Using compressed sensing to study sequence memory capacity in networked systems

Adam S. Charles, Han Lun Yap and Christopher J. Rozell
School of ECE, Georgia Institute of Technology, Atlanta, GA, USA (e-mail: acharles6,yaphanlun,crozell@gatech.edu)

Many interesting systems are modeled as networks with dynamic interactions between the nodes (e.g., neural networks, social networks, sensor networks). Recurrent interactions between these nodes create a type of short term memory (STM) where the transient network state collectively retains information about past inputs. Characterizing the fundamental limits of STM in networked systems is critical to understanding the computational abilities of these networks. For example, fundamental questions include determining the effects on memory capacity of network size, connectivity patterns, and input statistics. One canonical model for network interactions has the form:

$$\boldsymbol{x}[n] = f\left(\boldsymbol{W}\boldsymbol{x}[n-1] + \boldsymbol{z}s[n] + \widetilde{\boldsymbol{\epsilon}}[n]\right), \quad (1)$$

where $\boldsymbol{x}[n] \in \mathbb{R}^M$ are the network states at time n, $\boldsymbol{W}$ is the $(M \times M)$ recurrent (feedback) connectivity matrix, $s[n] \in \mathbb{R}$ is the input sequence at time n, $\boldsymbol{z}$ is the $(M \times 1)$ projection of the input into the network, $\widetilde{\boldsymbol{\epsilon}}[n]$ is a potential network noise source, and $f : \mathbb{R}^M \rightarrow \mathbb{R}^M$ is a possible pointwise nonlinearity. The general idea is that if $\boldsymbol{W}$ is rich enough (often taken to be random connections), a single input will reverberate in the network, thereby creating a "memory" of the past input in the current network states.

The STM capacity of the linear version of this network model (i.e., $f(\boldsymbol{x}) = \boldsymbol{x}$) has recently been extensively studied [1]–[4]. Existing analyses (using various assumptions and definitions of "capacity") derive STM capacity limits of $N \leq M$, meaning that the number of past inputs significantly recoverable by the current network state scales linearly with the number of nodes in the network. The main contribution of the work described in this abstract is to leverage the established guarantees of the compressed sensing (CS) literature to provide rigorous, non-asymptotic recovery error bounds for sparse input sequences that show network STM capacities can be significantly higher than the number of the nodes in the network (as hinted at in [1]). Our analysis characterizes the impact on STM capacity of the input sparsity level and sparsity basis, as well as the characteristics of the recurrent connectivity matrix. We provide both perfect recovery guarantees for finite inputs, as well as results on the recovery tradeoffs when the network has an infinitely long input sequence. The latter analysis highlights the fact that when the network has an infinitely long streaming input, the system has an optimal recovery length that balances errors due to omission and recall mistakes.

To be concrete, we assume that every length-N segment of the input sequence $\boldsymbol{s}[n]$ can be written using the basis $\boldsymbol{\Psi}$ with S non-zero coefficients. Next, we write the network dynamics as a CS measurement operation. The linear dynamics of Eqn (1) can be used to write the network state at time N in terms of the input signal and the iteratively applied connectivity matrix $\boldsymbol{x}[N] = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{\epsilon}$ where, $\boldsymbol{A}$ is a $M \times N$ matrix, the k^{th} column of $\boldsymbol{A}$ is $\boldsymbol{W}^{k-1}\boldsymbol{z}$, $\boldsymbol{s} = [\boldsymbol{s}[N], \ldots, \boldsymbol{s}[1]]^T$, the initial state of the system is $\boldsymbol{x}[0] = 0$, and $\boldsymbol{\epsilon}$ is the node activity not accounted for by the input stimulus (e.g. the sum of network noise terms $\boldsymbol{\epsilon} = \sum_{k=1}^{N} \boldsymbol{W}^{N-k}\widetilde{\boldsymbol{\epsilon}}[k]$). Interpreted as a CS problem, we see that if $\boldsymbol{A}$ satisfies the restricted isometry property (RIP) (i.e. $\boldsymbol{A}$ forms a tight frame for sparse signals) for the sparsity basis $\boldsymbol{\Psi}$, established error bounds from the CS literature [5] provide strong guarantees on recovering $\boldsymbol{s}$ from the current network states $\boldsymbol{x}[N]$ via ℓ_1 minimization.

The simplest network construction amenable to analysis arises when $\boldsymbol{W}$ is a random orthonormal matrix (as in [1], [2], [4]) and $\boldsymbol{z} = \frac{1}{\sqrt{M}}\boldsymbol{U}\mathbf{1}_M$, where $\mathbf{1}_M$ is a vector of M ones and $\boldsymbol{U}$ is the matrix of eigenvectors of $\boldsymbol{W}$. Under this construction above, our main technical contribution extends existing recent results [6] on randomly subsampled Discrete Time Fourier Transform matrices to show that $\boldsymbol{A}$ satisfies the RIP if the number of nodes M satisfies the inequality

$$M \geq C\frac{S}{\delta^2}\mu^2\left(\boldsymbol{\Psi}\right)\log^4\left(N\right)\log\left(\eta^{-1}\right), \quad (2)$$

where δ is the RIP conditioning of $\boldsymbol{A}$, N is the length of the recovered input signal, S is the input sequence sparsity, C is a constant, $\mu(\boldsymbol{\Psi})$ quantifies the similarity between the sparsity basis $\boldsymbol{\Psi}$ and a randomly sampled DTFT, and η is a small pre-determined probability of failure to satisfy RIP. The main consequence of this result is that the STM capacity of the network scales exponentially in the number of nodes when the sparse structure of the inputs is exploited (compared to linear scaling of existing results not exploiting signal structure).

The work described in this abstract also generalizes this basic result in two ways. First, we also consider the case of infinitely long input sequences (when the network has decay properties). In this setting, when one chooses the length of the signal to recover, we establish bounds on the recovery error for the decayed input signal. The main technical approach here treats older input signals as network noise and then applies the earlier recovery bounds established via the RIP. Interestingly, this analysis indicates that a given network has an optimal signal length that should be recovered because it balances errors due to omission and recall mistakes. Second, we also analyze several other network constructions that achieve other desirable properties (e.g., small world networks, block diagonal connectivity matrices, random input weights $\boldsymbol{z}$), showing that these networks also satisfy the RIP conditions in Eqn (2) (sometimes while incurring a reasonable penalty). The analytic results described here are also supported by illustrative simulations showing that the qualitative behavior matches quantitative characteristics of the theoretical guarantees.

REFERENCES

[1] Ganguli S, Sompolinsky H (2010) Short-term memory in neuronal networks through dynamical compressed sensing. *Conference on Neural Information Processing Systems*.
[2] Ganguli S, Huh D, Sompolinsky, H (2008) Memory traces in dynamical systems. *Proceedings of the National Academy of Sciences*.
[3] Jaeger H (2001) Short term memory in echo state networks. GMD Report 152 German National Research Center for Information Technology.
[4] White O L, Lee D D, Sompolinsky H (2004) Short-term memory in orthogonal neural networks. Physical Review Lett. 92:14:148102.
[5] Candes E J (2006) Compressive sampling. Proc. Int. Congr. Mathematicians 3:1433-1452.
[6] Rauhut H (2010) Compressive sensing and structured random matrices. Theoretical Found. and Numerical Methods for Sparse Recovery 9:1-92.