8.334 Final Project: Information Coding in 1D neural networks*

Felix Barber†
Department of Molecular and Cellular Biology, Harvard University.

Supervised: Prof. Haim Sompolinsky & Dr. Yu Hu
(Dated: May 16, 2015)

We consider renormalization and information coding properties of a 1D chain of connected neurons with a continuous variable (firing rate) at each site. Our implementation of renormalization largely follows [1]. A form for the entropy in the limit of $\omega = 0$ is obtained and it is demonstrated that the entropy decreases under renormalization close to a critical fixed point. A closed form for the Fisher information in frequency is obtained, and shown to exhibit regular behavior under renormalization at a critical fixed point.

I. INTRODUCTION

The concept of self organized criticality has been considered in the context of neural networks for some time. Studies have demonstrated power law scaling in the propagation of spontaneously generated local field potentials within cortical networks [2, 3]. More recently, the renormalization group has found application within the theory of deep learning [4]. Motivated to some extent by these studies, we consider a highly simplified system: a collection of neurons excited by an incident signal $S(t)$ and weights $\vec{w}$ (e.g. neurons in a chain are connected to one large and driven excitatory interneuron), subject to Langevin dynamics with *asymmetric* couplings between nearest neighbor neurons. Dynamics in systems of such a form have been studied extensively, e.g. [5]. We Fourier transform in time and implement decimation on this dynamical system, following [1]. One subject of interest is the information coding properties of the system close to a fixed point, and we quantify this by briefly studying the entropy and Fisher information of the system under renormalization.

II. RESULTS

A. Asymmetric coupling decimation

Consider a network of neurons with asymmetric nearest neighbor interactions. Neurons are labelled by a site index i with an associated firing rate ϕ_i. We describe this system using the Langevin equation 1 with a velocity term that due to the interaction asymmetry can no longer be described as the derivative of a Hamiltonian.

$$\frac{\partial \phi_i}{\partial t} = -a\phi_i + b\phi_{i+1} + c\phi_{i-1} + \eta_i(t) \quad (1)$$

Here $\vec{\eta}$ is a gaussian distributed noise with mean

$$\langle \eta_i(t) \rangle = 0 \quad (2)$$

and variance

$$\langle \eta_i(t)\eta_j(t') \rangle = 2D_{i,j}(t - t'). \quad (3)$$

We note that following [1] we have introduced noise couplings which are not necessarily limited to the same site $D_{i,j} = D_0\delta_{i,j} + D_1\delta_{i\pm1,j}$. The reason for this is that upon renormalization in real space via decimation of odd numbered sites, we naturally introduce non-zero coupling D_1 between nearest neighbors. We also assume that there may exist correlations in time. Similarly, this is necessitated by the appearance of frequency terms in D under renormalization.

Rather than recast this into a Fokker-Planck equation, a more general RG analysis of which was studied by Janssen [6], we follow the same procedure as [1]. We Fourier transform in time to obtain

$$\alpha\hat{\phi}_i(\omega) = b\hat{\phi}_{i+1} + c\hat{\phi}_{i-1} + \hat{\eta}_i(\omega). \quad (4)$$

Here $\alpha = a - \mathrm{i}\omega$. This can now be decimated over all sites $2i + 1$ to obtain

$$(\alpha^2 - 2bc)\hat{\phi}_{2i} = b^2\hat{\phi}_{2i+2} + c^2\hat{\phi}_{2i-2} + \alpha\hat{\xi}_{2i} \quad (5)$$

with $\hat{\xi}_{2i} = \hat{\eta}_{2i} + (b\hat{\eta}_{2i+1} + c\hat{\eta}_{2i-1})/\alpha$.

We rescale and renormalize:

$$\hat{\phi}_{2i}(\alpha, D) = \alpha A\hat{\phi}'_i(\alpha', D') \quad (6a)$$

* Thank you to Prof. Haim Sompolinsky and Dr. Yu Hu for supervising this project.
† felixbarber@g.harvard.edu

$$\hat{\xi}_{2i} = A\hat{\eta}'_i \tag{6b}$$

This system has recursion relations

$$\begin{aligned} \alpha' &= \alpha^2 - 2bc, \\ a' &= a^2 - 2bc, \\ \omega' &= 2\omega - \mathrm{i}\omega^2 \\ b' &= b^2, \\ c' &= c^2. \end{aligned}$$

yielding finite fixed points $(a, b, c) = (2, 1, 1)$ (i), $(1, 1, 0)$ (ii), $(1, 0, 1)$ (iii) and $(0, 0, 0)$ (iv) (infinite fixed points are not considered on biological grounds). Here (iv) is the trivial fixed point. We note that to allow relaxation of spatial Fourier modes in the limit of long times we require $a \geq b + c$. It is easy to show that this relation is preserved by these RG equations.
Note that under this rescaling we have different noise couplings D'_0 and D'_1 (hereafter relabeled as $\vec{D}$):

$$\langle \hat{\eta}'_i(\omega_1)\hat{\eta}'_j(\omega_2)\rangle = 2(2\pi)\hat{D}'_{i,j}(\omega_1)\delta(\omega_1 + \omega_2) \tag{7}$$

defined by by $\vec{D}'(\omega') = T(\omega)\vec{D}(\omega)$, where T is defined in the $\omega = 0$ limit by equation 8.

$$T = \frac{2a}{A^2}\begin{pmatrix} 1 + \frac{b^2+c^2}{a^2} & 2\frac{b+c}{a} \\ \frac{bc}{a^2} & \frac{b+c}{a} \end{pmatrix} \tag{8}$$

By considering the eigenvalues of T at each fixed point we can study the behavior of fluctuations under renormalization. We note that in order to constrain fluctuations from diverging under repeated decimation while retaining a fixed point for D we must specify $A = 2^{3/2}$ for fixed point (i). This choice, however, forces fluctuations to become progressively smaller at the fixed points (ii) or (iii), as shown in the following table of eigenvalues and eigenvectors.

	FP i	FP ii	FP iii
λ_1	1	1/2	1/2
λ_2	1/4	1/4	1/4
v_1	$(4, 1)$	$(1, 0)$	$(1, 0)$
v_2	$(-2, 1)$	$(-2, 1)$	$(-2, 1)$

The smaller eigenvalue λ_2 corresponds to an "irrelevant" direction in fluctuation space in both cases.

B. Information coding close to a fixed point

We aim to describe the system of equation 4, with addition of a signal $S(t)$ and weights w_i. Written in vector notation this becomes

$$-\mathrm{i}\omega\vec{\phi}(\omega) = -A\vec{\phi}(\omega) + \vec{\eta}(\omega) + \vec{w}S(\omega). \tag{9}$$

Here we have removed the $\hat{\phi}$ when denoting the time Fourier transform for notational convenience.

If $A - \mathrm{i}\omega I$ is invertible then this can be solved as

$$\vec{\phi}(\omega) = (A - \mathrm{i}\omega I)^{-1}(\vec{\eta}(\omega) + \vec{w}S(\omega)). \tag{10}$$

This solution clearly separates into deterministic and stochastic components. As the stochastic component is the sum of normally distributed random variables $\vec{\eta}$, we expect $\vec{\phi}$ to itself be gaussian, with mean $\langle\vec{\phi}(\omega)\rangle = (A - \mathrm{i}\omega I)^{-1}\vec{w}S(\omega)$. We calculate the variance of this distribution as

$$\begin{aligned} \langle(\phi_i(\omega) - \mu_i(\omega))(\phi_j(\omega') - \mu_j(\omega'))\rangle &= 2(2\pi)\delta(\omega + \omega') \\ &\times (A - \mathrm{i}\omega I)^{-1}_{ik}(A - \mathrm{i}\omega' I)^{-1}_{jl}D_{kl}(\omega) \\ &= 2(2\pi)\delta(\omega + \omega')[(A - \mathrm{i}\omega I)^{-1}D(A^T - \mathrm{i}\omega I)^{-1}]_{ij} \\ &= \Sigma_{ij}(\omega) \end{aligned}$$

We are now interested to calculate coding properties of these neurons close to a critical fixed point. There are a number of measures of information coding which could be studied, including the entropy, mutual information and Fisher information. The mutual information requires further assumptions about the distribution of signals $S(t)$, perhaps the simplest of which would involve a gaussian weight $(1/2)\int \mathrm{d}\tau[tS(\tau)^2 + K(\partial_\tau S(\tau))^2]$. The Fisher information has found application previously in similar contexts [7] as a means of capturing the extent to which the conditional distribution changes when the signal history changes. We shall treat the Fisher information later. For now, we simply compute the entropy of our gaussian distribution in firing rates, conditioned over the signal $S(t)$.

$$\begin{aligned} S &= -\int \mathrm{d}^N\vec{\phi}\, p(\vec{\phi}|\{S(\omega)\})\log(p(\vec{\phi}|\{S(\omega)\})) \\ \Rightarrow S &= \frac{1}{2}\log\left|(2\pi e)^N \det(\Sigma)\right| \end{aligned}$$

Note that the result is independent of the signal $S(t)$, so that (for this calculation) we do not need to consider the behavior of S under decimation. We can see that the determinant separates into the determinant of D and a matrix that depends solely on the variables a, b, c and ω: $\det(\Sigma) = (4\pi)^N\det(D)/\det(A - \mathrm{i}\omega I)^2$.

D and A are so-called circulant matrices, defined only by circularly shifted vectors $\vec{D}$ and $\vec{A}$. A general circulant matrix C defined by a vector $\vec{c}$ has eigenvalues

$$\lambda_j = c_0 + c_{N-1}\omega_j + c_{n-2}\omega_j^2 + ... + c_1\omega_j^{N-1} \quad (11)$$

with $\omega_j = \exp[2\pi \mathrm{i} j/N]$. This indicates that in the limit of zero frequency $\omega \to 0$ the relevant determinants are:

$$\det(D) = \prod_{j=0}^{N-1} (D_0 + 2D_1 \cos(2\pi j/N))$$
$$\det(A) = \prod_{j=0}^{N-1} \left(a - be^{2\pi \mathrm{i} j/N} - ce^{-2\pi \mathrm{i} j/N}\right)$$

It is therefore clear that $\det(A) = 0$ at critical fixed points (i), (ii) and (iii) due to the $j = 0$ eigenvalue and saturation of the finite time relaxation inequality $a \geq b + c$ at these points. The entropy diverges at a critical fixed point, and therefore must decrease upon application of the renormalization group.

To demonstrate how this decrease occurs, we consider the behavior of entropy under decimation close to the critical fixed point (i), with $a = a_c + \epsilon$ for $\epsilon > 0$, $b = c = 1$. We note that at a fixed point renormalization of an initial noise correlation $\vec{D} = u_1\vec{v_1} + u_2\vec{v_2}$ leads to $\vec{D}^{(n)} = \lambda_1 u_1 \vec{v_1} + \lambda_2 u_2 \vec{v_2}$. At (i) $\vec{v}_1$ is marginal under RG application, so that the stable fixed point for noise correlations obtained here will be $u_1\vec{v}_1$. With this in mind we also assume u_2 is initially very small, so that at (i) the rescaling of $\det(D)$ will be $1 + \mathcal{O}(\epsilon, u_2)$. In contrast, $\det(A)$ is rescaled by a positive factor ≥ 4 (shown below). This combined with the number reduction $N' = N/2$ clearly leads to an overall reduction in entropy upon renormalization close to the critical fixed point:

$$\epsilon' = 2a_c\epsilon + \mathcal{O}(\epsilon^2)$$
$$\Rightarrow a' - b'e^{2\pi \mathrm{i} j/N} - c'e^{-2\pi \mathrm{i} j/N} = a - 2\cos(2\pi j/N) + (2a_c - 1)\epsilon + \mathcal{O}(\epsilon^2)$$
$$\Rightarrow \det(A') \approx \det(A) \times \left(1 + (2a_c - 1)\epsilon \sum_{j=0}^{N-1} \frac{1}{2 + \epsilon - 2\cos(2\pi j/N)}\right)$$

The denominator above remains positive for all j, attaining its minimal value of ϵ for $j = 0$. This leads to a rescaling of $\det(A)$ by a factor ≥ 4, which compared with the magnitude of rescaling of $\det(D)$ results in an overall entropy decrease. However, note that this decrease in total entropy may be trivially true, given the reduction of site number $N' = N/2$. It would be interesting to derive the behavior of the entropy *per site* under RG action close to a fixed point, but that remains for further work.

We can also calculate the Fisher information in frequency $J(S(\omega'))$, defined by equation 12. As mentioned earlier, the Fisher information is generally used to extract information about a change in conditional distribution with changing signal history. Our result does not have this interpretation, as it represents the extent of induced change when a Fourier mode in the signal changes.

$$J(S(\omega')) = \left\langle \frac{\delta^2 \log(p(\vec{\phi}|\{S(\omega)\}))}{\delta S(\omega')^2} \right\rangle_{\{S(\omega)\}} \quad (12)$$

Here the subscript $\{S(\omega)\}$ indicates that the average should be taken over the probability distribution conditioned on $S(\omega)$. For the gaussian weights here this yields the comparatively simple result $J(S(\omega')) = \vec{w}^T D^{-1}(\omega')\vec{w}/(4\pi)$. Interestingly, this quantity remains well defined at a critical fixed point under renormalisation, with noise couplings tending towards a fixed point $\vec{D} = (4, 1)$ at which $\det(D)$ is non-zero.

III. DISCUSSION

We have considered an application of the renormalization group to an asymmetric chain of neurons. We studied the entropy of the system conditioned over an excitatory signal $S(t)$, and found that this diverged at a critical fixed point. This perhaps is to be expected, since the appearance of a zero eigenvalue in the coupling matrix at criticality is representative of a divergence in variance at a critical fixed point. The demonstration of entropy decrease upon renormalization close to a critical fixed point is interesting, but it seems that since the site number is rescaled by a factor of 1/2 this result is trivial. What would be more interesting is to obtain this result for the entropy per site, but work on this is ongoing.

Our study of the Fisher information in frequency demonstrated regular behavior at a critical fixed point, which is interesting by contrast to the highly irregular behavior of entropy at these points. The Fisher information in general remains finite at a critical fixed point in coupling

space (depending upon initial conditions for $\vec{D}$), and under renormalization approaches a fixed point in noise correlations at which the matrix D^{-1} is well defined.
Further work will implement specific weights for signals $S(t)$ to extract the mutual information of the system. We will also attempt to incorporate the finite but small ω case, treating this as a perturbation about the $T(\omega = 0)$ limit.
Here asymmetry in the site interactions was motivated by the observation that neuron excitation is not a purely symmetric process. Another aspect of neural networks which is of interest is that of *next* nearest neighbor interactions. We have attempted to treat this problem in a similar method to that above, beginning with the Langevin equation 13.

$$\frac{\partial \phi_i}{\partial t} = -a\phi_i + b\phi_{i\pm1} + c\phi_{i\pm2} + \eta_i(t) \tag{13}$$

However, under decimation, this chain produces non-zero interactions between *all* remaining neurons. One can adopt a pertubative RG approach by taking an expansion in powers of c, the next nearest neighbor coupling parameter, where one assumes $c \ll 1$. This RG approach has been taken to first order in c, and early results are promising.

IV. CONCLUSIONS

We considered a 1D chain of neurons with firing rate as the site variable. Decimation was implemented in a manner following [1]. The $\omega = 0$ form of the entropy was obtained, and was shown to diverge at the critical fixed points of the system. It was demonstrated in one specific case that this entropy decreases under renormalization close to a fixed point. We also obtained a simple closed form for the Fisher information in frequency, and studied this closed form at a critical fixed point in coupling space.

[1] Giacometti, A., Maritan, A., Toigo, F., & Banavar, J. R. (1995). Real-space renormalization group for Langevin dynamics in absence of translational invariance. *Journal of statistical physics*, 79(3-4), 649-668.

[2] Beggs, J. M., & Plenz, D. (2003). Neuronal avalanches in neocortical circuits. emphThe Journal of neuroscience, 23(35), 11167-11177.

[3] Shriki, O., Alstott, J., Carver, F., Holroyd, T., Henson, R. N., Smith, M. L., ... & Plenz, D. (2013). Neuronal avalanches in the resting MEG of the human brain. *The journal of Neuroscience*, 33(16), 7079-7090.

[4] Mehta, P., & Schwab, D. J. (2014). An exact mapping between the Variational Renormalization Group and Deep Learning. *arXiv preprint* arXiv:1410.3831.

[5] Buice, M. A., Cowan, J. D., & Chow, C. C. (2010). Systematic fluctuation expansion for neural network activity equations. Neural computation, 22(2), 377-426.

[6] Janssen, H. K. (1976). On a Lagrangean for classical field dynamics and renormalization group calculations of dynamical critical properties. *Zeitschrift für Physik B Condensed Matter*, 23(4), 377-380.

[7] Ganguli, S., Huh, D., & Sompolinsky, H. (2008). Memory traces in dynamical systems. Proceedings of the National Academy of Sciences, 105(48), 18970-18975.