Learning degenerate parameteric distributions of RNNs that solve tasks

Sean Bittner

February 16, 2019

1 Introduction

In neuroscientific studies, RNNs are often trained to execute dynamic computations via the performance of some task. This is done with the intention of comparing the trained system's activity with that measured in the brain. There are a variety of methods used to train RNNs, and how these learning methods bias the learned connectivities (and potentially the implemented algorithm) within the broader solution space remains poorly understood. An assessment of the degenerate parameterizations of RNNs that solve a given task would be valuable for characterizing learning algorithm biases, as well as other analyses. Recent work by (Matroguisseppe & Ostojic, 2018) allows us to derive statistical properties of the behavior of recurrent neural networks (RNNs) given a low-rank parameterization of their connectivity. This work builds on dynamic mean field theory (DMFT) for neural networks (Sompolinsky et al. 1988), which is exact in the limit of infinite neurons, but has been shown to yield accurate approximations for finite size networks. We provide some brief background regarding DMFT and the recent theoretical advancements that facilitate our examination of the solution space of RNNs performing computations.

2 Background

2.1 Dynamic mean field theory (DMFT) in neuroscience

Mean field theory (MFT) originated as a useful tool for physicists studying many-body problems, particularly interactions of many particles in proximity. Deriving an equation for the probability of configurations of such systems of particles in equilibrium requires a partition function, which is essentially the normalizing constant of the probability distribution. The partition function relies on the Hamiltonian, which is an expression for the total energy of the system. Many body problems in physics are usually pairwise interaction models, resulting in combinatoric growth issue in the calculation of the Hamiltonian. A mean field assumption that some degrees of freedom of the system have independent probabilities makes approximations to the Hamilton tractable. Importantly, when minimizing the free energy of the system (to find the equilibrium state), the mean field assumption allows the derivation of consistency equations. For a given system parameterization, we can solve the consistency equations using an off-the-shelf nonlinear system of equations solver.

Using the same modeling strategy as MFT, physicists developed dynamic mean field theory (DMFT) to describe dynamics of macroscopic spin glass properties. Later, this same formalism was used to describe dynamic properties of unstructured neural networks (Somp 88).

Add some text here about how those equations are set up...

2.2 Low rank RNNs

The network dynamics of neuron i's rate x evolve according to:

$$\dot{x}_i(t) = -x_i(t) + \sum_{j=1}^{N} J_{ij}\phi(x_j(t)) + I_i \tag{1}$$

where the connectivity is comprised of a random and structured component:

$$J_{ij} = g\chi_{ij} + P_{ij} \tag{2}$$

The random all-to-all component has elements drawn from $\chi_{ij} \sim \mathcal{N}(0, \frac{1}{N})$, and the structured component is a sum of r unit rank terms:

$$P_{ij} = \sum_{k=1}^{r} \frac{m_i^{(k)} n_j^{(k)}}{N} \tag{3}$$

We use this theory to compute $T(x)$ when training DSNs to learn maximum entropy distributions of network connectivities that solve a task. While the theory is currently used to design low-rank solutions to tasks, we are able to learn the full distribution of low-rank RNN parameterizations that solve a given task.

2.3 Derivation of important DMFT variables

3 DMFT solvers

3.1 Rank 1 sponatneous stationary solutions

Rank-1 vectors m and n have elements drawn

$$m_i \sim \mathcal{N}(M_m, \Sigma_m)$$

$$n_i \sim \mathcal{N}(M_n, \Sigma_n)$$

Parameters:

$$z = \begin{bmatrix} g & M_m & M_n & \Sigma_m \end{bmatrix}^\top$$

Consistency equations: (eq 83 of M & O)

$$\begin{aligned} \mu &= M_m M_n \langle [\phi_i] \rangle := F(\mu, \Delta_0) \\ \Delta_0 &= g^2 \langle \left[\phi_i^2\right] \rangle + \Sigma_m^2 M_n^2 \langle [\phi_i] \rangle^2 := G(\mu, \Delta_0) \end{aligned} \tag{4}$$

Solver:

$$\begin{aligned} \dot{\mu}(t) &= -\mu(t) + F(\mu(t), \Delta_0(t)) \\ \dot{\Delta_0}(t) &= -\Delta_0(t) + G(\mu(t), \Delta_0(t)) \end{aligned} \tag{5}$$

3.2 Rank1 sponatneous chaotic solutions

Rank-1 vectors m and n have elements drawn

$$m_i \sim \mathcal{N}(M_m, \Sigma_m)$$

$$n_i \sim \mathcal{N}(M_n, \Sigma_n)$$

Parameters:

$$z = \begin{bmatrix} g & M_m & M_n & \Sigma_m \end{bmatrix}^\top$$

Consistency equations: (eq 86 of M & O)

$$\begin{aligned} \mu = F(\mu, \Delta_0, \Delta_\infty) &= M_m M_n \int \mathcal{D}z \phi(\mu + \sqrt{\Delta_0} z) \\ \Delta_0 = G(\mu, \Delta_0, \Delta_\infty) &= [\Delta_\infty^2 + 2g^2 \{ \int \mathcal{D}z \Phi^2(\mu + \sqrt{\Delta_0} z) \\ - \int \mathcal{D}z [\int \mathcal{D}x \Phi(\mu + \sqrt{\Delta_0 - \Delta_\infty} x + \sqrt{\Delta_\infty} z)]^2 \} &+ M_n^2 \Sigma_m^2 \langle [\phi_i] \rangle^2 (\Delta_0 - \Delta_\infty)]^{\frac{1}{2}} \\ \Delta_\infty = H(\mu, \Delta_0, \Delta_\infty) &= g^2 \int \mathcal{D}z \left[\int \mathcal{D}x \Phi(\mu + \sqrt{\Delta_0 - \Delta_\infty} + \sqrt{\Delta_\infty} z \right]^2 + M_n^2 \Sigma_m^2 \langle [\phi_i] \rangle^2 \end{aligned} \tag{6}$$

Solver:

$$\begin{aligned} \dot{\mu}(t) &= -\mu(t) + F(\mu(t), \Delta_0(t), \Delta_\infty(t)) \\ \dot{\Delta_0}(t) &= \Delta_0(t) + G(\mu(t), \Delta_0(t), \Delta_\infty(t)) \\ \dot{\Delta_\infty}(t) &= -\Delta_\infty(t) + H(\mu(t), \Delta_0(t), \Delta_\infty(t)) \end{aligned} \tag{7}$$

3.3 Rank 1 with input chaotic solutions

Rank-1 vectors m and n have elements drawn

$$m_i \sim \mathcal{N}(M_m, \Sigma_m)$$

$$n_i \sim \mathcal{N}(M_n, \Sigma_n)$$

The current has the following statistics:

$$I = M_I + \frac{\Sigma_{mI}}{\Sigma_m} x_1 + \frac{\Sigma_{nI}}{\Sigma_n} x_2 + \Sigma_\perp h$$

where x_1, x_2, and h are standard normal random variables.

Parameters:

$$z = \begin{bmatrix} g & M_m & M_n & M_I & \Sigma_m & \Sigma_n & \Sigma_{mI} & \Sigma_{nI} & \Sigma_\perp \end{bmatrix}^\top$$

(expansion of 98 of M & O)
The $\ddot{\Delta}$ equation is broken into the equation for Δ_0 and Δ_∞ by the autocorrelation dynamics assertions.

$$\ddot{\Delta(\tau)} = -\frac{\partial V}{\partial \Delta}$$

$$\ddot{\Delta} = \Delta - \{ g^2 \langle [\phi_i(t) \phi_i(t + \tau)] \rangle + \Sigma_m^2 \kappa^2 + 2 \Sigma_{mI} \kappa + \Sigma_I^2 \}$$

We can write out the potential function by integrating the negated RHS.

$$V(\Delta, \Delta_0) = \int \mathcal{D}\Delta \frac{\partial V(\Delta, \Delta_0)}{\partial \Delta}$$

$$V(\Delta, \Delta_0) = -\frac{\Delta^2}{2} + g^2\langle[\Phi_i(t)\Phi_i(t+\tau)]\rangle + (\Sigma_m^2\kappa^2 + 2\Sigma_{mI}\kappa + \Sigma_I^2)\Delta + C$$

We assume that as time goes to infinity, the potential relaxes to a steady state.

$$\frac{\partial V(\Delta_\infty, \Delta_0)}{\partial \Delta} = 0$$

$$\frac{\partial V(\Delta_\infty, \Delta_0)}{\partial \Delta} = -\Delta + \{g^2\langle[\phi_i(t)\phi_i(t+\infty)]\rangle + \Sigma_m^2\kappa^2 + 2\Sigma_{mI}\kappa + \Sigma_I^2\} = 0$$

$$\Delta_\infty = g^2\langle[\phi_i(t)\phi_i(t+\infty)]\rangle + \Sigma_m^2\kappa^2 + 2\Sigma_{mI}\kappa + \Sigma_I^2$$

$$\Delta_\infty = g^2 \int \mathcal{D}z \left[\int \mathcal{D}x\phi(\mu + \sqrt{\Delta_0 - \Delta_\infty}x + \sqrt{\Delta_\infty}z\right]^2 + \Sigma_m^2\kappa^2 + 2\Sigma_{mI}\kappa + \Sigma_I^2$$

Also, we assume that the energy of the system is perserved throughout the entirety of its evolution.

$$V(\Delta_0, \Delta_0) = V(\Delta_\infty, \Delta_0)$$

$$-\frac{\Delta_0^2}{2} + g^2\langle[\Phi_i(t)\Phi_i(t)]\rangle + (\Sigma_m^2\kappa^2 + 2\Sigma_{mI}\kappa + \Sigma_I^2)\Delta_0 + C = -\frac{\Delta_\infty^2}{2} + g^2\langle[\Phi_i(t)\Phi_i(t)]\rangle + (\Sigma_m^2\kappa^2 + 2\Sigma_{mI}\kappa + \Sigma_I^2)\Delta_\infty + C$$

$$\frac{\Delta_0^2 - \Delta_\infty^2}{2} = g^2\left(\langle[\Phi_i(t)\Phi_i(t)]\rangle - \langle[\Phi_i(t)\Phi_i(t)]\rangle\right) + (\Sigma_m^2\kappa^2 + 2\Sigma_{mI}\kappa + \Sigma_I^2)(\Delta_0 - \Delta_\infty)$$

$$\frac{\Delta_0^2 - \Delta_\infty^2}{2} = g^2\left(\int \mathcal{D}z\Phi^2(\mu + \sqrt{\Delta_0}z) - \int \mathcal{D}z \int \mathcal{D}x\Phi(\mu + \sqrt{\Delta_0 - \Delta_\infty}x + \sqrt{\Delta_\infty}z)\right)$$
$$+(\Sigma_m^2\kappa^2 + 2\Sigma_{mI}\kappa + \Sigma_I^2)(\Delta_0 - \Delta_\infty)$$

Consistency equations:

$$\begin{aligned}
\mu &= F(\mu, \kappa, \Delta_0, \Delta_\infty) = M_m\kappa + M_I \\
\kappa &= G(\mu, \kappa, \Delta_0, \Delta_\infty) = M_n\langle[\phi_i]\rangle + \Sigma_{nI}\langle\left[\phi_i'\right]\rangle \\
\frac{\Delta_0^2 - \Delta_\infty^2}{2} &= H(\mu, \kappa, \Delta_0, \Delta_\infty) = g^2\left(\int \mathcal{D}z\Phi^2(\mu + \sqrt{\Delta_0}z) - \int \mathcal{D}z \int \mathcal{D}x\Phi(\mu + \sqrt{\Delta_0 - \Delta_\infty}x + \sqrt{\Delta_\infty}z)\right) \\
&\quad +(\Sigma_m^2\kappa^2 + 2\Sigma_{mI}\kappa + \Sigma_I^2)(\Delta_0 - \Delta_\infty) \\
\Delta_\infty &= L(\mu, \kappa, \Delta_0, \Delta_\infty) = g^2 \int \mathcal{D}z \left[\int \mathcal{D}x\phi(\mu + \sqrt{\Delta_0 - \Delta_\infty}x + \sqrt{\Delta_\infty}z\right]^2 + \Sigma_m^2\kappa^2 + 2\Sigma_{mI}\kappa + \Sigma_I^2
\end{aligned} \tag{8}$$

Solver:

$$\begin{aligned}
x(t) &= \frac{\Delta_0(t)^2 - \Delta_\infty(t)^2}{2} \\
\Delta_0(t) &= \sqrt{2x(t) + \Delta_\infty(t)^2} \\
\dot{\mu}(t) &= -\mu(t) + F(\mu(t), \kappa(t), \Delta_0(t), \Delta_\infty(t)) \\
\dot{\kappa}(t) &= -\kappa + G(\mu(t), \kappa(t), \Delta_0(t), \Delta_\infty(t)) \\
\dot{x}(t) &= -x(t) + H(\mu(t), \kappa(t), \Delta_0(t), \Delta_\infty(t)) \\
\dot{\Delta_\infty}(t) &= -\Delta_\infty(t) + L(\mu(t), \kappa(t), \Delta_0(t), \Delta_\infty(t))
\end{aligned} \tag{9}$$

3.4 Integration of a noisy stimulus

Parameters:

$$z = \begin{bmatrix} g & M_m & M_n & M_I & \Sigma_m & \Sigma_n & \Sigma_{mI} & \Sigma_{nI} & \Sigma_\perp \end{bmatrix}^\top$$

Behavior:

$$z = \begin{bmatrix} \kappa(M_{I,low}), \kappa(M_{I,high}), \Delta_T, \kappa(M_{I,low})^2, \kappa(M_{I,high})^2, \Delta_T^2 \end{bmatrix}^\top$$

3.5 Rank 2 networks have the following consistency equations for Δ_0 and Δ_∞

$$\rho_m = \langle m_i^{(1)} m_i^{(2)} \rangle$$

$$\frac{\Delta_0^2 - \Delta_\infty^2}{2} = H(\mu, \kappa, \Delta_0, \Delta_\infty) = g^2 \left(\int \mathcal{D}z \Phi^2(\mu + \sqrt{\Delta_0} z) - \int \mathcal{D}z \int \mathcal{D}x \Phi(\mu + \sqrt{\Delta_0 - \Delta_\infty} x + \sqrt{\Delta_\infty} z) \right)$$
$$+ (2\rho_m \kappa_1 \kappa_2 + {\Sigma_m^{(1)}}^2 \kappa_1^2 + {\Sigma_m^{(2)}}^2 \kappa_2^2 + \Sigma_I^2)(\Delta_0 - \Delta_\infty)$$

$$\Delta_\infty = L(\mu, \kappa, \Delta_0, \Delta_\infty) = g^2 \int \mathcal{D}z \left[\int \mathcal{D}x \phi(\mu + \sqrt{\Delta_0 - \Delta_\infty} x + \sqrt{\Delta_\infty} z \right]^2 + 2\rho_m \kappa_1 \kappa_2 + {\Sigma_m^{(1)}}^2 \kappa_1^2 + {\Sigma_m^{(2)}}^2 \kappa_2^2 + \Sigma_I^2 \tag{10}$$

3.6 Context-dependent discrimination

$$y_A \sim \mathcal{N}(0, \Sigma_{y_A} = 1.2)$$
$$y_B \sim \mathcal{N}(0, \Sigma_{y_B} = 1.2)$$
$$I_A \sim \mathcal{N}(0, \Sigma_{I_A} = 1.2)$$
$$I_B \sim \mathcal{N}(0, \Sigma_{I_B} = 1.2)$$
$$I_{ctx,A} \sim \mathcal{N}(0, \Sigma_{I_{ctx,A}} = 1)$$
$$I_{ctx,B} \sim \mathcal{N}(0, \Sigma_{I_{ctx,B}} = 1)$$

$$I(t) = c_A(t) I^A + c_B(t) I^B + \gamma_A I_{ctx,A} + \gamma_B I_{ctx,B}$$
$$m^{(1)} = y_A + \rho_m I_{ctx,A} + \beta_m w$$
$$n^{(1)} = I^A + \rho_n I_{ctx,A} + \beta_n w$$
$$m^{(2)} = y_B + \rho_m I_{ctx,B} + \beta_m w$$
$$n^{(2)} = I^B + \rho_n I_{ctx,B} + \beta_n w$$
$$y(t) = \beta_m (\kappa_1 + \kappa_2) \langle \left[\phi_i' \right] \rangle$$

$$\Sigma_I = c_A(t) \Sigma_{I_A} + c_B(t) \Sigma_{I_B} + \gamma_A \Sigma_{I_{ctx,A}} + \gamma_B \Sigma_{I_{ctx,B}}$$
$$\Sigma_m^{(1)} = \Sigma_{y_A} + \rho_m \Sigma_{I_{ctx,A}} + \beta_m \Sigma_w$$

$$\Sigma_m^{(2)} = \Sigma_{y_B} + \rho_m \Sigma_{I_{ctx,B}} + \beta_m \Sigma_w$$

Parameters:

$$z = \begin{bmatrix} g & \rho_m & \rho_n & \beta_m & \beta_n \end{bmatrix}^\top$$

Consistency equations:

$$\begin{aligned}
\kappa_1(t) &= F(\kappa_1(t), \kappa_2(t), \Delta_0(t), \Delta_\infty(t)) = \rho_m \rho_n \kappa_1 \langle \left[\phi_i'\right] \rangle + \beta_m \beta_n (\kappa_1 + \kappa_2) \langle \left[\phi_i'\right] \rangle + c_A \Sigma_I^2 + \rho_n \gamma_A \\
\kappa_2(t) &= F(\kappa_1(t), \kappa_2(t), \Delta_0(t), \Delta_\infty(t)) = \rho_m \rho_n \kappa_2 \langle \left[\phi_i'\right] \rangle + \beta_m \beta_n (\kappa_1 + \kappa_2) \langle \left[\phi_i'\right] \rangle + c_B \Sigma_I^2 + \rho_n \gamma_B \\
\frac{\Delta_0^2 - \Delta_\infty^2}{2} &= H(\mu, \kappa, \Delta_0, \Delta_\infty) = g^2 \left(\int \mathcal{D}z \Phi^2(\mu + \sqrt{\Delta_0} z) - \int \mathcal{D}z \int \mathcal{D}x \Phi(\mu + \sqrt{\Delta_0 - \Delta_\infty} x + \sqrt{\Delta_\infty} z) \right) \\
&+ ((\Sigma_w^2 + \beta_m^2)(\kappa_1^2 + \kappa_2^2) + \Sigma_I^2 (c_A^2 + c_B^2) + (\rho_m \kappa_1 + \gamma_A)^2 + (\rho_m \kappa_2 + \gamma_B)^2)(\Delta_0 + \Delta_\infty) \\
\Delta_\infty &= L(\mu, \kappa, \Delta_0, \Delta_\infty) = g^2 \int \mathcal{D}z \left[\int \mathcal{D}x \phi(\mu + \sqrt{\Delta_0 - \Delta_\infty} x + \sqrt{\Delta_\infty} z \right]^2 + \\
&(\Sigma_w^2 + \beta_m^2)(\kappa_1^2 + \kappa_2^2) + \Sigma_I^2 (c_A^2 + c_B^2) + (\rho_m \kappa_1 + \gamma_A)^2 + (\rho_m \kappa_2 + \gamma_B)^2
\end{aligned} \tag{11}$$

Solver:

$$\begin{aligned}
x(t) &= \frac{\Delta_0(t)^2 - \Delta_\infty(t)^2}{2} \\
\mu &= 0 \\
\Delta_0(t) &= \sqrt{2x(t) + \Delta_\infty(t)^2} \\
\dot{\kappa_1}(t) &= -\kappa_1 + F(\kappa_1(t), \kappa_2(t), \Delta_0(t), \Delta_\infty(t)) \\
\dot{\kappa_2}(t) &= -\kappa_2 + G(\kappa_1(t), \kappa_2(t), \Delta_0(t), \Delta_\infty(t)) \\
\dot{x}(t) &= -x(t) + H(\kappa_1(t), \kappa_2(t), \Delta_0(t), \Delta_\infty(t)) \\
\dot{\Delta_\infty}(t) &= -\Delta_\infty(t) + L(\kappa_1(t), \kappa_2(t), \Delta_0(t), \Delta_\infty(t))
\end{aligned} \tag{12}$$

Behavior:

$$z = \begin{bmatrix} y_1, y_2, y_3, y_4, \Delta_T, \text{sec moments...} \end{bmatrix}^\top$$

3.7 Chaotic limit cycles

$$\begin{aligned}
m^{(1)} &= \alpha x_1 + \rho y_1 \\
n^{(1)} &= \alpha x_2 + \rho y_2 \\
m^{(2)} &= \alpha x_3 + \rho y_2 + \gamma \rho y_1 \\
n^{(2)} &= \alpha x_4 - \rho y_1
\end{aligned}$$