Supplemental Material for

Sensory input to cortex encoded on low-dimensional periphery-correlated subspaces

Andrea K Barreiro[1], Cheng Ly[2], Prashant Raju[3], Shree Hari Gautam[3], Woodrow L Shew[3]

We consider the case of two populations of neurons whose responses to two stimuli, A and B, are correlated both within and across populations. We assume the responses to each stimulus can be described by a multivariate Gaussian, i.e.

$$\mathrm{P}(r_X,\, r_Y|S) = \mathrm{N}\left(\begin{bmatrix}\mu_{X,S}\\ \mu_{Y,S}\end{bmatrix}, \Sigma_S\right),$$

where $S = \{A, B\}$, $\mu_{X,S} \in R^m$, $\mu_{Y,S} \in R^n$, and Σ_S is a symmetric, positive-definite matrix of size $(m+n) \times (m+n)$. Here X and Y refer to two populations of neurons which are both responsive to A and B, containing m and n neurons respectively; for example, Y may be a cortical region and X a pre-cortical region which supplies afferent input to Y. Without loss of generality, we simplify notation by shifting the mean responses so that $\mu_{X,A} = 0,\ \mu_{Y,A} = 0$; thus, we can drop the stimulus subscript on the mean vectors and use $\mu_X = \mu_{X,B},\ \mu_Y = \mu_{Y,B}$. We will further assume that the stimulus-conditioned noise correlations are the same for each stimulus: i.e. that $\Sigma_A = \Sigma_B =: \Sigma$.

We seek to determine the *decision boundary*; the surface in R^m (or R^n) which divides the region for which $\mathrm{P}(A|r) > \mathrm{P}(B|r)$ from the region for which $\mathrm{P}(A|r) < \mathrm{P}(B|r)$. In this setting the optimal decision boundary is given by a hyperplane in R^m or R^n; equivalently, by a one-dimensional projection of the response vector. The decision boundary is given by (for example) $\mathrm{u} \in R^m$ such that

$$u^T \Sigma_X^{-1} \mu_X = \frac{1}{2}\mu_X^T \Sigma_X^{-1} \mu_X + \log\frac{P(B)}{P(A)},$$

(Here, Σ_X and Σ_Y are the marginal covariances in populations X and Y respectively.) Therefore, the projection vector must be the normal vector to this plane:

$$v_X = \Sigma_X^{-1}\mu_X;\ v_Y = \Sigma_Y^{-1}\mu_Y; \qquad (S1)$$

in populations X and Y respectively.

Alternatively, observing that $v^T\, r_X\, |S$ is a one-dimensional Gaussian with

$$\mathrm{E}[v^T r_X|S] = v^T \mu_{X,S};\ \mathrm{Var}[v^T r_X|S] = v^T \Sigma v \qquad (S2)$$

we can derive the same outcome by maximizing the signal-to-noise ratio; i.e. $v_X = \mathrm{argmin}\left(\frac{v^T \Sigma v}{v^T \mu_X}\right)$. Equivalently, from the perspective of *linear discriminant analysis*, this maximizes between-class (where “class”=stimulus identity) variability while minimizing within-class variability (Cunningham and Ghahramani, 2015).

Decoding using canonical correlation analysis

We now compute the projection directions associated with *canonical correlation analysis* (CCA). Given two sets of zero-mean observations from X and Y, the goal of CCA is to find the linear projections of the observations that are maximally correlated (Hotelling, 1936). This technique uses the full stimulus-averaged population response; however, we will show that under certain conditions, the maximally correlated direction from CCA coincides with the optimal decoder. Assuming $\mathrm{P}(A) = \mathrm{P}(\mathrm{B})$, the covariance structure within each population is

$$\Sigma_{XX} = \frac{1}{4}\mu_X\mu_X^T + \Sigma_X;\ \Sigma_{YY} = \frac{1}{4}\mu_Y\mu_Y^T + \Sigma_Y$$

While the stimulus-averaged covariance matrix between populations X and Y is

$$\Sigma_{XY} = \frac{1}{4}\mu_X\mu_Y^T + \Sigma_C$$

Here Σ_X, Σ_Y, and Σ_C are the covariances within and across populations, conditioned on stimulus; i.e.

$$\Sigma = \begin{bmatrix} \Sigma_X & \Sigma_C \\ \Sigma_C^T & \Sigma_Y \end{bmatrix}$$

Where $(\Sigma_X)_{jk} = \mathrm{E}\left[\left(r_{X,j} - E\left[r_{X,j} \mid S\right]\right)\left(r_{X,k} - E\left[r_{X,k} \mid S\right]\right) \mid S\right]$ for $1 \le j,\mathrm{k} \le m$ and $(\Sigma_C)_{jk} = \mathrm{E}\left[\left(r_{X,j} - E\left[r_{X,j} \mid S\right]\right)\left(r_{Y,k} - E\left[r_{Y,k} \mid S\right]\right) \mid S\right]$ for $1 \le j \le m,\ 1 \le k \le n$.

The projection directions for the two populations, X and Y, are given by the eigenvectors of D_X and D_Y respectively:

$$D_X = \Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{XY}^T;\ \ D_Y = \Sigma_{YY}^{-1}\Sigma_{XY}^T\Sigma_{XX}^{-1}\Sigma_{XY} \qquad (S3)$$

We will distinguish the *principal CCA direction,* or CC1, as the eigenvector associated with the largest eigenvalue, and denote them $v_{X,CC1}$, $v_{Y,CC1}$ respectively. We note that the cross-covariance matrix Σ_{XY} has two contributions, one reflecting *signal correlations* ($\frac{1}{4}\mu_X\mu_Y^T$) and the other *noise correlations* (Σ_C). The latter reflects trial-to-trial correlations which are not reflected in the mean response. We will now show that when noise correlations are absent ($\Sigma_C = 0$), the principal CCA direction coincides with the optimal decoding direction. Without loss of generality, we focus on D_X; parallel statements hold for D_Y.

Lemma 1: If $\Sigma_C = 0$, then D_X is a rank 1 matrix.

Proof: The *rank* of a matrix is the dimension of its column space; i.e. the dimension of the subspace of vectors that can be the outcome of matrix multiplication. It is well known that the rank of a matrix product is bounded above by the minimum rank of the matrices: i.e. $rank(AB) \leq \min\big(rank(A), rank(B)\big)$. When $\Sigma_C = 0$, Σ_{XY} is given by an outer product; i.e. it is rank 1. Therefore $rank(D_X) \leq 1$ as well.

Theorem 1: If $\Sigma_C = 0$, then the correlated (non-zero) eigenvector of D_X coincides with the projection direction which is optimal for decoding.

Proof: Because D_X is rank 1, it has at most 1 non-zero eigenvalue, with one corresponding eigenvector. This eigenvector *must* coincide with the single vector in a basis for the column space. The cross-population matrix Σ_{XY} is rank 1 and $\mathrm{range}(\Sigma_{XY}) = \mathrm{Span}\{\mu_X\}$; therefore $\mathrm{range}(D_X) = \mathrm{range}(\Sigma_{XX}^{-1}\Sigma_{XY}) = \mathrm{Span}\{\Sigma_{XX}^{-1}\mu_X\}$.

Next, we seek to write $\mathrm{Span}\{\Sigma_{XX}^{-1}\mu_X\}$ in terms of $\mathrm{Span}\{\Sigma_X^{-1}\mu_X\}$. Using the matrix determinant lemma (Sherman-Morrison formula), and noting that

$$\Sigma_{XX} = \Sigma_X + uu^T$$

where $u = \mu_X/2$,

$$\Sigma_{XX}^{-1} = \Sigma_X^{-1} - \frac{\Sigma_X^{-1}uu^T\Sigma_X^{-1}}{1 + u^T\Sigma_X^{-1}u} = \Sigma_X^{-1} - \Sigma_X^{-1}u\left(\frac{u^T\Sigma_X^{-1}}{1 + u^T\Sigma_X^{-1}u}\right)$$

The second term *already* maps into $\mathrm{Span}\{\Sigma_X^{-1}\mu_X\}$, regardless of what vector is multiplied on the right. Therefore

$$\Sigma_{XX}^{-1}\mu_X = \Sigma_X^{-1}\mu_X - \Sigma_X^{-1}u\left(\frac{u^T\Sigma_X^{-1}}{1 + u^T\Sigma_X^{-1}u}\right)\mu_X = \Sigma_X^{-1}\mu_X - \Sigma_X^{-1}\mu_X\left(\frac{u^T\Sigma_X^{-1}\mu_X/2}{1 + u^T\Sigma_X^{-1}u}\right) \propto \Sigma_X^{-1}\mu_X$$

Theorem 2: If $\Sigma_C = 0$, then any other eigenvector of D_X gives chance-level decoding.

Proof: If $v^T\mu_X = 0$, then $\Sigma_{XY}^T v = 0$ and therefore $D_X v = 0$. Therefore v is an eigenvector of D_X with eigenvalue 0. But then

$$E[v^T r_X | A] = E[v^T r_X | B] = 0$$

i.e. the stimuli A and B cannot be discriminated.

References

Cunningham J, Ghahramani Z: Linear Dimensionality Reduction: Survey, Insights, and Generalizations. *Journal of Machine Learning Research* 2015, **16**: 2859−2900.

H. Hotelling. Relations between two sets of variates. Biometrika, 28:321–377, 1936.