Coarse–graining, fixed points, and scaling in a large population of neurons

1 Introduction

The paper in analysis [1] uses data from mouse brain pulse correlations to find 'hints of scaling' - that is, through a phenomenological approach, find data indicating that this type of network is self-similar with respect to the change of scale. It does so by applying a Renormalization Group strategy to coarsen the variables, getting to scales with increasingly less details about the network.

The experiment that generated the network can be summarized as follows:

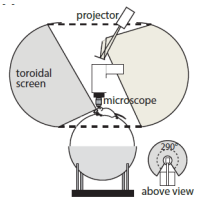

Figure 1: Schematic of the experiment.

The mouse has its brain genetically modified to produce a protein that fluoresces in the presence of Calcium. This is useful since, when there is a synaptic pulse, Calcium enters the intra-cellular medium and slowly goes back out. This allows for a microscope to get signal from brain activity in that region.

The mouse was then fixed in position by its head, and had a microscope reading signals in its hippocampus, in a region with particularly high concentration of place-cells, which are active only when the animal visits a particular spot in the environment. It was then placed over a Styrofoam ball that was free to move, with its rotations recorded by an optical sensor, which, by its turn, was connected to a virtual reality screen that was wrapped around the animal's field of view, making it feel like it was freely running around. The microscope would then record fluorescence marks of correlations of neuron activity, and this was the data that composed the network used in the paper we are analyzing.

With this data in hands, we can look at different strategies to find parameters that may indicate that the network shows self-similarity.

2 Derivations

2.1 PCA

The first phase of the RG process is a coarsening step, which is related to dimensionality reduction techniques in data analysis. A very prominent one is Principal Component Analysis (PCA). In this method, we take σ_i variables living in a vector space V, and find the Euclidean projection onto a subspace $S \subset V$ that simplifies the covariance matrix. This is done as follows:

First, we find the eigenvalues and normalized eigenvectors of the covariance matrix of the system:

$$C_{ij} = \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle$$

$$\sum_{j=1}^{N} C_{ij} u_{jr} = \lambda_r u_{ir}$$

Let's suppose that the eigenvalues are ordered: $\lambda_1 > \lambda_2 > \lambda_3 > \ldots$.

Now we define the projector operator onto the subspace associated with the K largest eigenvalues:

$$\hat{P}_{ij}(K) = \sum_{r=1}^{K} u_{ir} u_{jr}$$

Next, the coarse-graining step is about transforming the previous variable σ_i into a $\widetilde{\sigma}_i = \sum_{j=1}^{N} \hat{P}_{ij}(K) \sigma_j$.

The PCA works best as an approximation if the following condition holds:

$$\frac{\sum_{i=1}^{D} \lambda_i}{\sum_{j=1}^{N} \lambda_j} \to 1, \text{ with } D \ll N$$

Having understood the idea behind PCA, we notice that this technique, together with many other famous dimensionality reduction ones, are approaches to analyzing data, and not models, like RG is, but they can work as a first step of the RG process.

In the case of real space RG, the most used techniques are related to clustering on a hierarchical fashion, in which at each step, the variables are grouped together (usually by summing them up) so that a set of 'nodes' in the previous iteraction becomes one 'supernode' in the next one.

For momentum space RG, on the other hand, in translation-invariant systems, the Fourier components diagonalize the covariance matrix [see Appendix 1: Diagonalizing Translation Invariant Systems]. This provides us with an easy

way to find the eigenvalues and eigenvectors of our matrix, so that the coarse-graining step in momentum space can be solved as a PCA problem.

It is important to notice that this process generates, at each step, variables that are linear combinations of the original ones, and the more we iterate on it, the more original variables get packed into linear combinations to form the coarsened ones. If the correlations of the system decay fast with the distance, eventually this process creates a probability distribution that converges to a fixed Gaussian form. The interesting results, though, are when that distribution converges to fixed non-Gaussian curves. This shows that the correlation length approaches infinity, which represents a fixed point in the RG flow, and it is associated with scaling behaviors of the system's parameters.

The key aspect we are usually interested in when dealing with RG is to understand how the probability distribution varies as we change the threshold representing the size of the clusters (K). This is the same as varying the definition of what should be taken as 'important' and what should be regarded as 'details', and understanding the behavior of $P(K)$ is our goal for the next session.

2.2 Coarsening

Our strategy is to begin with two approaches to the coarse-graining step which are analogous to real space and momentum space, and then see hints of a fixed point in the joint probability distribution of the whole network, $P(\sigma_i)$.

Real Space

Usually, the coarse-graining in physical systems happen by averaging over a small neighborhood of each point in space. In our case, we have a neuronal network, composed of extended elements that can connect large regions, so this approach is not possible. What we can do, instead, is to look at the correlations between different neurons directly, and the coarse-graining step happens by grouping together the most correlated pairs.

First, let us describe the activity in neuron i at time t as $\sigma_i(t)$. Since the spatial coarsening we are conducting does not mix different moments in time, we can drop the index: $\sigma_i(t) \equiv \sigma_i$. This time dependency is later looked at in detail in the article. We can then define the correlation matrix for each pair of neurons:

$$c_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}} \tag{1}$$

The approach then is to use the correlation matrix formed by these correlation coefficients c_{ij} and extract its largest non-diagonal value. The pairs of

indices of neurons associated with this value will be named (i, j_*). Then the steps are:

- Obtain the coarse-grained variable $\sigma_i^{(2)} = (\sigma_i^{(1)} + \sigma_{j_*}^{(1)})$
- Remove the pair (i, j_*) and greedily search for the next maximally correlated pair
- This iteration ends when, out of the N neurons that we have started with, only $\lfloor N/2 \rfloor$ are left
- Repeat the process for the coarsened layer formed in the previous iteration
- Stop when the clusters have grown to the desired size.

Proceeding this way, the final size obtained for the clusters is K, which can only be a power of 2, by the way we have built the coarsening step.

Momentum Space

As noticed before, for systems showing translation invariance, the eigenvectors are the Fourier components, and so, to apply PCA to it, we order those components by their associate eigenvalues, and place a cutoff, keeping only the $\hat{K}$ largest eigenvalues.

In momentum space, we begin the process in the same way as in real space, by defining a covariance matrix:

$$C_{ij} = \left\langle \sigma_i^{(1)} \sigma_j^{(1)} \right\rangle - \left\langle \sigma_i^{(1)} \right\rangle \left\langle \sigma_j^{(1)} \right\rangle$$

This time, however, we will not use it to define a correlation matrix and proceed hierarchically by removing the pairs showing the largest correlations, as before. We will, instead, find the eigenvalues and eigenvectors of C (its Fourier components), order them, and build the projection operator, $\hat{P}_{ij}$, over the $\hat{K}$ largest ones, as explained in the introduction [see PCA].

The coarse-grained variables are, then, defined as:

$$\phi_{\hat{K}}(i) = z_i(\hat{K}) \sum_j \hat{P}_{ij}(\hat{K}) \left[\sigma_j^{(1)} - \left\langle \sigma_j^{(1)} \right\rangle \right]$$

Where we subtracted the mean for all the variables calculated in the first layer and used a normalizing factor $z_i(\hat{K})$ to ensure that $\left\langle \phi_{\hat{K}}^2(i) \right\rangle = 1$.

Moments of the Distribution

Before analyzing the probability distribution as a whole, we can first look at the dependence of the first and second moments on K.

First Moment: For real space, we defined a coarse-graining based on summing up maximally correlated variables from the previuos iteration. The final size of the clusters is K, and at each iteration, the size is $\frac{1}{2^n}K$, where n is the number of iterations remaining before reaching size K. For this reason, we can be sure that:

$$M_1(k) = \frac{1}{N_k}\sum_{i=1}^{N_k}\left\langle \sigma_i^{(k)} \right\rangle \quad \sim \quad K$$

This summation means that after k steps we have reached N_k clusters, each one having $K = 2^{k-1}$ of the original variables.

Second Moment: If all the variables in each iteration were completely independent from each other, the variance of the distribution would be proportional to K. This can be seen from the fact that the total number of neurons is always the same, for all iterations, and it is given by: $T = K_k \cdot N_k$, so that the number of clusters in each iteration is proportional to $\frac{1}{K_k}$, and given that the standard deviation of a system is proportional to $\frac{1}{\sqrt{N}}$, the variance must be proportional to $\left(\frac{1}{\sqrt{N}}\right)^2 = K$.

The other corner case is the configuration in which all the variables are perfectly correlated, so that the variance would be proportional to K^2.

Anything in between these two values could be an indication that the correlations have a self-similar structure.

Full Distribution

Now that we have a first criterion to check if the systems may show self-similar behavior, let us look into the whole probability distribution for the set of neurons. This probability can be split into the probability of complete silence of the system (all neurons inactive) and that of nonzero activity:

$$P(\sigma_i^{(k)}) = P_{silence}(K)\ \delta(\sigma_i^{(k)}, 0)\ +\ [1 - P_{silence}(K)]\ F_K(\sigma_i^{(k)}/K)$$

Here the term $\delta(\sigma_i^{(k)}, 0)$ is the Kronecker's delta, that picks only the configuration with all activity variables set to zero, and $\phi = \sigma_i^{(k)}/K$ is the normalized activity, that measures how many neurons are active in the cluster of size K at each moment.

3 Results

Real Space

Correlation matrix

In the real space coarsening (finding direct correlations between different neurons), we get the following evolution for the correlation matrix:

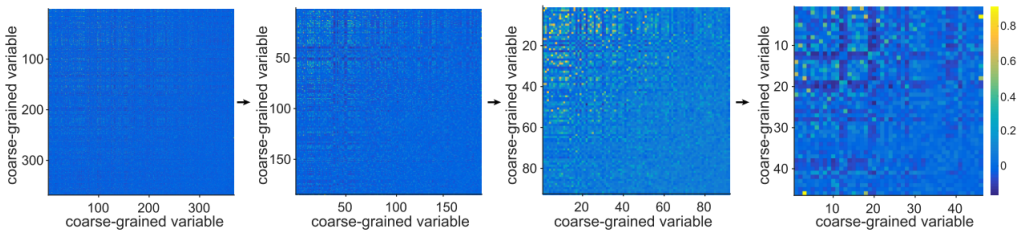

Figure 2: Correlation matrix direct coarsening evolving with K = 4, 8, 16, 32.

We can see that it clusters more and more. If we were to have finite correlation length in the system, the blocks would eventually become disjoint, but we see increasing correlations between the variables.

Variance

By plotting the variance with respect to K, as we discussed in the derivations section, exponents with values between 1 and 2 may be a hint of a scaling behavior. What we get is a scaling with $K^{\widetilde{\alpha}}$, $\widetilde{\alpha} = 1.40 \pm 0.06$:

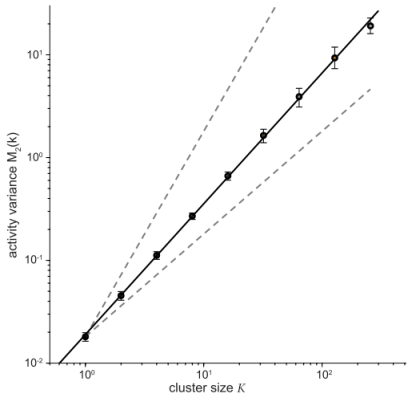

Figure 3: Variance of the system and the linear and square bounds.

Silence and activity

If the power-law dependence in the variance is a sign of self-similarity, this should also be present in other metrics. One of them is the full probability distribution, as described before. Assuming that the system is in equilibrium, we can write:

$$P_K\left(\sigma_i^{(1)}\right) = \frac{1}{Z_K} exp\left[-E\left(\sigma_i^{(1)}\right)\right]$$

In which the energy of the system can be Taylor-expanded into:

$$E\left(\sigma_i^{(1)}\right) = \sum_{i=1}^{K} h_i \sigma_i^{(1)} \; + \; \sum_{i,j=1}^{K} J_{ij} \sigma_i^{(1)} \sigma_j^{(1)} \; + \; \sum_{i,j,k=1}^{K} G_{ijk} \sigma_i^{(1)} \sigma_j^{(1)} \sigma_k^{(1)} + \ldots$$

By restricting the analysis to the probability of total silence in the cluster:

$$\sigma_i^{(1)} = 0 \; , \; \forall i \implies E = 0 \implies P = \frac{1}{Z_K}$$

From this we can build a free energy $F(K) = -ln\, Z_K$, and plot this function with respect to K, obtaining the following data, that fit well with the power law:

$$F = -\alpha K^{\widetilde{\beta}} \; , \; \widetilde{\beta} = 0.88 \pm 0.01$$

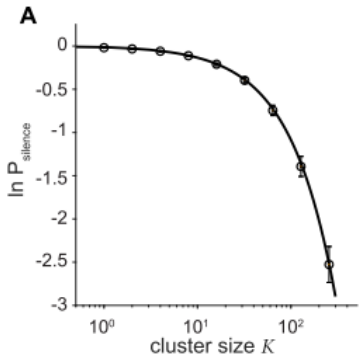

Figure 4: Probability of silence for the whole cluster with respect to the size K.

Therefore, the probability of silence in a population of K neurons does not fall exactly exponentially with K, as it would be expected for an uncorrelated network.

Momentum Space

In momentum space, as explained above, the way we can coarsen the network is by projecting the correlation matrix in the $\hat{K}$ eigenvectors associated with the largest eigenvalues. By varying $\hat{K}$, the article shows the probability density as a function of the normalized activity, and here again we see a convergence to a fixed form for the probability distribution which is non-Gaussian:

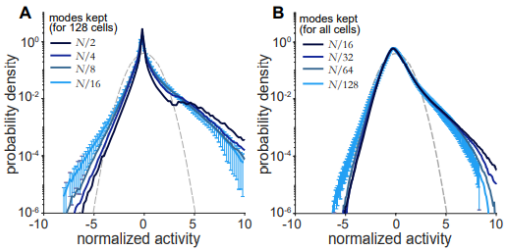

Figure 5: Probability density as a function of the activity for different values of $\hat{K}$. The curves vary less in the rightmost graph (more steps performed).

Dynamic Scaling

If we build temporal correlation functions, similar to the one we have defined previously, and plot them as functions of time normalized by a correlation time which is a function of K (cluster size), the correlation curves match each other almost perfectly, varying the cluster size from 2 to 256 neurons:

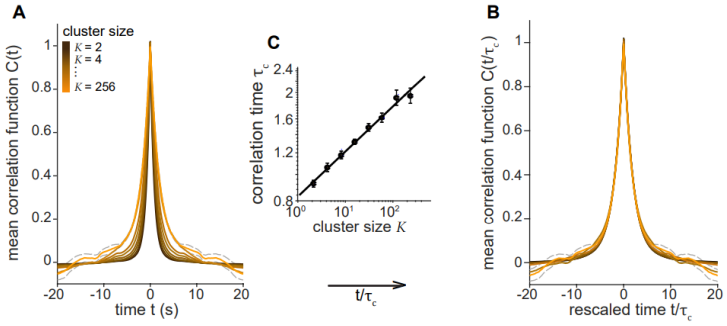

Figure 6: Mean temporal correlation between neurons as time passes. The leftmost graph show different curves that match very well when the x-axis gets normalized by the correlation time (depending only on K).

The paper shows other metrics, but the point they make is the same: showing indications that the network in analysis shows scaling behavior, and is, therefore, self similar across a considerably large band of values for K.

4 Appendix 1: Diagonalizing Translation Invariant Systems

Proposition: In systems with translation invariance, the matrix of correlations among microscopic variables is diagonalized by a Fourier transform.

Lemma: Suppose two matrices $\mathbf{M_1}$ and $\mathbf{M_2}$ each have a complete basis of eigenvectors. Then they share a common basis iff they commute: $\mathbf{M_1 M_2} = \mathbf{M_2 M_1}$.

Proof: Refer to page 14 of [3]

Def: The translation operator $\mathcal{T}_a$ is defined by the action over function f as:

$$\mathcal{T}_a f(t) = f(t+a)$$

Def: A translation invariant operator is one that commutes with $\mathcal{T}_a$, that is, an operator $\mathcal{O}$ for which:

$$\mathcal{T}_a \mathcal{O} f(t) = \mathcal{O} \mathcal{T}_a f(t)$$

In operators of this type, it follows from the Lemma that eigenfunctions that diagonalize the translation operator also diagonalize $\mathcal{O}$.

Theorem: The eigenfunctions of the translation operator are the Fourier basis functions, defined by $\mathbf{e_k(t)} = \frac{\mathbf{1}}{\sqrt{\mathbf{T}}} \mathbf{e}^{\mathbf{i}\frac{\mathbf{2\pi kt}}{\mathbf{T}}}$.

Proof:

$$\mathcal{T}_a e_k(t) = e_k(t+a) = \frac{1}{\sqrt{T}} e^{i\frac{2\pi k(t+a)}{T}} = \frac{1}{\sqrt{T}} e^{i\frac{2\pi ka}{T}} e^{i\frac{2\pi kt}{T}} = e^{i\frac{2\pi ka}{T}} e_k(t)$$

Therefore we arrive at the eigenvalue equation for the operator $\mathcal{T}_a$:

$$\mathcal{T}_a e_k = e^{i\frac{2\pi ka}{T}} e_k$$

Therefore, any operator that commutes with the translation operator (in other words, translation-invariant) is diagonal in the basis of eigenvectors of the Fourier transform, which are the e_k, as we wanted to prove in this session. ■

References

[1] Meshulam, L. et. at., *Coarse–graining and hints of scaling in a population of 1000+ neurons*, Phys. Rev. Lett. 123 (2019).

[2] `https://cims.nyu.edu/~cfgranda/pages/MTDS_spring19/notes/fourier.pdf`

[3] `https://ctn.zuckermaninstitute.columbia.edu/sites/default/files/content/Miller/math-notes-4.pdf`