

Compressed sensing in the brain: the role of sparseness in short term and long term memory.

Surya Ganguli, UCSF

Haim Sompolinsky, Harvard University and Hebrew University

One of the most exciting advances in signal processing is the field of compressed sensing (CS) [1]. In CS, sparse high-dimensional stimuli are represented by lower dimensional dense measurements, which are linear mixtures of the stimuli. CS shows that by using the computationally tractable L1 norm as a sparse prior, the high-dimensional stimuli (for example, human speech, natural movies, FMRI data) can be fully reconstructed from the compressed measurements.

In this work, we have extended CS theory and applied it to reveal new fundamental properties of neural learning and memory in the biologically relevant scenario of sparse coding.

Our first extension of CS addresses the capabilities of neuronal circuits to buffer temporal signals, subserving working memory. Working memory systems can store complex temporal sequences, such as speech, over many seconds even though the individual neurons involved, when isolated, forget their inputs rapidly over milliseconds. Previous work [2,3], which assumed normally distributed signals, has shown that a circuit of  $N$  neurons can store temporal signals of duration at most  $N$  (in units of the neuronal time constant). Furthermore, it was shown [4], that in the presence of noise, recurrent networks cannot outperform an equivalent feedforward network which functions as a delay line.

Here we show that sparse signals can be faithfully reconstructed from the instantaneous network activity even for a period of time that *exceeds* the number of neurons in the network. This enhanced capacity for storing sparse signals is realized by 'orthogonal' recurrent networks and not by feedforward networks. We compute analytically the memory capacity and distribution of errors in signal reconstruction as a function of network size, signal sparsity, and the distribution of nonzero components in the signal.

We also analyze the ability of neuronal networks to learn rules when the rule to be learned can be realized by a network with sparse connectivity. We show that in contrast to classical results concerning learning and generalization in neural networks, such networks can actually learn rules, and generalize correctly, using a remarkably small number of examples, *smaller* than the dimensionality of the input space.

Finally, we explore the properties of synaptic learning in sensory motor tasks which incorporate sparsity priors on synaptic connectivity patterns in the form of L1 norm. We analyze the dynamics of such learning models and yield predictions for the performance of the system which can be readily tested in experiments.

[1] E.J. Candes, M.B. Wakin, IEEE Sig.Proc.Mag. 2009.

[2] H. Jaeger GMD Rep. 148 2001.

[3] O.L.White, D.D. Lee, H. Sompolinsky, PRL 2004.

[4] S.Ganguli, D.Huh, H.Sompolinsky, PNAS 2008.