BIO-465 – Project
A network model of cortical surround suppression

Etienne Objois, Niels Vadot

EPFL, Switzerland

Abstract—We explore neuronal population models and use them to model cortical surround suppression. Cortical surround suppression is a phenomenon where the presence of nearby neighbours causes a neuron to have a lower activity level than if it were alone. We also expore the orientation tuning phenomenon, where the surround suppression is more or less strong, depending on the overlap between the external stimulus and the receptive fields. The source code is available at [1].

I. RATE MODELS OF NEURONAL POPULATIONS

A neuronal population is composed of multiple neurons, which can each be individually modeled. In the simple model from EQ. (1), a neuron is described by the voltage across its membrane $h(t)$, resistivity R to input currents $I(t) = I_{\text{ext}}(t) + I_{\text{network}}(t)$, a relaxation time constant τ, and an activity level $A(t)$ defined in terms of a filter $\alpha(s)$ and gain function $F(h)$ in the equation $A(t) = \int_0^\infty \alpha(s) F(h(t-s)) \, ds$.

$$\tau \dot{h}(t) = -h(t) + RI(t) \tag{1}$$

In a population, every neuron is modeled individually according to EQ. (1), and attributed an index k. We furthermore model the network currents linearly as in EQ. (2), by introducing a weight parameter W_{kn}, which describes the influence of neuron n on neuron k.

$$I_{\text{network},k}(t) = \sum_n W_{kn} A_n(t) \tag{2}$$

A further simplification can be made by letting $\alpha(s)$ be a very sharp filter, approximatively a Dirac delta. Then $A(t) \approx F(h(t))$.

We can rewrite the entire system as one equation in matrix form, where $\boldsymbol{h} = (h_1, \cdots, h_n)$, $\boldsymbol{R} = (R_1, \cdots, R_n)$, etc., and $\odot$ is the elementwise product :

$$\boldsymbol{\tau} \odot \dot{\boldsymbol{h}} = -\boldsymbol{h} + \boldsymbol{R} \odot (\mathsf{W}\boldsymbol{F}(\boldsymbol{h})) + \boldsymbol{R} \odot \boldsymbol{I}_{\text{ext}} \tag{3}$$

EQ. (3) models a population of neurons, but can also be interpreted to model an ensemble of populations. This is assumed from now on, and a more detailed proof and renormalization is given in subsubsection III-C1.

II. FIXED VS. RECURRENT INHIBITION

A. Excitatory population with self-coupling

We model a single population according to EQ. (3), and assume constant input current I_{ext}. The gain function is defined as $F(h) = \sigma(h) = \frac{1}{2}(1 + \tanh(h))$.

The self-coupling appears in the diagonal terms of W, and in this case there is only self-coupling, since W reduces to a scalar.
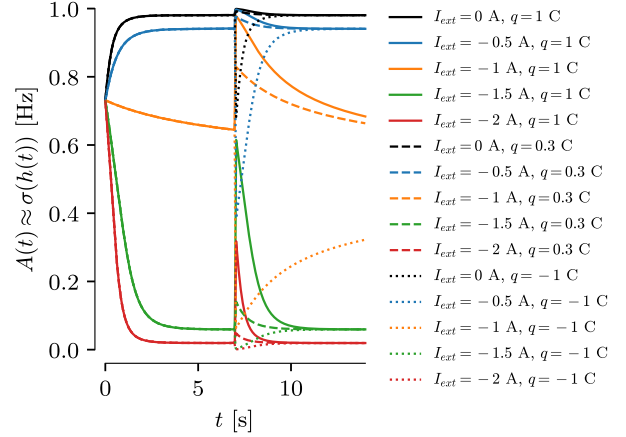

FIGURE 1: Single population with constant input and a delta spike of total charge q at $t = 7$. Parameters $W = 2$, $\tau = 0.6, R = 1, h(0) = 0.5$.

We vary the constant input current in FIG. 1 and study $t < 7$ (before the delta spike). The simulation is started at $h(0) \neq 0$ in order to break symmetry of the gain function. For all currents, the activity levels out over time. We can observe three regimes. (1) At $I_{\text{ext}} \lesssim -1$ A, there is too much negative current (i.e. too many negative charges arrive too fast, which depolarizes the population) and the self-coupling is not enough to maintain activity levels, which decay to nearly zero. (2) At $I_{\text{ext}} \gtrsim -1$ A, the opposite happens, and self-coupling manages to drive the activity levels to nearly one. (3) At $I_{\text{ext}} \approx -1$ A, the external current and self-coupling nearly cancel out, leaving the activity somewhere between zero and one.

The introduction of a delta spike at $t = 7$ makes the activity levels jump, which then recover and tend back to the same equilibrium as before the spike. For a same total charge, unsaturated neurons ($0 \lesssim A \lesssim 1$) jump a larger amplitude and take more time to recover than saturated neurons ($A \approx 0$ or $A \approx 1$). The jump is asymmetric with respect to charge sign : if $A > \frac{1}{2}$, then for $q < 0$ the perturbation is greater than for $q > 0$ (and vice-versa). For total charges of the same sign, the size of the jump is largest for the largest absolute charge.

In order to get insight into the behavior of the coupling strength, FIG. 2 shows a 1D phase plot of the differential equation. Fixed points solutions to $\dot{h} = 0$. I_{ext} controls the vertical offset of the curves but more interestingly W controls the appearance of two local extrema, and there-
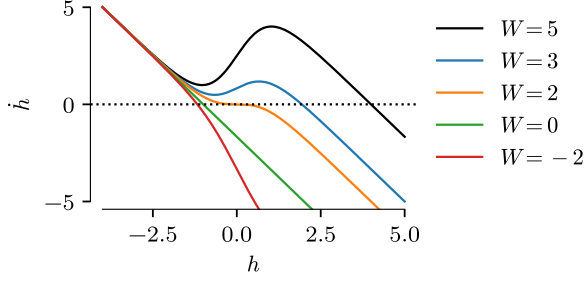
FIGURE 2: 1D phase plot of a single population. Parameters $\tau = 0.6, R = 1, I_{\text{ext}} = -1$. Critical values are $W_{\text{crit}} = 2$ and $I_{\text{crit}} = -1$.
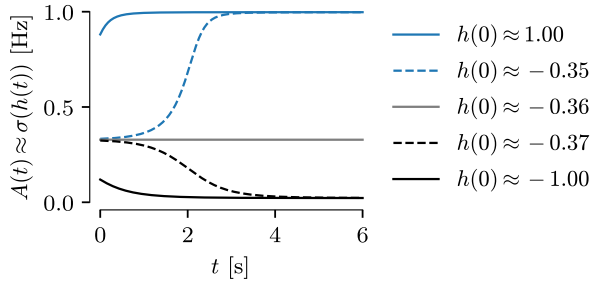

FIGURE 3: Simulation of a single population with three fixed points. Parameters $W = 5, R = 1, I_{\text{ext}} = -2$.

fore the possibility of two other solutions. This happens when the curvature changes sign, and critically when the derivative is null. Solving $\partial_h^2 h = 0$ and $\partial_h \dot{h} = 0$ for h and W, we find $h_{\text{crit}} = 0$ and $W_{\text{crit}} = 2/R$. When $W < W_{\text{crit}}$, only one stable solution exists. When $W > W_{\text{crit}}$, three solutions can exist (two stable, and one unstable at $h = 0$) depending on the value of I_{ext}.

If we furthermore impose $\dot{h} = 0$ at the critical value, then we get $I_{\text{crit}} = -1/R$. When $I < I_{\text{crit}}$, the three solutions can exist (provided the current is not too negative), whereas when $I > I_{\text{crit}}$ only one solution exists.

Intuitively, the current has to be depolarizing enough but the self-coupling strong enough to compensate, and create a second solution. FIG. 3 demonstrates a set of parameters for which there are two stable solutions, and one unstable solution.

B. Linear rate networks

We specialize EQ. (3) in the case of a linear rate network, that is $F(h) = \text{Id}(h) = h$. Assuming external current to be independant of time, this equation is of the form $\dot{h} = \Lambda h + b$, where $\Lambda = (1/\tau) \odot (R \odot W - \text{Id})$ and $b = (1/\tau) \odot (R \odot I_{\text{ext}})$. This can be solved analytically :

$$\dot{h} = \Lambda h + b, \text{ changing variables } y = \Lambda h + b$$
$$\implies \dot{y} = \Lambda y \implies \dot{y} = \exp(\Lambda t) y_0$$

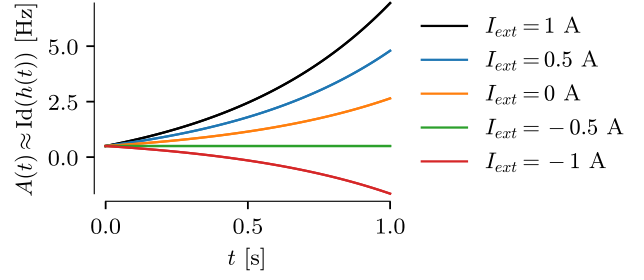We can change back variables, denoting the Moore-

FIGURE 4: Simulation of a linear single population. Parameters $W = 2, R = 1, \tau = 0.6, h(0) = 0.5$.

Penrose inverse Λ^+, and get the general solution EQ. (4).

$$h(t) = \begin{cases} \Lambda^+ (\exp(\Lambda t)(\Lambda h_0 + b) - b) & \text{if } \Lambda \neq 0 \\ bt & \text{if } \Lambda = 0 \end{cases} \quad (4)$$

The stability of the solution (in the case $\Lambda \neq 0$) becomes clear by rewriting $\Lambda h(t) = \exp(\Lambda t) v - b$, where $v = \Lambda h_0 + b$. Recall that the eigenspace of A is invariant under matrix exponentiation.

Components of v that lie in an eigenspace associated to $\lambda < 0$ decay exponentially, diverge if $\lambda > 0$, and behave linearly if $\lambda = 0$. This is demonstrated in FIG. 4, where depending on I_{ext}, the activity either stays constant (linear behavior) or diverges to $\pm\infty$ exponentially.

In a linear rate network, the presence of a inhibitory current is not sufficient to ensure stabilization of a single population, whereas is was sufficient for a nonlinear rate network (in fact, in the case $F(h) = \sigma(h)$ the activity is bounded in $[0, 1]$ so the network is always stable.).

C. Inhibition-stabilized network (ISN)

We add another population to the network. The weights of the outbound currents (second column of W) of the new population are negative, such that is can be seen as the inhibitory (I) population. The original population has positive outbound weights (first column of W), becoming the excitatory (E) population.

To study stability of the network, we compute the eigenvalues of Λ from subsection II-B, and find a condition such that they are all negative.

$$\Lambda = \begin{pmatrix} \frac{W_{00}-1}{\tau_0} & \frac{W_{01}}{\tau_0} \\ \frac{W_{10}}{\tau_1} & \frac{W_{11}-1}{\tau_1} \end{pmatrix}.$$

We find that both eigenvalues change sign at

$$W_{00}^\star = \frac{W_{01}W_{10} + W_{11} - 1}{W_{11} - 1}.$$

Notice that since $W_{11} < 0$, we have $W_{11} \neq 1$ and a stability threshold exists for any W_{10}, W_{01}, W_{11}.

In this case, we can substitute $W_{10} = 4, W_{01} = -4, W_{11} = -7$ and find the system is stable when $W_{00} < W_{00}^\star = 3$. It diverges for $W_{00} > W_{00}^\star$ and evolves linearly when $W_{00} = W_{00}^\star$ (both eigenvalues are null with the same condition).
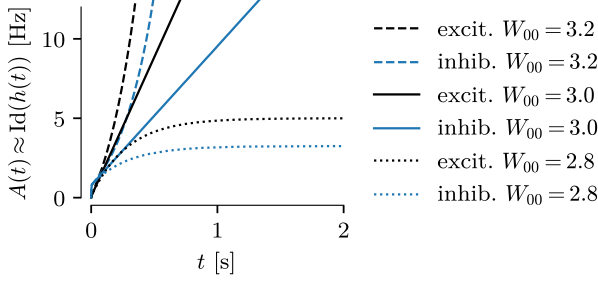
FIGURE 5: Simulation of a linear ISN. Parameters $W_{10} = 4, W_{01} = -4, W_{11} = -7, R = 1, \boldsymbol{\tau} = (0.06, 0.012), \boldsymbol{h}(0) = (0, 0), \boldsymbol{I}_{\text{ext}} = (4.0, 6.0)$.
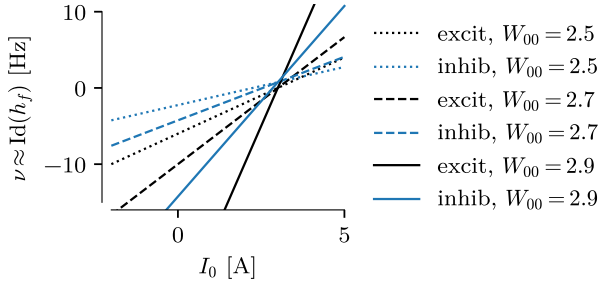

FIGURE 6: Simulation of a linear ISN. Parameters $W_{10} = 4, W_{01} = -4, W_{11} = -7, R = 1, \boldsymbol{\tau} = (0.06, 0.012), \boldsymbol{I}_{\text{ext}} = (4.0, 6.0)$.

FIGURE 7: Simulation of surround suppression using an ISN. Parameters $W_{10} = 4, W_{01} = -4, W_{11} = -7, R = 1, \boldsymbol{\tau} = (0.06, 0.012), \boldsymbol{I}_{\text{ext}} = (4.0, 1.6)$ if $t < 1.5$, else $(4.0, 3.8)$.

The numerical simulations (setting $W_{10} = 4$, $I_0 = 4.0$, $I_1 = 6.0$) shown in FIG. 5 confirm the analytical derivations. In the converging case $W_{00} = 2.8 < W_{00}^\star = 3.0$, the excitatory activity converges to $A = 5$ Hz, and the inhibitory activity converges to $A = 3.25$ Hz.

We denote $\boldsymbol{h_f} = \Lambda^+ \boldsymbol{b}$ (see subsection A) the value to which $\boldsymbol{h}$ converges. It depends on W, $\boldsymbol{I}_{\text{ext}}$, and more surprisingly $\boldsymbol{\tau}$. In the case of one population, there was no "competition" and the system can converge at any speed. In the case of two populations, the competition means that now the rate of convergence matters. $\boldsymbol{h_f}$ is of the form, where cst terms are functions of W and $\boldsymbol{\tau}$:

$$\boldsymbol{h_f} = \begin{pmatrix} \text{cst} \cdot I_0 + \text{cst} \cdot I_1 \\ \text{cst} \cdot I_0 + \text{cst} \cdot I_1 \end{pmatrix}.$$

The f-I curve shown in FIG. 6 is linear, as expected from the form of $\boldsymbol{h_f}$. Stronger excitatory self-couplings of the E population are associated to a steeper line. Compared to the linear excitatory-only network which (nearly) always diverges exponentially, the addition of an inhibitory population allows for the existence of stable regimes. This shows that constant inhibitory current is not enough to stabilize the system, it needs to scale with the activity of the excitatory population, as it does here.

This linear analysis extends to the sigmoidal gain function $F(h) = \sigma(h)$ in the case of a small potential h (after some rescaling), since $\sigma(h) = \frac{1}{2} + \frac{h}{2} + \mathcal{O}(h^3) \approx \text{Id}(\frac{h+1}{2})$.
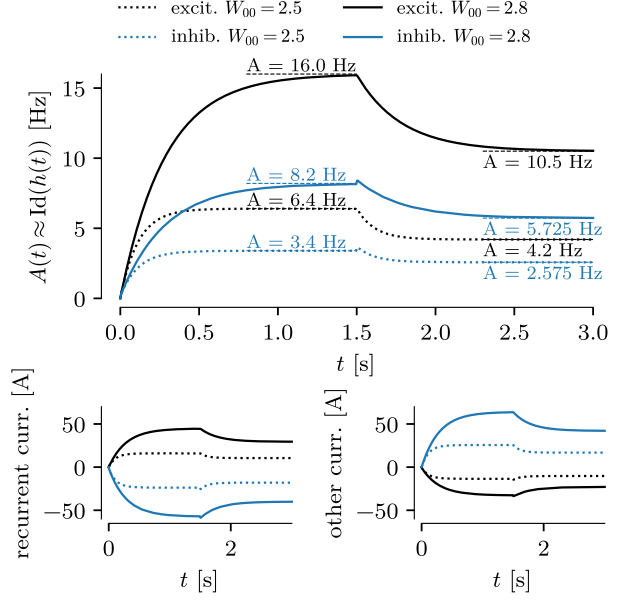
III. MODELING SURROUND SUPPRESSION

A. Network mechanisms of surround suppression

During surround stimulus, neighbouring neurons add negative current contributions to the observed neuron. We can therefore model surround suppression by adding an inhibitory population which adds a negative current to excitatory population. Therefore by increasing the current to the inhibitory population, we are modeling exciting more of the surrounding population.

We set W_{00} such that the system is always stable, and converges (exponentially) to the stationary solution independantly of the initial condition, as derived in subsection II-C. FIG. 7 simulates surround suppression using an ISN, showing activity of the neurons as well as input currents.

The recurrent currents scale with the activity of the population. As the activity the excitatory population grows, so does the input current to the inhibitory population, which in term increases the negative current contribution to the excitatory population.

Lowering W_{00} results in convergence to a lower activity level for both populations. This is expected as from the analytical derivation, as asymptotically, and as shown in FIG. 8 for the simulation parameters of FIG. 7, we have

$$\boldsymbol{h_f} = \begin{pmatrix} \frac{2.1}{3 - W_{00}} \\ \frac{2.475 - 0.475 W_{00}}{3 - W_{00}} \end{pmatrix} \sim \begin{pmatrix} \frac{1}{W_{00}^\star - W_{00}} \\ \frac{1 - W_{00}}{W_{00}^\star - W_{00}} \end{pmatrix}.$$

B. Orientation tuning of surround suppression

Orientation tuning is a consequence of the difference in neuronal receptive fields. For example, some have a recep-
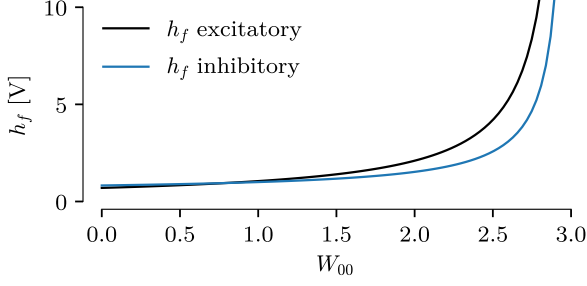
FIGURE 8: Dependency of h_f on excitatory self-coupling. Parameters $W_{10} = 4, W_{01} = -4, W_{11} = -7, R = 1, \boldsymbol{\tau} = (0.06, 0.012), \boldsymbol{I}_{\text{ext}} = (4.0, 3.6)$.

tive field which matches vertical stripes, thus responding strongly to them. If another, different excitation is applied (i.e. horizontal stripes) which doesn't match the receptive field, then the response isn't as strong.

In the brain, neurons are often arranged such that neurons with similar receptive fields are spatially close.

In our excitatory-inhibitory population model, we can model sensitivity to different stimuli as an increase (or decrease) of the input current to the neuron. Surround suppression for cross-oriented stimuli (of the surround population) is therefore modeled as an lower input current I_1 than in for iso-oriented stimuli. This has already been simulated in subsection III-A : the first half of the simulation has a lower input current (cross-oriented simuli) to the inhibitory population, and the second half has a higher current (iso-oriented simuli), resulting in an activity of the excitatory neuron higher in the first than in the latter.

C. Surround suppression in networks with bio-plausible connectivity

1) Population model renormalization: We denote $[K]$ the ensemble of neurons belonging to the population (multi-)index K, and $|K|$ the number of neurons in population K

The voltage, activity and external currents of the "mean field" population (indexed by K) composed of neurons $k \in [K]$ can be expressed as

$$H_K(t) = \frac{1}{|K|} \sum_{k \in [K]} h_k(t)$$

$$A_K(t) = \frac{1}{|K|} \sum_{k \in [K]} a_k(t)$$

$$I_{\text{ext},K} = \frac{1}{|K|} \sum_{k \in [K]} i_{\text{ext},k}(t)$$

and modeled by

$$T_K \dot{H}_K(t) = -H_K(t) \\ + R_K \sum_N W_{KN} A_N(t) + R_K I_{\text{ext},K}(t).$$

In the mean-field approximation, we model all neuron k in population K as being the same. Denoting $\mathcal{P}(k) = K$

as the population multi-index K corresponding to neuron k, we substitute to find the equation for the "mean neuron" (all terms in $\sum_{k \in [K]}$ are identical) :

$$T_{K(k)} \dot{h}_k(t) = -h_k(t) \\ + R_{K(k)} \sum_N W_{\mathcal{P}(k),N} \frac{1}{|N|} \sum_{n \in [N]} a_n(t) \\ + R_K i_{\text{ext},k}(t)$$

Noting $\sum_n = \sum_N \sum_{n \in [N]}$, we simplify

$$T_{\mathcal{P}(k)} \dot{h}_k(t) = -h_k(t) \\ + R_{\mathcal{P}(k)} \sum_n \frac{1}{|N|} W_{\mathcal{P}(k),\mathcal{P}(n)} a_n(t) \\ + R_K i_{\text{ext},k}(t).$$

Imposing the following constraints completes the renormalization :

$$\tau_k = T_{\mathcal{P}(k)}, \quad r_k = R_{\mathcal{P}(k)}, \quad w_{kn} = \frac{1}{|N|} W_{\mathcal{P}(k),\mathcal{P}(n)}$$

This result matches our intuition, as rescaling should leave the inputs to a neuron/population unchanged.

2) Homogeneous renormalization: We verify the results from subsubsection III-C1 by simulating the full population $h_k(t)$ and the corresponding renormalized population $H_{\mathcal{P}(k)}(t)$. FIG. 9 shows the difference between the potential and its renormalized counterpart in the single population case. The difference is very small, but nonzero due to machine precision. The same verification is performed by repeating the experiment FIG. 7, and results are plotted in FIG. 10. We note that in this experiment, the system is much more sensitive to large integration timesteps : using the same timestep as for the single population renormalization, the difference creeped up to 10^{-12} V at the start of the simulation, before quickly dropping down to around 10^{-15} V.

Since the weight matrix is now a homogeneous block matrix, every neuron in a population is expected to behave exactly in the same way. This has also been asserted numerically.

3) Heterogeneous renormalization: Instead of renormalizing homogeneously by transforming W into a block matrix, we can make each block random, such that the expectation $\mathbb{E}(w_{kn}) = \frac{1}{|N|} W_{\mathcal{P}(k),\mathcal{P}(n)} =: \mu$.

In this case, we sample each block entry w_{kn} following a lognormal distribution (which experimentally models the distribution of synaptic weights) with mean μ and standard deviation $\alpha\mu$ (in this case, we use $\alpha = 2.75$). Furthermore to capture the sparsity of connections in biological networks, we set each weight to zero with probability $p = 0.95$. In order to conserve the correct expectation, we map $\mu \mapsto \mu/p$.

We repeat the same experiment as in FIG. 1 with renormalized heterogeneous weights, represented in FIG. 11. We observe many weights are nearly zero, and only a few neurons have large weights. Simulation results are presented in FIG. 12, and show that the simulated neuronal population follows the mean-field results on average,
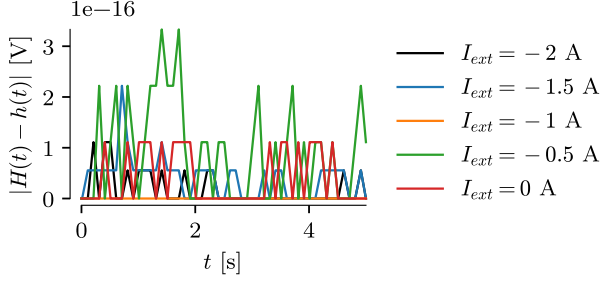
FIGURE 9: Renormalization check, repeating the same experiment as FIG. 1.
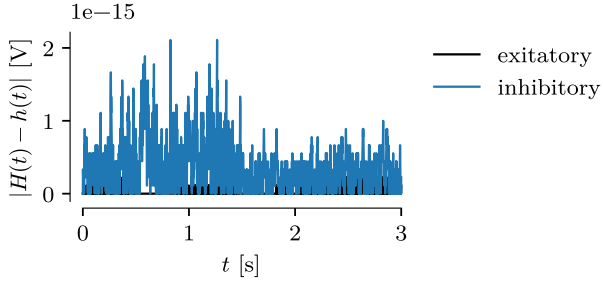

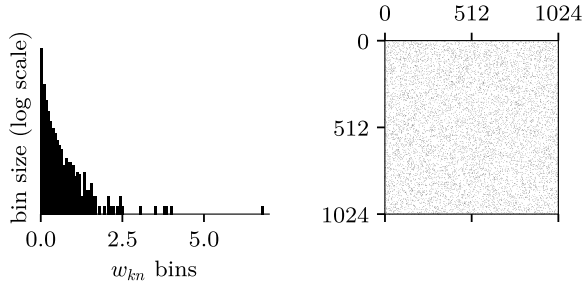FIGURE 10: Renormalization check, repeating the same experiment as FIG. 7.

FIGURE 11: Distribution of nonzero entries and visualization of the heterogeneous weights, for a single population renormalization.

despite a relatively broad standard deviation. The case of $I_{\text{ext}} = -1$ A has the most difference between models. This is a consequence of the gradient of the sigmoid gain function peaking at $h = 0$, which is the regime in $I_{\text{ext}} = -1$ is located. The further away h is from zero, the tighter the standard deviation will appear.
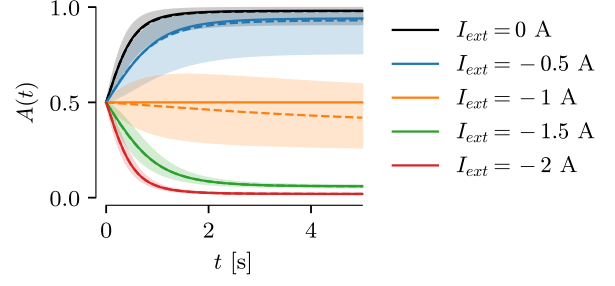

FIGURE 12: Renormalized heterogeneous weights, with same parameters as FIG. 1. Solid lines are results in the mean-field approximation, dotted lines represent the mean of the simulated neuron population, and shaded areas the standard deviation.

REFERENCES

[1] E. Objois and N. Vadot, "BIO-465 project source code," May 2022. [Online]. Available: https://github.com/ninivert/BIO465-cortical-surround-suppression

APPENDIX

A. *Full expression of h_f in the linear ISN*

$$h_{f0} = -\frac{I_0(W_{11}\tau_0 - \tau_0)}{\tau_0(W_{00}W_{11} - W_{00} - W_{01}W_{10} - W_{11} + 1)}$$
$$+ \frac{I_1 W_{01}}{W_{00}W_{11} - W_{00} - W_{01}W_{10} - W_{11} + 1}$$
$$h_{f1} = \frac{I_0 W_{10}}{W_{00}W_{11} - W_{00} - W_{01}W_{10} - W_{11} + 1}$$
$$- \frac{I_1(W_{00}\tau_1 - \tau_1)}{\tau_1(W_{00}W_{11} - W_{00} - W_{01}W_{10} - W_{11} + 1)}$$