BIOLOGICAL DETAIL

1. Missing detail

- Dale's principle.
  $\approx$ *Neurons* are either excitatory or inhibitory.
- No bias current.
  Unclear what the biological correlate of J^{bias} is.
- Conductance-based synapses.
  Two consecutive spikes do not have the same effect.
- Neural dynamics may dominate synaptic dynamics.

! Detail added to the NEF should connect biology to behaviour

2. Dale's Principle/Eliminating Bias Currents

- Not all weights w_{ij} for a pre-neuron j are either positive or negative
- We know $\mathbf{W} = \mathbf{ED}$.
- Problem: Hard to figure out signs in $\mathbf{W}$
- Directly solve for *full* $\mathbf{W}$ with constraints

! No decoding/re-encoding

2.1 Solving for weights in current-space

- **Idea:** Solve for a *current decoder* $\vec{w}_i$ for each post-neuron i (rows in W)
- Current-based synapse model:

$$\hat{J}_i = \langle \vec{a}(\vec{x}), \vec{w}_i \rangle$$

- NEF encoding equation:

$$J_i = \alpha_i \langle \vec{x}, \vec{e}_i \rangle + J_i^{\text{bias}}$$

- Want to find $\vec{w}_i$ that minimizes $(\hat{J}_i - J_i)^2$

$$\vec{w}_i = \arg\min_{\vec{w}_i} \sum_{k=1}^{N} \left(\big\langle \vec{a}(\vec{x}_k), \vec{w}_i \big\rangle - \alpha_i \big\langle f(x_k), \vec{e}_i \big\rangle + J_i^{\text{bias}} \right)^2 .$$

- In matrix form:

$$\vec{w}_i = \arg\min_{\vec{w}_i} \left\| \mathbf{A}\vec{w}_i - \mathbf{J}^{\text{tar}} \right\|^2$$

$\Rightarrow$ Least-squares problem

$\Rightarrow$ Directly decodes post-synaptic current; $\alpha_i, \vec{e}_i, J_i^{\text{bias}}$ are now merely *normative*

2.2 Accounting for Dale's Principle

- Split pre-neurons into excitatory (+) and inhibitory (−) set
- Two sets of connection weights $\vec{w}^+$, $\vec{w}^-$, pre-activities $\mathbf{A}^+$, $\mathbf{A}^-$
- New optimization problem:

$$\vec{w}_i^+, \vec{w}_i^- = \arg\min_{\vec{w}_i^+, \vec{w}_i^-} \left\| \mathbf{A}^+ \vec{w}_i^+ - \mathbf{A}^- \vec{w}_i^- - \mathbf{J}^{\mathrm{tar}} \right\|^2 , \quad \text{with respect to} \quad \vec{w}_i^+, \vec{w}_i^- \geq 0 \,.$$

- Rearrange:

$$\vec{w}_i' = \left(\vec{w}_i^-, \vec{w}_i^+\right) = \arg\min_{\vec{w}_i} \left\| \left(\mathbf{A}^+, -\mathbf{A}^-\right) \left(\vec{w}_i^-, \vec{w}_i^+\right)^T - \mathbf{J}^{\mathrm{tar}} \right\|^2 = \arg\min_{\vec{w}_i} \left\| \mathbf{A}' \left(\vec{w}_i'\right)^T - \mathbf{J}^{\mathrm{tar}} \right\|^2 ,$$

w.r.t. $\vec{w}_i' \geq 0$.

- Non-negative least square (NNLS) problem

3. Conductance-based synapses

- Second of two consecutive spikes induces a smaller current ⇒ Synapses are not linear

! Central assumption of the NEF!

- **Question:** Can we still get things to work *and* explain behaviour/computation in nervous systems?
- **Idea:** Multi-dimensional neuron model

$$\begin{aligned} \mathcal{G}[g_{\mathrm{E}}, g_{\mathrm{I}}] &\overset{\text{decompose}}{=} G[H(g_{\mathrm{E}}, g_{\mathrm{I}})] \\ H^{\mathrm{cond}}(g_{\mathrm{E}}, g_{\mathrm{I}}) &= \frac{b_1 g_{\mathrm{E}} + b_2 g_{\mathrm{I}})}{a_0 + a_1 g_{\mathrm{E}} + a_2 g_{\mathrm{I}}} \\ \hat{J}_i &= H[\langle \vec{a}(\vec{x}), \vec{w}_i^{\mathrm{E}} \rangle, \langle \vec{a}(\vec{x}), \vec{w}_i^{\mathrm{I}} \rangle] \end{aligned}$$

- Use above approach to solve for H