Benchmarking Optimization Methods for Quantum Recurrent Neural Networks

Erik Connerty Michael Stewart

December 7, 2025

Abstract

Quantum Machine Learning (QML) remains an early-stage research area within computer science, and the training and deployment of Quantum Recurrent Neural Networks (QRNNs)[1] are not yet well understood. This milestone develops and benchmarks two gradient- approximation methods—Simultaneous Perturbation Stochastic Approximation (SPSA)[2] and Finite-Difference Stochastic Approximation (FDSA)—for training QRNNs. We implement both methods in a PyTorch–Qiskit stack, target a Lorenz- system cross-prediction task, and measure runtime and accuracy. SPSA delivers an order-of-magnitude faster training time with comparable accuracy, making it the pragmatic choice for QRNN forecasting workloads.

1 Introduction

QML offers the promise of leveraging quantum circuits for learning tasks, yet practical training methodologies remain underexplored. QRNNs combine parameterized quantum circuits with classical recurrence, making them a natural fit for time-series prediction but posing unique optimization challenges. This work targets two objectives: (i) provide a working QRNN training framework supporting SPSA and FDSA, and (ii) benchmark the two methods to establish best practices for QRNN training.

2 Methods

2.1 Model and Framework

We implemented an 8-qubit QRNN as a custom PyTorch layer, with Qiskit Aer executing the quantum circuit per time step. Custom backward functions provide gradient estimates via SPSA or FDSA, enabling a unified training loop in PyTorch.

2.2 Optimization Algorithms

SPSA uses two circuit evaluations per update, independent of parameter count, yielding $O(N)$ scaling for N steps. **FDSA** uses coordinate-wise finite differences, offering more precise gradients at the cost of $O(PN)$ evaluations for P parameters. Parameter-shift was not used due to non-unitary operations in the circuit; FDSA served as the higher-fidelity baseline.

[1]QRNN: Quantum Recurrent Neural Network
[2]SPSA: Simultaneous Perturbation Stochastic Approximation

3 Experimental Setup

Learning Task: Lorenz-system cross prediction, inputting the x component and forecasting y and z one step ahead. **Data and Training**: Adam optimizer with small learning rates; metric RMSE. **Hardware**: CPU-based Qiskit Aer; FDSA additionally tested on 2×A100 GPUs. **Goal**: Compare runtime and accuracy trade-offs of SPSA vs. FDSA.

4 Results

4.1 Runtime

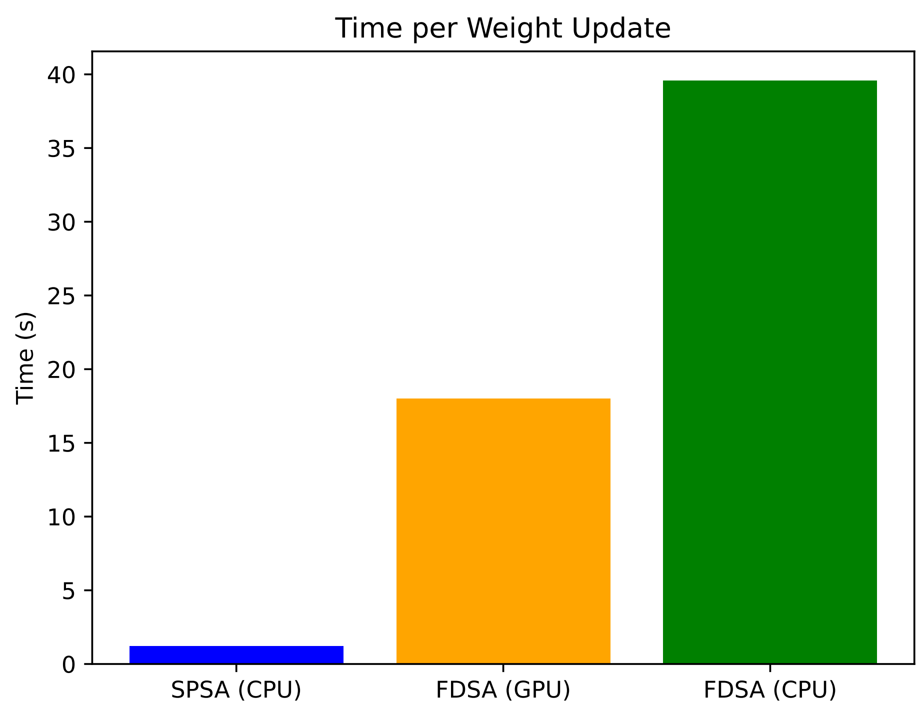
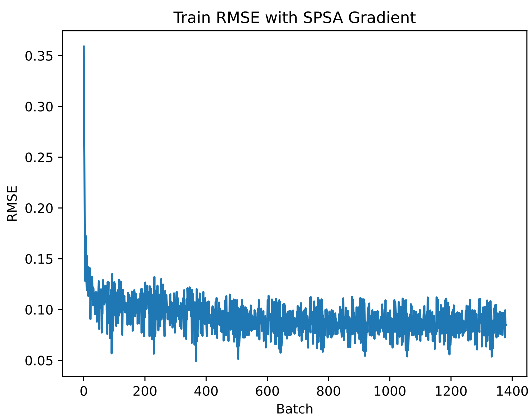


Figure 1: Per-epoch wall-clock runtime comparison for SPSA vs. FDSA.

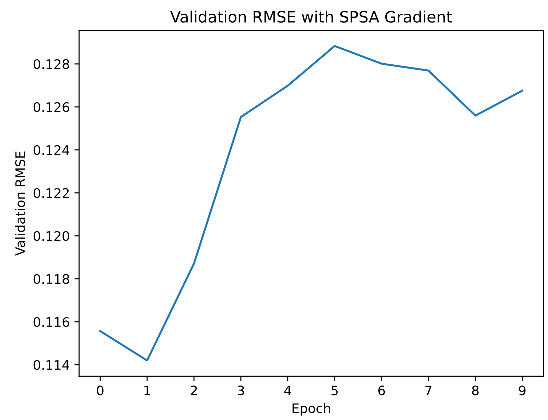
Method / Hardware	Epoch time	Notes
SPSA (CPU)	$\sim$ 1 hr/epoch	Fastest end-to-end
FDSA (CPU)	$\sim$ 8 hrs/epoch	Higher cost per gradient
FDSA (2x A100)	$\sim$ 14.5 hrs/epoch	GPU overhead dominated

Table 1: Per-epoch training times for the QRNN.

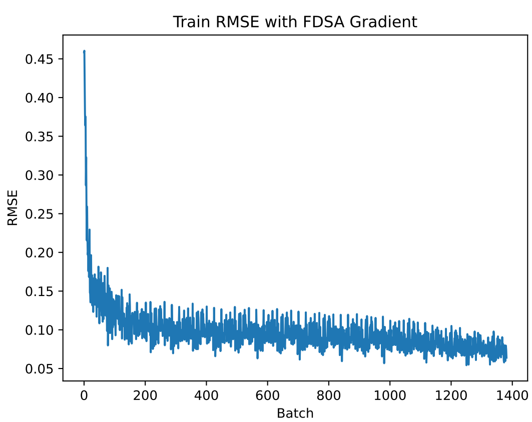
4.2 Accuracy



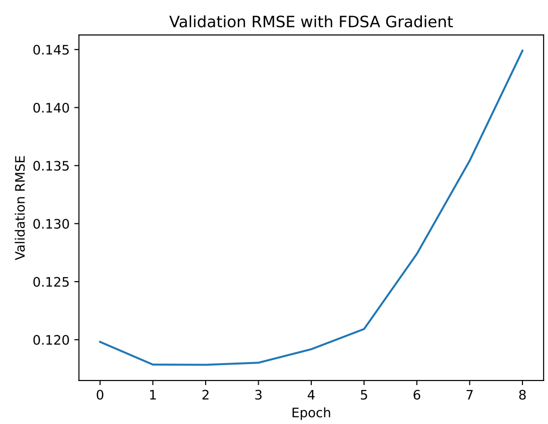
(a) SPSA — Training RMSE



(b) SPSA — Validation RMSE



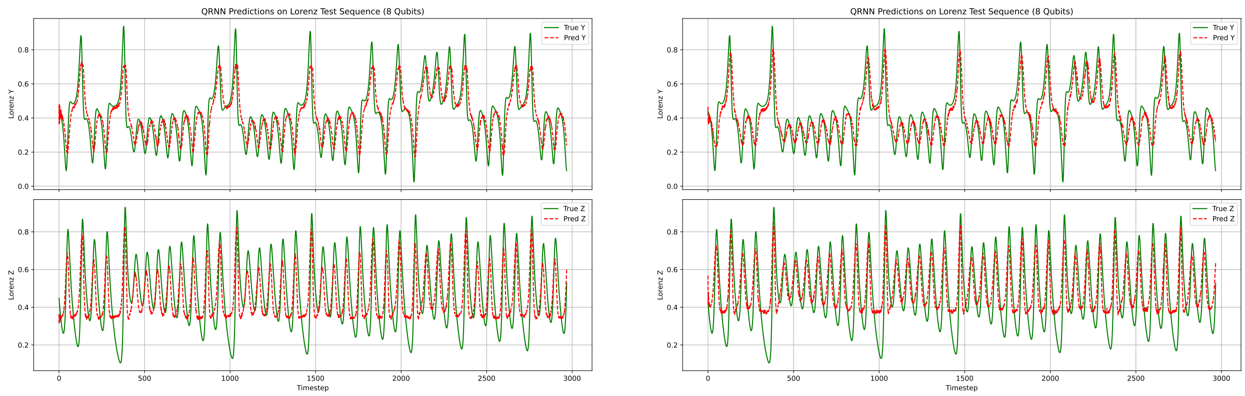
(c) FDSA — Training RMSE



(d) FDSA — Validation RMSE

Figure 2: RMSE diagnostics: SPSA (top), FDSA (middle), combined comparison (bottom).

4.3 Predictions



(a) SPSA — Prediction vs. Ground Truth

(b) FDSA — Prediction vs. Ground Truth

Figure 3: Example Lorenz-system predictions for SPSA and FDSA.

5 Discussion

SPSA's stochastic gradients provided practical convergence with minimal circuit evaluations, making it the clear choice for rapid iteration. FDSA offered marginally delayed overfitting but at substantial runtime cost. For QRNN forecasting workloads, SPSA provides superior efficiency with comparable accuracy.

6 Conclusion

We delivered a PyTorch–Qiskit QRNN framework supporting SPSA and FDSA, targeting an 8-qubit Lorenz cross-prediction task. SPSA achieved an order- of-magnitude faster training time with competitive accuracy. FDSA provided slightly improved stability but at prohibitive computational expense. Future work includes GPU kernel optimization, scaling to larger chaotic systems, and validation on hardware quantum backends.

References

[1] E. L. Connerty, E. N. Evans, G. Angelatos, and V. Narayanan. Quantum observers: A NISQ hardware demonstration of chaotic state prediction using quantum echo-state networks. *arXiv preprint arXiv:2505.06799*, 2025.

[2] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii. Quantum circuit learning. *Physical Review A*, 98(3):032309, 2018.