

Quantum Tilted Loss in Variational Quantum Optimization: Theory and Applications

Yixian Qiu¹ Josep Lumbrecas^{1,2} Xiufan Li¹ Patrick Rebentrost^{1,3}

¹ Centre for Quantum Technologies, National University of Singapore, Singapore

² Nanyang Quantum Hub, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore

³ Department of Computer Science, National University of Singapore, Singapore

Introduction.

Variational quantum algorithms (VQAs) prepare a parameterized quantum state $\rho(\theta)$, estimate a scalar objective from measurements, and update the parameters using a classical optimizer. This hybrid template includes VQE and QAOA, and is one of the most practical routes for using near-term quantum devices [1–3]. A major obstacle is trainability: for expressive ansätze or global observables, the standard expectation-value objective $L_0(O, \rho(\theta)) = \text{Tr}[O\rho(\theta)]$ can have exponentially small gradients, producing barren plateaus in the optimization landscape [4, 5].

The problem is not only geometric. In practice, the optimizer sees finite-shot estimates of losses and gradients, not the exact landscape. Thus a useful objective should both strengthen the optimization signal and remain estimable from a feasible number of measurements. Existing modifications, such as local costs and CVaR objectives, change which part of the measurement distribution is emphasized [5, 6]. What is missing is a simple framework that exposes both the landscape reshaping and its sampling cost.

We introduce the Quantum Tilted Loss (QTL), a quantum analogue of entropic risk. QTL applies exponential tilting to the measurement-outcome distribution of an observable. The tilt is controlled by one parameter γ : $\gamma = 0$ recovers the expectation value, $\gamma < 0$ emphasizes low-energy outcomes, and $\gamma > 0$ emphasizes high-energy outcomes. Larger tilts can amplify useful gradient signals, but they also increase the shot cost of estimating nonlinear objectives and gradients. QTL should therefore be viewed not as a universal solution to barren plateaus, but as a principled loss-design tool for studying the trade-off between trainability and estimability. Our contributions:

- **A tunable quantum loss framework.** We define QTL as an outcome-level entropic tilt of a measured observable. It recovers the expectation value at zero tilt, approaches extremal spectral outcomes at large tilt, and preserves the global minimizers of the target Hamiltonian.
- **Connections to existing objectives.** We relate QTL to lower-tail CVaR and Gibbs objectives, showing that several tunable losses can be understood through the same structure.

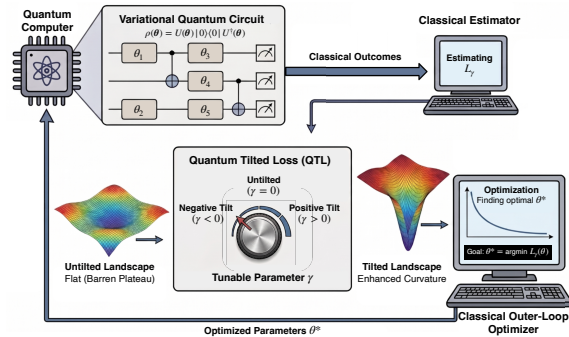


FIG. 1. Quantum-classical loop for QTL optimization.

- **Trainability versus estimability.** We prove finite-shot estimation and gradient identities, analyze how QTL reshapes local curvature, and show in projector and QAOA benchmarks that moderate or scheduled tilts can help, whereas overly large tilts become sampling-limited.

Definition and properties.

Let $O = \sum_i o_i \Pi_i$ be a Hermitian observable and let ρ be a quantum state. Measuring O on ρ induces the Born distribution $p_i = \text{Tr}(\rho \Pi_i)$ over outcomes o_i . For a tilt parameter $\gamma \in \mathbb{R}$, we define the Quantum Tilted Loss (QTL) as

$$L_\gamma(O, \rho) = \frac{1}{\gamma} \log \text{Tr}(e^{\gamma O} \rho), \quad \gamma \neq 0, \quad (1)$$

and $L_\gamma(O, \rho) = \text{Tr}(O\rho)$ when $\gamma = 0$. Thus QTL is the entropic tilt of the measurement-outcome distribution. For minimization problems, negative γ emphasizes low-energy outcomes; positive γ emphasizes high-energy outcomes; and $\gamma \rightarrow 0$ recovers the usual expectation value.

Faithfulness. If $o_{\min}$ is the smallest eigenvalue of O and $\Pi_{\min}$ is the projector onto its ground eigenspace, then for every $\gamma \neq 0$, $\min_\rho L_\gamma(O, \rho) = o_{\min}$. Therefore QTL changes the geometry of the loss landscape, but not the target solution. If the ansatz can represent a ground state, minimizing $L_\gamma(O, \rho(\theta))$ has the same global optimum as minimizing the standard energy.

Relation to other objectives. QTL also clarifies the relationship between several objective functions already used in variational optimization such as CVaR introduced in [6]. Lower-tail CVaR averages the best α fraction of measured energies, whereas QTL smoothly reweights all outcomes. For $\gamma < 0$, we prove the corrected entropic bound $\text{CVaR}_\alpha(\theta) \geq L_\gamma(O, \rho(\theta)) + \frac{1}{\gamma} \log(\frac{1}{\alpha})$, which shows that QTL is a smooth entropic counterpart of lower-tail CVaR [6]. Also, QTL provides a single lens on expectation-value training, tail-sensitive objectives, and Gibbs-type objectives [7, 8]. Unlike local costs, QTL does not replace the target observable; unlike CVaR, it gives a smooth log-partition loss; and unlike Gibbs objectives, it is defined directly from the measured spectral distribution of the observable.

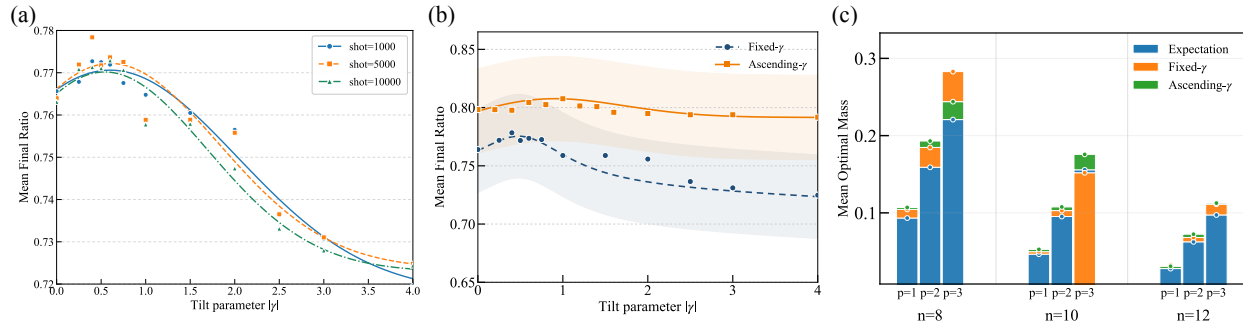


FIG. 2. (a) Fixed tilts show a non-monotonic dependence on $|\gamma|$. (b) An ascending-tilt schedule improves stability in the finite-shot regime. (c) Tilting can increase the probability mass assigned to optimal bitstrings.

Main results: trainability–estimability trade-off.

The benefit of QTL is geometric: it can make selected spectral outcomes more visible to the optimizer. The cost is statistical: nonlinear exponential weights can increase the variance of finite-shot estimators.

Result 1: finite-shot estimation. For diagonal Hamiltonians, $O = \sum_{z \in \{0,1\}^n} E(z) |z\rangle\langle z|$, computational-basis measurements give energy samples $X = E(z)$. The tilted loss is estimated by $\hat{L}_{\gamma,m} = \frac{1}{\gamma} \log \left(\frac{1}{m} \sum_{s=1}^m e^{\gamma X_s} \right)$. If $X \in [\lambda_{\min}, \lambda_{\max}]$ and $\Delta = \lambda_{\max} - \lambda_{\min}$, then it is sufficient to use

$$m = \tilde{O} \left(\frac{(e^{|\gamma| \Delta} - 1)^2}{\gamma^2 \epsilon^2} \log \frac{1}{\delta} \right)$$

samples to estimate L_γ to additive error ε with failure probability at most δ . Hence a stronger tilt can sharpen the objective, but the worst-case sampling cost grows with $|\gamma|\Delta$.

Result 2: gradient estimation. Let $Z_\gamma(\theta) = \text{Tr}(e^{\gamma O} \rho(\theta))$. If parameter θ_k enters through a gate generated by an operator with two eigenvalues $\pm v_k$, then QTL admits the exact parameter-shift identity $\partial_{\theta_k} L_\gamma(\theta) = \frac{v_k}{\gamma} \frac{Z_\gamma(\theta_k^+) - Z_\gamma(\theta_k^-)}{Z_\gamma(\theta)}$. Unlike the expectation-value case, this is not a central difference of tilted losses. It is a normalized difference of tilted partition functions, so finite-shot errors are amplified by both the exponential weights and the normalization.

Result 3: landscape reshaping. The Hessian decomposes as $\nabla^2 L_\gamma = (\gamma Z_\gamma)^{-1} \nabla^2 Z_\gamma - \gamma \nabla L_\gamma \nabla L_\gamma^\top$. For $\gamma < 0$, the rank-one term increases curvature along directions aligned with the gradient; for $\gamma > 0$, it suppresses this curvature. Thus QTL does not merely rescale the landscape. It reshapes it anisotropically. The key question is whether this reshaped geometry can be resolved within the available shot budget [9].

Application: Structured benchmark and tilted-QAOA.

We first illustrate QTL in a solvable projector benchmark. Let the target state be $|0\rangle^{\otimes n}$, use the product ansatz $V(\theta) = \bigotimes_{j=1}^n e^{i\theta_j \sigma_x^{(j)}/2}$, and take $O_G = \mathbb{I} - |0\rangle\langle 0|^{\otimes n}$. Under this benchmark model, QTL preserves the global minimizers while changing the landscape geometry. In this model, fixed small tilt keeps the usual exponential gradient-variance decay, whereas the explicit negative schedule $\gamma(n) \sim n$ gives the polynomial lower bound $\text{Var}[\partial_{\theta_k} L_{\gamma(n)}] = \Omega(1/n^2)$. This is a structured benchmark result rather than a general barren-plateau avoidance theorem: it relies on the observable, ansatz and an n -dependent tilt.

We also test QTL in QAOA for MaxCut problem. For a graph $G = (V, E)$, we use the diagonal cost Hamiltonian $H_C = \sum_{(i,j) \in E} \frac{1}{2} (Z_i Z_j - \mathbb{I})$, and replace the expectation-value objective by QTL. Since H_C is diagonal, QTL is estimated directly from sampled bitstring energies.

Figure 2 summarizes the tilting effect. On Erdős-Rényi MaxCut instances, the best fixed tilts occur at moderate negative values. Larger tilts are less reliable because the exponentially weighted estimator has higher finite-shot variance. A linear continuation strategy, starting from $\gamma = 0$ and gradually increasing it, is more stable under limited shots and often places more probability on optimal bitstrings.

The reason why we cannot go to arbitrary large tilts is captured by the noise-to-signal ratio $\mathcal{R}_\gamma = \frac{\sigma_{\text{shot}}^2}{N \|\nabla L_\gamma(\theta)\|^2}$. Gradient information is resolvable only when $\mathcal{R}_\gamma \lesssim 1$. Thus negative tilting can improve trainability by increasing the gradient signal, but may reduce estimability due to the finite-shot sample-complexity bound (Result 1). This formalizes the estimability–trainability trade-off for non-linear losses.

Conclusion and outlook.

We introduced the QTL as a tunable objective for VQAs. It applies exponential tilting to the measurement-outcome distribution of an observable, recovering the standard expectation value at zero tilt while emphasizing selected spectral regions for nonzero tilt. The framework preserves the global minimizers of the target Hamiltonian, connects naturally to CVaR and Gibbs objectives, and gives a clear mechanism for reshaping landscapes. Its main lesson is a trainability–estimability trade-off: negative tilts can strengthen optimization signals and improve finite-shot performance in structured settings, but stronger tilts also increase sampling variance. Thus QTL should be treated as a limited resource, chosen according to the observable structure, ansatz, and measurement budget.

[1] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*,

- 5:4213, 2014.
- [2] E. Farhi, J. Goldstone, and S. Gutmann. A quantum approximate optimization algorithm. arXiv:1411.4028, 2014.
 - [3] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles. Variational quantum algorithms. *Nature Reviews Physics*, 3:625–644, 2021.
 - [4] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9:4812, 2018.
 - [5] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles. Cost-function-dependent barren plateaus in shallow parametrized quantum circuits. *Nature Communications*, 12:1791, 2021.
 - [6] P. K. Barkoutsos, G. Nannicini, A. Robert, I. Tavernelli, and S. Woerner. Improving variational quantum optimization using CVaR. *Quantum*, 4:256, 2020.
 - [7] L. Li, M. Fan, M. Coram, P. Riley, and S. Leichenauer. Quantum optimization with a novel Gibbs objective function and ansatz architecture search. *Physical Review Research*, 2:023074, 2020.
 - [8] T. Li, A. Beirami, M. Sanjabi, and V. Smith. On tilted losses in machine learning: theory and applications. *Journal of Machine Learning Research*, 24(142):1–79, 2023.
 - [9] K. Ito, W. Mizukami, and K. Fujii. Universal noise-precision relations in variational quantum algorithms. *Physical Review Research*, 5:023025, 2023.
 - [10] I. Kolotouros and P. Wallden. Evolving objective function for improved variational quantum optimization. *Physical Review Research*, 4:023225, 2022.