Fixed point analysis

We sought to understand the model behavior by analyzing the neural dynamics during the successful completion of a task. Following Driscoll et al., 2024, we used the Fixed Point Finder package [https://github.com/mattgolub/fixed-point-finder] to search for stationary points corresponding to correct model output. We hypothesized that an arrangement of these features in state space could characterize the trained models in a way invariant to the specific solutions (weight matrices) found during training. However, we did not find evidence of fixed points in our analysis. In five models with different random seeds at the beginning of training, we observed that the minimum q values were not sufficiently small given the timescale of our task to be considered fixed points. They were >> 1 / T^2. To confirm that the state vector had velocity while the model output remained constant by our criteria, we calculated the magnitude of the velocity vector over time from our hidden states. One can see an appreciable velocity on the scale of sqrt(q). Cosine similarity between successive velocity vectors revealed ballistic motion during the beginning of the task followed by diffusive motion dominated by the intrinsic noise we include in our equations of motion. So while the hidden state vector does not settle into a stable fixed point as we hypothesized, it does fluctuate around a constant position. Given that the hidden state dimension is much greater than the output dimension, the output matrix is underdetermined and multiple hidden states map to (by our criteria) the same output state.