Quantum Neural Networks Are Lipschitz Smooth

Christopher Kverne *
Florida International University

1 Introduction

1.1 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, in turn, implies that the squared amplitudes of the final state vector are also L-smooth functions.

1.2 L-smoothness Definition

A function $f(\theta)$ is L-smooth if its gradient is Lipschitz continuous. This condition is satisfied if the norm of its Hessian matrix is bounded for all θ:

$$||\nabla f(\theta_1) - \nabla f(\theta_2)|| \leq L||\theta_1 - \theta_2|| \tag{1}$$

A sufficient condition to prove L-smoothness is to show that the Hessian is bounded:

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

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. The state, parameterized by a vector θ, is represented as a linear combination of basis states:

$$|\psi(\theta)\rangle = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_{2^n} \end{bmatrix} = \sum_{i=0}^{2^n-1} c_i(\theta)\,|i\rangle \tag{3}$$

where $c_i(\theta)$ are the complex amplitudes, which are functions of the parameters θ.

2.2 The QNN Output

The output of a QNN is typically the expectation value of a Hermitian operator (an observable) M. The output $f(\theta)$ is given by:

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

Given that the elements of M are constants, this expression can be expanded in terms of the amplitudes:

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

*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

Preprint.

We aim to show that for any chosen observable the final output is L-Smooth, or in other words that all amplitdue functions squared are L-Smooth.

2.3 The Gate Set

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

- **Parameterized Gates:** These include single-qubit rotations ($R_x(\theta), R_y(\theta), R_z(\theta)$) and multi-qubit parameterized gates ($R_{xx}(\theta), R_{yy}(\theta), R_{zz}(\theta), CR_x(\theta), CR_y(\theta), CR_z(\theta)$).
- **Constant Gates:** These include single-qubit gates ($H, X, Y, Z, S, T, \dots$) and multi-qubit entangling gates ($CNOT, CZ, SWAP, \dots$).

We aim to show that for any combination of these gates, the final output is L-Smooth.

3 Core Proof: Amplitudes as Trigonometric Polynomials

We prove by induction on the number of gates in the circuit that all amplitudes of the state vector are multivariate trigonometric polynomials of the circuit parameters.

3.1 Hypothesis P(k)

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

3.2 Base Case (k=0)

For $k = 0$, the circuit contains no gates. The state is the initial state, typically $|\psi_0\rangle = |00\dots0\rangle$. The amplitudes are constants (e.g., 1 and 0), which are trivial trigonometric polynomials of degree zero. Thus, P(0) holds.

3.3 Inductive Step

We assume that P(k-1) is true: for the state $|\psi_{k-1}(\boldsymbol{\theta}_{1:k-1})\rangle$, all amplitudes $c_j(\boldsymbol{\theta}_{1:k-1})$ are trigonometric polynomials of the first $k-1$ parameters. We now 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 determined by the matrix-vector product:

$$c'_i = \sum_j (U_k)_{ij} c_j$$

We analyze the two categories of gates from our universal gate set.

Case 1: U_k is a parameterized gate, $U_k(\theta_k)$. The matrix elements of any standard rotational gate (e.g., $R_y(\theta_k)$, $R_{xx}(\theta_k)$) are, by definition, trigonometric functions of the parameter θ_k, such as $\cos(\theta_k/2)$ or $\sin(\theta_k/2)$. These elements are therefore trigonometric polynomials of θ_k. The new amplitude $c'_i(\boldsymbol{\theta}_{1:k})$ is a sum of terms, where each term is a product of a matrix element $(U_k(\theta_k))_{ij}$ and an old amplitude $c_j(\boldsymbol{\theta}_{1:k-1})$.

By our inductive hypothesis, c_j is a trigonometric polynomial of the first $k-1$ parameters. Since the set of multivariate trigonometric polynomials is closed under both multiplication and addition, the resulting sum of products, c'_i, is also a multivariate trigonometric polynomial of the full parameter set $\boldsymbol{\theta}_{1:k}$.

Case 2: U_k is a constant gate. The matrix for any constant gate (e.g., CNOT, H, SWAP) consists of constant complex values. The new amplitudes $c'_i(\boldsymbol{\theta}_{1:k-1})$ are therefore linear combinations of the previous amplitudes $c_j(\boldsymbol{\theta}_{1:k-1})$ with constant coefficients. A finite linear combination of trigonometric polynomials is itself a trigonometric polynomial, so this operation preserves the property.

In both cases, if the amplitudes of $|\psi_{k-1}\rangle$ are trigonometric polynomials, so are the amplitudes of $|\psi_k\rangle$. Therefore, by the principle of induction, P(k) holds for any number of gates from the specified set.

4 L-smoothness of the QNN Output

4.1 The Output Function is a Trigonometric Polynomial

As established, the final amplitudes $c_i(\theta)$ are multivariate trigonometric polynomials. The QNN output, $f(\theta) = \sum_{i,j} c_i^*(\theta) M_{ij} c_j(\theta)$, is a sum of products of these polynomials and constant matrix elements M_{ij}. Therefore, $f(\theta)$ is also a multivariate trigonometric polynomial.

4.2 Bounded Hessian

The Hessian matrix of $f(\theta)$ has elements given by $H_{ij} = \frac{\partial^2 f}{\partial \theta_i \partial \theta_j}$. Since $f(\theta)$ is a trigonometric polynomial, its second partial derivatives are also trigonometric polynomials. A trigonometric polynomial is a finite sum of sine and cosine terms, which are globally bounded functions. Thus, every element of the Hessian matrix is bounded.

4.3 Conclusion

Since all elements of the Hessian are bounded, the matrix norm $||\nabla^2 f(\theta)||$ is also globally bounded. A function with a bounded Hessian is, by definition, L-smooth. This completes the proof that the output of a QNN composed of standard rotational and constant gates is an L-smooth function of its parameters.

5 When L-smoothness Can Break

5.1 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:

$$U(\theta_k) = R_x(g(\theta_k)) \quad \text{where} \quad g(\theta_k) = \begin{cases} 0 & \text{if } \theta_k < \tau \\ \theta_k & \text{if } \theta_k \geq \tau \end{cases} \tag{6}$$

This is analogous to a ReLU activation function in classical neural networks.

5.2 Non-Differentiability

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.

6 Practical Considerations

While we have shown that QNNs are theoretically L-smooth, this property can be obscured in practice. On current Noisy Intermediate-Scale Quantum (NISQ) hardware, device noise makes gradient calculations estimations rather than exact measurements. This can impact the observed smoothness and the performance of gradient-based optimization algorithms.

7 Finding the L-Bound

SKIP THIS SECTION. GO TO SECTION 8 AS THIS ATTEMPT TO FIND A BOUND FAILED BUT THE SECTION 8 ATTEMPT WORKED SOMEWHAT ALTHOUGH LOOSELY This is the most difficult part of the paper as QNNs exhibit strange properties when it comes to how the state function grows.

- Find explicit formula of Hessian where $f(\theta) = \sum_{i,j} c_i^*(\theta) M_{ij} c_j(\theta)$
- Choose matrix norm
- Find maximum value of this norm (this seems very difficult)

7.1 Defining the Hessian and 2nd Derivative

The final output is defined as:

$$f(\theta) = \sum_{i,j} c_i^*(\theta) M_{ij} c_j(\theta) \tag{7}$$

To define a general term for the Hessian we will start by finding the second derivative of the output. I.e. we want to find a general term for:

$$H_{kl} = \frac{\partial f}{\partial \theta_k \partial \theta_l} \tag{8}$$

We start by taking the first derivative:

$$\frac{\partial f}{\partial \theta_k} = \frac{\partial}{\partial \theta_k} \sum_{i,j} M_{ij} c_i^* c_j = \sum_{i,j} M_{ij} \frac{\partial}{\partial \theta_k} (c_i^* c_j) = \sum_{i,j} M_{ij} \left[\frac{\partial c_i^*}{\partial \theta_k} c_j + \frac{\partial c_j}{\partial \theta_k} c_i^* \right] \tag{9}$$

Now to differentiate w.r.t the second parameter θ_l we start with each term:

$$\frac{\partial}{\partial \theta_l} \left[\frac{\partial c_i^*}{\partial \theta_k} c_j \right] = \frac{\partial^2 c_i^*}{\partial \theta_l \partial \theta_k} c_j + \frac{\partial c_i^*}{\partial \theta_k} \frac{\partial c_j}{\partial \theta_l} \tag{10}$$

$$\frac{\partial}{\partial \theta_l} \left[\frac{\partial c_j}{\partial \theta_k} c_i^* \right] = \frac{\partial^2 c_j}{\partial \theta_l \partial \theta_k} c_i^* + \frac{\partial c_j}{\partial \theta_k} \frac{\partial c_i^*}{\partial \theta_l} \tag{11}$$

Substituting this gives us a full expression for the elements in the Hessian:

$$H_{kl} = \sum_{i,j} M_{ij} \left[\frac{\partial^2 c_i^*}{\partial \theta_l \partial \theta_k} c_j + \frac{\partial c_i^*}{\partial \theta_k} \frac{\partial c_j}{\partial \theta_l} + \frac{\partial^2 c_j}{\partial \theta_l \partial \theta_k} c_i^* + \frac{\partial c_j}{\partial \theta_k} \frac{\partial c_i^*}{\partial \theta_l} \right] \tag{12}$$

7.2 Bounding this Term

How can we even attempt to bound this term or find an upper bound of L? Sure we differentiate ci and cj once or twice based on all 2 combinations of parameters and since ci and cj are trigonometric polynomials differentiating them once or twice shouldn't be an issue and should be bounded like we show above. But how to proceed is very difficult? Maybe we need to simplify the Hkl term we derived (although I don't think that's possible). I will therefore try a more general approach below.

8 Finding an Explicit L-Bound

While the proof that Quantum Neural Networks (QNNs) are L-smooth is established by demonstrating that their output is a trigonometric polynomial, deriving a tight and explicit bound for the constant L directly from the Hessian's amplitude-based formulation is intractable for general circuits. Here, we present a more direct approach using an operator-based formalism to derive a clear upper bound for L.

8.1 Operator-Based Formulation of the Hessian

We begin with the definition of the QNN output as the expectation value of a Hermitian observable M:

$$f(\theta) = \langle \psi(\theta) | M | \psi(\theta) \rangle \tag{13}$$

The state $|\psi(\theta)\rangle$ is produced by applying a sequence of unitary gates to an initial state $|\psi_0\rangle$, such that $|\psi(\theta)\rangle = U(\theta)|\psi_0\rangle$. Let the circuit consist of P parameterized gates, where each unitary

$U_k(\theta_k)$ is generated by a Hermitian operator G_k according to the Schrödinger equation, $U_k(\theta_k) = \exp(-i\theta_k G_k)$. The first partial derivative of $f(\theta)$ with respect to a parameter θ_k is:

$$\frac{\partial f}{\partial \theta_k} = \frac{\partial \langle \psi |}{\partial \theta_k} M |\psi\rangle + \langle \psi | M \frac{\partial |\psi\rangle}{\partial \theta_k} \tag{14}$$

Since $f(\theta)$ is a real-valued function, this simplifies to:

$$\frac{\partial f}{\partial \theta_k} = 2\text{Re}\left(\langle \psi | M \frac{\partial |\psi\rangle}{\partial \theta_k} \right) \tag{15}$$

To find the elements of the Hessian matrix, $H_{kl} = \frac{\partial^2 f}{\partial \theta_k \partial \theta_l}$, we differentiate a second time. This yields:

$$H_{kl} = \frac{\partial^2 \langle \psi |}{\partial \theta_l \partial \theta_k} M |\psi\rangle + \frac{\partial \langle \psi |}{\partial \theta_k} M \frac{\partial |\psi\rangle}{\partial \theta_l} + \frac{\partial \langle \psi |}{\partial \theta_l} M \frac{\partial |\psi\rangle}{\partial \theta_k} + \langle \psi | M \frac{\partial^2 |\psi\rangle}{\partial \theta_l \partial \theta_k} \tag{16}$$

Again, because the output is real, this can be written more compactly as:

$$H_{kl} = 2\text{Re}\left(\frac{\partial \langle \psi |}{\partial \theta_k} M \frac{\partial |\psi\rangle}{\partial \theta_l} + \langle \psi | M \frac{\partial^2 |\psi\rangle}{\partial \theta_l \partial \theta_k} \right) \tag{17}$$

This expression for the Hessian elements is more suitable for bounding than the formulation based on amplitudes.

8.2 Bounding the State Vector Derivatives

The key to bounding H_{kl} is to bound the norm of the state's partial derivatives. The derivative of a unitary gate $U_k = \exp(-i\theta_k G_k)$ is $\frac{\partial U_k}{\partial \theta_k} = -iG_k U_k$. The first partial derivative of the state vector is:

$$\frac{\partial |\psi\rangle}{\partial \theta_k} = U_P(\theta_P) \cdots \frac{\partial U_k(\theta_k)}{\partial \theta_k} \cdots U_1(\theta_1)|\psi_0\rangle = U_P \cdots (-iG_k U_k) \cdots U_1 |\psi_0\rangle \tag{18}$$

Since all U_j are unitary and norm-preserving, the norm of this derivative vector is:

$$\left\| \frac{\partial |\psi\rangle}{\partial \theta_k} \right\| = \| -iG_k U_k \cdots U_1 |\psi_0\rangle \| = \| G_k |\psi_{k-1}\rangle \| \tag{19}$$

where $|\psi_{k-1}\rangle$ is the state just before gate U_k. Using the property of an induced matrix norm, $\|Ax\| \le \|A\| \|x\|$, and the fact that the state vector is normalized ($\| |\psi_{k-1}\rangle \| = 1$):

$$\left\| \frac{\partial |\psi\rangle}{\partial \theta_k} \right\| \le \|G_k\| \tag{20}$$

Let's denote the spectral norm (2-norm) of the generator as $g_k = \|G_k\|_2$. For standard single-qubit rotations, the generator is of the form $\frac{1}{2}P$ (where P is a Pauli matrix), so $g_k = \|\frac{1}{2}P\|_2 = \frac{1}{2}$. Next, we bound the second partial derivative.

- **Case 1: Mixed partial derivative ($k \neq l$).** Assume $l > k$ without loss of generality.

$$\frac{\partial^2 |\psi\rangle}{\partial \theta_l \partial \theta_k} = U_P \cdots (-iG_l U_l) \cdots (-iG_k U_k) \cdots U_1 |\psi_0\rangle \tag{21}$$

The norm is bounded by the product of the generator norms:

$$\left\| \frac{\partial^2 |\psi\rangle}{\partial \theta_l \partial \theta_k} \right\| = \| G_l U_l \cdots U_{k+1} G_k |\psi_{k-1}\rangle \| \le \|G_l\| \|G_k\| = g_l g_k \tag{22}$$

- **Case 2: Unmixed partial derivative ($k = l$).**

$$\frac{\partial^2|\psi\rangle}{\partial\theta_k^2} = U_P \cdots \frac{\partial^2 U_k}{\partial\theta_k^2} \cdots U_1|\psi_0\rangle = U_P \cdots (-iG_k)^2 U_k \cdots U_1|\psi_0\rangle \tag{23}$$

The norm is bounded by the square of the generator norm:

$$\left\| \frac{\partial^2|\psi\rangle}{\partial\theta_k^2} \right\| = \| - G_k^2|\psi_{k-1}\rangle\| \leq \|G_k^2\| \leq \|G_k\|^2 = g_k^2 \tag{24}$$

8.3 Establishing the Final Bound for L

We can now bound the magnitude of each Hessian element $|H_{kl}|$:

$$|H_{kl}| = \left| 2\text{Re}\left(\frac{\partial\langle\psi|}{\partial\theta_k} M \frac{\partial|\psi\rangle}{\partial\theta_l} + \langle\psi|M \frac{\partial^2|\psi\rangle}{\partial\theta_l\partial\theta_k} \right) \right| \leq 2 \left(\left| \frac{\partial\langle\psi|}{\partial\theta_k} M \frac{\partial|\psi\rangle}{\partial\theta_l} \right| + \left| \langle\psi|M \frac{\partial^2|\psi\rangle}{\partial\theta_l\partial\theta_k} \right| \right) \tag{25}$$

Using the Cauchy-Schwarz inequality and the bounds on the derivative norms:

$$|H_{kl}| \leq 2 \left(\left\| \frac{\partial|\psi\rangle}{\partial\theta_k} \right\| \|M\| \left\| \frac{\partial|\psi\rangle}{\partial\theta_l} \right\| + \|\langle\psi|\|\|M\| \left\| \frac{\partial^2|\psi\rangle}{\partial\theta_l\partial\theta_k} \right\| \right) \tag{26}$$

$$|H_{kl}| \leq 2\|M\| \left((g_k)(g_l) + (1)(g_l g_k) \right) = 4\|M\| g_k g_l \tag{27}$$

This bound holds for both $k \neq l$ and $k = l$. To satisfy the L-smoothness condition, $||\nabla^2 f(\theta)|| \leq L$, we can use the Frobenius norm for the Hessian matrix:

$$L = ||H||_F = \left(\sum_{k,l=1}^{P} |H_{kl}|^2 \right)^{1/2} \tag{28}$$

Substituting our bound for each element:

$$||H||_F \leq \left(\sum_{k,l=1}^{P} (4\|M\| g_k g_l)^2 \right)^{1/2} = 4\|M\| \left(\sum_{k,l=1}^{P} g_k^2 g_l^2 \right)^{1/2} \tag{29}$$

The double summation can be factored:

$$\sum_{k,l=1}^{P} g_k^2 g_l^2 = \left(\sum_{k=1}^{P} g_k^2 \right) \left(\sum_{l=1}^{P} g_l^2 \right) = \left(\sum_{k=1}^{P} g_k^2 \right)^2 \tag{30}$$

This gives a final bound for the Frobenius norm of the Hessian:

$$||\nabla^2 f(\theta)||_F \leq 4\|M\| \left(\sum_{k=1}^{P} g_k^2 \right) \tag{31}$$

Therefore, an upper bound for the L-smoothness constant is:

$$\boxed{L_{bound} = 4\|M\|_2 \sum_{k=1}^{P} \|G_k\|_2^2} \tag{32}$$

Where $\|M\|_2$ is the spectral norm of the measurement observable, P is the number of parameterized gates, and $\|G_k\|_2$ is the spectral norm of the generator of the k-th parameterized gate.