Relating Observability and Compressed Sensing of Time-Varying Signals in Recurrent Linear Networks

MohammadMehdi Kafashan^a, Anirban Nandi^a, ShiNung Ching^{a,b,*}

^a*Department of Electrical and Systems Engineering, Washington University in St. Louis,
One Brookings Drive, Campus Box 1042, MO 63130*

^b*Division of Biology and Biomedical Sciences, Washington University in St. Louis, One
Brookings Drive, Campus Box 1042, MO 63130*

Abstract

In this paper, we study how the dynamics of recurrent networks, formulated as general dynamical systems, mediate the recovery of sparse, time-varying signals. Our formulation resembles the well-described problem of compressed sensing, but in a dynamic setting. We specifically consider the problem of recovering a high-dimensional network input, over time, from observation of only a subset of the network states (i.e., the network output). Our goal is to ascertain how the network dynamics may enable recovery, even if classical methods fail at each time instant. We are particularly interested in understanding performance in scenarios where both the input and output are corrupted by disturbance and noise, respectively. Our main results consist of the development of analytical conditions, including a generalized observability criterion, that ensure exact and stable input recovery in a dynamic, recurrent network setting.

Keywords: Recurrent networks, linear dynamical systems, over-actuated systems, sparse input, l_1 minimization

*Corresponding author

Email addresses: kafashan@ese.wustl.edu (MohammadMehdi Kafashan),
nandia@ese.wustl.edu (Anirban Nandi), shinung@ese.wustl.edu (ShiNung Ching)

1. Introduction

1.1. Background, Motivation and Contributions

The notion of efficient coding dictates that information processing in neuronal networks is optimal according to a cost function that penalizes excessive neuronal activation [1–3]. Given the temporally punctate nature of neuronal spikes, sparsity has been frequently considered within the schema of efficient coding [4, 5]. Notable within the sparse coding framework [6–8] are analyses that leverage the theory of compressed sensing (CS) [9–11] for underdetermined linear systems (e.g., [12–15]). CS outlines conditions under which a linear system admits exact and stable recovery (i.e., decoding) of a sparse input [16–19]. Classical CS, however, does not typically consider temporal dynamics within the recovery problem.

Since networks in the brain are inherently recurrent, interest has also grown around so-called *dynamic CS* and, specifically, on the recovery of signals subjected to transformation via a dynamical system (or, network). In this context, sparsity has been formulated in three ways: 1) In the network states (state sparsity) [20–22]; 2) In the structure/parameters of the network model (model sparsity) [23, 24]; and 3) In the inputs to the network (input sparsity) [25–35]. In this paper, we focus on this latter category of recovery problems. Our motivation is to understand how three stages of a generic network architecture – an afferent stage, a recurrent stage, and an output stage (see Figure 1) – interplay in order to enable an observer, sampling the output, to recover the (sparse) input in the presence of both disturbance and noise. Such an architecture is pervasive in sensory networks in the brain wherein a large number of sensory neurons, receiving excitation from the periphery, impinge on an early recurrent network layer that transforms the afferent excitation en route to higher brain regions [36, 37]. Moreover, beyond neural contexts, understanding network characteristics for dynamic CS may aid in the analysis of systems for efficient processing of naturally sparse time-varying signals [27]; and in the design of resilient cyber-physical systems [25, 26, 28], wherein extrinsic perturbations are sparse and

Figure 1: Schematic of the considered network architecture. We study how the afferent, recurrent and output stages of this architecture interplay in order to enable accurate estimation of the input $u(t)$ from $y(t)$ in the presence of both disturbance and noise.

time-varying.

Specifically, we consider a network that receives an input u_t and generates observations (network outputs), y_t via its recurrent dynamics, i.e.,

$$r_{t+1} = f(r_t, u_t, d_t) \quad (1)$$

$$y_t = g(r_t, e_t) \quad (2)$$

where, here, r_t are the network states, d_t is the disturbance and e_t is the noise. Our focus is on how the network dynamics, embedded in $f(\cdot), g(\cdot)$, impact the extent to which u_t can be inferred from y_t in the case where the dimensionality of the latter is substantially less than that of the former. We will focus exclusively

on the case where these dynamics are linear. Considered at face value, the problem of sparse recovery in a dynamic setting could be broken down into a sequential problem wherein classical CS is performed on the system frozen at each time instant. However, this approach would require the strong condition that the system admits classical CS pointwise. In a dynamic network, recovery may be possible even if the conditions for (static) CS fail at each time instant. For instance, in a linear dynamical system, if the number of observations is smaller than the dimension of the state space, then it is not possible to simply ‘invert’ the observations to recover the states at any particular time. However, the notion of observability provides conditions under which such recovery is possible, as long as those observations are allowed

to accrue through the system dynamics over some period of time.

Here, we postulate, then derive, a set of analytical conditions for dynamic CS that involve observability, but are in general less stringent than necessitating CS at each time instant. Associated with these conditions, we also derive bounds on the l_2 -norm of the reconstruction error over a finite time horizon, which can be defined in terms of both the disturbance and the noise. Salient interpretations of the conditions, such as a basic trade-off between the ability of a network to simultaneously reject input disturbances while still admitting stable recovery, are highlighted and demonstrated via simulation.

1.2. Relationship to Prior Results in Sparse Time-Varying Input Recovery

The sparse input recovery problem for linear dynamical systems can be formulated in both the spatial and temporal dimension. Our contributions are related to the former. As mentioned above, in this context, previous work has considered recovery of spatially sparse inputs for network resilience [25, 26] and encoding of inputs with sparse increments [27]. In [28], conditions for exact sparse recovery are formulated in terms of a coherence-based observability criterion for intended applications in cyber-physical systems. Our contributions herein provide a general set of analytical results, including performance bounds, pertaining to exact and stable sparse input recovery of linear systems in the presence of both noise and disturbance.

A second significant line of research in sparse input recovery problems pertains to the temporal dimension. There, the goal is to understand how a temporally sparse signal (i.e., one that takes the value of zero over a nontrivial portion of its history) can be recovered from the state of the network at a particular instant in time. This problem has been used to describe ‘memory’ in dynamical models of brain networks [29–34]. In particular, in [35] the problem of ascertaining memory is related to CS performed on the network states, over a receding horizon of a scalar-valued input signal. In contrast to these works, we consider spatial sparsity of vector-valued inputs with explicit regard for both disturbance and an overt observation equation, i.e., states not to be directly sampled, but

are transformed to an (in general lower-dimensional) output.

2. Detailed Problem Formulation

We consider a discrete-time linear network, formulated in the typical form of a linear dynamical system, i.e.,:

$$\begin{aligned}\mathbf{r}_{k+1} &= \mathbf{A}\mathbf{r}_k + \mathbf{B}\mathbf{u}_k + \mathbf{d}_k \\ \mathbf{y}_k &= \mathbf{C}\mathbf{r}_k + \mathbf{e}_k,\end{aligned}\tag{3}$$

where k is an integer time index, $\mathbf{r}_k \in \mathbb{R}^n$ is the activity of the network nodes (e.g., in the case of a rate-based neuronal network [38–44], the firing rate of each neuron), $\mathbf{u}_k \in \mathbb{R}^m$ is the extrinsic input, $\mathbf{d}_k \in \mathbb{R}^n$ is the input disturbance, $\mathbf{e}_k \in \mathbb{R}^p$ is the measurement noise independent from $\mathbf{d}_k$, and $\mathbf{y}_k \in \mathbb{R}^p$ is the observation at time k . The matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ describes connections between nodes in the network, $\mathbf{B} \in \mathbb{R}^{n \times m}$ contains weights between input and output and $\mathbf{C} \in \mathbb{R}^{p \times n}$ is the measurement matrix. Such a model is, of course, quite general and can be used to describe recurrent dynamics in neuronal networks [45–49], and machine learning applications such as pattern recognition and data mining [50–53].

We consider the case of bounded disturbance and noise, i.e., $\|\mathbf{e}_k\|_{\ell_2} \leq \epsilon$, $\|\mathbf{d}_k\|_{\ell_2} \leq \epsilon'$. Since m , the number of input nodes, is larger than n , the number of network nodes, $\mathbf{B}$ takes the form of a “wide” matrix. We assume that at each time at most s input nodes are active (s -sparse input), leading to an ℓ_0 constraint to (3) at each time point:

$$\|\mathbf{u}_k\|_{\ell_0} \leq s.\tag{4}$$

In the absence of disturbance and noise, recovering the input of (3) with the ℓ_0 constraint (4) amounts to the optimization problem:

$$\begin{aligned}(P0) \quad & \underset{(\mathbf{r}_k)_{k=0}^K, (\mathbf{u}_k)_{k=0}^{K-1}}{\text{minimize}} && \sum_{k=0}^{K-1} \|\mathbf{u}_k\|_{\ell_0} \\ & \text{subject to} && \mathbf{r}_{k+1} = \mathbf{A}\mathbf{r}_k + \mathbf{B}\mathbf{u}_k \\ & && \mathbf{y}_k = \mathbf{C}\mathbf{r}_k.\end{aligned}\tag{5}$$

It is clear that Problem (P0) is a non-convex discontinuous problem, which is
 90 not numerically feasible and is NP-Hard in general [54]. For static cases, such
 ℓ_0 optimization problems fall into the category of combinatorial optimization
 which require exhaustive search to find the solution [55].

Thus, throughout this paper, we follow the typical relaxation methodology
 used for such problems wherein the ℓ_0 norm is relaxed to the ℓ_1 norm, resulting
 in the problem:

$$\begin{aligned}
 (P1) \quad & \underset{(\mathbf{r}_k)_{k=0}^K, (\mathbf{u}_k)_{k=0}^{K-1}}{\text{minimize}} && \sum_{k=0}^{K-1} \|\mathbf{u}_k\|_{\ell_1} \\
 & \text{subject to} && \mathbf{r}_{k+1} = \mathbf{A}\mathbf{r}_k + \mathbf{B}\mathbf{u}_k \\
 & && \mathbf{y}_k = \mathbf{C}\mathbf{r}_k.
 \end{aligned} \tag{6}$$

In the case that either input disturbance, or measurement noise, or both exist,
 we solve the following convex optimization Problem (P2):

$$\begin{aligned}
 (P2) \quad & \underset{(\mathbf{r}_k)_{k=0}^K, (\mathbf{u}_k)_{k=0}^{K-1}}{\text{minimize}} && \sum_{k=0}^{K-1} \|\mathbf{u}_k\|_{\ell_1} \\
 & \text{subject to} && \mathbf{r}_{k+1} = \mathbf{A}\mathbf{r}_k + \mathbf{B}\mathbf{u}_k \\
 & && \|\mathbf{y}_k - \mathbf{C}\mathbf{r}_k\|_{\ell_2} \leq \epsilon'',
 \end{aligned} \tag{7}$$

where ϵ'' is the ℓ_2 norm of a surrogate parameter that aggregates the effects of
 disturbance and noise. In P2, the objective function is exactly the same as P1.
 95 However, due to noise and disturbance, the last constraints in P2 are quadratic
 inequalities instead of affine constraints in P1. In the case of noisy measurement
 with no disturbance $\epsilon'' = \epsilon$. In the next section, we show conditions for the
 network (3) under which Problems (P1) and (P2) result in exact and stable
 solutions.

100 3. Results

We will begin by presenting a general theory for problems (P1) and (P2)
 for underdetermined linear dynamic networks. We will then briefly show how
 this reduces exactly to the case of sequential application of classical CS when

the system is fully determined. Finally, we present an analysis to characterize
105 the ability of a network to reject input disturbances while still admitting stable
recovery in the presence of disturbance and noise, simultaneously.

3.1. Preliminaries

We begin by recalling some basic matrix notation and matrix norm properties that will be used throughout this paper. Given normed spaces $(\mathbb{R}^{n_1}, \|\cdot\|_{\ell_2})$ and $(\mathbb{R}^{n_2}, \|\cdot\|_{\ell_2})$, the corresponding induced norm or operator norm denoted by $\|\cdot\|_{i,2}$ over linear maps $\mathbf{D} : \mathbb{R}^{n_1} \rightarrow \mathbb{R}^{n_2}$, $\mathbf{D} \in \mathbb{R}^{n_2 \times n_1}$ is defined by

$$\begin{aligned} \|\mathbf{D}\|_{i,2} &= \sup\left\{\frac{\|\mathbf{D}\mathbf{r}\|_{\ell_2}}{\|\mathbf{r}\|_{\ell_2}} \mid \mathbf{r} \in \mathbb{R}^{n_1}\right\} \\ &= \max\{\sqrt{\lambda} \mid \lambda \in \sigma(\mathbf{D}^T \mathbf{D})\}, \end{aligned} \quad (8)$$

where $\sigma(\mathbf{M})$ is the set of eigenvalues of $\mathbf{M}$ (or the spectrum of $\mathbf{M}$).

Definition 1: A vector is said to be s -sparse if $\|\mathbf{c}\|_{\ell_0} \leq s$, in other words it has
110 at most s nonzero entries.

It is well known that in the static case (standard CS), exact and stable recovery of sparse inputs can be obtained under the restricted isometry property (RIP) [16–19, 56], defined as:

Definition 2: The restricted isometry constant δ_s of a matrix $\Phi \in \mathbb{R}^{n \times m}$ is defined as the smallest number such that for all s -sparse vectors $\mathbf{c} \in \mathbb{R}^m$ the following equation holds

$$(1 - \delta_s)\|\mathbf{c}\|_{\ell_2}^2 \leq \|\Phi\mathbf{c}\|_{\ell_2}^2 \leq (1 + \delta_s)\|\mathbf{c}\|_{\ell_2}^2. \quad (9)$$

It is known that many types of random matrices with independent and identically distributed entries or sub-Gaussian matrices satisfy the RIP condition
115 (9) with overwhelming probability [57–59].

3.2. A Generalized Observability Criteria for Time-Varying Sparse Input Recovery

In this case, we consider (3) in the absence of input disturbance ($\mathbf{d}_k = 0$)
120 with the linear map $\mathbf{C} : \mathbb{R}^n \rightarrow \mathbb{R}^p$. We will focus on the case where $p < n$ so

Figure 2: The matrix $\mathbf{B}_k^s$ is the $n \times s$ matrix corresponding the active columns of the full matrix $\mathbf{B}$ at time step k .

that direct inversion of $\mathbf{C}$ is not possible. For any positive K , we define the standard linear observability matrix as

$$\mathbf{O}_K \equiv \begin{pmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^K \end{pmatrix}. \quad (10)$$

If $\text{rank}(\mathbf{O}_K) = n$, then the system (3) is observable in the classical sense¹. Here, we do not assume any knowledge of the input other than the fact that it is s -sparse at each time. Note that if we iterate the output equation in (3) for $K + 1$ time steps and exploit the fact that the input vector is s -sparse as shown in Figure 2, we obtain:

$$\begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_K \end{pmatrix} = \mathbf{O}_K \mathbf{r}_0 + \mathbf{J}_K^s \begin{pmatrix} \mathbf{u}_0^s \\ \mathbf{u}_1^s \\ \vdots \\ \mathbf{u}_{K-1}^s \end{pmatrix}, \quad (11)$$

¹The system is said to be observable if, for any initial state and for any known sequence of input there is a positive integer K such that the initial state can be recovered from the outputs $\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_K$.

where $\mathbf{u}_k^s$ represents s nonzero elements of $\mathbf{u}_k$, and $\mathbf{J}_K^s$ is as follows:

$$\mathbf{J}_K^s = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{CB}_0^s & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{CAB}_0^s & \mathbf{CB}_1^s & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{CA}^{K-1}\mathbf{B}_0^s & \mathbf{CA}^{K-2}\mathbf{B}_1^s & \cdots & \mathbf{CB}_{K-1}^s \end{pmatrix}, \quad (12)$$

where $\mathbf{B}_i^s$ is the $n \times s$ matrix corresponding the active columns of $\mathbf{B}$ (corresponding to nonzero input entries) at time step i (see Figure 2). In general, we do not know where the active columns of $\mathbf{B}$ are located at each time *a priori*. Thus, $\mathcal{J}_K^s$ is defined as the set of all possible matrices satisfying the structure in (12), where the cardinality of this set is $\binom{m}{s}^K$.

We first establish conditions under which a one to one correspondence exists between sparse input and observed output for the system (3).

Lemma 1: Suppose that the sequence $(\mathbf{y}_k)_{k=0}^K$ from noiseless measurements is given, and $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are known. Assume $\text{rank}(\mathbf{O}_K) = n$ and the matrix $\mathbf{CB}$ satisfies the RIP condition (9) with isometry constant $\delta_{2s} < 1$. Further, assume

$$\text{rank}([\mathbf{O}_K \quad \mathbf{J}_K^{2s}]) = n + \text{rank}(\mathbf{J}_K^{2s}), \quad \forall \mathbf{J}_K^{2s} \in \mathcal{J}_K^{2s}. \quad (13)$$

Then, there is a unique s -sparse sequence of $(\mathbf{u}_k)_{k=0}^{K-1}$ and a unique sequence of $(\mathbf{r}_k)_{k=0}^K$ that generate $(\mathbf{y}_k)_{k=0}^K$.

Proof. See Appendix A □

The rank condition (13) implies that all columns of the observability matrix must be linearly independent of each other (i.e., the network is observable in the classical sense) and of all columns of $\mathbf{J}_K^s$. Since the exact location of the nonzero elements of the input vector are not known *a priori*, this condition is specified over all matrices in $\mathcal{J}_K^{2s}$. Thus, (13) is a combinatorial condition. From our simulation studies, we have observed that this condition holds for random Gaussian matrices almost always (see also Examples 1 and 2 in Section 4).

Having established the existence of a unique solution, we now proceed to study the convex problems (P1) and (P2) that recover these solutions for this case. First, we provide theoretical results for the stable recovery of the input where measurements are noisy i.e., (P2).

Theorem 1: (Noisy recovery) Assume $\text{rank}(\mathbf{O}_K) = n$ and the matrix $\mathbf{CB}$ satisfies the RIP condition (9) with $\delta_{2s} < \sqrt{2} - 1$. assume (13) holds. Suppose that the sequence $(\mathbf{y}_k)_{k=0}^K$ is given and generated from sequences $(\bar{\mathbf{r}}_k)_{k=0}^K$ and s -sparse $(\bar{\mathbf{u}}_k)_{k=0}^{K-1}$ based on

$$\begin{aligned}\bar{\mathbf{r}}_{k+1} &= \mathbf{A}\bar{\mathbf{r}}_k + \mathbf{B}\bar{\mathbf{u}}_k, \quad k = 0, \dots, K-1 \\ \mathbf{y}_k &= \mathbf{C}\bar{\mathbf{r}}_k + \mathbf{e}_k, \quad k = 0, \dots, K,\end{aligned}\tag{14}$$

where $(\|\mathbf{e}_k\|_{\ell_2} \leq \epsilon)_{k=0}^K$ and $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are known. Then, the solution to Problem (P2) obeys

$$\sum_{k=0}^{K-1} \|\mathbf{u}_k^* - \bar{\mathbf{u}}_k\|_{\ell_2} \leq C_s \epsilon,\tag{15}$$

where

$$C_s = 2\alpha C_0 K(1 - \rho)^{-1}.\tag{16}$$

C_0 , ρ , α are given explicitly below:

$$\begin{aligned}C_0 &= \left(1 + \sqrt{\frac{K \sigma_{\max}(\mathbf{C}^T \mathbf{C}) \sigma_{\max}(\mathbf{A}^T \mathbf{A}) \sigma_{\max}\left((\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})^T (\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})\right)}{\sigma_{\min}(\mathbf{O}_K^T \mathbf{O}_K) \sigma_{\min}\left((\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})^T (\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})\right)}}}\right), \\ \alpha &= \frac{2\sqrt{1 + \delta_{2s}}}{1 - \delta_{2s}}, \\ \rho &= \frac{\sqrt{2}\delta_{2s}}{1 - \delta_{2s}}, \\ \mathbf{P}_{\mathbf{J}_K^{2s}} &= \mathbf{J}_K^{2s} (\mathbf{J}_K^{2sT} \mathbf{J}_K^{2s})^{-1} \mathbf{J}_K^{2sT}.\end{aligned}\tag{17}$$

Proof. Assume that the the sequences $(\mathbf{r}_k^*)_{k=0}^K$ and sparse $(\mathbf{u}_k^*)_{k=0}^{K-1}$ are the solutions of Problem (P2). First we derive the bound for the $\|\mathbf{r}_k^* - \bar{\mathbf{r}}_k\|_{\ell_2}$ in the following Lemma.

Lemma 2: Suppose that the sequence $(\mathbf{y}_k)_{k=0}^K$ is given and generated from sequences $(\bar{\mathbf{r}}_k)_{k=0}^K$ and s -sparse $(\bar{\mathbf{u}}_k)_{k=0}^{K-1}$ based on (14), where $(\|\mathbf{e}_k\|_{\ell_2} \leq \epsilon)_{k=0}^K$ and $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are known. Then, any solution $\mathbf{r}_k^*$ to Problem (P2) obeys

$$\|\mathbf{r}_k^* - \bar{\mathbf{r}}_k\|_{\ell_2} \leq 2\epsilon \sqrt{\frac{K\sigma_{\max}\left((\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})^T(\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})\right)}{\sigma_{\min}\left(\mathbf{O}_K^T \mathbf{O}_K\right) \sigma_{\min}\left((\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})^T(\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})\right)}}} \quad (18)$$

Proof. See Appendix B. $\square$

From Lemma 2, and the equation $\mathbf{y}_k = \mathbf{CA}\bar{\mathbf{r}}_{k-1} + \mathbf{CB}\bar{\mathbf{u}}_{k-1} + \mathbf{e}_k$ we can derive a bound for $\|\mathbf{CB}(\mathbf{u}_k^* - \bar{\mathbf{u}}_k)\|_{\ell_2}$ as

$$\begin{aligned} \|\mathbf{CB}(\mathbf{u}_k^* - \bar{\mathbf{u}}_k)\|_{\ell_2} &= \|(\mathbf{e}_{k+1}^* + \mathbf{e}_{k+1}) + \mathbf{CA}(\bar{\mathbf{r}}_k - \mathbf{r}_k^*)\|_{\ell_2} \\ &\leq \|\mathbf{e}_{k+1} + \mathbf{e}_{k+1}^*\|_{\ell_2} + \|\mathbf{CA}(\bar{\mathbf{r}}_k - \mathbf{r}_k^*)\|_{\ell_2} \\ &\leq 2\epsilon \left(1 + \sqrt{\frac{K\sigma_{\max}(\mathbf{C}^T \mathbf{C}) \sigma_{\max}(\mathbf{A}^T \mathbf{A})}{\sigma_{\min}(\mathbf{O}_K^T \mathbf{O}_K) \sigma_{\min}\left((\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})^T(\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})\right)}}} \right) \\ &= 2C_0\epsilon. \end{aligned} \quad (19)$$

Now, denote $\mathbf{u}_k^* = \bar{\mathbf{u}}_k + \mathbf{h}_k$ where $\mathbf{h}_k$ can be decomposed into a sum of vectors $\mathbf{h}_{k,T_0(k)}, \mathbf{h}_{k,T_1(k)}, \mathbf{h}_{k,T_2(k)}, \dots$ for each k , each of sparsity at most s . Here, $T_0(k)$ corresponds to the location of non-zero elements of $\bar{\mathbf{u}}_k$, $T_1(k)$ to the location of s largest coefficients of $\mathbf{h}_{k,T_0^c(k)}$, $T_2(k)$ to the location of the next s largest coefficients of $\mathbf{h}_{k,T_0^c(k)}$, and so on. Also, let $T_{01}(k) \equiv T_0(k) \cup T_1(k)$. Extending the technique in [19, 56], it is possible to obtain a cone constraint for linear dynamical systems.

Lemma 3: (Cone constraint) The optimal solution for the input in Problem (P2) satisfies

$$\sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_{01}^c(1)}\|_{\ell_2} \leq \sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_0(1)}\|_{\ell_2}. \quad (20)$$

Proof. See Appendix C. $\square$

We can further establish a bound for the right hand side of (20):

Lemma 4: The optimal solution for the input in Problem (P2) satisfies the following constraint

$$\sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_{01}(1)}\|_{\ell_2} \leq K(1-\rho)^{-1}\alpha C_0\epsilon. \quad (21)$$

160 *Proof.* See Appendix D. □

Finally, based on Lemma 3 and Lemma 4, it is easy to see that

$$\begin{aligned} \sum_{k=0}^{K-1} \|\mathbf{h}_k\|_{\ell_2} &\leq \sum_{k=0}^{K-1} (\|\mathbf{h}_{k,T_{01}(1)}\|_{\ell_2} + \|\mathbf{h}_{k,T_{01}^c(1)}\|_{\ell_2}) \\ &\leq 2 \sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_{01}(1)}\|_{\ell_2} \\ &\leq 2\alpha C_0 K(1-\rho)^{-1}\epsilon = C_s\epsilon. \end{aligned} \quad (22)$$

□

Theorem 2: (Noiseless recovery) Assume $\text{rank}(\mathbf{O}_K) = n$ and the matrix $\mathbf{CB}$ satisfies the RIP condition (9) with $\delta_{2s} < \sqrt{2} - 1$. Further, assume (13) holds. Suppose that the sequence $(\mathbf{y}_k)_{k=0}^K$ is given and generated from sequences $(\bar{\mathbf{r}}_k)_{k=0}^K$ and s -sparse inputs $(\bar{\mathbf{u}}_k)_{k=0}^{K-1}$ based on dynamical equation (14), where $\epsilon = 0$ and $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are known. Then the sequences $(\bar{\mathbf{r}}_k)_{k=0}^K$ and $(\bar{\mathbf{u}}_k)_{k=0}^{K-1}$ are the unique minimizer to Problem (P1).
165

Proof. It can be concluded from Lemma 1 and Theorem 1 that with the assumption stated in the theorem and $\epsilon = 0$ the sequences $(\bar{\mathbf{r}}_k)_{k=0}^K$ and $(\bar{\mathbf{u}}_k)_{k=0}^{K-1}$ are the unique minimizer to Problem (P1).
170 □

The results in the above Theorems hinge on an RIP criteria involving the combined matrix $\mathbf{CB}$, which can be interpreted in the context of an overcomplete dictionary in the classical compressed sensing literature [60, 61]. In this sense, the $\mathbf{B}$ matrix (i.e., the connectivity of the afferent stage in Figure 1) can be interpreted as a dictionary that transforms the sparse input $\mathbf{u}$ onto the
175

recurrent network states. The output matrix $\mathbf{C}$ enabled ‘readout’ of the states. Thus, the properties of $\mathbf{C}$ and $\mathbf{B}$ individually do not have direct bearing on dynamic CS. However, since these matrices dictate the signal flow to and from the recurrent stage, dynamic CS depends on whether these two matrices are
180 matched in the sense of satisfying the RIP condition presented above.

3.3. Proper Reduction to Sequential CS

In the previous section, we focused on the case where the matrix $\mathbf{C}$ precludes direct inversion of the system states from the observations. It is worth noting that the results remain valid even without this assumption and, in fact, reduce
185 to the case of sequential application of classical compressed sensing.

This reduction is most easily seen by assuming that $\mathbf{C}$ is square and full rank, so that the network state can be obtained directly via inversion at each instant in time. It is intuitive that in this case, the cost function in (7) is minimized by application of standard CS at each time instant, depending exclusively on the
190 RIP properties of the matrix $\mathbf{B}$. Specifically, under the assumption that $\mathbf{C}$ is full rank, Theorem 1 reduces exactly to:

Theorem 3: Assume that the matrix $\mathbf{B}$ satisfies the RIP condition (9) with $\delta_{2s} < \sqrt{2} - 1$. Suppose that the sequence $(\mathbf{y}_k)_{k=0}^K$ is given and generated from sequences $(\tilde{\mathbf{r}}_k)_{k=0}^K$ and s -sparse $(\tilde{\mathbf{u}}_k)_{k=0}^{K-1}$ based on (14), where $(\|\mathbf{e}_k\|_{\ell_2} \leq \epsilon)_{k=0}^K$ and $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathcal{N}(\mathbf{C}) = \{0\}$ are known. Then, the solution to Problem (P2) obeys

$$\sum_{k=0}^{K-1} \|\mathbf{u}_k^* - \tilde{\mathbf{u}}_k\|_{\ell_2} \leq C_s \epsilon, \quad (23)$$

where

$$C_s = 2\alpha C_0 K(1 - \rho)^{-1}. \quad (24)$$

C_0 , ρ , α are given explicitly below:

$$\begin{aligned} C_0 &= \frac{1}{\sqrt{\sigma}} \left(1 + \sqrt{\frac{\sigma_{\max}(\mathbf{C}^T \mathbf{C}) \sigma_{\max}(\mathbf{A}^T \mathbf{A})}{\sigma_{\min}(\mathbf{C}^T \mathbf{C})}} \right), \\ \alpha &= \frac{2\sqrt{1 + \delta_{2s}}}{1 - \delta_{2s}}, \\ \rho &= \frac{\sqrt{2}\delta_{2s}}{1 - \delta_{2s}}, \\ \sigma_{\min}(\mathbf{C}^T \mathbf{C}) &< \sigma < \sigma_{\max}(\mathbf{C}^T \mathbf{C}). \end{aligned} \tag{25}$$

Proof. This Theorem is a special case of Theorem 1 and the proof can be obtained readily by following the proof of Theorem 1. $\square$

An analogous reduction for exact recovery in the noiseless case ($\epsilon = 0$) can also be made.

3.4. Recovery in the Presence of Disturbance and Noise

Finally, we show how eigenstructure of the network implies a fundamental trade-off between stable recovery and rejection of disturbance (i.e., corruption).

Here, for ease of presentation, we will use of the results of Theorem 3, though the concepts generalize. It is easy to see from (23) that the upper-bound of the recovery error is reduced by decreasing the maximum singular value of $\mathbf{A}$. Thus, we can use the upper-bound of the input recovery error as a comparative measure of performance. In the absence of both disturbance and noise, the best error performance is achieved when $\mathbf{A} = \mathbf{0}$, i.e., the network is static, which is intuitive since in this scenario any temporal effects would smear potentially salient parts of the signal.

On the other hand, having dynamics in the network should improve the error performance in the presence of the disturbance. To demonstrate this, consider (3) with $\mathbf{d}_k$ nonzero. When $\mathbf{A} = \mathbf{0}$, i.e., a static network, the disturbance can be exactly transformed to the measurement equation resulting in $\mathbf{C}\mathbf{d}_k + \mathbf{e}_k$ as a surrogate measurement noise with

$$\begin{aligned} \|\mathbf{C}\mathbf{d}_k + \mathbf{e}_k\|_{\ell_2} &\leq \sqrt{\sigma'}\epsilon' + \epsilon, \\ \sigma_{\min}(\mathbf{C}^T \mathbf{C}) &< \sigma' < \sigma_{\max}(\mathbf{C}^T \mathbf{C}). \end{aligned} \tag{26}$$

In this case, the error upper-bound can be obtained by exploiting the result of Theorem 3 as

$$\begin{aligned} \sum_{k=0}^{K-1} \|\mathbf{u}_k^* - \bar{\mathbf{u}}_k\|_{\ell_2} &\leq C'_s(\sqrt{\sigma'}\epsilon' + \epsilon), \\ C'_s &= \frac{2}{\sqrt{\sigma}}\alpha K(1-\rho)^{-1}. \end{aligned} \quad (27)$$

When $\mathbf{A}$ is nonzero, it is not possible to exactly map the disturbance to the output as above. Nevertheless, it is straightforward to approximate the relative improvement in performance. For instance, consider a system with a symmetric $\mathbf{A}$ where the disturbance and input are in displaced frequency bands. Then it is a direct consequence of linear filtering that the power spectral density of the disturbance can be attenuated according to

$$\mathcal{S}_{\mathbf{d}^{\text{Filt}}}(e^{j\omega}) = s_{\mathbf{d}}(e^{j\omega})(e^{j\omega}\mathbf{I}_n - \mathbf{A})^{-1}(e^{-j\omega}\mathbf{I}_n - \mathbf{A})^{-1}, \quad (28)$$

where ω is the frequency of the disturbance. So, for instance, if $\omega = \pi$,

$$\begin{aligned} \text{Tr}\{\mathcal{S}_{\mathbf{d}^{\text{Filt}}}(e^{j\pi})\} &= \text{Tr}\{s_{\mathbf{d}}(e^{j\pi})(\mathbf{I}_n + \mathbf{A})^{-2}\} \\ &= s_{\mathbf{d}}(e^{j\pi}) \sum_{i=1}^n (1 + \lambda_i(\mathbf{A}))^{-2}, \end{aligned} \quad (29)$$

where $\lambda_i(\mathbf{A})$ is the i^{th} eigenvalue of matrix $\mathbf{A}$. Assuming the input is sufficiently displaced in frequency from the disturbance, the error upper-bound can be then readily approximated using the results of Theorem 3 as follows

$$\begin{aligned} \sum_{k=0}^{K-1} \|\mathbf{u}_k^* - \bar{\mathbf{u}}_k\|_{\ell_2} &\leq C_s \left(\frac{\sqrt{\sigma''}}{n} \sum_{i=1}^n (1 + \lambda_i(\mathbf{A}))^{-2} \epsilon' + \epsilon \right), \\ \sigma_{\min}(\mathbf{C}^T \mathbf{C}) &< \sigma'' < \sigma_{\max}(\mathbf{C}^T \mathbf{C}). \end{aligned} \quad (30)$$

By comparing (27) and (30), it is evident that the presence of dynamics in $\mathbf{A}$ can reduce the error upper-bound by at least a factor of two, assuming σ' and σ'' are close to each other. In the examples below, we will show that, in fact, performance in many cases can exceed this bound considerably.

4. Examples

In this section, we present two brief examples that demonstrate the key aspects of the results. For solving our convex optimization problems, we used

Figure 3: The recovered input for (A) $p = 35$ and (B) $p = 45$ for both static (middle images) and dynamic (left images) CS where $n = 45$ and $m = 68$. Original input is in the right hand side denoted as true digit.

CVX with a MATLAB interface [62, 63]. To speed up the computations, alternative ℓ_1 minimization algorithms could potentially be utilized [64, 65], though we do not explore these herein.

4.1. Example 1: Recovery in Recurrent Networks

In this example, we construct a recurrent network of the form in Figure 1, which satisfies conditions in Theorem 1. Specifically, we consider a random network so that the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are constructed with entries that are drawn from an independent, identically distributed normal distribution. The dimensionality considered is $n = 45$, $m = 68$, $p = 35$.

The sparse input is defined as the image of a digit, shown in Figure 3 where each pixel takes the value of 0 or 1. Here, the horizontal axis is treated as time, i.e., column k of the image is the input to the system at time k with $K = 68$.

Since $p < n$, it is clear that sequential application of static CS is not possible in this case (i.e., since $\mathbf{C}$ cannot be inverted). However, for this network setup, we verify numerically that the conditions of Theorem 2 hold generically, and thus recovery of the sparse input is expected after K time steps by solving (P1) in one step over the entire time horizon, i.e., one-step dynamic CS. Figure 3A confirms this result, wherein exact recovery of the input is shown. Moreover, we confirm that sequential application of CS does not lead to correct recovery.

Figure 4: (A) MSE versus the maximum singular value of $\mathbf{A}$, for several random realization of $\mathbf{A}$ with noise and in the absence of disturbance. (B) MSE versus the maximum singular value of $\mathbf{A}$, for several random diagonal $\mathbf{A}$ with noise and disturbance.

As a point of comparison, we also considered the case wherein $\mathbf{C}$ is full rank (i.e., $p = 45$) (Figure 3B), in which case sequential application of CS does indeed lead to recovery.

4.2. Example 2: Recovery in the Presence of Disturbance and Noise

In this example we highlight the competing influence of disturbance and noise in a recurrent setting, and specifically our derived error bounds. We construct networks of the same dimension as in Example 1, but for simplicity, now assume that $\mathbf{C}$ is full rank.

We begin by considering the effects of noise only, wherein $\mathbf{e}_k$ is assumed to follow an uniform distribution $\mathcal{U}(-0.5, 0.5)$ while $\mathbf{d}_k = \mathbf{0}$. Figure 4A shows the mean square error (MSE) versus the maximum singular value of $\mathbf{A}$, for several random realization of $\mathbf{A}$. It can be seen from this figure that by increasing $\sqrt{\sigma_{\max}(\mathbf{A}^T \mathbf{A})}$, the recovery performance is degraded, as we expect based on the derived bound for the error in (23).

However, based on our discussion in Section 3.4, we also know that dynamics embedded in $\mathbf{A}$ are beneficial for the rejection of disturbances. Thus, to contrast Figure 4A, we consider the case when disturbance is added to the input. In Figure 4B, we show the MSE versus $\sqrt{\sigma_{\max}(\mathbf{A}^T \mathbf{A})}$ for several random $\mathbf{A}$ when $\mathbf{e}_k \sim \mathcal{U}(-0.5, 0.5)$ and $\mathbf{d}_k \sim \mathcal{N}(0, 1)$. It is seen that increasing $\sqrt{\sigma_{\max}(\mathbf{A}^T \mathbf{A})}$ allows disturbance rejection and improves performance.

Figure 5: (A) Four noiseless frames of a movie. Recovery via (B) static CS and (C) dynamic CS in the presence of disturbance.

Figure 6: PSNR of the recovered frame versus frame number for both CS with and without dynamics.

Taken together, Figure 4 emphasizes the basic trade-off in recovery performance due to noise and disturbance. Thus, in the presence of both sources of corruption, an effective network must exhibit a relatively balanced spectrum.

To provide an intuitive visual demonstration of this principle, we construct an input consisting of a smoothly changing sequence of 64 frames (images). Each frame, is corrupted with disturbance at each time, and the difference between two consecutive frames is considered as the sparse input to the network. The

Figure 7: MSE as a function of (A) $\log(1/\epsilon)$ and (B) $\log(1/\epsilon')$ for the reconstructed input with $n = 50$, $m = 100$, $p = 40$, evaluated over 10 time steps for 100 random trials (different random matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ in each trial).

260 disturbance $\mathbf{d}_k$ is assumed to be a random variable drawn from a Gaussian distribution, $\mathcal{N}(0, 0.2)$, passed through a fifth-order Chebyshev high pass filter. Each frame has $m = 400$ pixels and $K = 64$. This input is passed through a random network of dimension $n = p = 200$, i.e. random Gaussian matrices for $\mathbf{B}$ and $\mathbf{C}$.

265 We designed the matrix $\mathbf{A}$ to balance the performance bound (23) and the ability to reject the disturbance as per Section 3.4. Figure 5A shows the original frames at different times. We assumed that the first frame is known exactly. Frames recovered from the output of a static network, i.e., $\mathbf{A} = \mathbf{0}$ are depicted in Figure 5B. In contrast, frames recovered from the output of the designed dynamic network are shown in Figure 5C. It is clear from the figure that the quality of recovery is better in the latter case. Figure 6 illustrates the PSNR, defined as $10 \log(\frac{1}{MSE})$ as a function of frame number with and without dynamics. It can be concluded from this figure that having recurrent dynamics embedded in $\mathbf{A}$ results in recovery that is more robust to disturbance, while 270 without dynamics, error propagates over time, and the reconstruction quality is degraded.

We conducted simulation experiments to examine the effect of the noise and disturbance strength on the reconstruction error when $\mathbf{C}$ is not full. Figure 7 shows the average MSE for the reconstructed input versus $\log(1/\epsilon)$ and

280 $\log(1/\epsilon')$, respectively with $n = 50$, $m = 100$, $p = 40$, evaluated over 10 time steps for 100 random trials (different random matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ in each trial) with 4-sparse and 10-sparse inputs. This figure shows that the reconstruction error decreases as a function of noise energy.

5. Discussion

285 In this paper, we presented analytical conditions that are sufficient to ensure the recovery of time-varying signals processed via a recurrent linear network, wherein the network output is allowed to accrue prior to recovery. The conditions amount to a generalized observability criteria and are less stringent than the approach of performing sparse recovery pointwise, i.e., at each instant in time. Further, we derived several bounds in terms of the system dynamics for recovery performance in the presence of both input disturbance and observation noise. We showed that dynamics can play opposing roles in mediating accurate recovery with respect to these two different sources of corruption.

Potential Applicability: The analysis developed herein is suited to the treatment of multivariate time-series that are modeled as the output of a network with linear dynamics. The key results are specific to contexts where extrinsic perturbations to such networks are sparse and time-varying. Both engineering and scientific impacts of these results can be postulated. In the former case, the goal may be to *design* networks to process time-varying signals that are naturally sparse, such as high-dimensional neural data. In the latter (scientific) case, the goal is to understand how the naturally occurring architectures of networks, such as those in the brain, confer advantages for processing of afferent signals. Such a paradigm may be useful in the theoretical formulation of sensory neural coding mechanisms [15]. Other application domains such as video processing [27], medical imaging [66], and resilient cyber-physical systems [25, 26] may also be considered. In either the scientific or application-driven cases, a precursor to further study are a set of verifiable conditions that overtly link network characteristics/dynamics to sparse input processing, i.e., the main

thrust of the present work.

310 *Assumptions and Limitations:* A key assumption in instantiating our analysis results is that the network dynamical matrices ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$) are known and/or specified. As well, reconstruction of inputs requires knowledge of noise bounds (i.e., upper-bounds on the magnitude of both disturbance, ϵ , and noise, ϵ' , as per Section 3). In practice, several data-driven algorithms, such as expectation
315 maximization [67, 68], could be used in the estimation of such bounds, though we did not pursue this here. Importantly, there is no need to know the actual noise time-series in absolute terms, which would have restricted functionality. Treatment of systems with nonlinear dynamics, including a more detailed examination of networks with prescribed topologies (using the developed theory),
320 are left as subjects for future work.

Appendix A. Proof of Lemma 1

We start the proof using contradiction. Let us assume that the sequence of $(\mathbf{u}_k)_{k=0}^{K-1}$ and $(\mathbf{r}_k)_{k=0}^K$ is not unique and there is another sequence of s -sparse $(\hat{\mathbf{u}}_k)_{k=0}^{K-1}$ and $(\hat{\mathbf{r}}_k)_{k=0}^K$ which satisfies the system (3) with noiseless measurements.
325 Note that $\mathbf{u}_k - \hat{\mathbf{u}}_k$ has at most $2s$ nonzero elements. Similar to the depiction in Figure 2, we can rewrite (11) based on $2s$ columns of $\mathbf{B}$ corresponding to $2s$ active non-zero elements of $\mathbf{u}_k - \hat{\mathbf{u}}_k$ as

$$\begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_K \end{pmatrix} = \mathbf{O}_K \mathbf{r}_0 + \mathbf{J}_K^{2s} \begin{pmatrix} \mathbf{u}_0^{2s} \\ \mathbf{u}_1^{2s} \\ \vdots \\ \mathbf{u}_{K-1}^{2s} \end{pmatrix}, \quad (\text{A.1})$$

$$\begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_K \end{pmatrix} = \mathbf{O}_K \hat{\mathbf{r}}_0 + \mathbf{J}_K^{2s} \begin{pmatrix} \hat{\mathbf{u}}_0^{2s} \\ \hat{\mathbf{u}}_1^{2s} \\ \vdots \\ \hat{\mathbf{u}}_{K-1}^{2s} \end{pmatrix}. \quad (\text{A.2})$$

By subtracting the above equations from each other we have

$$\mathbf{O}_K(\mathbf{r}_0 - \hat{\mathbf{r}}_0) + \mathbf{J}_K^{2s} \begin{pmatrix} \mathbf{u}_0^{2s} - \hat{\mathbf{u}}_0^{2s} \\ \mathbf{u}_1^{2s} - \hat{\mathbf{u}}_1^{2s} \\ \vdots \\ \mathbf{u}_{K-1}^{2s} - \hat{\mathbf{u}}_{K-1}^{2s} \end{pmatrix} = \mathbf{0}. \quad (\text{A.3})$$

Based on assumptions that $\text{rank}(\mathbf{O}_K) = n$ and $\text{rank}([\mathbf{O}_K \ \mathbf{J}_K^{2s}]) = n + \text{rank}(\mathbf{J}_K^{2s})$, $\forall \mathbf{J}_K^{2s} \in \mathcal{J}_K^{2s}$, all columns of the observability matrix must be linearly independent of each other, and of all columns of the $\mathbf{J}_K^{2s}$ matrix. Hence, the vector $\mathbf{r}_0 - \hat{\mathbf{r}}_0 = \mathbf{0}$. Having $\mathbf{r}_0 = \hat{\mathbf{r}}_0$ and the matrix $\mathbf{CB}$ satisfying the RIP condition (9) with isometry constant $\delta_{2s} < 1$, it is easy to see that $\mathbf{u}_k = \hat{\mathbf{u}}_k$ and therefore there exists unique state and s -sparse input sequences.

Appendix B. Proof of Lemma 2

Lets assume that the the sequences $(\mathbf{r}_k^*)_{k=0}^K$ and s -sparse $(\mathbf{u}_k^*)_{k=0}^{K-1}$ are the solutions of Problem (P2). In this case (A.3) can be rewritten as

$$\mathbf{O}_K(\mathbf{r}_0^* - \bar{\mathbf{r}}_0) + \mathbf{J}_K^{2s} \begin{pmatrix} \mathbf{u}_0^{*2s} - \bar{\mathbf{u}}_0^{2s} \\ \mathbf{u}_1^{*2s} - \bar{\mathbf{u}}_1^{2s} \\ \vdots \\ \mathbf{u}_{K-1}^{*2s} - \bar{\mathbf{u}}_{K-1}^{2s} \end{pmatrix} = \begin{pmatrix} \mathbf{e}_0 \\ \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_K \end{pmatrix} + \begin{pmatrix} \mathbf{e}_0^* \\ \mathbf{e}_1^* \\ \vdots \\ \mathbf{e}_K^* \end{pmatrix}, \quad (\text{B.1})$$

where $\|\mathbf{e}_k^*\|_{\ell_2} \leq \epsilon$. Based on assumptions that $\text{rank}(\mathbf{O}_K) = n$ and $\text{rank}([\mathbf{O}_K \ \mathbf{J}_K^{2s}]) = n + \text{rank}(\mathbf{J}_K^{2s})$, $\forall \mathbf{J}_K^{2s} \in \mathcal{J}_K^{2s}$ we can project the above equation using the projection $(\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})$ where $\mathbf{P}_{\mathbf{J}_K^{2s}} = \mathbf{J}_K^{2s}(\mathbf{J}_K^{2sT} \mathbf{J}_K^{2s})^{-1} \mathbf{J}_K^{2sT}$. It is straightforward to verify that $(\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}}) \mathbf{J}_K^{2s} = \mathbf{0}$ and therefore we have

$$(\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}}) \mathbf{O}_K(\mathbf{r}_0^* - \bar{\mathbf{r}}_0) = (\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}}) \begin{pmatrix} \mathbf{e}_0 \\ \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_K \end{pmatrix} + (\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}}) \begin{pmatrix} \mathbf{e}_0^* \\ \mathbf{e}_1^* \\ \vdots \\ \mathbf{e}_K^* \end{pmatrix}. \quad (\text{B.2})$$

Finally, we can derive the error bound for the state error from (B.2) as

$$\|\mathbf{r}_0^* - \bar{\mathbf{r}}_0\|_{\ell_2} \leq 2\epsilon \sqrt{\frac{K\sigma_{max}\left((\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})^T(\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})\right)}{\sigma_{min}\left(\mathbf{O}_K^T \mathbf{O}_K\right)\sigma_{min}\left((\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})^T(\mathbf{I} - \mathbf{P}_{\mathbf{J}_K^{2s}})\right)}} \quad (\text{B.3})$$

335 In a similar fashion we have an exact same bound for the state error at a different time k .

Appendix C. Proof of Lemma 3

For each $j \geq 2$ and $k = 0, \dots, K-1$ we have

$$\|\mathbf{h}_{k,T_j(k)}\|_{\ell_2} \leq s^{1/2}\|\mathbf{h}_{k,T_j(k)}\|_{\ell_\infty} \leq s^{-1/2}\|\mathbf{h}_{k,T_{j-1}(k)}\|_{\ell_1}, \quad (\text{C.1})$$

and thus

$$\begin{aligned} \sum_{j \geq 2} \|\mathbf{h}_{k,T_j(k)}\|_{\ell_2} &\leq s^{-1/2}(\|\mathbf{h}_{k,T_1(k)}\|_{\ell_1} + \|\mathbf{h}_{k,T_2(k)}\|_{\ell_1} + \dots) \\ &\leq s^{-1/2}\|\mathbf{h}_{k,T_0^c(k)}\|_{\ell_1}. \end{aligned} \quad (\text{C.2})$$

Therefore, we have the following equation

$$\begin{aligned} \|\mathbf{h}_{k,T_{01}^c(k)}\|_{\ell_2} &= \left\| \sum_{j \geq 2} \mathbf{h}_{k,T_j(k)} \right\|_{\ell_2} \leq \sum_{j \geq 2} \|\mathbf{h}_{k,T_j(k)}\|_{\ell_2} \\ &\leq s^{-1/2}\|\mathbf{h}_{k,T_0^c(k)}\|_{\ell_1}. \end{aligned} \quad (\text{C.3})$$

Since $(\mathbf{u}^*)_k$ minimizes the cost function in Problem (P2),

$$\begin{aligned} \sum_{k=0}^{K-1} \|\bar{\mathbf{u}}_k\|_{\ell_1} &\geq \sum_{k=0}^{K-1} \|\mathbf{u}_k^*\|_{\ell_1} = \sum_{k=0}^{K-1} \|\bar{\mathbf{u}}_k + \mathbf{h}_k\|_{\ell_1} \\ &= \sum_{k=0}^{K-1} \left(\sum_{i \in T_0(k)} |\bar{\mathbf{u}}_{k,i} + \mathbf{h}_{k,i}| + \sum_{i \in T_0^c(k)} |\bar{\mathbf{u}}_{k,i} + \mathbf{h}_{k,i}| \right) \\ &\geq \sum_{k=0}^{K-1} (\|\bar{\mathbf{u}}_{k,T_0(k)}\|_{\ell_1} - \|\mathbf{h}_{k,T_0(k)}\|_{\ell_1} + \|\mathbf{h}_{k,T_0^c(k)}\|_{\ell_1} \\ &\quad + \|\bar{\mathbf{u}}_{k,T_0^c(k)}\|_{\ell_1}) \end{aligned} \quad (\text{C.4})$$

$\bar{\mathbf{u}}_0, \dots, \bar{\mathbf{u}}_{K-1}$ are non-zero for $T_0(0), \dots, T_0(K-1)$, respectively. Therefore,

$$\|\bar{\mathbf{u}}_{0,T_0^c(0)}\|_{\ell_1} = \dots = \|\bar{\mathbf{u}}_{K-1,T_0^c(K-1)}\|_{\ell_1} = 0 \quad (\text{C.5})$$

which gives

$$\sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_0^c(k)}\|_{\ell_1} \leq \sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_0(k)}\|_{\ell_1}. \quad (\text{C.6})$$

Considering

$$\|\mathbf{h}_{k,T_0(k)}\|_{\ell_1} \leq s^{1/2} \|\mathbf{h}_{k,T_0(k)}\|_{\ell_2}, \quad (\text{C.7})$$

and substituting it into (C.3) and (C.6) we have

$$\sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_{01}^c(k)}\|_{\ell_2} \leq \sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_0(k)}\|_{\ell_2}. \quad (\text{C.8})$$

Appendix D. Proof of Lemma 4

To find the bound for $\sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2}$, we start with

$$\mathbf{CBh}_k = \mathbf{CBh}_{k,T_{01}(k)} + \sum_{j \geq 2} \mathbf{CBh}_{k,T_j(k)}, \quad (\text{D.1})$$

which gives

$$\|\mathbf{CBh}_{k,T_{01}(k)}\|_{\ell_2}^2 = \langle \mathbf{CBh}_{k,T_{01}(k)}, \mathbf{CBh}_k \rangle - \langle \mathbf{CBh}_{k,T_{01}(k)}, \sum_{j \geq 2} \mathbf{CBh}_{k,T_j(k)} \rangle. \quad (\text{D.2})$$

From (19) and the RIP condition for $\mathbf{CB}$,

$$\begin{aligned} |\langle \mathbf{CBh}_{k,T_{01}(k)}, \mathbf{CBh}_k \rangle| &\leq \|\mathbf{CBh}_{k,T_{01}(k)}\|_{\ell_2} \|\mathbf{CBh}_k\|_{\ell_2} \\ &\leq 2\epsilon C_0 \sqrt{1 + \delta_{2s}} \|\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2}, \end{aligned} \quad (\text{D.3})$$

and, moreover, application of the parallelogram identity for disjoint subsets $T_0(k)$ and $T_j(k), j \geq 2$ results in

$$|\langle \mathbf{CBh}_{k,T_0(k)}, \mathbf{CBh}_{k,T_j(k)} \rangle| \leq \delta_{2s} \|\mathbf{h}_{k,T_0(k)}\|_{\ell_2} \|\mathbf{h}_{k,T_j(k)}\|_{\ell_2}. \quad (\text{D.4})$$

Inequality (D.4) holds for T_1 in place of T_0 . Since T_0 and T_1 are disjoint

$$\|\mathbf{h}_{k,T_0(k)}\|_{\ell_2} + \|\mathbf{h}_{k,T_1(k)}\|_{\ell_2} \leq \sqrt{2} \|\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2}, \quad (\text{D.5})$$

which results in

$$\begin{aligned} (1 - \delta_{2s}) \|\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2}^2 &\leq \|\mathbf{CBh}_{k,T_{01}(k)}\|_{\ell_2}^2 \\ &\leq \|\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2} (2C_0\epsilon\sqrt{1 + \delta_{2s}} + \sqrt{2}\delta_{2s} \sum_{j \geq 2} \|\mathbf{h}_{k,T_j(k)}\|_{\ell_2}). \end{aligned} \quad (\text{D.6})$$

It follows from (C.2) and (D.6) that

$$\|\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2} \leq \alpha C_0 \epsilon + \rho s^{-1/2} \|\mathbf{h}_{k,T_0^c(k)}\|_{\ell_2}. \quad (\text{D.7})$$

Now, using (C.6) and (D.7) we can conclude that

$$\begin{aligned} \sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2} &\leq K \alpha C_0 \epsilon + \rho s^{-1/2} \sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_0^c(k)}\|_{\ell_2} \\ &\leq K \alpha C_0 \epsilon + \rho \sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_0(k)}\|_{\ell_2} \\ &\leq K \alpha C_0 \epsilon + \rho \sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2}, \end{aligned} \quad (\text{D.8})$$

which means

$$\sum_{k=0}^{K-1} \|\mathbf{h}_{k,T_{01}(k)}\|_{\ell_2} \leq K(1 - \rho)^{-1} \alpha C_0 \epsilon. \quad (\text{D.9})$$

Acknowledgments

340 We would like to thank Professor Humberto Gonzalez (WUSTL) for helpful
input and discussions. ShiNung Ching holds a Career Award at the Scientific
Interface from the Burroughs-Wellcome Fund. This work was partially sup-
ported by AFOSR 15RT0189, NSF ECCS 1509342 and NSF CMMI 1537015,
from the US Air Force Office of Scientific Research and the US National Science
345 Foundation, respectively.

References

1. Srinivasan MV, Laughlin SB, Dubs A. Predictive coding: a fresh view of
inhibition in the retina. *Proceedings of the Royal Society of London B:
Biological Sciences* 1982;216(1205):427–59.
- 350 2. Baddeley R, Abbott LF, Booth MC, Sengpiel F, Freeman T, Wakeman EA,
Rolls ET. Responses of neurons in primary and inferior temporal visual
cortices to natural scenes. *Proceedings of the Royal Society of London B:
Biological Sciences* 1997;264(1389):1775–83.

3. Graham DJ, Chandler DM, Field DJ. Can the theory of "whitening" explain the center-surround properties of retinal ganglion cell receptive fields?
355 *Vision Research* 2006;46(18):2901–13.
4. Olshausen BA, et al. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 1996;381(6583):607–9.
5. Blättler F, Hahnloser RH. An efficient coding hypothesis links sparsity and
360 selectivity of neural responses. *PloS one* 2011;6(10):e25506.
6. Mairal J, Bach F, Ponce J, Sapiro G. Online dictionary learning for sparse coding. In: *Proceedings of the 26th Annual International Conference on Machine Learning*. 2009:689–96.
7. Mairal J, Bach F, Ponce J, Sapiro G. Online learning for matrix factor-
365 ization and sparse coding. *The Journal of Machine Learning Research* 2010;11:19–60.
8. Lee H, Battle A, Raina R, Ng AY. Efficient sparse coding algorithms. In: *Advances in Neural Information Processing Systems*. 2006:801–8.
9. Candès EJ, Wakin MB. An introduction to compressive sampling. *IEEE
370 Signal Process Mag* 2008;25(2):21–30.
10. Eldar YC, Kutyniok G. Compressed sensing: theory and applications. Cambridge University Press; 2012.
11. Haupt J, Bajwa WU, Rabbat M, Nowak R. Compressed sensing for networked data. *IEEE Signal Process Mag* 2008;25(2):92–101.
12. Petrantonakis PC, Poirazi P. A compressed sensing perspective of hippocampal function. *Frontiers in Systems Neuroscience* 2014;8.
13. Ganguli S, Sompolinsky H. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annual review of neuroscience* 2012;35:485–508.

- 380 14. Barranca VJ, Kovačič G, Zhou D, Cai D. Network dynamics for optimal compressive-sensing input-signal recovery. *Phys Rev E* 2014;90(4):042908.
15. Barranca VJ, Kovačič G, Zhou D, Cai D. Sparsity and compressed coding in sensory systems. *PLoS Computational Biology* 2014;10(8).
16. Donoho DL. For most large underdetermined systems of linear equations
385 the minimal l_1 norm solution is also the sparsest solution. *Comm Pure Appl Math* 2006;59(6):797–829.
17. Candès EJ, Romberg J, Tao T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans Inf Theory* 2006;52(2):489–509.
- 390 18. Candès EJ, Tao T. Decoding by linear programming. *IEEE Trans Inf Theory* 2005;51(12):4203–15.
19. Candès EJ, Romberg JK, Tao T. Stable signal recovery from incomplete and inaccurate measurements. *Comm Pure Appl Math* 2006;59(8):1207–23.
20. Vaswani N. Kalman filtered compressed sensing. In: *Proc. 15th IEEE International Conference on Image Processing*. 2008:893–6.
395
21. Charles A, Asif MS, Romberg J, Rozell C. Sparsity penalties in dynamical system estimation. In: *Proc. 45th IEEE Annual Conference on Information Sciences and Systems (CISS)*,. 2011:1–6.
22. Wakin MB, Sanandaji BM, Vincent TL. On the observability of linear
400 systems from random, compressive measurements. In: *Proc. 49th IEEE Conference on Decision and Control (CDC)*. 2010:4447–54.
23. Sanandaji BM, Vincent TL, Wakin MB, Tóth R, Poolla K. Compressive system identification of lti and ltv arx models. In: *Proc. 50th IEEE Conf. on Decision and Control and European Control Conference (CDC-ECC)*.
405 2011:791–8.

24. Napoletani D, Sauer TD. Reconstructing the topology of sparsely connected dynamical networks. *Phys Rev E* 2008;77(2):026103.
25. Shoukry Y, Nuzzo P, Puggelli A, Sangiovanni-Vincentelli AL, Seshia SA, Tabuada P. Secure state estimation under sensor attacks: A satisfiability modulo theory approach. *arXiv preprint arXiv:1412.4324* 2014;.
26. Fawzi H, Tabuada P, Diggavi S. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Trans Automatic Control* 2014;59(6):1454–67.
27. Ba DE, Babadi B, Purdon PL, Brown EN. Exact and stable recovery of sequences of signals with sparse increments via differential l_1 minimization. In: *Proc. Advances in Neural Information Processing Systems (NIPS)*. 2012:2636–44.
28. Sefati S, Cowan NJ, Vidal R. Linear systems with sparse inputs: Observability and input recovery. In: *Proc. IEEE American Control Conference (ACC)*. 2015:5251–7.
29. Jaeger H. Short term memory in echo state networks. *GMD report, German National Research Center for Information Technology* 2001;.
30. White OL, Lee DD, Sompolinsky H. Short-term memory in orthogonal neural networks. *Phys Rev Lett* 2004;92(14):148102.
31. Ganguli S, Huh D, Sompolinsky H. Memory traces in dynamical systems. *Proceedings of the National Academy of Sciences* 2008;105(48):18970–5.
32. Hermans M, Schrauwen B. Memory in linear recurrent neural networks in continuous time. *Neural Networks* 2010;23(3):341–55.
33. Wallace E, Maei HR, Latham PE. Randomly connected networks have short temporal memory. *Neural Comput* 2013;25(6):1408–39.

34. Ganguli S, Sompolinsky H. Short-term memory in neuronal networks through dynamical compressed sensing. In: *Proc. Advances in neural information processing systems*. 2010:667–75.
35. Charles AS, Yap HL, Rozell CJ. Short-term memory capacity in networks via the restricted isometry property. *Neural Comput* 2014;26(6):1198–235.
36. Ito I, Ong RCY, Raman B, Stopfer M. Sparse odor representation and olfactory learning. *Nat Neurosci* 2008;11(10):1177–84.
37. Raman B, Joseph J, Tang J, Stopfer M. Temporally diverse firing patterns in olfactory receptor neurons underlie spatiotemporal neural codes for odors. *J Neurosci* 2010;30(6):1994–2006.
38. Ostojic S, Brunel N, Hakim V. How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. *J Neurosci* 2009;29(33):10234–53.
39. Kafashan M, Palanca BJ, Ching S. Bounded-observation kalman filtering of correlation in multivariate neural recordings. In: *Proc. 36th Annu. Int. Conf. Eng. Med. Biol.* 2014:.
40. Abbott LF. Lapicque’s introduction of the integrate-and-fire model neuron (1907). *Brain Res Bull* 1999;50(5):303–4.
41. Dayan P, Abbott L. Theoretical neuroscience: computational and mathematical modeling of neural systems. Comput. Neurosci.; Cambridge, MA, USA: Massachusetts Institute of Technology Press; 2005.
42. Izhikevich EM, et al. Simple model of spiking neurons. *IEEE Trans Neural Netw* 2003;14(6):1569–72.
43. Kafashan M, Lepage KQ, Ching S. Node selection for probing connections in evoked dynamic networks. In: *Proc. 53rd IEEE Annu. Conf. Decision and Control (CDC)*. 2014:.

44. Kafashan M, Ching S. Optimal stimulus scheduling for active estimation of evoked brain networks. *J Neural Eng* 2015;12(6):066011.
45. Douglas RJ, Martin KA. Recurrent neuronal circuits in the neocortex. *Curr Biol* 2007;17(13):R496–500.
46. Kohonen T, Oja E. Fast adaptive formation of orthogonalizing filters and associative memory in recurrent networks of neuron-like elements. *Biol Cybern* 1976;21(2):85–95.
47. Seung HS, Lee DD, Reis BY, Tank DW. Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron* 2000;26(1):259–71.
48. Güler NF, Übeyli ED, Güler İ. Recurrent neural networks employing lyapunov exponents for EEG signals classification. *Expert Syst Appl* 2005;29(3):506–14.
49. Omlin CW, Giles CL. Extraction of rules from discrete-time recurrent neural networks. *Neural Netw* 1996;9(1):41–52.
50. Ghanem B, Ahuja N. Sparse coding of linear dynamical systems with an application to dynamic texture recognition. In: *Proc. 20th IEEE Int. Conf. Pattern Recog. (ICPR)*. 2010:987–90.
51. Wei X, Shen H, Kleinsteuber M. An adaptive dictionary learning approach for modeling dynamical textures. In: *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*. IEEE; 2014:3567–71.
52. Li L. Fast algorithms for mining co-evolving time series. Tech. Rep. CMU-CS-11-127.; Carnegie Inst. Tech., Dept. Computer Science; 2011.
53. Tao Y, Faloutsos C, Papadias D, Liu B. Prediction and indexing of moving objects with unknown motion patterns. In: *Proc. ACM SIGMOD Int. Conf. Management of data*. 2004:611–22.

54. Muthukrishnan S. Data streams: Algorithms and applications. New Brunswick, NJ, USA: Now Publishers Inc; 2005.
- 485 55. Parker RG, Rardin RL. Discrete optimization. San Diego, CA, USA: Academic Press Professional, Inc.; 1988.
56. Candes EJ. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique* 2008;346(9):589–92.
57. Candes EJ, Tao T. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans Inf Theory* 2006;52(12):5406–
490 25.
58. Baraniuk R, Davenport M, DeVore R, Wakin M. A simple proof of the restricted isometry property for random matrices. *Constr Approx* 2008;28(3):253–63.
- 495 59. Mendelson S, Pajor A, Tomczak-Jaegermann N. Uniform uncertainty principle for bernoulli and subgaussian ensembles. *Constr Approx* 2008;28(3):277–89.
60. Candes E, Demanet L, Donoho D, Ying L. Fast discrete curvelet transforms. *Multiscale Modeling & Simulation* 2006;5(3):861–99.
- 500 61. Candès EJ, Donoho DL. New tight frames of curvelets and optimal representations of objects with piecewise c2 singularities. *Commun Pure Appl Math* 2004;57(2):219–66.
62. CVX Research I. CVX: Matlab software for disciplined convex programming, version 2.0. 2012.
- 505 63. Grant MC, Boyd SP. Graph implementations for nonsmooth convex programs. In: *Recent Advances in Learning and Control*. Springer; 2008:95–110.

64. Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences* 2009;2(1):183–
510 202.
65. Bioucas-Dias JM, Figueiredo MA. A new twist: two-step iterative shrinkage/thresholding algorithms for image restoration. *Image Processing, IEEE Transactions on* 2007;16(12):2992–3004.
66. Jung H, Ye JC. Motion estimated and compensated compressed sensing
515 dynamic magnetic resonance imaging: What we can learn from video compression techniques. *International Journal of Imaging Systems and Technology* 2010;20(2):81–98.
67. Ghahramani Z, Hinton GE. Parameter estimation for linear dynamical systems. *Technical Report CRG-TR-96-2, University of Toronto, Department
520 of Computer Science* 1996;.
68. Shumway RH, Stoffer DS. Time series analysis and its applications: with R examples. Springer Science & Business Media; 2010.