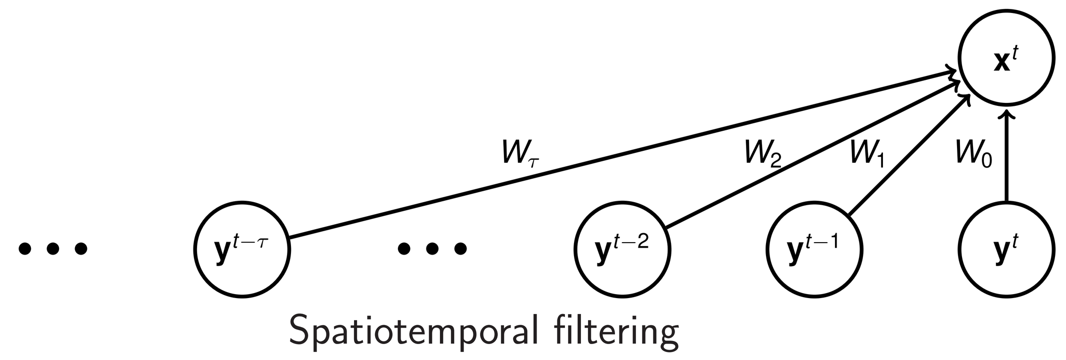
Learning visual motion in recurrent neural networks

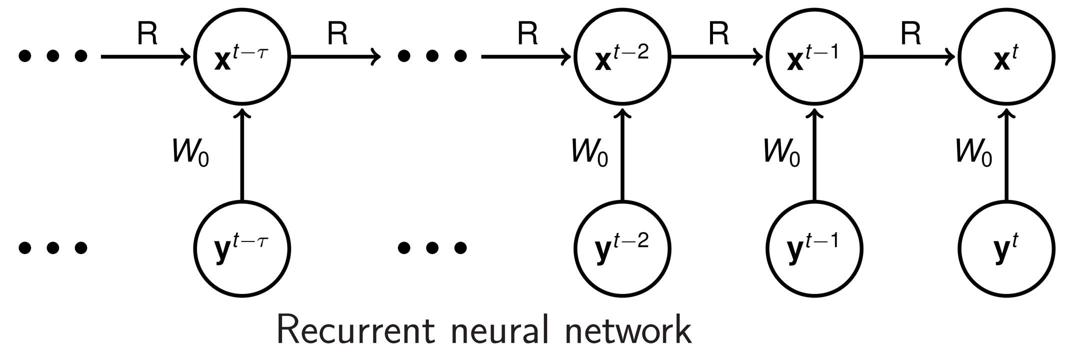
Marius Pachitariu, Maneesh Sahani
Gatsby Computational Neuroscience Unit, UCL

Space for tablet

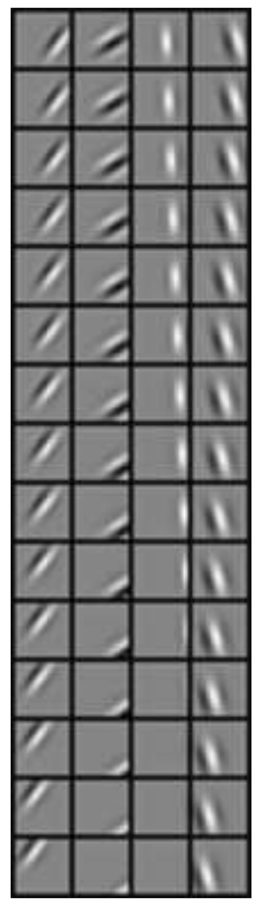
RNN can reparametrize spatiotemporal features



Spatiotemporal filtering



Recurrent neural network



$$\mathbf{x}^t = \sum_{\tau=0}^{\infty} \mathbf{W}_\tau \ \mathbf{y}^{t-\tau}$$

$$\mathbf{W}_\tau = (\mathbf{R})^\tau \ \mathbf{W}_0$$

$$\mathbf{x}^t = \mathbf{W}_0 \ \mathbf{y}^t + \mathbf{R} \sum_{\tau=0}^{\infty} (\mathbf{R})^\tau \ \mathbf{W}_0 \ \mathbf{y}^{t-1-\tau}$$

$$\mathbf{x}^t = \mathbf{W}_0 \ \mathbf{y}^t + \mathbf{R} \ \mathbf{x}^{t-1}$$

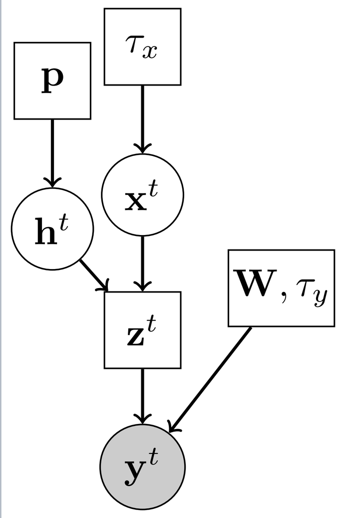
Can solve for $\mathbf{R}$ by linear regression

$$\mathbf{W}_1 \approx \mathbf{R} \ \mathbf{W}_0$$
$$\mathbf{W}_2 \approx \mathbf{R} \ \mathbf{W}_1$$
$$\dots$$
$$+ \text{L1-regularization}$$

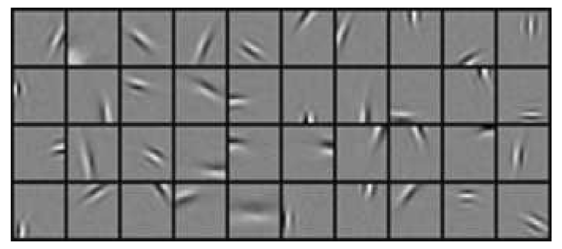
Advantages of RNN

- Does not require copies of the past
  - → less memory usage
  - → the brain has short timescales + bottleneck in LGN
  - → no evidence for true delay lines in cortex
- Fewer parameters
  - → important for learning and generalization
- Reduced computational complexity
  - → good for high bandwidth data
- Can integrate over long time periods
  - → natural visual motion can be slow and noisy

Binary-gated Gaussian sparse coding

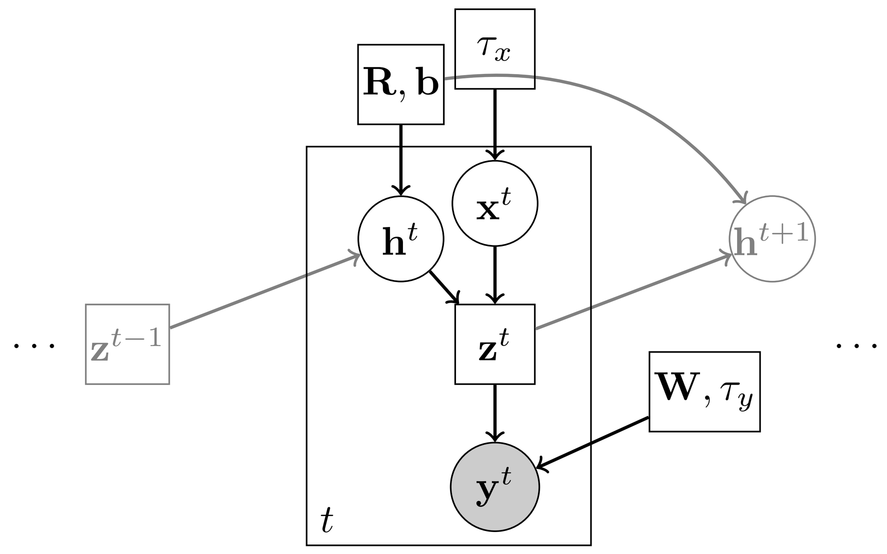


$$\mathbf{h}_k^t = \text{Bernoulli}(\mathbf{p}_k)$$
$$\mathbf{x}^t = \mathcal{N}\left(0, \tau_x^2 \cdot \mathbf{I}\right)$$
$$\mathbf{z}^t = \mathbf{h}^t \circ \mathbf{x}^t$$
$$\mathbf{y}^t = \mathcal{N}\left(\mathbf{W} \cdot \mathbf{z}^t, \tau_y^2\right)$$



Binary-gated Gaussian recurrent neural network

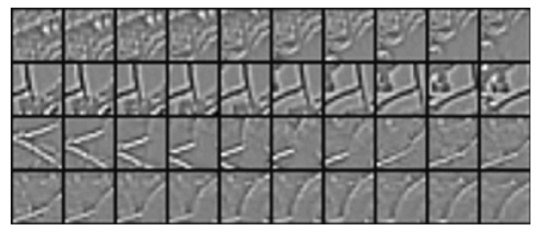
$$\mathbf{P}\left(\mathbf{x}^{t+1}, \mathbf{h}^{t+1} | \mathbf{x}^t, \mathbf{h}^t\right) = \mathbf{P}\left(\mathbf{x}^{t+1}\right) \mathbf{P}\left(\mathbf{h}^{t+1} | \mathbf{h}^t \circ \mathbf{x}^t\right)$$
$$\mathbf{x}^{t+1} = \mathcal{N}\left(\mathbf{0}, \tau_x^2 \mathbf{I}\right)$$
$$\mathbf{P}\left(\mathbf{h}^{t+1} | \mathbf{h}^t \circ \mathbf{x}^t\right) = \sigma\left(\mathbf{R} \cdot \left(\mathbf{h}^t \circ \mathbf{x}^t\right) + \mathbf{b}\right)$$
$$\sigma(x) = 1/(1 + \exp(-x))$$



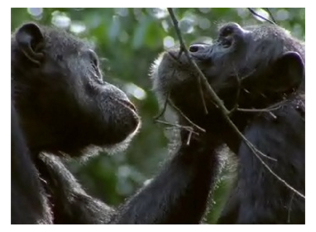
$$\mathcal{L}_{\text{bgG-RNN}} = \sum_t \mathcal{L}_{\text{bgG-RNN}}^t$$

$$\mathcal{L}_{\text{bgG-RNN}}^t = \text{const} - \|\mathbf{y}^t - \mathbf{W}(\mathbf{x}^t \circ \mathbf{h}^t)\|^2/2\tau_y^2 - \|\mathbf{x}^t\|^2/2\tau_x^2 + $$
$$+ \sum_{j=1}^{N} h_j^t \log \sigma\left(\mathbf{R}\left(\mathbf{h}^{t-1} \circ \mathbf{x}^{t-1}\right) + \mathbf{b}\right)_j$$
$$+ \sum_{j=1}^{N} (1 - h_j^t) \log\left(1 - \sigma\left(\mathbf{R}\left(\mathbf{h}^{t-1} \circ \mathbf{x}^{t-1}\right) + \mathbf{b}\right)_j\right)$$

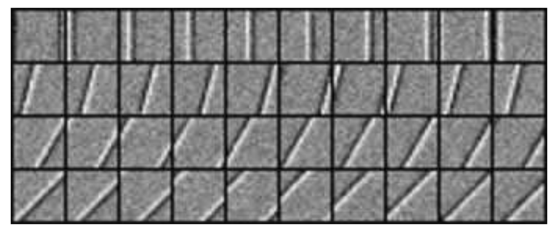
Natural movies and artificial stimuli



Training data (ZCA whitened)

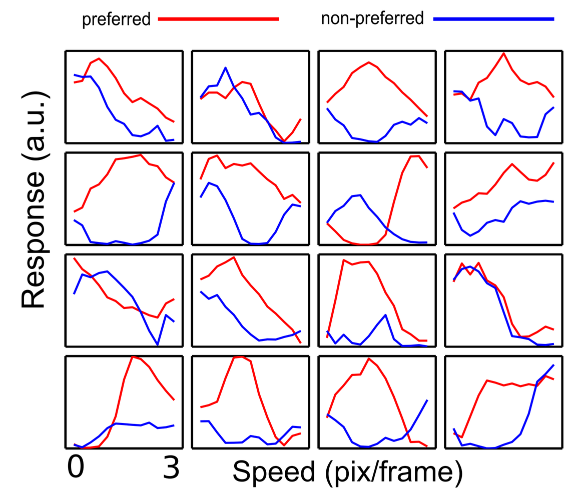


Full single frame of training data

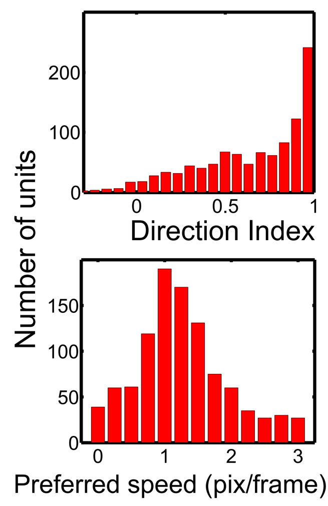


Test data (whitened)

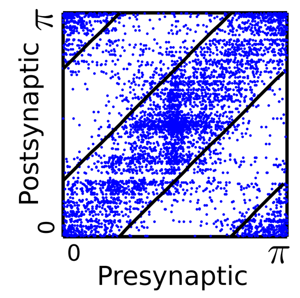
Results



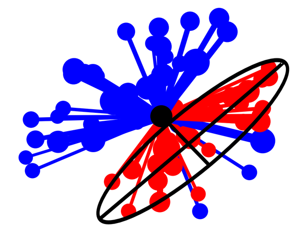
Rectified neural responses ($h_j \circ x_j$) to drifting square gratings. Compare with [1].



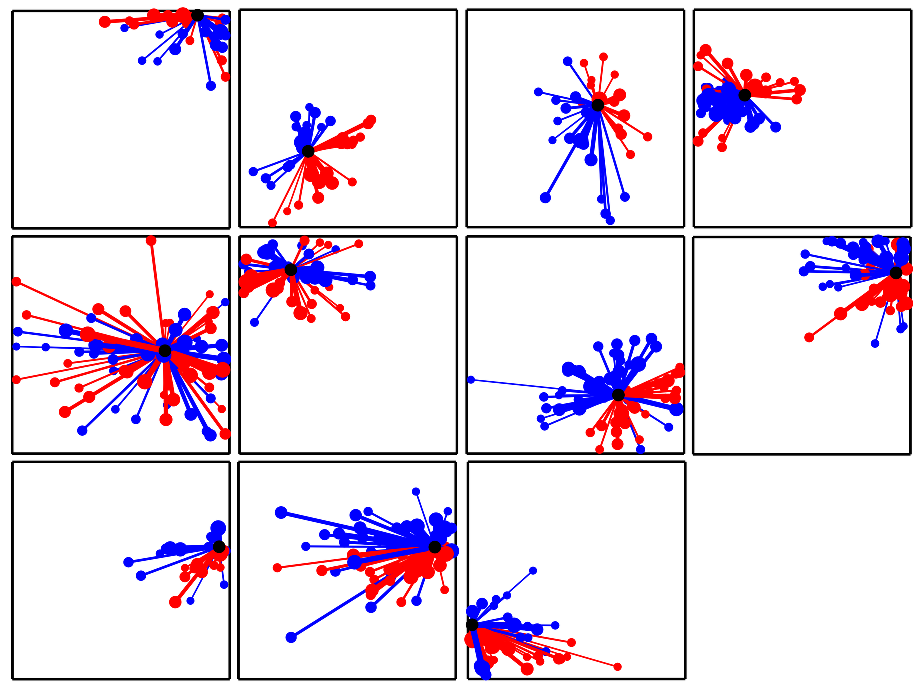
Direction indices (DI, top) and preferred speeds (bottom). $\text{DI} = 1 - r_{\text{non-pref}}/r_{\text{pref}}$



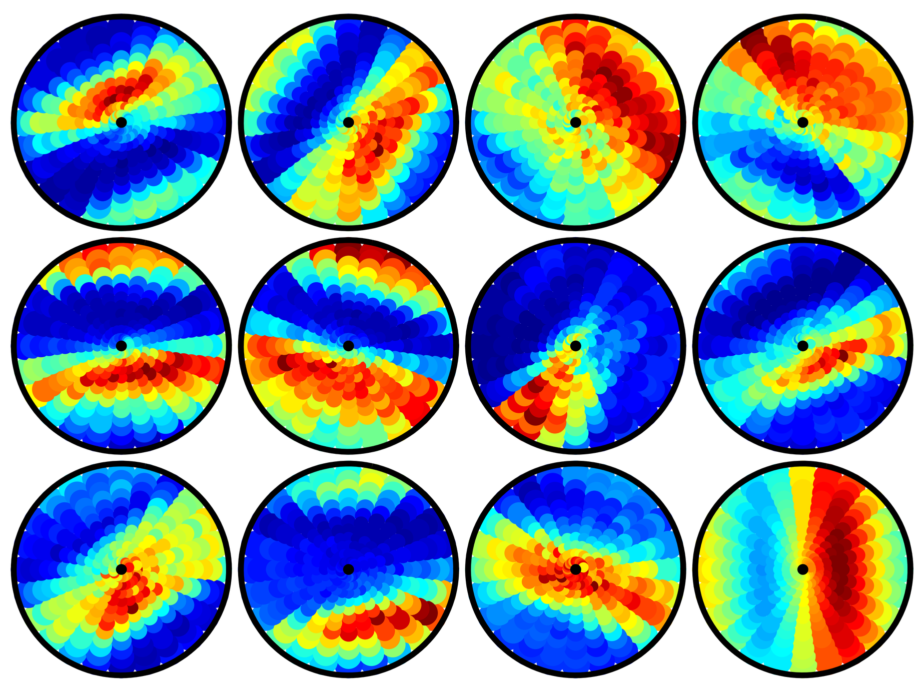
Strongest connections between units with similar orientation tuning



Excitation/inhibition in the preferred/non-preferred direction



Connections of 15 randomly selected units. Dots are centers of Gabor fits to spatial receptive field.

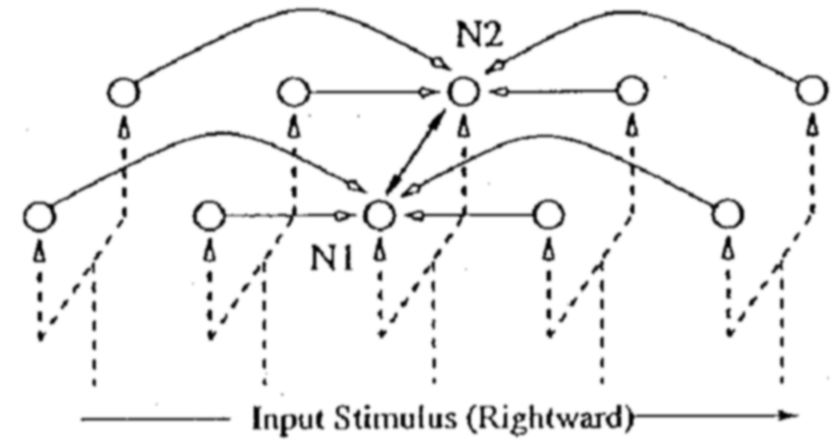


Polar plots of responses to small drifting Gabors. Same units as previous plot. Last polar plot is the average across the population, aligned by preferred direction.

Online inference and learning

- learning: optimize the full joint likelihood, EM style
- inference: set values of $\mathbf{x}^t, \mathbf{h}^t$ that (ideally) maximize the likelihood.
- greedy filtering: assuming we have already set $\hat{\mathbf{x}}^t, \hat{\mathbf{h}}^t$ for $t = 1$ to T, we propose to obtain $\hat{\mathbf{x}}^{T+1}, \hat{\mathbf{h}}^{T+1}$ exclusively from $\hat{\mathbf{x}}^T, \hat{\mathbf{h}}^T$ and the new image frame $\mathbf{y}^{T+1}$.
- inference at each time point becomes a sparse coding problem. For tractability, we use another greedy algorithm for sparse coding inference: matching pursuit. Starting from $\mathbf{h}^t = \mathbf{x}^t = \mathbf{0}$, sequentially choose the next $\mathbf{h}_k^t$ to turn on with its corresponding $\mathbf{x}_k^t$ such that the greatest increase in likelihood is achieved.

Neural sequence learning via STDP



Sequence learning forms the basis of an earlier simple toy but biophysically realistic model [2] based on STDP at the lateral synapses of a recurrently connected network of neurons. The gradient of the likelihood in bgG-RNN results in a similar STDP-like rule.

$$\frac{\partial \mathcal{L}_{\text{bgG-RNN}}^t}{\partial R_{jk}} = \left(h_k^{t-1} x_k^{t-1}\right) \cdot \left(h_j^t - \sigma\left(\mathbf{R}\left(\mathbf{h}^{t-1} \circ \mathbf{x}^{t-1}\right) + \mathbf{b}\right)_j\right)$$

Conclusions

- Recurrent neural networks can analyze visual motion in an online fashion without delayed inputs.
- Formulating a generative model allows learning the recurrent connections via an STDP rule.
- As a model of V1, the RNN makes testable predictions about the lateral connectivity of neurons. Responses to stimuli may however be similar to those of spatiotemporal filters.

References

1. GA Orban, H Kennedy and J Bullier. Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey: influence of eccentricity. *Journal of Neurophysiology*, 56(2):462-480, 1986.
2. RPN Rao and TJ Sejnowski. Predictive sequence learning in recurrent neocortical circuits. *Advances in Neural Information Processing Systems*, 13, 2000.

This work was funded by the Gatsby Charitable Foundation.

Contact: marius@gatsby.ucl.ac.uk