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

LABORATORY FOR NEUROENGINEERING Georgia Tech

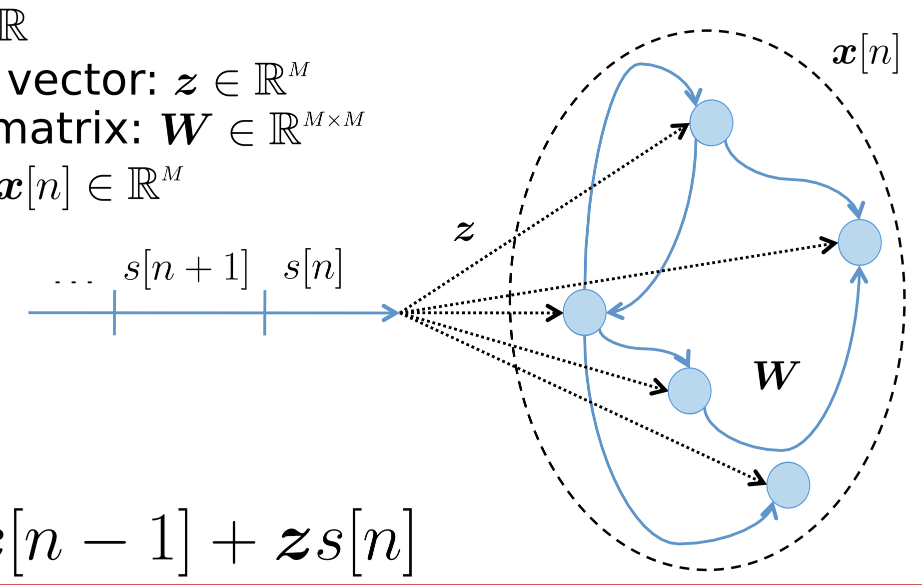
Introduction

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?

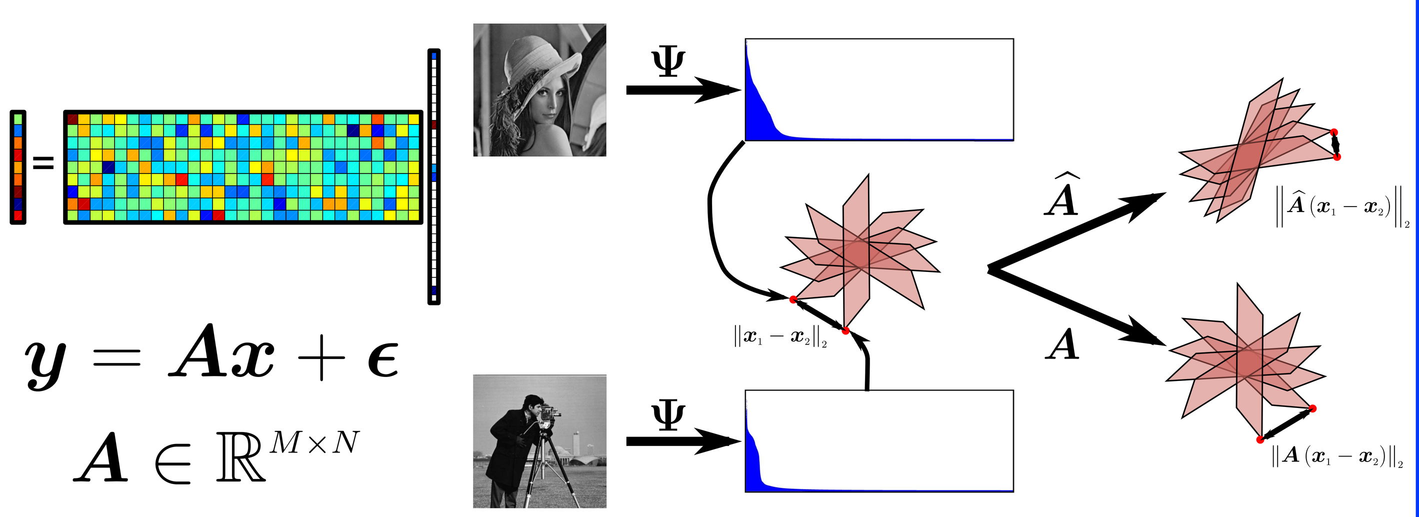
- Previous work capped STM length as $N \leq M$ [3-4].
- More recent work suggests $N > M$ when the stimulus has **low dimensional structure** [5].

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

Compressive Sensing

Compressive Sensing (CS):

- Allows robust recovery of undersampled signals [6].
- Relies on distances preservation for compressible signals.
- Many biologically relevant signals are compressible [7].



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

'Distance preservation' ⟷ Restricted Isometry Property (RIP)
$\boldsymbol{A}$ satisfies RIP with K, δ if

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

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

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

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

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

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

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

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

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

Showing The RIP

Can we analyze ESN dynamics using the RIP?

System Model:

Following Ganguli et al.,

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

Eigenvalue decomposition: $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{-1}$

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

Guarantees on System Model:

If we choose:

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

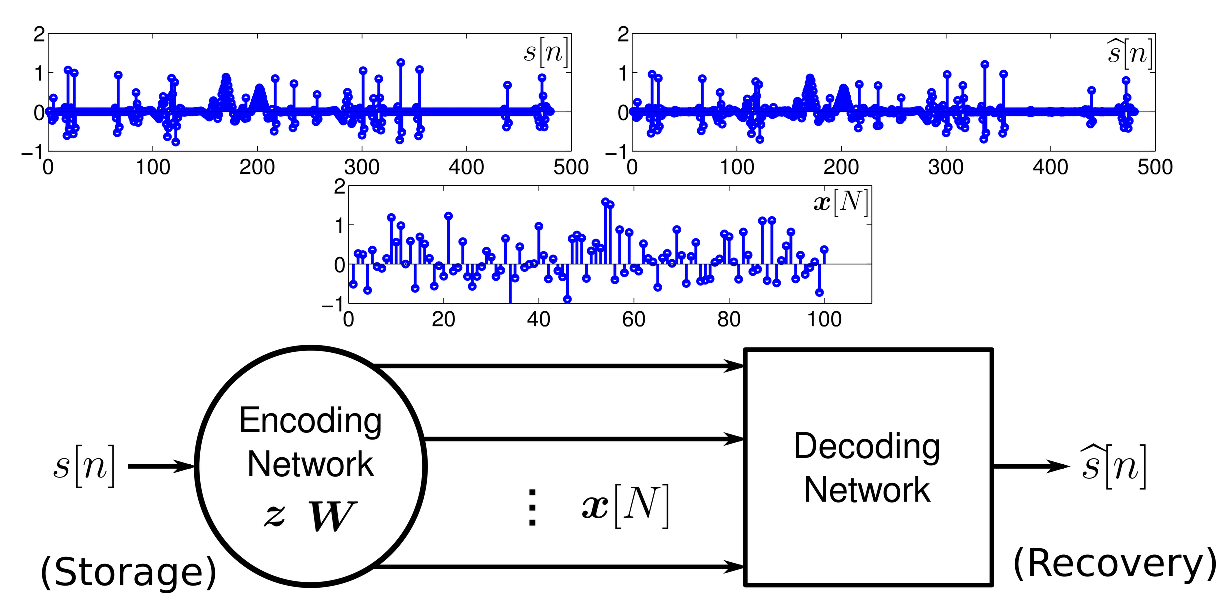
then we can show:

- $\boldsymbol{F}$ is a subsampled Fourier transform
- RIP holds with high probability $1-\eta$ if
$$M \geq CK\mu^2(\boldsymbol{\Psi})\delta^{-2}\log^4(N)\log(\eta^{-1})$$
with coherence to the set of all sinusoids.
- Recovery using
$$\widehat{\boldsymbol{s}} = \arg\min_{\boldsymbol{s}} \|\boldsymbol{x}[N] - \boldsymbol{U}\boldsymbol{Z}\boldsymbol{F}\boldsymbol{s}\|_2^2 + \lambda\|\boldsymbol{\Psi}\boldsymbol{s}\|_1$$
gives the error
$$\|\widehat{\boldsymbol{s}} - \boldsymbol{s}\|_2 \leq \alpha\|\boldsymbol{\epsilon}\|_2 + \beta\frac{\|\boldsymbol{\Psi}\boldsymbol{s} - [\boldsymbol{\Psi}\boldsymbol{s}]_K\|_1}{\sqrt{K}}$$

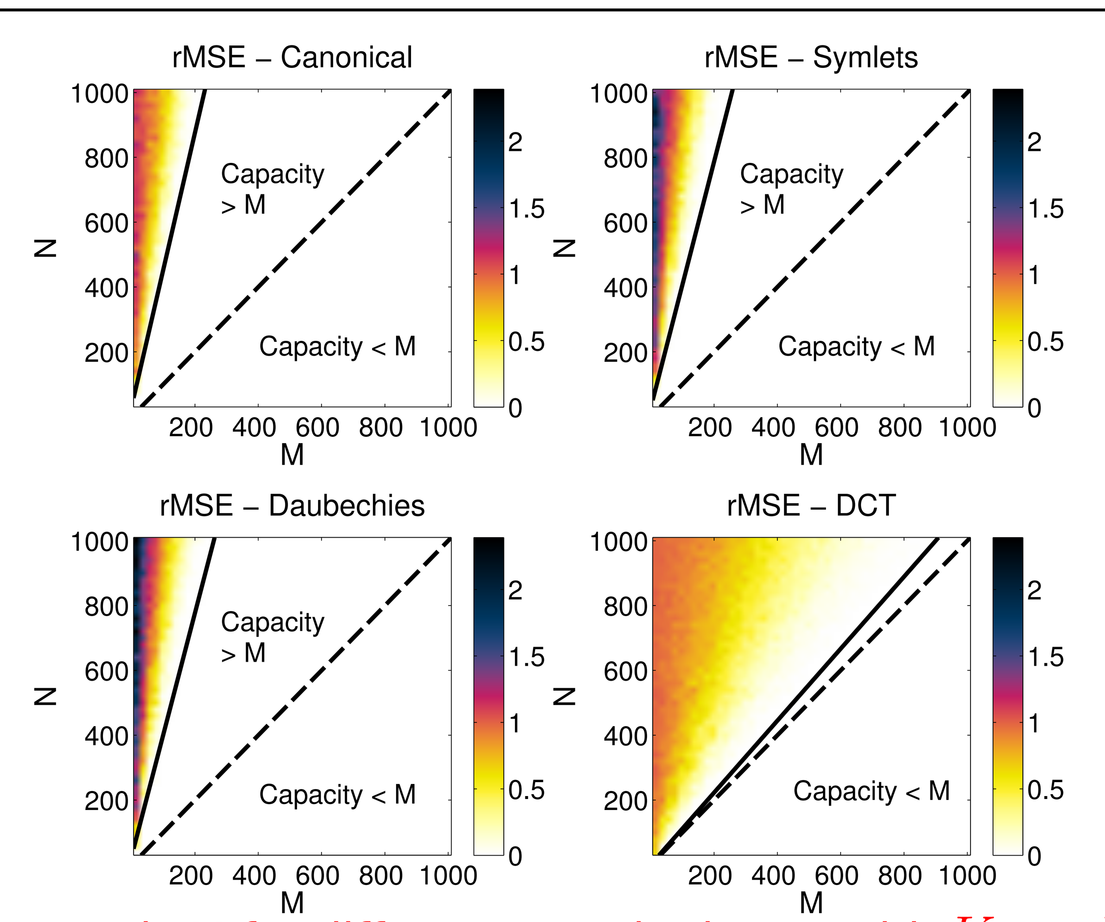
Note:

- 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].

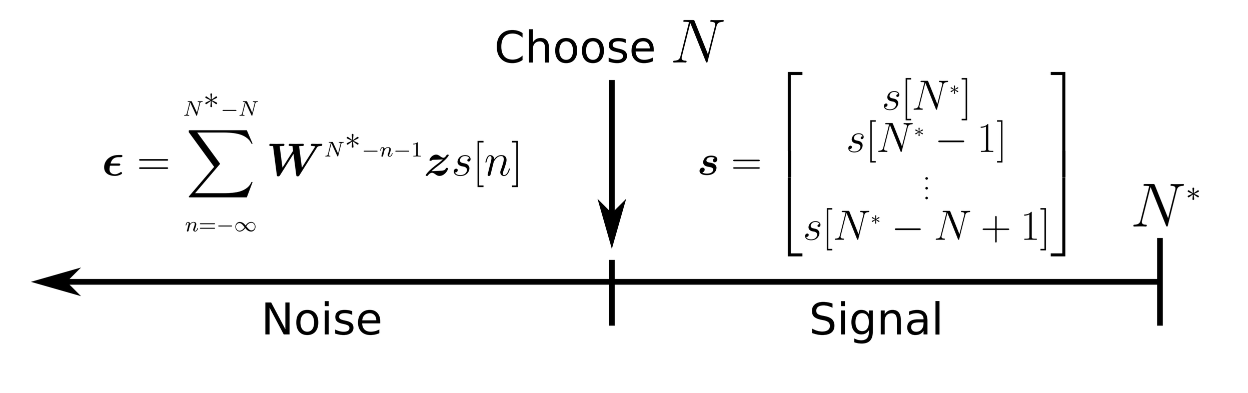


The STM exceeds M for different sparsity bases with $K = \rho N$ so long as the coherence is low.

Infinite Length Inputs

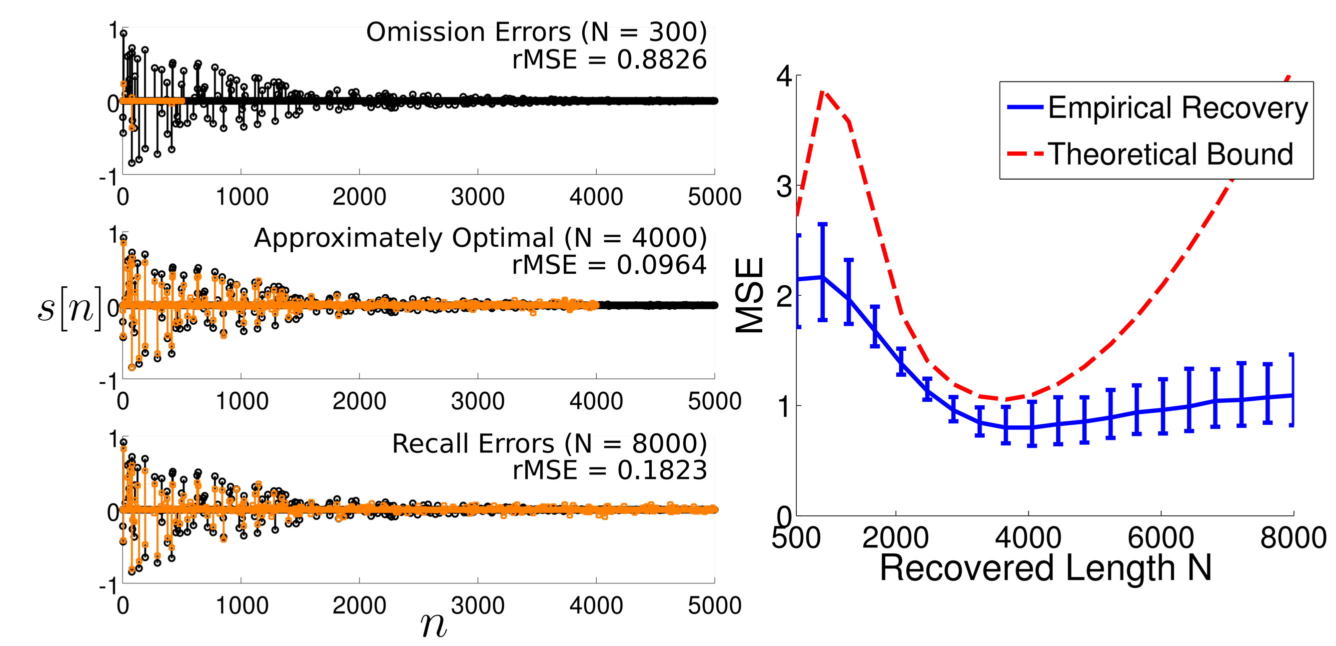
Infinite length inputs necessitate

- Eigenvalues of $\boldsymbol{W}$ have magnitude $q < 1$
- Regular decay can be isolated as
$$\boldsymbol{x}[N] = \boldsymbol{U}\boldsymbol{Z}\boldsymbol{F}\boldsymbol{Q}\boldsymbol{s} + \boldsymbol{\epsilon} \qquad \boldsymbol{Q} = \text{diag}([1\ \ q\ \ \ldots\ \ q^{N-1}])$$
- Effects of older stimuli are noise



Using the RIP condition, we can bound the error by

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



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 subgroups 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 factor:

$$M \geq CK\mu^2(\boldsymbol{\Psi})\delta^{-2}\log^5(N)\log(\eta^{-1})$$

If feed-forward vector is mis-aligned with the eigenvectors, 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

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