : \Phi^\dagger(Y) \geq A\}$
$\alpha = \sup_{X,W \geq 0} \{\mathrm{Tr}[AX] : B - \Phi(X) = W\}$	Introduce slack variables	$\beta = \inf_{Y,Z \geq 0} \{\mathrm{Tr}[BY] : \Phi^\dagger(Y) - A = Z\}$
$\alpha(c) = \sup_{X,W \geq 0} \left\{\mathrm{Tr}[AX] - c\Vert B - \Phi(X) - W\Vert_2^2\right\}$ $\alpha = \lim_{c\to\infty} \alpha(c)$	Unconstrained forms (introduce penalty terms)	$\beta(c) = \inf_{Y,Z \geq 0} \left\{\mathrm{Tr}[BY] + c\left\Vert \Phi^\dagger(Y) - A - Z\right\Vert_2^2\right\}$ $\beta = \lim_{c\to\infty} \beta(c)$
$\alpha(c) = \sup_{\substack{\lambda,\mu \geq 0, \\ \rho,\sigma \in \mathcal{D}}} \left\{\lambda \,\mathrm{Tr}[A\rho] - c\Vert B - \lambda\Phi(\rho) - \mu\sigma\Vert_2^2\right\}$ $\alpha = \lim_{c\to\infty} \alpha(c)$	Write positive semi-definite matrices as scaled density matrices	$\beta(c) = \inf_{\substack{\kappa,\nu \geq 0, \\ \tau,\omega \in \mathcal{D}}} \left\{\kappa \,\mathrm{Tr}[B\tau] + c\left\Vert \kappa\Phi^\dagger(\tau) - A - \nu\omega\right\Vert_2^2\right\}$ $\beta = \lim_{c\to\infty} \beta(c)$
$\tilde{\alpha}(c) = \sup_{\substack{\lambda,\mu \geq 0, \\ \theta_1,\theta_2 \in \Theta}} \left\{\lambda \,\mathrm{Tr}[A\rho(\theta_1)] - c\Vert B - \lambda\Phi(\rho(\theta_1)) - \mu\sigma(\theta_2)\Vert_2^2\right\}$ $\lim_{c\to\infty} \tilde{\alpha}(c) \leq \alpha$	Parameterize density matrices	$\tilde{\beta}(c) = \inf_{\substack{\kappa,\nu \geq 0, \\ \theta_3,\theta_4 \in \Theta}} \left\{\kappa \,\mathrm{Tr}[B\tau(\theta_3)] + c\left\Vert \kappa\Phi^\dagger(\tau(\theta_3)) - A - \nu\omega(\theta_4)\right\Vert_2^2\right\}$ $\beta \leq \lim_{c\to\infty} \tilde{\beta}(c)$

Estimate the objective function using a quantum computer + Perform the optimization using a classical optimizer

Key Theoretical Insight of QSlack

- QSlack bounds on the optimal SDP values α and β:

$$\tilde{\alpha} \leq \alpha \leq \beta \leq \tilde{\beta}$$

where

$$\tilde{\alpha} := \sup_{\substack{\lambda, \mu \geq 0, \\ \theta_1, \theta_2 \in \Theta}} \left\{ \lambda \operatorname{Tr}[A\rho(\theta_1)] : B - \lambda \Phi(\rho(\theta_1)) = \mu \sigma(\theta_2) \right\},$$

$$\tilde{\beta} := \inf_{\substack{\kappa, \nu \geq 0, \\ \theta_3, \theta_4 \in \Theta}} \left\{ \kappa \operatorname{Tr}[B\tau(\theta_3)] : \kappa \Phi^\dagger(\tau(\theta_3)) - A = \nu \omega(\theta_4) \right\},$$

- In the ideal case in which the optimal $\tilde{\alpha}$ and $\tilde{\beta}$ can be found, then QSlack bounds the optimal SDP values α and β from above and below

How to Parameterize n-Qubit Mixed States?

- Purification ansatz: generate parameterized pure state on $2n$ qubits and trace over half of them:

$$\rho(\theta) = \mathrm{Tr}_R[U(\theta)|0\rangle\langle 0|_{RS} U^\dagger(\theta)]$$

- Convex combination ansatz (CCA): sample x randomly from a parameterized neural net & then input $|x\rangle$ to a parameterized unitary:

$$\rho(\varphi, \gamma) = \sum_x p_\varphi(x) U(\gamma) |x\rangle\langle x| U^\dagger(\gamma)$$

- Advantage of CCA is that only n qubits are needed

Executing QSlack on a Quantum Computer

- Suppose A, B, and Φ correspond to efficiently measurable observables
- Then a q. computer can measure all objective-function terms efficiently (using expectations of observables or destructive SWAP tests)
- In general, it is not clear how to do so on a classical computer
- Hope is that QSlack might give a q. advantage over classical methods

Performing Optimization

- Estimate objective function and its gradient ∇_θ using q. computer and parameter-shift rule (can also use SPSA)
- Then employ gradient descent or its variants

Example: Ground-State Energy

- Recall the ground-state energy problem:

$$\lambda_{\min}(H) = \inf_{\rho \in \mathscr{D}} \operatorname{Tr}[H\rho] = \sup_{\eta \in \mathbb{R}} \left\{ \eta : \eta I \leq H \right\}$$

- For the minimization, apply the variational quantum eigensolver (VQE)
- VQE upper bounds $\lambda_{\min}(H)$, but quality of bound is unclear
- For the maximization, apply QSlack to lower bound $\lambda_{\min}(H)$ (called Dual-VQE for this application)

Example: Ground-State Energy (Ctd.)

- In more detail, Dual-VQE method:

$$\begin{aligned}
\sup_{\eta\in\mathbb{R}} \left\{\eta : \eta I \le H\right\} &= \sup_{\eta\in\mathbb{R}, W\ge 0} \left\{\eta : H - \eta I = W\right\} \\
&= \sup_{\eta \in \mathbb{R}, \nu \ge 0, \omega \in \mathscr{D}} \left\{\eta : H - \eta I = \nu\omega\right\} \\
&= \lim_{c\to\infty} \sup_{\eta \in \mathbb{R}, \nu \ge 0, \omega \in \mathscr{D}} \left\{\eta - c \left\| H - \eta I - \nu\omega \right\|_2^2\right\} \\
&\ge \lim_{c\to\infty} \sup_{\eta \in \mathbb{R}, \nu \ge 0, \theta \in \Theta} \left\{\eta - c \left\| H - \eta I - \nu\omega(\theta) \right\|_2^2\right\}
\end{aligned}$$

Example: Ground-State Energy (Ctd.)

- Expanding the objective function gives

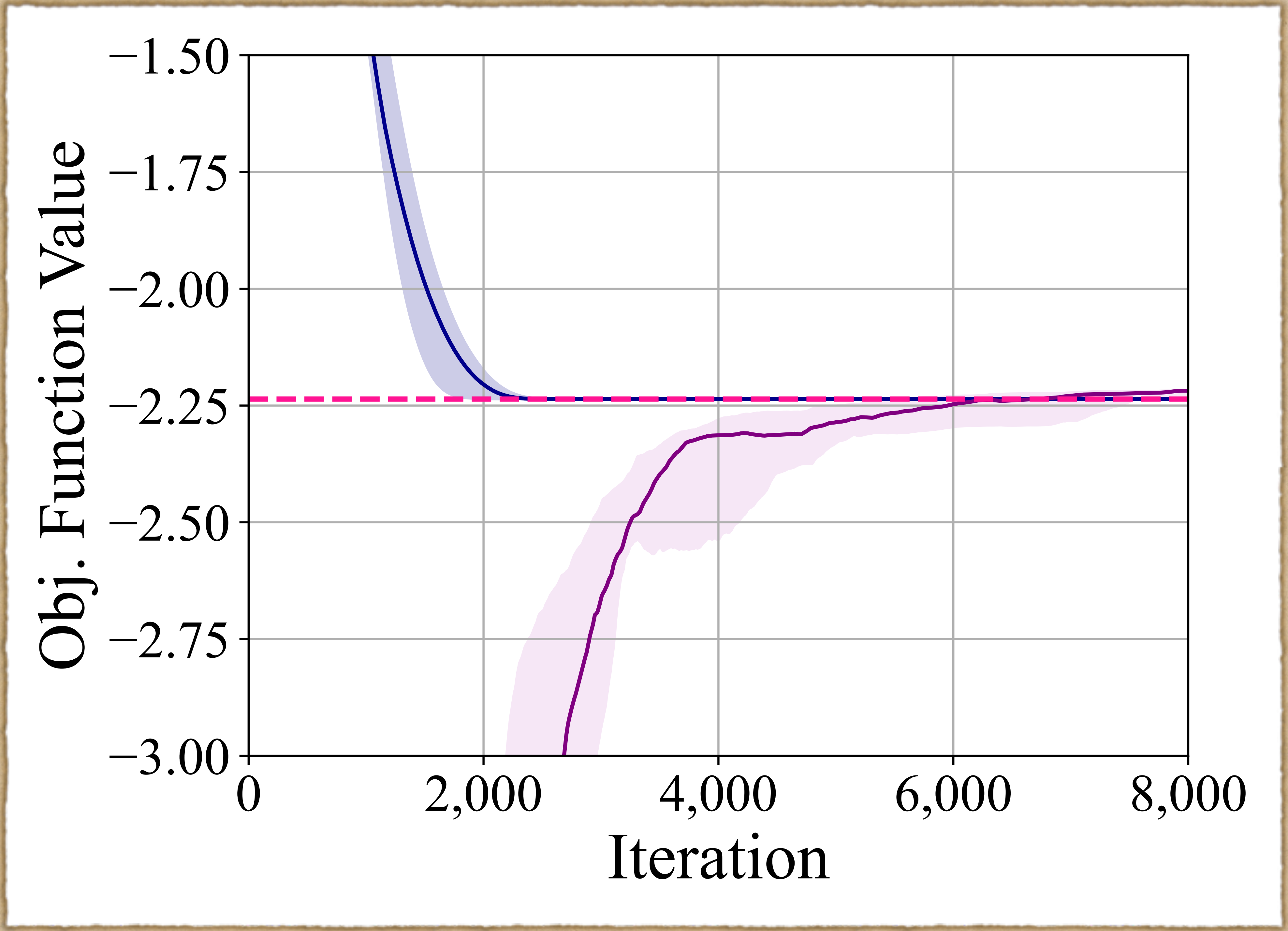
$$\sup_{\eta\in\mathbb{R},\nu\geq 0,\theta\in\Theta} \eta - c\Big(\mathrm{Tr}[H^2] + \eta^2 2^n + \nu^2\,\mathrm{Tr}[\omega(\theta)^2] - 2\eta\,\mathrm{Tr}[H] - 2\nu\,\mathrm{Tr}[H\omega(\theta)] + 2\eta\nu\Big)$$

- Various terms can be precomputed or easily calculated, the term $\mathrm{Tr}[\omega(\theta)^2]$ can be estimated by classical sampling, and the term $\mathrm{Tr}[H\omega(\theta)]$ can be estimated by a quantum computer

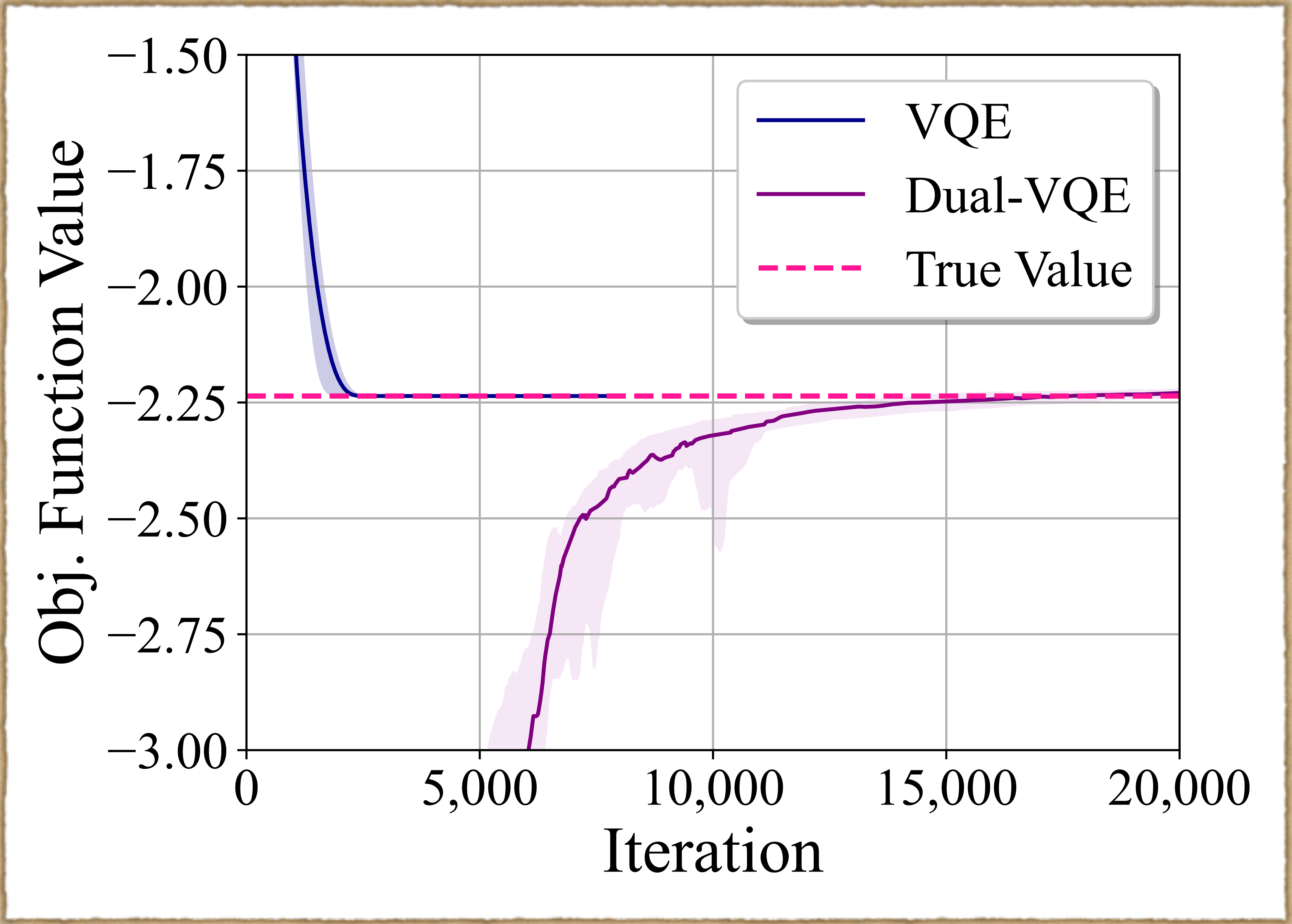
- Overhead of using quantum computer is the same as VQE

Simulation Results

Two-qubit, transverse-field Ising model



Purification ansatz



Convex combination ansatz

Comparing VQE and Dual-VQE

- In simulations, VQE converges faster than Dual-VQE does
- Possible reasons: Dual-VQE objective function is more complex, non-linear in $\omega(\theta)$, involves global term $\mathrm{Tr}[\omega(\theta)^2]$, and optimization over mixed states
- Differences in optimization landscapes can lead to different resource requirements for VQE and Dual-VQE

Obstacle: Barren Plateau Problem

- The barren plateau problem plagues variational quantum algorithms, and it can affect QSlack also
- In short, gradients of cost functions vanish exponentially with problem size
- Most promising approach (arXiv:2208.13673) seems to be finding a good initial state via tensor networks (classical simulation) and then continuing after this starting point with the usual hybrid quantum-classical algorithm
- This should work for Dual-VQE and constrained Hamiltonian optimization, but more work is needed to figure out how to apply it to other QSlack problems that have q. states as input

Conclusion

- QSlack is a general method for bounding the optimal value of an SDP from above and below
- All objective-function terms can be efficiently estimated on a q. computer if SDP inputs are efficiently measurable observables
- Showcased the method on the ground-state energy problem

Future Directions

- Determine behavior of QSlack for finite values of the penalty constant c
- Scale up simulations to larger numbers of qubits and incorporate error mitigation
- Modify QSlack to be an interior-point method