Training an RNN with a Low-Rank Structure

Aniket Deshpande

CONTENTS

Date: November 25, 2025

footnote

1. Training an RNN with a Low-Rank Structure

Setup. The model is a continuous time recurrent neural network

$$\tau \dot{h}(t) = -h(t) + J\phi(h(t)) + Bu(t), \quad y(t) = w^T \phi(h(t))$$

where J is the recurrent weight matrix, ϕ is ReLU or $\tanh$, $u(t)$ is a low-dimensional input (one or two OU stimuli), and $y(t)$ is a 1-2D readout. The task being learned is tracking/denoising the driving OU signals (and possibly switch between two stimuli), i.e. make $y(t)$ match a target trace $y^*(t)$ generated by the OU signals. The loss function would be the mean-squared error over a time window (with light regularization)

$$\mathscr{L} = \frac{1}{T}\int_0^T \left\| y(t) - y^*(t) \right\|^2 dt + \lambda_W \|J\|_F^2 + \lambda_{\text{bal}} \underbrace{\left\| \frac{1}{N} \mathbf{1}^T J \right\|_2^2}_{\text{row/col balance}}$$

The balance term nudges J towards zero row/column sums so that the random bulk stays centered. Our optimizer will be standard gradient descent through time (BPTT) on J (and often the readout w, occasionally B). We can backpropagate the MSE through the RNN unrolled over time. We should monitor the spectrum of $J - I$ (or the linearized Jacobian) to see whether/when an outlier separates from the circular bulk. Additionally, some simple weight stats (row/col means, within/between clusters) to visualize emerging structure, also task error.

Goal. The goal of this side project is to supplement the original DMFT analysis. Our current work analyzes fixed low-rank perturbations $J = gW - b/N\mathbf{1}\mathbf{1}^T + muv^T$ and their dynmaical consequences. The training goal is to learn: if we *do not* impose muv^T structure, but instead train the RNN to do the OU-tracking task, does learning spontaneously add a low-rank component that looks like uv^T? How does the magnitude m and alignment compare with the task geometry?

After training, we can project the learned change $\Delta J = J - J_0$ onto a low-rank model. We fit $\Delta J \approx \sum_{r=1}^{R} m_r u_r v_r^T$ (e.g. SVD with sign/scale conventions). Check whether $R = 1$ or 2 suffices for most of the variance. We also should compare spectra: the bulk stays roughly circular, and we should see an outlier emerge and move with training, exactly the same object we analyzed in the DMFT (the $\Re\lambda_{\text{out}}$ crossing).