Simple low-dimensional models for complex data

Christian K. Machens[1], Ranulfo Romo[2], and Carlos D. Brody[3]

[1]Ludwigs-Maximilians-University Munich, Department Biologie II, Gorsshaderner Str. 2, 82152 Planegg-Martinsried Germany, [2] Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, 04510, México D.F., México, [3]Princeton University, Department of Molecular Biology, Schultz Lab, Washington Rd., Princeton, NJ 08544

During short-term memory maintenance, different neurons in prefrontal cortex (PFC), recorded under identical conditions, show a wide variety of temporal dynamics and response properties [1]. These data are a specific example of the more general finding that neural recordings from frontal cortices often reveal that different neurons have very different response characteristics. Modeling this complexity of responses has been difficult. Most commonly, some features of the responses are focused on, and models that fit those reduced features are built (e.g., [2]). But can the full complexity of responses be easily captured?

We have previously reported that the complex responses in PFC during short-term memory can be summarized in 5 dimensions (i.e., 5 parameters suffice to capture most of the variance in the data across neurons; Machens et al., COSYNE '06). Olasagasti, Goldman, and colleagues have described a method to fit experimentally-obtained steady-state firing rates (that is, no dynamics) in a network model of persistent activity (Olasagasti et al., COSYNE '05, COSYNE '06). We now combine and extend these two approaches, and show how a simple linear fitting procedure leads to a model that describes the data in few dimensions yet captures most of the complexity and dynamics of the neural responses.

Let us assume we have observed, experimentally, M timepoints in the firing rates of N neurons– a total of $M \cdot N$ data points. Let us model this data in a recurrent network of N neurons, with full connectivity. Such a network will have N^2 weights (i.e., as yet undetermined connection strengths). If $N > M$ we have more unknowns than data points, and we could in principle solve the system exactly, reproducing all of the measured neural firing rates. The fitting procedure we use to achieve this requires the inversion of a matrix D representing all the data. To avoid overfitting the data, we use the singular value decomposition to represent, and then easily invert, the data matrix D: setting small singular values to zero corresponds to reducing the dimensionality of the model, which avoids overfitting. For the PFC data during short-term memory that we have previously analyzed, we find, in accordance with our previous results, that five dimensions suffice to describe the data (Machens et al., COSYNE '06). The current approach now maps these data directly onto a neural network model, reproducing the dynamics of the data with most of their experimentally-observed richness and variety.

Acknowledgments
We thank Mark Goldman for helpful discussions. This work was supported by a Swartz Fellowship and an Emmy-Noether-Grant of the DFG to CKM and by NIH grant R01-MH067991 to CDB.

References

[1] Timing and Neural Encoding of Somatosensory Parametric Working Memory in Macaque Prefrontal Cortex. C.D. Brody, A. Hernandez, A. Zainos, and R. Romo, *Cereb. Cortex* 13:1196-1207, 2003.

[2] Flexible control of mutual inhibition: a neural model of two-interval discrimination. C.K. Machens, R. Romo, and C.D. Brody, *Science*, 307:1121-1124, 2005.