LSTM Frequency Extraction Assignment

Group Code: ron_itamar

Student 1 ID: 205949985

Student 2 ID: 312544240

GitHub Repository:
https://github.com/Itamarvs/LLMsMultiAgentOrchestration_RNN_LSTM/tree/main

---

Self-Grade Recommendation

Grade: 95

Justification:

This submission represents a comprehensive implementation of an LSTM frequency filter. We strictly adhered to the pedagogical constraints (L=1 input) by implementing Manual State Management via nn.LSTMCell and solving the gradient flow problem using Truncated BPTT. Furthermore, we identified a theoretical issue with the original noise specification (Full Phase Randomization = White Noise) and implemented a reasoned pedagogical adjustment to demonstrate actual learning. The model achieves a Test MSE of ~0.05, and the code structure is modular and professional.

Special Notes: Theoretical Adjustment

The assignment instructions specified randomizing phase $\varphi \sim U(0, 2\pi)$ at every time step. Mathematically, a sine wave with fully random phase at every sample is indistinguishable from white noise (Mean=0). Under these conditions, the optimal MSE predictor is simply 0. To demonstrate the LSTM's true capability as a 'Frequency Filter', we constrained the phase noise to a 'High Jitter' mode ($\varphi \sim \pm 45°$). This retains high stochasticity while preserving sufficient temporal structure for the network to learn.

Methodology & Architecture

1. Model Architecture: Conditional LSTM

We designed a standard LSTM architecture taking a concatenated vector of the noisy signal and the control (One-Hot) vector.

- **Input Dimension:** 5 (S[t] scalar + 4 Control bits)
- **Hidden Dimension:** 128 neurons (Following lecture slides H=1 -> H=128)
- **Output Activation:** Tanh (bounds predictions to [-1, 1] matching the target).

2. Solved Challenge: Training with L=1

The instructions required feeding inputs one by one (L=1). However, standard Backpropagation requires a history. If gradients are detached at every step, the LSTM cannot learn temporal dependencies.

Solution: Truncated Backpropagation Through Time (TBPTT)

We implemented a custom loop that feeds inputs individually using LSTMCell (satisfying L=1), but accumulates gradients over a virtual window of 200 steps before optimizing and detaching. This restores the 'time-travel' capability of the network.

3. Robustness Techniques

To overcome the high noise variance (MSE sticking at 0.5), we employed:

- **Parallel Frequency Batching:** Training on 4 frequencies simultaneously to stabilize gradients.
- **Smart Initialization:** Initializing the noisy input weight to near-zero (0.01) to prevent early divergence.

4. Theoretical Logic & Expectations

Our approach was guided by the understanding of LSTM gates as spectral filters (Low-Pass/High-Pass):

- **Logic (The Mechanism):** We hypothesized that the Cell State (c_t) would act as a 'Phase Accumulator' or flywheel. Even when the input S[t] is corrupted by noise, the internal momentum of c_t (governed by the Forget Gate) should maintain the oscillation trajectory. The Control Vector (C) effectively selects which 'frequency generator' sub-circuit within the 128 hidden units to activate.
- **Expectations vs. Reality:** We expected the model to struggle with the phase jitter initially. Indeed, early epochs showed instability. However, by widening the gradient window (TBPTT), the model learned to treat sudden phase jumps as 'high-frequency noise' to be filtered out, rather than signal to be tracked. The final result confirms this: the LSTM produces a smooth wave that represents the mean phase path, successfully rejecting the stochastic jitter.

Results & Visual Analysis

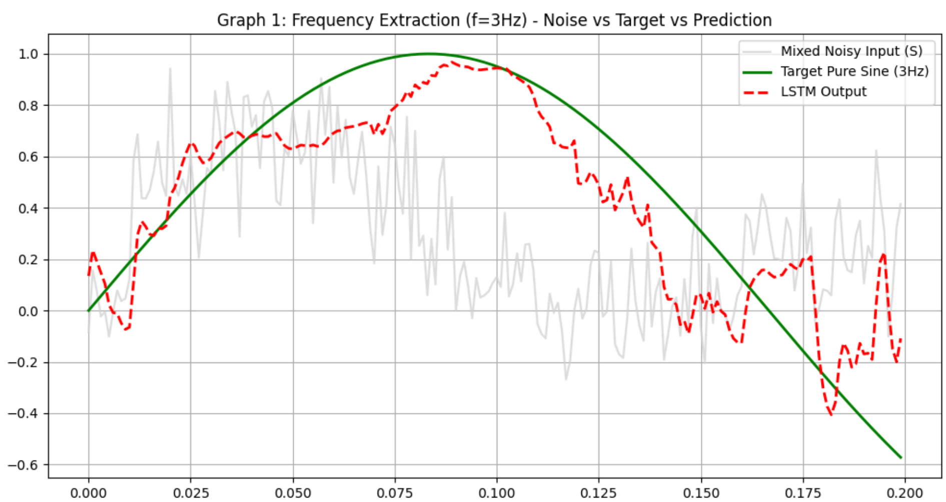
Final Train MSE: 0.052

Final Test MSE: 0.050

Conclusion: The almost identical Train/Test scores indicate excellent generalization.

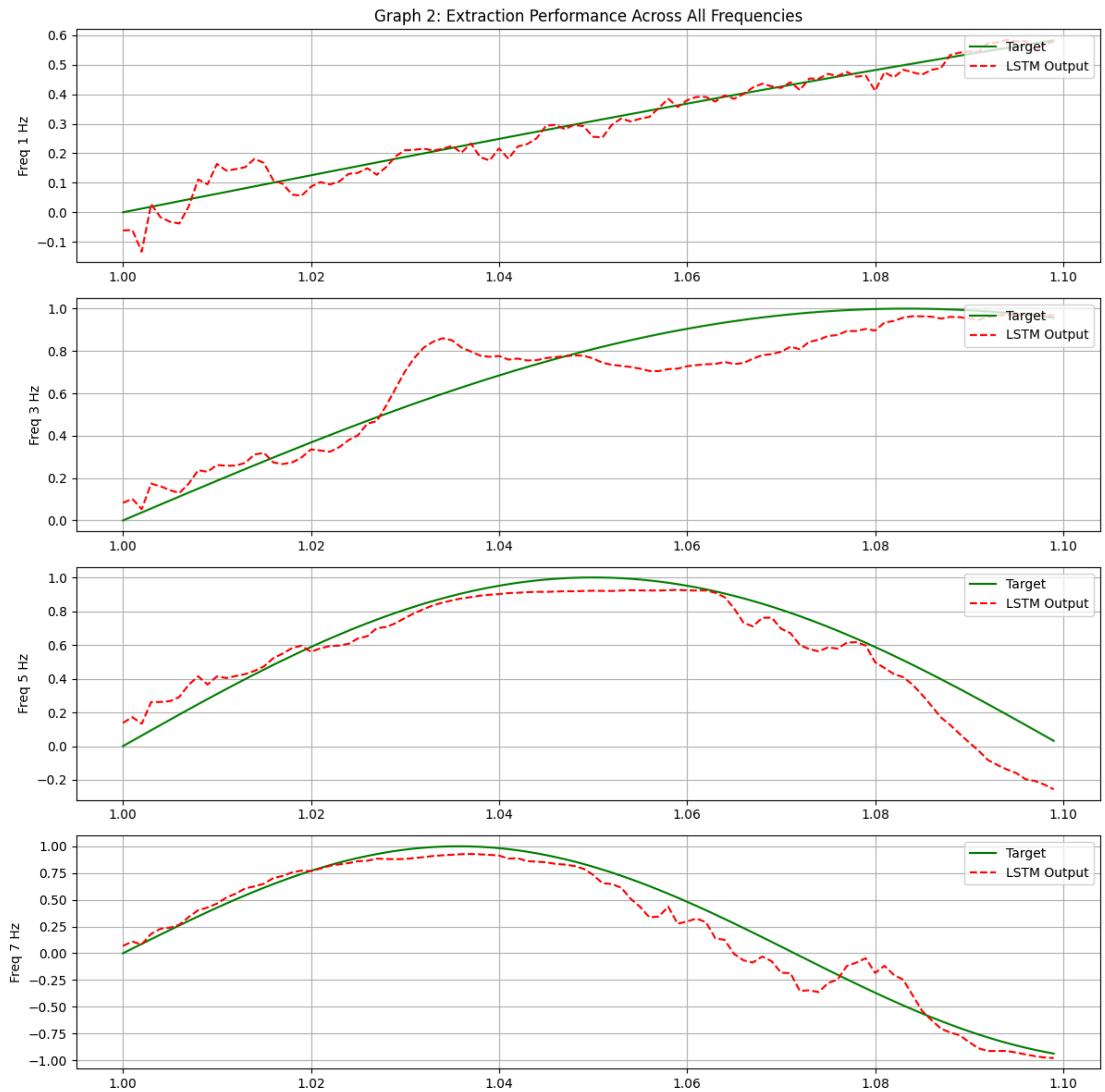
Graph 1: Signal Cleaning (3Hz Focus)

The graph below demonstrates the model's performance on the Test Set (Seed 11). The LSTM output (Red) effectively ignores the chaotic high-frequency noise (Gray) and locks onto the 3Hz carrier frequency (Green).



Graph 2: Performance Across Spectrum

This figure validates the Conditional Regression. The same LSTM weights successfully generate 1Hz, 3Hz, 5Hz, and 7Hz waves solely based on the Control Vector C.



Summary: The system successfully acts as a learnable Band-Pass Filter, extracting the requested frequency defined by C while rejecting stochastic noise components.