

Inferring low-rank network models from neural activity

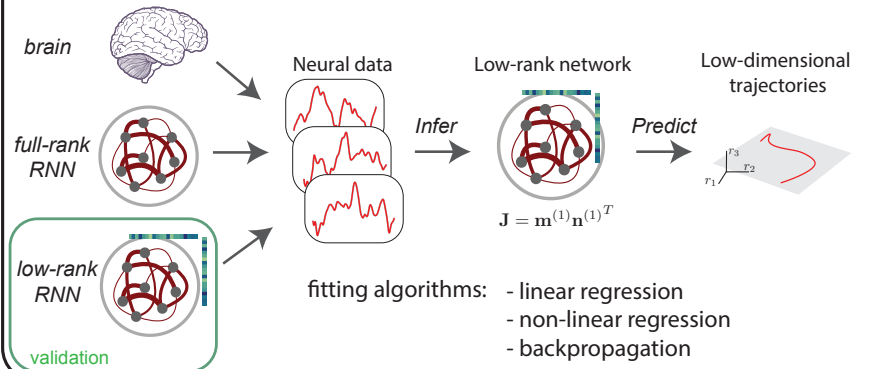
Adrian Valente¹, Jonathan Pillow^{2,3}, Srdjan Ostojic¹

1. Group for Neural Theory, LNC2, DEC, École Normale Supérieure, Paris 2. Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA
3. Department of Psychology, Princeton University, Princeton, NJ, USA

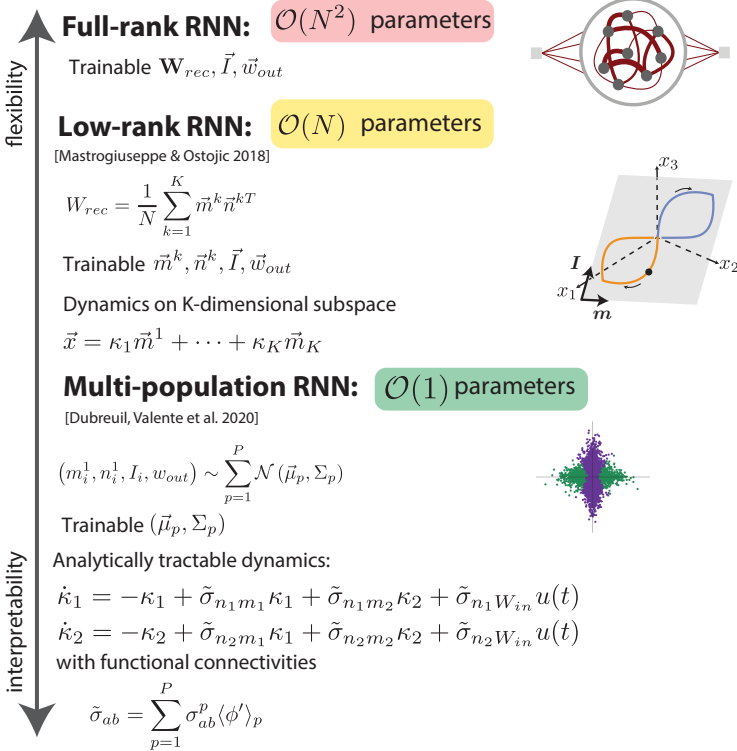
Abstract

- We propose a new latent variable model based on low-rank RNNs [Mastrogiuseppe & Ostojic 2018]
- We train low-rank RNNs to reproduce neural activity neuron-by-neuron, with three different algorithms
- We validate our approach by generating low-rank trajectories from low-rank RNNs and retrieving their connectivity with our method.
- We also train low-rank RNNs to reproduce activity of a pre-trained full-rank RNN. This leads to deep insights into the mechanism used by the full-rank RNN, and to predictions on the role of different neurons
- Finally, we show how our approach is very robust to subsampling, which is an experimental constraint.

Approach: fitting low-rank RNNs to neural recordings

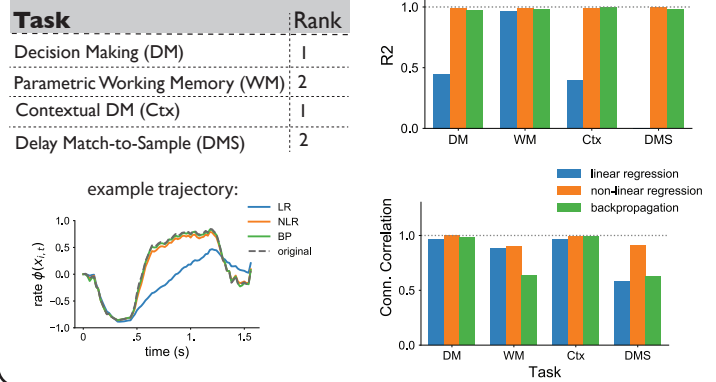


Classes of networks



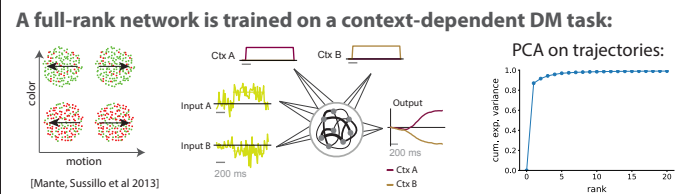
Validation of the approach

- We generate trajectories from low-rank RNNs pre-trained on 4 tasks. Both activity and connectivity can be retrieved.

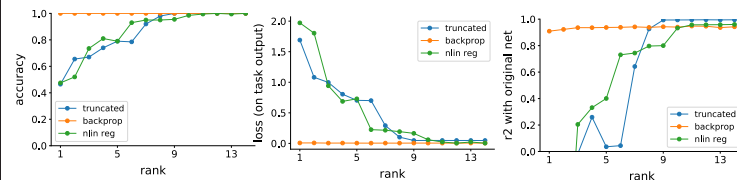


"Opening the black box" of a full-rank network

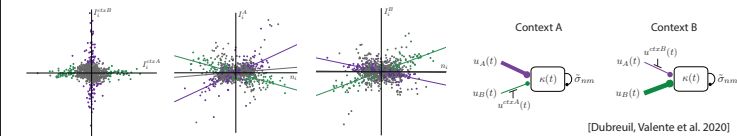
- This method can give insights into the mechanism used by a full-rank network



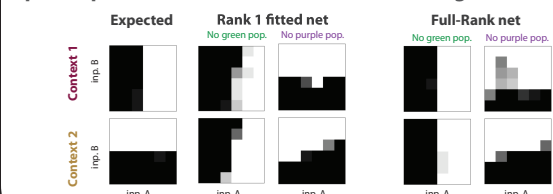
With backprop, a rank-1 network suffices to capture neural activity:



Reverse-engineering rank-1 connectivity shows role for two populations:

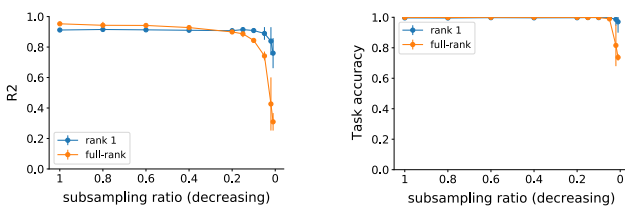


Specific predictions for inactivations on the original network can be done:



Neuron subsampling

- This method is very robust to subsampling
- Results for networks fitted to subsampled trajectories of a full-rank, 1000 neurons network



Take-home messages

- Low-rank networks are a model that is both mechanistic, interpretable, and trainable, directly linking connectivity to dynamics.
- They have a natural interpretation in terms of latent variables.
- They can finally help reverse-engineer black-box mechanisms in full-rank RNNs, and help reduce their complexity ("distill" them).