Quantum Neural Networks Are Lipschitz Smooth

Anonymous Author(s)
Affiliation
Address
email

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:

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

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.

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

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 ($H, X, Y, Z, S, T, \ldots$) and multi-qubit entangling gates ($CNOT, CZ, SWAP, \ldots$).

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 $|\psi_0\rangle = |00\ldots0\rangle$. 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: $\boldsymbol{U_k}$ is a parameterized gate, $\boldsymbol{U_k(\theta_k)}$. 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: $\boldsymbol{U_k}$ 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.