VARIATIONAL QUANTUM ALGORITHMS ARE LIPSCHITZ SMOOTH

Anonymous authors
Paper under double-blind review

ABSTRACT

Variational Quantum Algorithms (VQAs) are a cornerstone of near-term quantum computing, yet a full understanding of their optimization landscape is still developing. A key property for guaranteeing the convergence of gradient-based optimizers is L-smoothness, which implies a Lipschitz continuous gradient. In this work, we first formally prove that the objective function of any Quantum Neural Network (QNN) composed of standard parameterized and fixed gates is an L-smooth function of its parameters. We then derive a tight and explicit upper bound for the smoothness constant, L, using an operator-based formalism. This bound, $L \leq 4 \|M\|_2 \sum_{k=1}^{P} \|G_k\|_2^2$, depends directly on the observable M and the generators G_k of the parameterized gates, providing valuable theoretical insight for designing and optimizing VQAs.

1 INTRODUCTION

1.1 MOTIVATION AND GOAL

The primary objective of this work is to demonstrate that for a general Quantum Neural Network (QNN) architecture, the expectation value of an observable is an L-smooth function of the circuit parameters. This property is crucial as it provides a theoretical guarantee on the stability of the loss landscape, which is fundamental for the performance of gradient-based optimization algorithms commonly used to train QNNs.

1.2 DEFINITION OF L-SMOOTHNESS

A differentiable function $f(\boldsymbol{\theta})$ is said to be L-smooth if its gradient is Lipschitz continuous with a constant $L \geq 0$. This is formally stated as:

$$\|\nabla f(\boldsymbol{\theta}_1) - \nabla f(\boldsymbol{\theta}_2)\| \leq L \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| \tag{1}$$

For a twice-differentiable function, a sufficient condition for L-smoothness is that the norm of its Hessian matrix is globally bounded:

$$\left\|\nabla^2 f(\boldsymbol{\theta})\right\| \leq L \tag{2}$$

2 THE GENERAL QNN MODEL

2.1 THE QUANTUM STATE

An n-qubit quantum state is a vector in a 2^n-dimensional Hilbert space. In a QNN, this state is parameterized by a vector $\boldsymbol{\theta} \in \mathbb{R}^P$, and can be written as a linear combination of computational basis states $|i\rangle$:

$$|\psi(\boldsymbol{\theta})\rangle = \sum_{i=0}^{2^n-1} c_i(\boldsymbol{\theta}) |i\rangle \tag{3}$$

where the complex amplitudes $c_i(\boldsymbol{\theta})$ are functions of the circuit parameters $\boldsymbol{\theta}$.

2.2 THE QNN OUTPUT

The output of a QNN, which serves as the objective function in a VQA, is the expectation value of a Hermitian operator M (an observable) with respect to the final state:

$$f(\boldsymbol{\theta}) = \langle M \rangle = \langle \psi(\boldsymbol{\theta}) | M | \psi(\boldsymbol{\theta}) \rangle \tag{4}$$

This can be expanded in terms of the state amplitudes as:

$$f(\boldsymbol{\theta}) = \sum_{i,j} c_i^*(\boldsymbol{\theta}) M_{ij} c_j(\boldsymbol{\theta}) \tag{5}$$

2.3 THE GATE SET

A QNN circuit is constructed from a sequence of unitary quantum gates. For this proof, we consider a universal gate set, which can be broadly categorized as follows:

- **Parameterized Gates:** Single-qubit rotations ($R_x(\theta), R_y(\theta), R_z(\theta)$) and two-qubit parameterized rotations ($R_{xx}(\theta), R_{yy}(\theta), R_{zz}(\theta)$, etc.).
- **Constant Gates:** Fixed single-qubit gates (H, X, Y, Z, S, T) and entangling gates ($CNOT, CZ, SWAP$).

3 CORE PROOF: QNN OUTPUTS AS TRIGONOMETRIC POLYNOMIALS

We prove by induction on the number of gates, k, that all amplitudes of the state vector are multivariate trigonometric polynomials of the circuit parameters $\boldsymbol{\theta}$.

3.1 INDUCTIVE HYPOTHESIS: P(K)

After applying k gates from the universal set, all amplitudes $c_i(\boldsymbol{\theta})$ of the state vector $|\psi_k(\boldsymbol{\theta})\rangle$ are multivariate trigonometric polynomials of the parameters $\boldsymbol{\theta}$ on which they depend.

3.2 BASE CASE: K=0

For $k = 0$, the state is the initial state, typically $|\psi_0\rangle = |0...0\rangle$. The amplitudes are constants (one is 1, the rest are 0), which are trivial trigonometric polynomials of degree zero. Thus, P(0) holds.

3.3 INDUCTIVE STEP

Assume P($k-1$) is true: the amplitudes $c_j(\boldsymbol{\theta}_{1:k-1})$ of the state $|\psi_{k-1}\rangle$ are trigonometric polynomials. Now, we apply the k-th gate, U_k, to produce the new state $|\psi_k\rangle = U_k |\psi_{k-1}\rangle$. The new amplitudes c'_i are a linear combination of the old amplitudes: $c'_i = \sum_j (U_k)_{ij} c_j$.

Case 1: U_k **is a parameterized gate,** $U_k(\theta_k)$ The matrix elements of rotational gates are trigonometric functions of θ_k (e.g., $\cos(\theta_k/2)$, $\sin(\theta_k/2)$). The new amplitudes c'_i are sums of products of these elements and the old amplitudes c_j. Since the set of multivariate trigonometric polynomials is closed under multiplication and addition, each c'_i is also a multivariate trigonometric polynomial of $\boldsymbol{\theta}_{1:k}$.

Case 2: U_k **is a constant gate** The matrix elements of U_k are constant complex numbers. The new amplitudes are therefore finite linear combinations of the previous amplitudes. A linear combination of trigonometric polynomials is itself a trigonometric polynomial, so the property is preserved.

By the principle of induction, the amplitudes of the final state vector of any QNN are trigonometric polynomials of the parameters.

4 L-SMOOTHNESS OF THE QNN OUTPUT

4.1 THE OUTPUT FUNCTION AS A TRIGONOMETRIC POLYNOMIAL

The QNN output, $f(\boldsymbol{\theta}) = \sum_{i,j} c_i^*(\boldsymbol{\theta}) M_{ij} c_j(\boldsymbol{\theta})$, is a sum of products of the amplitudes (which are trigonometric polynomials) and constants M_{ij}. Therefore, $f(\boldsymbol{\theta})$ is also a multivariate trigonometric polynomial.

4.2 BOUNDED HESSIAN AND CONCLUSION

The Hessian matrix of $f(\boldsymbol{\theta})$ contains second partial derivatives, $H_{ij} = \frac{\partial^2 f}{\partial\theta_i \partial\theta_j}$. Since derivatives of trigonometric polynomials are also trigonometric polynomials, and trigonometric polynomials (being finite sums of sines and cosines) are globally bounded, every element of the Hessian is bounded. A matrix with all bounded elements has a bounded norm. Thus, $\left\|\nabla^2 f(\boldsymbol{\theta})\right\|$ is globally bounded, which proves that the QNN output is an L-smooth function of its parameters.

5 FINDING AN EXPLICIT L-BOUND

While the proof via trigonometric polynomials establishes L-smoothness, it does not easily yield a practical bound. Here, we derive an explicit bound for L using an operator-based formalism.

5.1 OPERATOR-BASED FORMULATION OF THE HESSIAN

Let the unitary evolution of the QNN be $U(\boldsymbol{\theta}) = U_P(\theta_P) \cdots U_1(\theta_1)$, where each parameterized gate is given by $U_k(\theta_k) = \exp(-i\theta_k G_k)$ for a Hermitian generator G_k. The objective function is given by:

$$f(\boldsymbol{\theta}) = \langle\psi_0| U^\dagger(\boldsymbol{\theta}) M U(\boldsymbol{\theta}) |\psi_0\rangle = \langle\psi(\boldsymbol{\theta})| M |\psi(\boldsymbol{\theta})\rangle . \tag{6}$$

The first partial derivative of $f(\boldsymbol{\theta})$ with respect to a parameter θ_k is:

$$\frac{\partial f}{\partial \theta_k} = \frac{\partial\langle\psi|}{\partial\theta_k} M|\psi\rangle + \langle\psi|M\frac{\partial|\psi\rangle}{\partial\theta_k} = 2\text{Re}\left(\langle\psi|M\frac{\partial|\psi\rangle}{\partial\theta_k}\right) \tag{7}$$

Thus the elements of the Hessian matrix can be expressed as:

$$H_{kl} = \frac{\partial^2 f}{\partial\theta_k \partial\theta_l} = 2\text{Re}\left(\frac{\partial\langle\psi(\boldsymbol{\theta})|}{\partial\theta_k} M \frac{\partial|\psi(\boldsymbol{\theta})\rangle}{\partial\theta_l} + \langle\psi(\boldsymbol{\theta})| M \frac{\partial^2|\psi(\boldsymbol{\theta})\rangle}{\partial\theta_l\partial\theta_k}\right) \tag{8}$$

5.2 BOUNDING THE STATE VECTOR DERIVATIVES

The key is to bound the norms of the state's partial derivatives. The first partial derivative is:

$$\frac{\partial|\psi(\boldsymbol{\theta})\rangle}{\partial\theta_k} = U_P \cdots U_{k+1}(-iG_k U_k) U_{k-1} \cdots U_1 |\psi_0\rangle \tag{9}$$

When taking the norm of $\|U_P \cdots U_{k+1}(-iG_k U_k) U_{k-1} \cdots U_1 |\psi_0\rangle\|$ we can eliminate the $U_P \cdots U_{k+1}$ term since products of unitary operators are themselves unitary and preserve norms. Additionally we can simplify $U_k U_{k-1} \cdots U_1 |\psi_0\rangle$ to $|\psi_k\rangle$ and $-i$ to 1. We then have the norm of $\|G_k |\psi_k\rangle\| \leq \|G_k\| \cdot \||\psi_k\rangle\|$ following the Cauchy-Schwarz inequality. Since $\||\psi_k\rangle\| = 1$ we get:

$$\left\|\frac{\partial|\psi(\boldsymbol{\theta})\rangle}{\partial\theta_k}\right\| \leq \|G_k\| \triangleq g_k \tag{10}$$

Doing similar simplifications, the norms of the second partial derivatives are bounded as follows:

- **Mixed partial derivative ($k \neq l$):**
  In this example the the second partial derivative is given by:

$$\frac{\partial^2 |\psi\rangle}{\partial \theta_l \partial \theta_k} = U_P \cdots U_{l+1}(-iG_l U_l) \cdots (-iG_k U_k) \cdots U_1 |\psi_0\rangle \tag{11}$$

  When taking the norm of $\|U_P \cdots U_{l+1}(-iG_l U_l) \cdots (-iG_k U_k) \cdots U_1 |\psi_0\rangle\|$ we can remove the term $U_P \cdots U_{l+1}$ and simplify $U_k \cdots U_1 |\psi_0\rangle$ to $|\psi_k\rangle$. This gives:

$$\left\| \frac{\partial^2 |\psi\rangle}{\partial \theta_l \partial \theta_k} \right\| = \|G_l U_l \cdots U_{k+1} G_k |\psi_k\rangle\| \leq \|G_l U_l \cdots U_{k+1} G_k\| \cdot \| |\psi_k\rangle \| \tag{12}$$

  Where $\| |\psi_k\rangle \| = 1$. By using the the submultiplicative property of norms we get:

$$\|G_l U_l \cdots U_{k+1} G_k\| \leq \|G_l\| \cdot \|U_l\| \ldots \|U_{k+1}\| \cdot \|G_k\| \tag{13}$$

  Since the norm of each unitary operator $\|U_i\| = 1$ we get the final inequality

$$\left\| \frac{\partial^2 |\psi\rangle}{\partial \theta_l \partial \theta_k} \right\| \leq \|G_l\| \|G_k\| \triangleq g_l g_k \tag{14}$$

- **Unmixed partial derivative ($k = l$):**
  The second derivative is defined as

$$\frac{\partial^2 |\psi\rangle}{\partial \theta_k^2} = U_P \cdots (-iG_k)^2 U_k \cdots U_1 |\psi_0\rangle \tag{15}$$

  Following the same logic when taking the norm, $U_P \ldots U_{k+1}$ can be removed. $(-iG_k)^2 = -G_k^2$ and $U_k \ldots U_1 |\psi_0\rangle = |\psi_k\rangle$ whose norm is one which gives:

$$\left\| \frac{\partial^2 |\psi\rangle}{\partial \theta_k^2} \right\| \leq \|G_k^2\| \triangleq g_k^2 \tag{16}$$

5.3 Establishing the Final Bound for L

We can now bound the magnitude of each Hessian element $|H_{kl}|$ using the triangle inequality and the Cauchy-Schwarz inequality:

$$\begin{aligned} |H_{kl}| &\leq 2 \left(\left| \frac{\partial \langle \psi(\boldsymbol{\theta})|}{\partial \theta_k} M \frac{\partial |\psi(\boldsymbol{\theta})\rangle}{\partial \theta_l} \right| + \left| \langle \psi(\boldsymbol{\theta})| \, M \frac{\partial^2 |\psi(\boldsymbol{\theta})\rangle}{\partial \theta_l \partial \theta_k} \right| \right) && (17) \\ &\leq 2 \left(\left\| \frac{\partial |\psi(\boldsymbol{\theta})\rangle}{\partial \theta_k} \right\| \|M\| \left\| \frac{\partial |\psi(\boldsymbol{\theta})\rangle}{\partial \theta_l} \right\| + \| |\psi(\boldsymbol{\theta})\rangle \| \, \|M\| \left\| \frac{\partial^2 |\psi(\boldsymbol{\theta})\rangle}{\partial \theta_l \partial \theta_k} \right\| \right) && (18) \end{aligned}$$

Substituting our derivative bounds gives:

$$|H_{kl}| \leq 2 \|M\| \left(g_k g_l + 1 \cdot g_l g_k \right) = 4 \|M\| \, g_k g_l \tag{19}$$

To find a bound for L, we can use the Frobenius norm of the Hessian:

$$L \leq \|H\|_F = \left(\sum_{k,l=1}^{P} |H_{kl}|^2 \right)^{1/2} \leq \left(\sum_{k,l=1}^{P} (4 \|M\| \, g_k g_l)^2 \right)^{1/2} \tag{20}$$

This simplifies to our final result:

$$\boxed{L_{bound} = 4 \|M\|_2 \sum_{k=1}^{P} \|G_k\|_2^2} \tag{21}$$

5.4 THE NORM OF THE GATE GENERATORS (G_k)

In modern VQAs, the parameterized gates are almost always single or multi-qubit Pauli rotations. For all of these standard gates, the spectral norm of their generator is constant:

- **Single-Qubit Rotations (R_i):** For gates like $R_x(\theta), R_y(\theta)$, or $R_z(\theta)$, the generator is $G_k = \frac{1}{2}P_i$, where $P_i \in \{X, Y, Z\}$ is a Pauli matrix . The spectral norm of any Pauli matrix is $\|P_i\|_2 = 1$. Consequently, the generator norm is always:

$$\|G_k\|_2 = \|\frac{1}{2}P_i\|_2 = \frac{1}{2}\|P_i\|_2 = \frac{1}{2}\ .$$

- **Two-Qubit Rotations (R_{ii}):** For entangling gates like $R_{XX}(\theta), R_{YY}(\theta)$, or $R_{ZZ}(\theta)$, the generator is $G_k = \frac{1}{2}P_i \otimes P_j$. Using the tensor product property of the spectral norm ($\|A \otimes B\|_2 = \|A\|_2\|B\|_2$), the norm is again:

$$\|G_k\|_2 = \|\frac{1}{2}P_i \otimes P_j\|_2 = \frac{1}{2}\|P_i\|_2\|P_j\|_2 = \frac{1}{2}(1)(1) = \frac{1}{2}.$$

Therefore, for any standard single or two-qubit Pauli rotation gate, the generator's spectral norm is precisely $\frac{1}{2}$. Using this to simplify the bound we get:

$$\boxed{L_{bound} = P\,\|M\|_2} \tag{22}$$

5.5 THE NORM OF THE OBSERVABLE (M)

The norm of the observable, $\|M\|_2$, is problem-dependent and establishes the scale of the objective function's landscape. It is **not guaranteed to be 1**.

- **Case 1: Simple Pauli Observables.** If the goal is to measure a simple property, the observable might be a single Pauli operator, like $M = Z_i$ (measuring the spin of the i-th qubit). In this common scenario, the norm is indeed $\|M\|_2 = \|Z_i\|_2 = 1$.
- **Case 2: General Hamiltonians.** In many important applications, such as in chemistry (VQE) or physics, the observable M is a Hamiltonian representing the system's energy. Such a Hamiltonian is typically a weighted sum of Pauli strings, e.g., $M = \sum_j w_j S_j$ where S_j are Pauli strings (like $X_0Y_1Z_2$) and w_j are real coefficients. The norm $\|M\|_2 = \|\sum_j w_j S_j\|_2$ depends on the coefficients and the structure of the Hamiltonian, and is generally not equal to 1.

SIMPLIFIED BOUND FOR A NORMALIZED, COMMON SCENARIO

Let's consider a widespread VQA architecture with P parameterized Pauli rotation gates. For every such gate, we have established that $\|G_k\|_2 = 1/2$. If we are in the specific, common case where the observable is a single Pauli operator or has been normalized such that $\|M\|_2 = 1$, our bound from Eq. (14) simplifies dramatically:

$$L_{\text{bound}} = P \tag{23}$$

This remarkably simple result, $L \leq P$, provides an intuitive limit for this normalized VQA setup: the smoothness constant is bounded directly by the number of parameterized gates. For the general case, the bound remains $L \leq \|M\|_2 \cdot P$, directly scaling with the norm of the chosen observable.

6 WHEN L-SMOOTHNESS BREAKS

6.1 CONDITIONAL GATES

The L-smoothness property is not guaranteed if the QNN architecture includes conditional gates that introduce discontinuities. For example, consider a gate whose rotation angle depends on a non-smooth function g:

$$U(\theta_k) = R_x(g(\theta_k)) \quad \text{where} \quad g(\theta_k) = \begin{cases} 0 & \text{if } \theta_k < \tau \\ \theta_k & \text{if } \theta_k \geq \tau \end{cases} \tag{24}$$

This is analogous to a ReLU activation function in classical neural networks. Such conditional logic makes the function non-differentiable at the threshold point $\theta_k = \tau$, which breaks the L-smoothness property. It is important to note, however, that most standard QNN architectures do not use such gates as QNNs exhibit non-linearity inherently through its gate set.