Supplementary Material for "Probabilistic Modeling of Dependencies Among Visual Short-Term Memory Representations"

A. Emin Orhan **Robert A. Jacobs**
Department of Brain & Cognitive Sciences
University of Rochester
Rochester, NY 14627
{eorhan,robbie}@bcs.rochester.edu

Derivation of Equation 5

This section provides the derivation of Equation 5 in the paper.

We consider the problem of encoding N stimuli ($\mathbf{s} = [s_1, \ldots, s_N]$) in a correlated population of K neurons with Gaussian noise:

$$p(\mathbf{r}|\mathbf{s}) = \frac{1}{\sqrt{(2\pi)^K \det Q(\mathbf{s})}} \exp[-\frac{1}{2}(\mathbf{r} - \mathbf{f}(\mathbf{s}))^T Q^{-1}(\mathbf{s})(\mathbf{r} - \mathbf{f}(\mathbf{s}))] \tag{1}$$

where $\mathbf{r}$ is a vector containing the firing rates of the neurons in the population, $\mathbf{f}(\mathbf{s})$ represents the tuning functions of the neurons and Q represents the specific covariance structure chosen. We assume a 'limited range correlation structure' for Q that has been analytically studied several times in the literature [1-8]. In a neural population with limited range correlations, the covariance between the firing rates of the k-th and l-th neurons (the kl-th cell of the covariance matrix) is assumed to be a monotonically decreasing function of the distance between their preferred stimuli:

$$Q_{kl}(\mathbf{s}) = a f_k(\mathbf{s})^\alpha f_l(\mathbf{s})^\alpha \exp(-\frac{||\mathbf{c}^{(k)} - \mathbf{c}^{(l)}||}{L}) \tag{2}$$

where $\mathbf{c}^{(k)}$ and $\mathbf{c}^{(l)}$ are the tuning function centers of the neurons. Under the assumptions we make in the paper (assumptions (i)-(iii) on page 6), Equation 2 is equal to: $Q_{kl} = \rho^{|k-l|} a f_k^\alpha f_l^\alpha$ with $\rho = \exp(-1/(L\eta))$ where L is a length parameter determining the spatial extent of the correlations and η is the density of the tuning functions.

We are interested in deriving the FIM, $J(\mathbf{s})$, for our correlated neural population encoding the stimuli $\mathbf{s}$. The ij-th cell of the FIM is defined as:

$$J_{ij}(\mathbf{s}) = -E[\frac{\partial^2}{\partial s_i \partial s_j} \log p(\mathbf{r}|\mathbf{s})] \tag{3}$$

It is easy to show that for a multivariate Gaussian distribution as in Equation 1, the ij-th element of the Fisher information matrix is equal to (see, e.g. ref. [3]):

$$J_{ij}(\mathbf{s}) = \frac{\partial \mathbf{f}^T}{\partial s_i} Q^{-1} \frac{\partial \mathbf{f}}{\partial s_j} + \frac{1}{2}\text{Tr}[\frac{\partial Q}{\partial s_i} Q^{-1} \frac{\partial Q}{\partial s_j} Q^{-1}] \tag{4}$$

Following [8], we denote the first term in Equation 4 as J_{mean} and the second term as J_{cov} (this is because the first term represents the amount of information encoded in changes in the mean firing rates of the neurons, whereas the second term represents the amount of information encoded in changes in the covariance of the firing rates of the neurons.)

1

In order to derive J_{mean} and J_{cov}, we need to invert Q. For the limited range correlation structure we chose (Equation 2), it is easy to show that the kl-th element of Q^{-1} can be written as (see, e.g. ref. [5]):

$$(Q^{-1})_{kl} = \frac{1}{a f_k^\alpha f_l^\alpha} \frac{1+\rho^2}{1-\rho^2} [\delta_{kl} - \frac{\rho}{1+\rho^2}(\delta_{k+1,l} + \delta_{k-1,l})] \tag{5}$$

where δ is the delta function. This last equation makes it clear that Q^{-1} is a banded matrix, which considerably simplifies the calculation of J_{mean} and J_{cov}.

Carrying out the multiplication in the first term of Equation 4, we get:

$$J_{mean} = \frac{\partial \mathbf{f}^T}{\partial s_i} Q^{-1} \frac{\partial \mathbf{f}}{\partial s_j} = -\sum_{k=2}^{K} \frac{1}{a f_{k-1}^\alpha f_k^\alpha} \frac{\rho}{1-\rho^2} \frac{\partial f_{k-1}}{\partial s_i} \frac{\partial f_k}{\partial s_j} + \sum_{k=1}^{K} \frac{1}{a f_k^\alpha f_k^\alpha} \frac{1+\rho^2}{1-\rho^2} \frac{\partial f_k}{\partial s_i} \frac{\partial f_k}{\partial s_j}$$

$$- \sum_{k=1}^{K-1} \frac{1}{a f_k^\alpha f_{k+1}^\alpha} \frac{\rho}{1-\rho^2} \frac{\partial f_k}{\partial s_i} \frac{\partial f_{k+1}}{\partial s_j}. \tag{6}$$

$$= \frac{1}{a} (\frac{1+\rho^2}{1-\rho^2} \sum_{k=1}^{K} h_k^{(i)} h_k^{(j)} - \frac{2\rho}{1-\rho^2} \sum_{k=1}^{K-1} h_k^{(i)} h_{k+1}^{(j)}) \tag{7}$$

where $h_k^{(i)} = \frac{1}{f_k^\alpha} \frac{\partial f_k}{\partial s_i}$. In Equation 6, we used the fact that Q^{-1} is a banded matrix.

We now move on to the derivation of J_{cov}. We first derive $(\frac{\partial Q}{\partial s_i})_{kl}$:

$$(\frac{\partial Q}{\partial s_i})_{kl} = a \frac{\partial f_k^\alpha}{\partial s_i} f_l^\alpha \exp(-\frac{||\mathbf{c}^{(k)} - \mathbf{c}^{(l)}||}{L}) + a \frac{\partial f_l^\alpha}{\partial s_i} f_k^\alpha \exp(-\frac{||\mathbf{c}^{(k)} - \mathbf{c}^{(l)}||}{L}) \tag{8}$$

$$= \alpha(\frac{1}{f_k} \frac{\partial f_k}{\partial s_i} + \frac{1}{f_l} \frac{\partial f_l}{\partial s_i}) Q_{kl} \tag{9}$$

$$= \alpha(g_k^{(i)} + g_l^{(i)}) Q_{kl} \tag{10}$$

where $g_k^{(i)} = \frac{1}{f_k} \frac{\partial f_k}{\partial s_i}$. Using Equation 10, we get:

$$J_{cov} = \frac{1}{2} \text{Tr}[\frac{\partial Q}{\partial s_i} Q^{-1} \frac{\partial Q}{\partial s_j} Q^{-1}] \tag{11}$$

$$= \frac{\alpha^2}{2} \sum_{k=1}^{K} \sum_{l=1}^{K} \sum_{m=1}^{K} \sum_{n=1}^{K} Q_{kl}(g_k^{(i)} + g_l^{(i)})(Q^{-1})_{lm} Q_{mn}(g_m^{(j)} + g_n^{(j)})(Q^{-1})_{nk} \tag{12}$$

$$= \alpha^2 \sum_{k,l=1}^{K} Q_{kl} g_k^{(i)} g_k^{(j)} (Q^{-1})_{lk} + \alpha^2 \sum_{k,l=1}^{K} Q_{kl} g_k^{(i)} g_l^{(j)} (Q^{-1})_{lk} \tag{13}$$

where in Equation 12, we used the definition of the trace operator and Equation 13 results from the expansion and re-arrangement of Equation 12. To derive a more explicit expression for Equation 13, we note that:

$$Q_{kl}(Q^{-1})_{lk} = a\rho^{|k-l|} f_k^\alpha f_l^\alpha \frac{1}{a f_k^\alpha f_l^\alpha} \frac{1+\rho^2}{1-\rho^2} [\delta_{lk} - \frac{\rho}{1+\rho^2}(\delta_{l+1,k} + \delta_{l-1,k})] \tag{14}$$

$$= \rho^{|k-l|} \frac{1+\rho^2}{1-\rho^2} [\delta_{lk} - \frac{\rho}{1+\rho^2}(\delta_{l+1,k} + \delta_{l-1,k})] \tag{15}$$

If we plug this last expression into Equation 13, the first term in Equation 13 becomes $\alpha^2 \sum_{k=1}^{K} g_k^{(i)} g_k^{(j)}$, because $\sum_{k=1}^{K} Q_{kl}(Q^{-1})_{lk} = 1$. For the second term of Equation 13, we have:

$$\sum_{k,l=1}^{K} g_k^{(i)} g_l^{(j)} \rho^{|k-l|} \frac{1+\rho^2}{1-\rho^2} [\delta_{lk} - \frac{\rho}{1+\rho^2}(\delta_{l+1,k} + \delta_{l-1,k})] = -\sum_{k=2}^{K} g_{k-1}^{(i)} g_k^{(j)} \frac{\rho^2}{1-\rho^2} + \sum_{k=1}^{K} g_k^{(i)} g_k^{(j)} \frac{1+\rho^2}{1-\rho^2}$$

$$- \sum_{k=1}^{K-1} g_k^{(i)} g_{k+1}^{(j)} \frac{\rho^2}{1-\rho^2} = \alpha^2 \sum_{k=1}^{K} g_k^{(i)} g_k^{(j)} \frac{1+\rho^2}{1-\rho^2} - \alpha^2 \sum_{k=1}^{K-1} g_k^{(i)} g_{k+1}^{(j)} \frac{2\rho^2}{1-\rho^2}. \tag{16}$$

2

Combining this with the first term in Equation 13, we get:

$$J_{cov} = \alpha^2 \left(\frac{2}{1-\rho^2} \sum_{k=1}^{K} g_k^{(i)} g_k^{(j)} - \frac{2\rho^2}{1-\rho^2} \sum_{k=1}^{K-1} g_k^{(i)} g_{k+1}^{(j)} \right) \tag{17}$$

Finally, combining J_{mean} (Equation 7) and J_{cov} (Equation 17) together, the ij-th element of the Fisher information matrix is given by:

$$
\begin{aligned}
J_{ij}(\mathbf{s}) = J_{mean} + J_{cov} &= \frac{1}{a} \left(\frac{1+\rho^2}{1-\rho^2} \sum_{k=1}^{K} h_k^{(i)} h_k^{(j)} - \frac{2\rho}{1-\rho^2} \sum_{k=1}^{K-1} h_k^{(i)} h_{k+1}^{(j)} \right) \\
&\quad + \alpha^2 \left(\frac{2}{1-\rho^2} \sum_{k=1}^{K} g_k^{(i)} g_k^{(j)} - \frac{2\rho^2}{1-\rho^2} \sum_{k=1}^{K-1} g_k^{(i)} g_{k+1}^{(j)} \right).
\end{aligned} \tag{18}
$$

where $h_k^{(i)} = \frac{1}{f_k^\alpha} \frac{\partial f_k}{\partial s_i}$ and $g_k^{(i)} = \frac{1}{f_k} \frac{\partial f_k}{\partial s_i}$. Equation 18 is identical to Equation 5 in the paper.

An alternative mean firing rate function (f)

In the paper, for convenience, we assumed that the neurons can be divided into N groups where in each group the tuning functions are a function of the feature value of only one of the stimuli, i.e. $f_k(\mathbf{s}) = f_k(s_n)$ for neurons in group n, so that the effects of other stimuli on the mean firing rates of neurons in group n are negligible. Here, we relax this assumption and consider a mean firing rate function of the form:

$$f_k(\mathbf{s}) = \frac{1}{N} \sum_{n=1}^{N} g_{k,n} \exp\left(-\frac{(s_n - c_k)^2}{\sigma^2} \right) \tag{19}$$

where $g_{k,n}$ represents the effect of the n-th stimulus on the mean firing rate of the k-th neuron. In this case, all stimuli influence the mean firing rate of each neuron, but in varying degrees. In our application, there are $N = 2$ stimuli, so Equation 19 takes the form:

$$f_k(\mathbf{s}) = \frac{1}{2} \left[g_{k,1} \exp\left(-\frac{(s_1 - c_k)^2}{\sigma^2} \right) + g_{k,2} \exp\left(-\frac{(s_2 - c_k)^2}{\sigma^2} \right) \right] \tag{20}$$

We again divide the neurons into two groups and assume that the mean firing rates of the neurons in the first group are affected more by the first stimulus, i.e. $g_{k,1} > g_{k,2}$ for the first group of neurons and the mean firing rates of the neurons in the second group are affected more by the second stimulus, i.e. $g_{k,2} > g_{k,1}$ for the second group of neurons. This is a reasonable assumption in our case, because the two stimuli have different vertical locations in our experiment and they can be expected to influence the mean firing rates of the neurons that have similar preferred vertical locations more than the firing rates of the neurons that have less similar preferred vertical locations. Figure 1 shows the simulation results for this alternative mean firing rate function (compare with Figure 4 in the paper). It is clear that this generalization of the mean firing rate function we used in the paper can also explain the pattern of correlations we observed in our experiment. However, with this mean firing rate function (i.e. Equation 20), the specific pattern of correlations predicted by the model depends strongly on the correlation scale length parameter L. Figure 2 shows predicted mean correlation coefficients as a function of $|s_1 - s_2|$ for different L values. For large L values, the predicted pattern of correlations is qualitatively similar to the correlation pattern we observed in our experiment. However, for small L values, the predicted correlation pattern is reversed such that small $|s_1 - s_2|$ values produce strong negative correlations in the stimulus estimates. As $|s_1 - s_2|$ is increased, correlations increase as well.

References

[1] Abbott, L.F. & Dayan, P. (1999) The effect of correlated variability on the accuracy of a population code. *Neural Computation* 11:91-101.

[2] Shamir, M. & Sompolinsky, H. (2004) Nonlinear population codes. *Neural Computation* 16:1105-1136.

[3] Wilke, S.D. & Eurich, C.W. (2001) Representational accuracy of stochastic neural populations. *Neural Computation* 14:155-189.

3

(a)   (b)
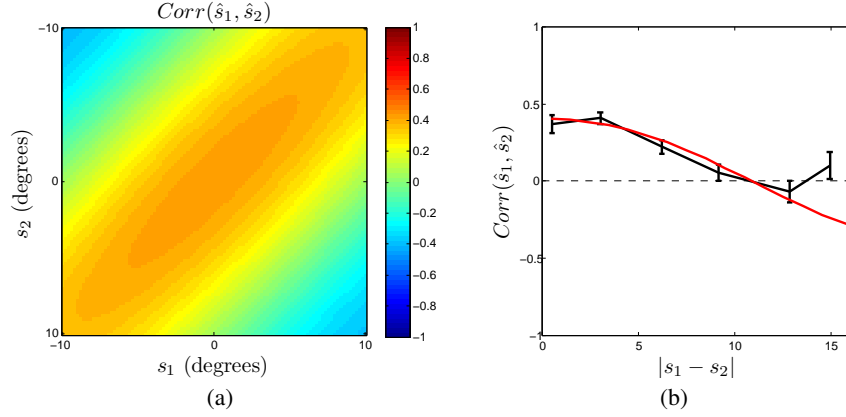
Figure 1: (a) Correlation coefficients estimated from the inverse of the FIM for all stimuli pairs s_1, s_2. (b) Mean correlation coefficients as a function of $|s_1 - s_2|$ (red: model's prediction; black: collapsed data from all 4 subjects in the experiment with $N = 2$). Parameter values: $g_{k,1} = 50$, $g_{k,2} = 30$ for the first group of neurons and $g_{k,1} = 30$, $g_{k,2} = 50$ for the second group of neurons, $K = 500$ (250 neurons in each group), $\sigma = 12.0$, $L = 0.165$, $\alpha = 0.5$, $a = 1$. Only σ and L parameters were optimized.
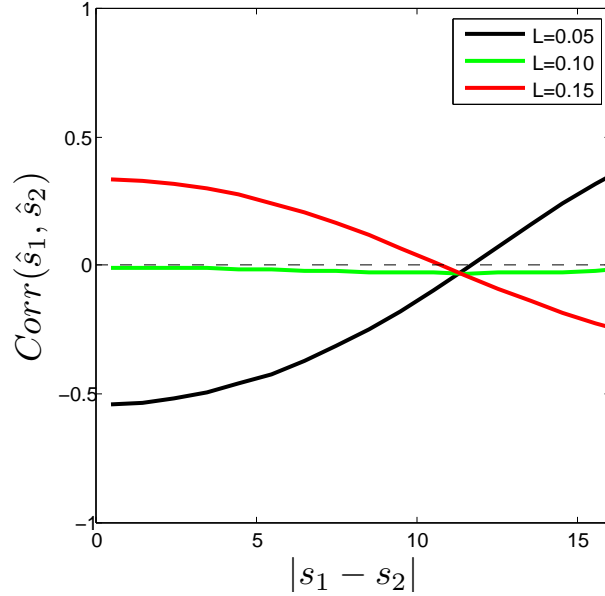
Figure 2: Predicted mean correlation coefficients as a function of $|s_1 - s_2|$ for different L values. Other parameter values were the same as in the previous figure.

4

[4] Berens, P., Ecker, A.S., Gerwinn, S., Tolias, A.S. & Bethge, M. (2011) Reassessing optimal neural population codes with neurometric functions. *PNAS* **108**(11): 44234428.

[5] Snippe, H.P. & Koenderink, J.J. (1992) Information in channel-coded systems: correlated receivers. *Biological Cybernetics* 67: 183-190.

[6] Sompolinsky, H., Yoon, H., Kang, K. & Shamir, M. (2001) Population coding in neural systems with correlated noise. *Physical Review E* 64: 051904.

[7] Josić, K., Shea-Brown, E., Doiron, B. & de la Rocha, J. (2009) Stimulus-dependent correlations and population codes. *Neural Computation* 21:2774-2804.

[8] Ecker, A.S., Berens, P., Tolias, A.S. & Bethge, M. (2011) The effect of noise correlations in populations of diversely tuned neurons. *Journal of Neuroscience* **31** (40):1427214283.

5