

Memory Traces in Normal Neural Networks

Surya Ganguli^{*1}, Dongsung Huh^{*2}, and Haim Sompolinsky³

¹ Sloan-Swartz Center for Theoretical Neurobiology, UCSF. ² Computational Neurobiology Program, UCSD. ³ Racah Institute of Physics, ICNC Hebrew University, and Center for Brain Science, Harvard University.

* These authors contributed equally to this work.

Critical cognitive phenomena such as planning and decision making rely on the ability of the brain to hold information in short-term memory. Many proposals exist for the maintenance of such short-term memories in persistent activity that arises from stable fixed point attractors in the dynamics of recurrent neural networks. However such fixed points are incapable of storing temporal sequences of recent events. An alternate, and relatively less explored paradigm, is the storage of arbitrary temporal input sequences in the transient responses of a recurrent neural network. In this paradigm, short-term memory reconstructs the past input history from the network's current dynamical state¹⁻³. Such a paradigm raises important theoretical questions about the memory capacity of a generic recurrent network to store information about past inputs. How does this capacity scale with the number of neurons N ? How is it affected by noise? How does the memory for past inputs degrade as a function of time? Prior work has addressed these theoretical questions primarily within the restricted class of linear neural networks whose connectivity is given by an orthogonal matrix³. Here we address these questions for a more general class of matrices known as *normal matrices*, i.e., matrices that have an orthogonal basis of eigenvectors. This class of matrices includes symmetric and antisymmetric as well as orthogonal matrices.

We develop a general mean field theory to compute the memory capacity of a linear recurrent network with a normal connectivity matrix. Focusing on fully connected random symmetric matrices, we find that the memory trace for past inputs decays exponentially as function of time. The rate of decay is inversely related to the degree of stability of the network. At the boundary between stability and instability, this memory trace decays as a power law. The predicted power-law is verified in numerical simulations. The total capacity is, however, sensitive to noise. The strength of the noise must be exponentially small in N in order for the capacity to be extensive, i.e. scale linearly with N . This result contrasts with the orthogonal case in which extensive capacity only requires the strength of noise to scale as $\frac{1}{N}$. We have also studied numerically short-term memory properties of classes of nonrandom connectivity matrices including asymmetric sparse matrices. This work highlights the architectural constraints imposed on recurrent networks that subserve short term memory functions, yields a theoretical understanding of the decay of memory traces in them, and provides a basis for further studies of more biologically realistic dynamical models of short term memory in cortical networks.

Acknowledgments

HS thanks O. White for very helpful discussions. Supported by the Swartz and Israel Science Foundations.

References

[1] W. Maass, T. Natschlag, and H. Markram, *Neural Comput.* 14 2531, 2002. [2] H. Jaeger, GMD Report 152, German National Research Center for Information Technology, 2001. [3] O. White, D. Lee and H. Sompolinsky, *Phys. Rev. Lett.* 92(14) 1481102, 2004.