Short-Term Memory Capacity in Recurrent Networks via Compressed Sensing

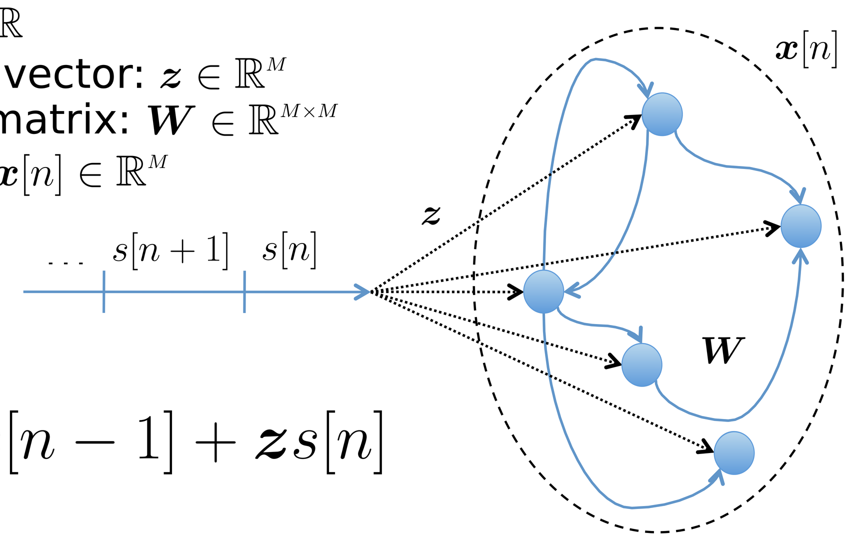
Adam Charles[1], Han Lun Yap[1], Christopher J. Rozell[1]
[1]Georgia Institute of Technology

Introduction

Short Term Memory (STM) in neural systems plays a vital role in a number of biologically important tasks:
- Prediction
- Classification
- Working memory in biological networks [1]

To understand STM, use **Echo State Networks** as a proxy. ESNs use random network connectivity [2].

Stimuli: $s[n] \in \mathbb{R}$
Feed-Forward vector: $\boldsymbol{z} \in \mathbb{R}^M$
Connectivity matrix: $\boldsymbol{W} \in \mathbb{R}^{M\times M}$
Neural state: $\boldsymbol{x}[n] \in \mathbb{R}^M$



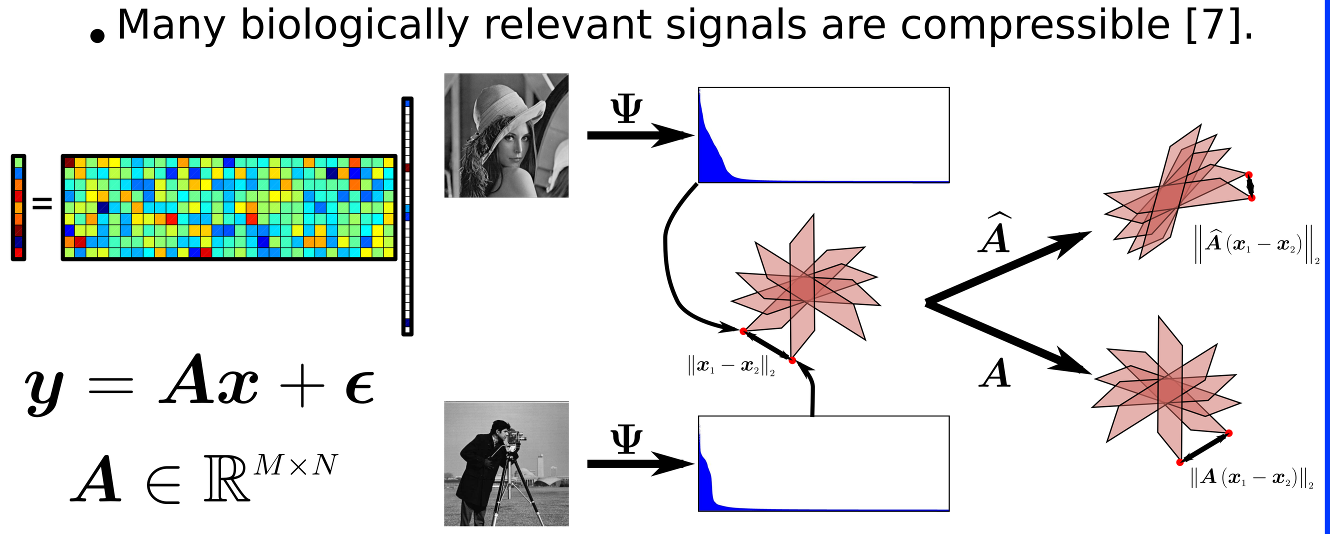
$$\boldsymbol{x}[n] = \boldsymbol{W}\boldsymbol{x}[n-1] + \boldsymbol{z}s[n]$$

How much of the stimulus content is encoded in the network?
- Previous work capped STM length as $N \leq M$ [3-4].
- More recent work suggests $N > M$ when the stimulus has **low dimensional structure** [5].

We show increased STM for sparse input patterns using compressive sensing techniques.

Compressive Sensing

Compressive Sensing (CS):
- Allows robust recovery of undersampled signals from generalized linear measurements [6].
- Relies on distances preservation for compressible signals (signal energy concentrates under a linear transform).
- Many biologically relevant signals are compressible [7].



$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon}$$
$$\boldsymbol{A} \in \mathbb{R}^{M\times N}$$

The 'distance preservation' in CS is the Restricted Isometry Property (RIP). $\boldsymbol{A}$ satisfies RIP with K,δ if

$$1-\delta \leq \left\|\boldsymbol{A}\boldsymbol{\Psi}^T\boldsymbol{x}\right\|_2 / \left\|\boldsymbol{x}\right\|_2 \leq 1+\delta$$

for all K-sparse $\boldsymbol{x}$. CS results state that K-sparse $\boldsymbol{x}$ are robustly recoverable from $\boldsymbol{y}$ if $\boldsymbol{A}$ satisfies RIP-($2K,\delta$) by solving

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\Psi}^T\boldsymbol{x}\right\|_2^2 + \lambda\left\|\boldsymbol{x}\right\|_1$$

λ trades off sparsity and data fidelity. The solution error is

$$\left\|\boldsymbol{\Psi}^T\widehat{\boldsymbol{x}} - \boldsymbol{\Psi}^T\boldsymbol{x}\right\|_2 \leq \alpha\left\|\boldsymbol{\epsilon}\right\|_2 + \beta\frac{\left\|\boldsymbol{x}-\boldsymbol{x}_K\right\|_1}{\sqrt{K}}$$

for constants α and β, and $\boldsymbol{x}_K$ is the best K term approximation to $\boldsymbol{x}$.

Showing The RIP

RIP usually shown to hold with high probability for a random matrix e.g. Random Gaussian matrices satisfy RIP with high probability if

$$M \geq CK\mu^2\left(\boldsymbol{\Psi}\right)\delta^{-2}\log\left(N\right)$$

Coherence $\mu(\boldsymbol{\Psi})$ measures similarity of $\boldsymbol{\Psi}$ and $\boldsymbol{A}$ [8].
Structured systems need more measurements for comparable results

Can we analyze ESN dynamics using the RIP?

System Model:

Following Ganguli et al.,

$$\boldsymbol{x}[N] = [\boldsymbol{z}\ \ \boldsymbol{W}\boldsymbol{z}\ \ \ldots\ \ \boldsymbol{W}^{N-1}\boldsymbol{z}]\begin{bmatrix} s[N] \\ s[N-1] \\ \vdots \\ s[1]\end{bmatrix} = \boldsymbol{A}\boldsymbol{s}$$

We use the Eigenvalue decomposition $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{-1}$ to extract the exact effect of the elements of the ESN:

$$\boldsymbol{x}[N] = \boldsymbol{U}\boldsymbol{Z}\boldsymbol{F}\boldsymbol{s} \qquad \boldsymbol{F}_{i,j} = \boldsymbol{D}_{i,i}^j$$
$$\boldsymbol{Z} = \mathrm{diag}(\boldsymbol{U}^{-1}\boldsymbol{z})$$

Guarantees on System Model:

If we choose:
- Random orthogonal $\boldsymbol{W}$: $\boldsymbol{W}^T\boldsymbol{W} = \boldsymbol{I}_M$
- $\boldsymbol{z} = \boldsymbol{U}\boldsymbol{1}_M$

then we can show that
- $\boldsymbol{F}$ is a subsampled Fourier transform
- RIP holds with high probability if

$$M \geq CK\mu^2\left(\boldsymbol{\Psi}\right)\delta^{-2}\log^4(N)$$

with coherence to the set of all sinusoids.
- Recovery using

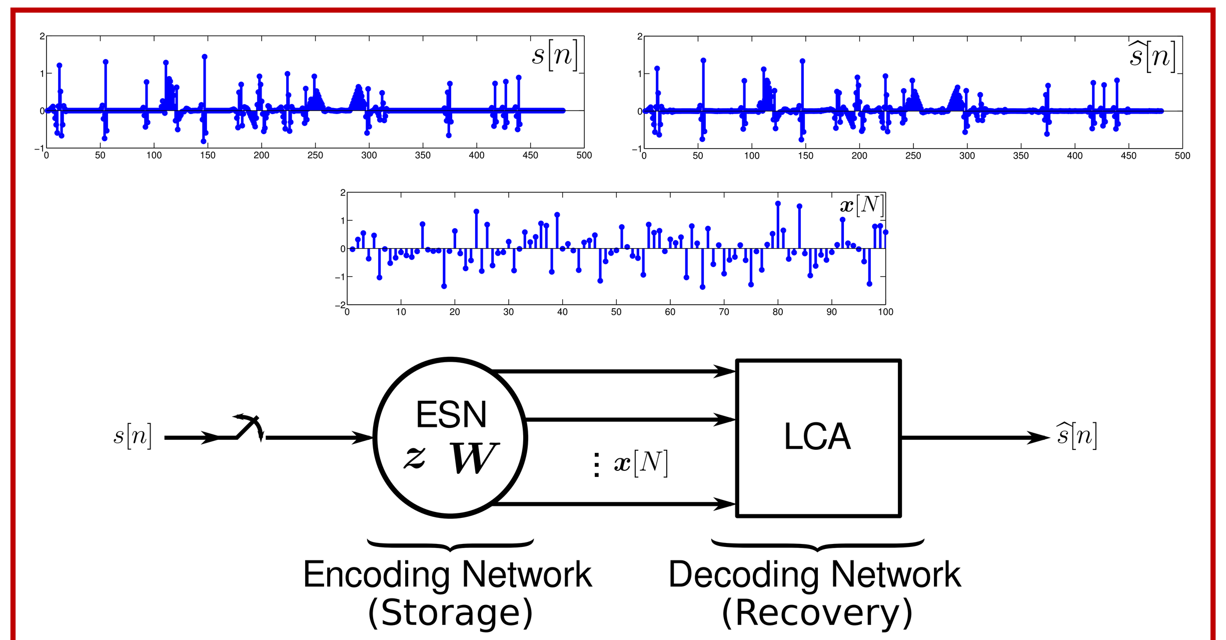
$$\widehat{\boldsymbol{s}} = \arg\min_{\boldsymbol{s}} \left\|\boldsymbol{x}[N] - \boldsymbol{U}\boldsymbol{Z}\boldsymbol{F}\boldsymbol{s}\right\|_2^2 + \lambda\left\|\boldsymbol{\Psi}\boldsymbol{s}\right\|_1$$

gives the error

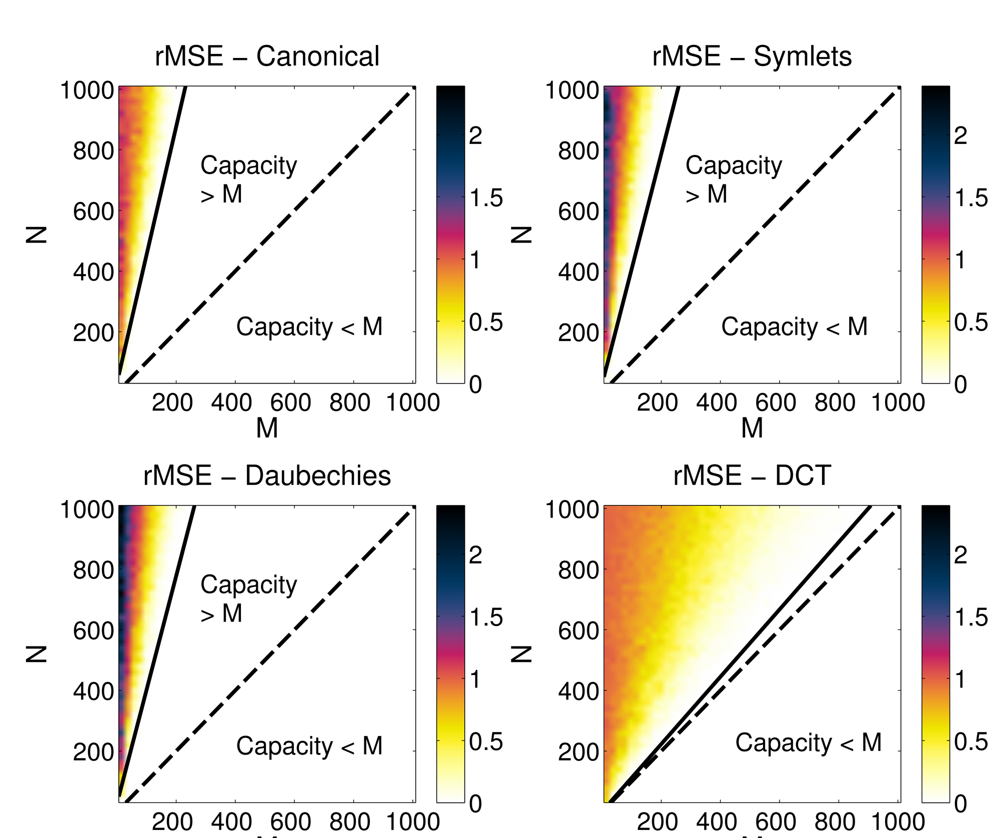
$$\left\|\widehat{\boldsymbol{s}} - \boldsymbol{s}\right\|_2 \leq \alpha\left\|\boldsymbol{\epsilon}\right\|_2 + \beta\frac{\left\|\boldsymbol{\Psi}\boldsymbol{s} - [\boldsymbol{\Psi}\boldsymbol{s}]_K\right\|_1}{\sqrt{K}}$$

Note:
- M linear with K, logarithmic with N
- No error, K-sparse inputs → Perfect recovery

Finite Length Inputs



The RIP is satisfied by the ESN for N > M. Thus, compressible signals are recoverable. Secondary circuits such as the Locally Competitive Algorithm (LCA) could decode the network nodes [9].



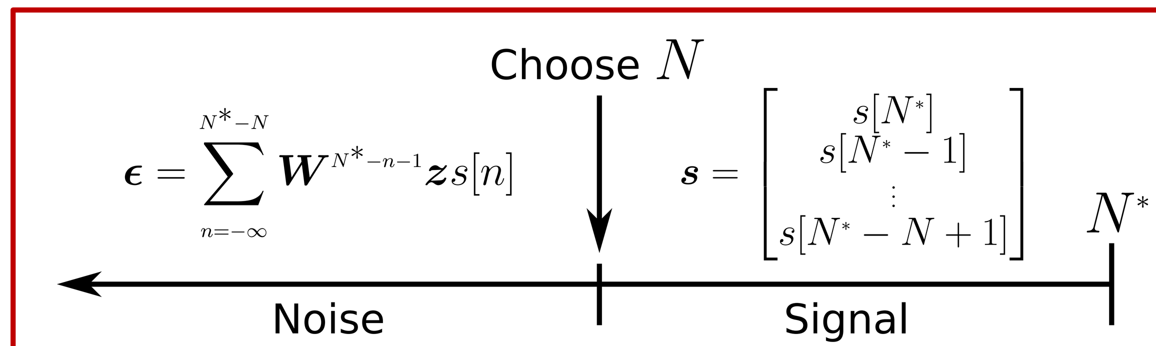
The STM exceeds M for different sparsity bases with $K = \rho N$ so long as the coherence is low.

Infinite Length Inputs

Infinite length inputs necessitate
- Eigenvalues of $\boldsymbol{W}$ have magnitude $q < 1$
- Regular decay can be isolated as

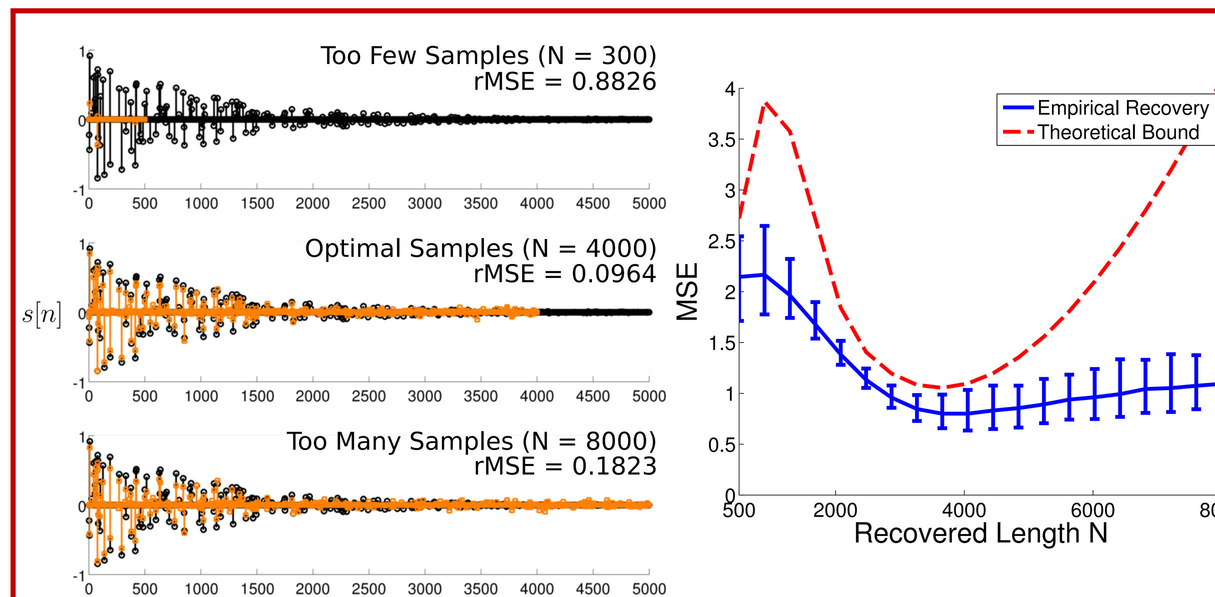
$$\boldsymbol{x}[N] = \boldsymbol{U}\boldsymbol{Z}\boldsymbol{F}\boldsymbol{Q}\boldsymbol{s} + \boldsymbol{\epsilon} \qquad \boldsymbol{Q} = \mathrm{diag}([1\ \ q\ \ \ldots\ \ q^{N-1}])$$

- Effects of older stimuli are noise



Using the RIP condition, we can bound the error by

$$\left\|\boldsymbol{Q}\widehat{\boldsymbol{s}} - \boldsymbol{Q}\boldsymbol{s}\right\| \leq \alpha s_{\max}\left\|\boldsymbol{U}\right\|_2\left|\frac{q^N - q^{N+\rho N^*}}{1-q}\right| + \frac{\beta s_{\max}}{\sqrt{S^*}}\left(\frac{q^{S^*} - q^{\rho N}}{1-q}\right)$$



The error bound in the infinite case has a minimum. Recalling too much or too little history results in errors

Network Types

Orthogonal networks can have different topologies:
- Fully Connected
- Modular (disjoint fully connected subgroups)
- Small World (sparsely connected groups of fully connected neurons)

Non-orthogonal $\boldsymbol{W}$ (with same eigenvalue properties) changes robustness, but not noiseless guarantees.

Feed-forward vector can be chosen at random - adds a log-squared factor:

$$M \geq CK\mu^2\left(\boldsymbol{\Psi}\right)\delta^{-2}\log^6\left(N\right)$$

If feed-forward vector is mis-aligned with the eigen-vectors, the effective network nodes decreases.

Conclusions

We analyze the exact dynamics for ESNs using tools from compressive sinsing. In short:
- RIP shows that stimuli for ESNs are recoverable
- Tractable recovery algorithm (even neural solvers)
- Many bases possible in finite case
- Infinite case demonstrates an optimal recovery length (best STM length)
- Can account for some deviations from basic assumptions

Future directions:
- Better understand the role of eigenvalue decay
- Extend results to general low-dimensional time series embedding

References:

[1] D. V. Buonomano and W. Maass. State-dependent computations: spatiotemporal processing in cortical networks. Nat Rev Neurosci, 10(2):113-125, Feb. 2009.

[2] H. Jaeger. The ``echo state'' approach to analysing and training recurrent neural networks. Technical Report GMD Report 148, German National Research Center for Information Technology, 2001.

[3] O. L. White, D. D. Lee, and H. Sompolinsky. Short-term memory in orthogonal neural networks. Phys. Rev. Lett., 92(14):148102, 2004.

[4] S. Ganguli, D. Huh, and H. Sompolinsky. Memory traces in dynamical systems. Proc. Natl. Acad. Sci. USA, 105(48):18970–18975, Dec. 2008.

[5] S. Ganguli and H. Sompolinsky. Short-term memory in neuronal networks through dynamical compressed sensing. Proc. Conf. Advances Neural Inform. Process. Syst., 2010.

[6] E. J. Candès. Compressive sampling. Proc. Int. Congr. Mathematicians, 3:1433–1452. 2006.

[7] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 381(6583):607–609, 1996.

[8] H. Rauhut. Compressive sensing and structured random matrices. Theoretical Found. and Numerical Methods for Sparse Recovery, 9:1–92, 2010.

[9] C. J. Rozell, D. H. Johnson, R. G. Baraniuk, and B. A. Olshausen. Sparse coding via thresholding and local competition in neural circuits. Neural Computation, 20(10):2526–2563, Oct. 2010.

Acknowledgements

This work was partially supported by NSF grant CCF-0830456 and DSO National Laboratories, Singapore.

Contact

A.S.C., H.L.Y. and C.J.R are at the Georgia Institute of Technology in Atlanta Georgia: {acharles6,yaphanlun,crozell}@gatech.edu