

# Linear Systems with Sparse Inputs: Observability and Input Recovery

Shahin Sefati<sup>1</sup>, Noah J. Cowan<sup>2</sup> and René Vidal<sup>1</sup>

**Abstract**—In this work, we introduce a new class of linear time-invariant systems for which, at each time instant, the input is sparse with respect to an overcomplete dictionary of inputs. Such systems may be appropriate for modeling a system which exhibits multiple discrete behaviors orchestrated by the sparse input. Although the input is assumed to be *unknown*, we show that the additional structure imposed on the input allows us to recover both the initial state and the sparse, but unknown, input from output measurements alone. For this purpose, we derive sufficient observability and sparse recovery conditions that integrate classical observability conditions for linear systems with incoherence conditions for sparse recovery. We also propose a convex optimization algorithm for jointly estimating the initial condition and recovering the sparse input.

## I. INTRODUCTION

Linear dynamical systems (LDSs) are widely used for modeling and classification of time-series data, including dynamic textures [1], [2], surgical video data [3], [4], and human movements [5]–[7]. In human gait classification, for example, time series of simple human gaits, such as walking or running, are modeled as the output of a stationary LDS [5]. Since two LDSs corresponding to the same gait (e.g., walking) are expected to be closer to each other (in the space of LDSs) than two LDSs corresponding to different gaits (e.g., walking and running), gait classification can be achieved by defining a suitable distance in the space of LDSs, such as the Martin distance [8], Binet-Cauchy Kernels [9], or the align distance [10]. In practice, however, complex motor behaviors are categorically non-stationary. Moreover, many different motor behaviors are performed by the same human. Therefore, a better model may be to assume that the dynamics of the human are the same across multiple behaviors. In this case, however, how do we capture the fact that the same human can perform different behaviors?

Sparse representation theory is a very powerful approach to capturing multiple classes in data. Sparsity embodies the notion that very complex (and high-bandwidth) signals can often be represented as a linear combination of surprisingly few basis vectors. To model multiple classes, one can assume that signals from different classes activate different sets of basis vectors; thus classification can be done by looking at which basis vectors are activated.

This material is based upon work supported by the National Science Foundation under Grant No. 1335035 (R.V.) and a James S. McDonnell Complex Systems Scholar Award (S.S., R.V. and N.J.C.).

<sup>1</sup>S. Sefati and R. Vidal are with the Center for Imaging Science, Johns Hopkins University, Baltimore, MD 21218 USA. {shahin, rvidal}@cis.jhu.edu

<sup>2</sup>N. J. Cowan is with Department of Mechanical Engineering, Johns Hopkins University, Baltimore, MD 21218 USA. ncowan@jhu.edu

These ideas from sparse representation theory has found widespread applications in data acquisition [11], machine learning [12], computer vision [13], and neuroscience [14]. Unfortunately, ideas from sparse representation theory have not yet been exploited in dynamical systems theory for the purpose of modeling and recognition of complex behaviors. Some ideas from sparse representation theory have been explored in the context of observer design for linear systems [15]–[17] (see section II-C for more details). However, such approaches assume that the states, initial conditions, or innovations are sparse. While promising, such notions of sparsity for linear systems may limit the class of systems that can be described. For example, if the state  $x_k$  is sparse, a state transition to  $x_{k+1} = Ax_k$  need not give a sparse vector  $x_{k+1}$ , unless  $A$  satisfies strong conditions (such as being a generalized permutation matrix).

### A. Paper Contributions

In this paper, we propose an alternative approach to incorporating notions from sparse representation theory into linear systems theory. In particular, we introduce a new class of linear time invariant (LTI) systems with *sparse inputs*. The input at *each time step* is assumed to be sparse with respect to an overcomplete dictionary of inputs. This dictionary of inputs is meant to encode the flexibility necessary to drive a wide variety of possible behaviors of the system. At each time instant, a small subset of dictionary atoms (“behaviors”) is selected by the sparse input. In this way, a wide range of complex behaviors can be orchestrated by choosing an appropriate sequence of sparse inputs. In contrast to LDSs with stationary inputs (e.g. ARMA models), this approach may facilitate modeling of complex and inherently non-stationary biological movements and their control.

In general, without any particular assumptions on the parameters of the dynamical model, the structure of the deterministic (or distribution of the stochastic) inputs, or the relative dimension of the dynamical system, input and measured output, the problem of recovering the initial condition of an LDS with *unknown* inputs is an ill-posed problem. The problem of designing Luenberger-type observers for linear systems with unknown inputs or disturbances has been widely studied [18]–[20] (and the references therein). The prior works have established few necessary and sufficient conditions on the parameters of the system for the existence of a stable observer and some of the observer design procedures provided an estimation of the unknown input as well.

In this paper, in addition to introducing the new modeling framework, we demonstrate that it is possible to perfectly recover both the initial state of the system as well as the

sparse input from measurements of the output alone. For this purpose, we derive sufficient conditions for the observability of the initial state and the recovery of sparse input. In addition, we propose a convex optimization formulation to jointly recover the initial condition and the sequence of unknown, but sparse, inputs.

## B. Paper Outline

The remainder of this paper is organized as follows. Section II-A reviews basic concepts from sparse representation theory and sparse recovery algorithms. Section II-B reviews the classical observability and state estimation problems for LTI systems with known inputs. Section II-C reviews existing notions of sparsity for LDSs and relates them to our approach. Section III introduces a new class of discrete-time LDSs with sparse inputs. It proposes a convex optimization formulation for the joint recovery of the initial condition and unknown sparse inputs and derives conditions for the correct recovery of both the initial state and the sparse inputs. Simulation results and discussions are included in Section IV, followed by a brief conclusion in Section V.

## II. BACKGROUND ON SPARSITY AND LINEAR DYNAMICAL SYSTEMS

### A. Sparse Representation and Sparse Recovery

Compressive sensing (CS) and sparse signal recovery have gained significant attention in recent years. The theory of CS states that some sparse signals can be exactly and robustly recovered from an underdetermined and possibly noisy set of measurements. Specifically, consider a real-valued signal  $x \in \mathbb{R}^m$ . The signal  $x$  can be represented in a basis consisting of  $m$  vectors in  $\mathbb{R}^m$ :  $\Psi = [\psi_1, \psi_2, \dots, \psi_m] \in \mathbb{R}^{m \times m}$  as:

$$x = \sum_{i=1}^m \psi_i s_i = \Psi s, \quad (1)$$

where  $s = [s_1, s_2, \dots, s_m]^T$ . Here,  $s$  represents the same signal relative to the basis defined by the columns of  $\Psi$ .

Sparse representation theory is predicated on the idea that most real signals admit a sparse representation with respect to some properly chosen basis. More specifically, given a signal  $x \in \mathbb{R}^m$ , a basis (or dictionary)  $\Psi$ ,  $x$  can be represented as  $x = \Psi s$ , where the signal  $s$  is  $S$ -sparse, namely it has at most  $S$  non-zero elements,  $\|s\|_0 \leq S \ll m$ . To fix notation, the support of  $s$ ,  $\text{supp}(s) = \{i | s_i \neq 0\}$ , is the set of indices corresponding to the non-zero entries and  $\|s\|_0 = |\text{supp}(s)|$ . Consider a linear measurement of the entries of the signal  $x$ :  $y = \phi^T x$ . This measurement can be viewed as the inner product of the measurement vector  $\phi$  and the signal  $x$ :  $y = \langle \phi, x \rangle$ . Assuming we have  $p$  measurement vectors ( $\{\phi_i\}_{i=1}^p, p < m$ ), the underdetermined sensing matrix,  $\Phi$ , is constructed by taking the  $\phi_i^T$ 's as its rows:

$$y = \Phi x = \Phi \Psi s = \Theta s. \quad (2)$$

Equation (2) is an underdetermined system of equations and in general there exists infinite number of solutions. However, under some remarkably general conditions on the

sensing matrix  $\Phi$ , it turns out that one can recover  $s$  exactly from far fewer samples ( $p \ll m$ ) of the signal than predicted by the Nyquist sampling theorem. Concretely, the sparse solution to the underdetermined system of equations  $y = \Theta s$  can be found by solving the following optimization problem:

$$P_0 : \min_{s \in \mathbb{R}^m} \|s\|_0 \quad \text{subject to} \quad y = \Theta s. \quad (3)$$

However, in general this is an NP hard, combinatorial and non-convex optimization problem. There are two main approaches to solving such a non-convex optimization problem:

- 1) Heuristic greedy algorithms such as *Orthogonal Matching Pursuit* (OMP) [21], attempt to directly solve  $P_0$  by finding the columns of  $\Theta$  (known as the atoms of the dictionary in a sparse overcomplete representation) that have the highest correlation with the measurement.
- 2) Convex relaxation methods such as *Basis Pursuit* (BP) [22], which solve the convex optimization problem:

$$P_1 : \min_{s \in \mathbb{R}^m} \|s\|_1 \quad \text{subject to} \quad y = \Theta s. \quad (4)$$

Three main categories of theoretical guarantees establish that under appropriate assumptions, the convex problem  $P_1$  is equivalent to the non-convex problem  $P_0$ : exact recovery condition [23]; restricted isometry property (RIP) [24]; and mutual coherence [25]. Among these, mutual coherence is the easiest one to compute. Specifically, the mutual coherence of a matrix  $\Theta$  is defined as the maximum absolute value of the coherence between normalized columns:

$$\mu(\Theta) = \max_{i,j,i \neq j} \frac{|\theta_i^T \theta_j|}{\|\theta_i\| \|\theta_j\|}. \quad (5)$$

It is shown in [25] that if the matrix  $\Theta$  satisfies

$$\mu(\Theta)(2S - 1) < 1, \quad (6)$$

then the  $S$ -sparse solution to  $P_0$  can be obtained by solving the convex problem  $P_1$ . In other words, the angle between the normalized columns of the matrix  $\Theta$  should be above some threshold.

The standard techniques in CS hold for signals that are sparse in the standard basis ( $\Psi = I$ ) or in some proper orthonormal basis. Rauhut et al. [26] showed that similar techniques could be applied to recover signals that are sparse with respect to an overcomplete dictionary. More recently, Candès et al. [27] provided theoretical guarantees adopted to overcomplete and redundant (coherent) dictionaries.

### B. State Estimation for Deterministic LTI Systems

Consider an LTI system of the form:

$$\begin{aligned} x_{k+1} &= Ax_k + \Psi u_k, & x_k &\in \mathbb{R}^n, u_k \in \mathbb{R}^m \\ y_k &= Cx_k, & y_k &\in \mathbb{R}^p. \end{aligned} \quad (7)$$

Here,  $A \in \mathbb{R}^{n \times n}$  is called the dynamic matrix,  $C \in \mathbb{R}^{p \times n}$  is the observation matrix,  $x_k$  is the state,  $y_k$  is the output, and  $u_k$  in the input signal.

For the linear system defined in (7), let

$$Y_N = [y_0^T, y_1^T, y_2^T, \dots, y_{N-1}^T]^T \in \mathbb{R}^{Np} \quad (8)$$

denote the vector of all outputs up to time  $N$ . Likewise, let

$$U_N = [u_0^T, u_1^T, u_2^T, \dots, u_{N-2}^T]^T \in \mathbb{R}^{(N-1)m} \quad (9)$$

denote the vector of all inputs up to time  $N$ . One can easily show that

$$Y_N = \mathcal{O}_N x_0 + \Gamma_N U_N, \quad (10)$$

where  $\mathcal{O}_N$  is the observability matrix:

$$\mathcal{O}_N = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{N-1} \end{bmatrix} \quad (11)$$

and

$$\Gamma_N = \begin{bmatrix} 0 & 0 & \dots & 0 \\ C\Psi & 0 & \dots & 0 \\ CA\Psi & C\Psi & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{N-2}\Psi & CA^{N-3}\Psi & \dots & C\Psi \end{bmatrix}. \quad (12)$$

State estimation is one of the fundamental problems in LDS theory. The state estimation problem typically boils down to recovering the sequence of states from inputs and outputs over time, i.e.,  $U_N$  and  $Y_N$ . Indeed a system is typically defined to be *observable* if one can recover any  $x_0$  from measurements of the inputs and outputs. It can be easily verified from (10) that if the input and output sequences are known, the initial condition  $x_0$  can be recovered if and only if the rank of the observability matrix  $\mathcal{O}_N$  is  $n$ .

When the inputs are unknown, as assumed in this paper, if there are fewer outputs than inputs, i.e.  $p < m$ , recovering the inputs is an ill-posed problem. In this case it is not clear if and how to reduce this to a standard state estimation problem, which requires knowing the inputs to solve (10).

### C. Prior Work on Sparsity in Linear Dynamical Systems

Several attempts have been made to incorporate sparsity in the context of LDSs. Most attempts can be categorized depending on how or where sparsity is imposed, e.g., sparse parameters, sparse states, or sparse inputs. In this section we provide a review of these trends.

1) *LDSs with Sparse Parameters and Sparsity in System Identification*: In this setting, sparsity in the parameter space is assumed and exploited for system identification and model order reduction (i.e., approximating a given complex system with a *simple* system of lowest order). Often sparsity inducing norms or constraints are used in an optimization framework. One of the earliest examples includes [21], [28] where system identification and model *order* reduction are posed as the problem of choosing an efficient representation (*fewest* number of coefficients) of the frequency response of an LDS in the so-called rational wavelet basis. Some other examples in this category include [29]–[32]. Topology identification of large-scale sparsely connected dynamical networks can also be facilitated with a sparsity inducing regularizer [31], [32].

2) *LDSs with Sparse States*: Recently, a number of filtering and smoothing algorithms have been proposed for recovering time-varying sparse signals whose temporal evolution can be modeled by an LDS [15], [17], [33]. These algorithms are mostly targeted at compressive sensing of time-varying sparse signals and they usually require limiting assumptions in the temporal evolution model such as slow changing sparsity patterns of the signal. In a different approach, Wakin et al. [16] studied the observability of linear systems with sparse high-dimensional initial state and randomized compressive measurements. Overall, while enforcing sparsity on the states results in promising state estimation algorithms of sparse time-varying signals, we believe this notion of sparsity is fundamentally limited. For example, if we consider an LDS  $x_{k+1} = Ax_k$ , then if  $x_k$  is sparse,  $Ax_k$  need not be sparse except for a special matrix  $A$  (e.g., a permutation).

3) *LDSs with Sparse Inputs*: Sparse input models have mostly been used for modeling spike trains, i.e., signals that are sparse in time. For instance, blind deconvolution with  $\ell_1$  regularization for recovering spike trains has long been applied in the context of seismic signal processing [34]. However, such approaches and more recent ones have been limited to single-input single-output systems, usually with finite impulse response [34], [35]. More recently, in [36] a similar approach has been proposed for modeling multivariate time series of human actions as the output of an LDS driven by a one dimensional spike train. This work also proposes an alternating minimization framework for jointly recovering the train of input spikes and the LDS model parameters. This approach is an example of blind system identification or deconvolution. The main limitation of this approach is that the input class may be too restrictive, and may not be able to generate complex dynamic behavior. In contrast, in our proposed LDS with sparse inputs much more complex dynamics can be modeled through the use of a rich input dictionary. Some other works (e.g., [17]) consider sparse inputs as noise rather than stimulus, i.e., as an undesired signal that should be suppressed. To the best of our knowledge, imposing sparsity at the input of MIMO LDSs while including an overcomplete dictionary for the input (thus allowing a rich class of inputs) is a novel approach that has not appeared in the literature so far.

## III. STATE ESTIMATION AND INPUT RECOVERY FOR LINEAR SYSTEMS WITH SPARSE INPUTS

Sparse representation theory applies to the recovery of *static* sparse signals from a small number of measurements, while observability theory focuses on the estimation of the system state (or initial conditions) of a *dynamical* system. How can these two ideas be combined to estimate both the initial conditions and the input signals? In this section, we show how to extend sparse representation theory to a dynamical context. We first propose a new class of LDSs with sparse inputs. We then propose a framework to jointly recover the initial condition and the sparse inputs and derive conditions for the correctness of the recovery. Next we assume that the initial condition is known (or without loss

of generality is set to zero) and discuss the step-by-step recovery of sparse inputs.

### A. Linear Dynamical Systems with Sparse Inputs

In this subsection, we introduce a new class of LDSs whose input  $u_k$  at *each time step* is sparse with respect to an overcomplete dictionary of inputs  $\Psi$ , namely,  $\|u_k\|_0 \leq S$  ( $S \ll m$ ). Formally, a *linear dynamical system with sparse inputs* (LDS-SI) is a system of the form:

$$\begin{aligned} x_{k+1} &= Ax_k + \Psi u_k, \quad x_k \in \mathbb{R}^n, u_k \in \mathbb{R}^m, \|u_k\|_0 \leq S \\ y_k &= Cx_k, \quad y_k \in \mathbb{R}^p, \end{aligned} \quad (13)$$

where  $n$  and  $m$  are possibly large. Notice that the support and value of the input are not constrained over time in our model. Sparsity of the input at each time step means that, at any given time, only a few columns of the basis  $\Psi$  get excited, but in as few as  $m/S$  steps, all of the columns could be excited, depending on how the support varies over time. Notice also that the overall behavior of LDS-SI systems is nonlinear. In fact, we can think of the proposed LTI systems with  $S$ -sparse inputs of dimension  $m$  as an LTI system with switched inputs of dimension  $S$ , hence a switched LTI system, where the number of discrete states is  $\binom{m}{S}$ .

### B. Joint Recovery of the Initial Condition and Sparse Inputs

In the classical linear state estimation setting, one assumes that the parameters of the model ( $\Gamma_N$  and  $\mathcal{O}_N$ ) and the input and output sequences ( $U_N$  and  $Y_N$ ) are known. Then, the estimation of the initial conditions boils down to inverting the matrix,  $\mathcal{O}_N$ , which can be done if and only if the observability matrix  $\mathcal{O}_N$  has rank  $n$ .

In this paper, we assume that the sequence of inputs,  $U_N$ , is *also unknown*. If  $p < m$  (fewer measurements than inputs), then (10) is an under-determined problem and in general an infinite number of solutions exist: there are  $Nm + n$  unknowns and only  $Np$  measurements ( $Nm + n > Np$ ). To address this issue, in this paper we consider a special set of inputs to the LDS, namely input signals,  $u_k$ , that are sparse at each time step ( $k = 0, 1, 2, \dots, N-1$ ). In other words we study the inputs that are either sparse in the standard basis or have a sparse representation with respect to an overcomplete dictionary,  $\Psi$  [25]. In this case, a simple counting argument suggests that a necessary condition for a solution to (10) to be unique is that  $pN \geq n + NS$ , where  $S$  is the sparsity level of inputs at each time step,  $p$  is the number of measurements at each step, and  $n$  is the dimension of dynamical system. Therefore, we expect the smallest number of steps to be on the order of  $N \geq n/p + NS/p$ . This is very intuitive since, loosely speaking,  $n/p$  is the minimum number of steps to recover  $x_0$  and  $NS/p$  is the minimum number of steps to recover  $U_N$ .

The above counting argument suggests that it may be possible to recover both  $U_N$  and  $x_0$  under some conditions. However, it does not tell us how to recover  $U_N$  and  $x_0$ . As discussed in Section II-A, in general, finding a sparse solution involves  $l_0$  minimization (see (3)), which is a non-convex and NP-hard optimization problem. To overcome

this, by analogy to the  $l_1$  relaxation approach to recover sparse signals, we propose the following convex optimization problem to jointly recover the sparse inputs and non-sparse initial condition:

$$\min_{U_N, x_0} \|U_N\|_1 \quad \text{subject to} \quad Y_N = \mathcal{O}_N x_0 + \Gamma_N U_N, \quad (14)$$

where

$$\|U_N\|_1 = \sum_{k=0}^{N-1} \|u_k\|_1 \quad (15)$$

and  $x_0$  and  $U_N$  are both unknown. This formulation is new in that it proposes a means by which to recover the state and unknown, but sparse, inputs. For the sake of simplicity, here we have chosen to work with the  $l_1$ -norm. However, prior knowledge about the structure of the sparsity patterns can lead to more effective structured sparsity-inducing norms in the optimization formulation.

A comparison between the proposed optimization problem in (14), the standard state estimation problem in (10) and the standard sparse recovery problem in (4) sheds light on the conditions required for perfect joint recovery. If in (14) we assume that the input and output sequences are known, then a necessary and sufficient condition to recover the initial condition  $x_0$  is that the rank of the observability matrix be equal to  $n$ . On the other hand, if in (14) we assume that the initial condition and the sequence of outputs are known, then a sufficient condition for stable and exact recovery of the unknown sparse inputs is that the matrix  $\Gamma_N$  satisfies conditions such as mutual coherence (see Equation (5)). Note, however, that the proposed optimization problem in (14) is slightly different from the standard sparse recovery problem since the initial condition,  $x_0$ , is also unknown. One may be tempted to concatenate the matrices  $\mathcal{O}_N$  and  $\Gamma_N$  to form  $\mathcal{M}_N = [\mathcal{O}_N \quad \Gamma_N]$ , and verify the mutual coherence condition on  $\mathcal{M}_N$  for the perfect joint recovery of  $x_0$  and  $U_N$ . However, notice that the  $l_1$  norm is imposed only on  $U_N$  and the initial condition  $x_0$  is generally not sparse. Therefore, the sufficient condition for the joint state estimation and sparse recovery in (14) does not follow directly from the incoherence conditions applied directly to  $\Gamma_N$  or  $\mathcal{M}_N$ .

The above discussion motivates the need for deriving new conditions for joint state estimation and sparse recovery. The following proposition gives a sufficient conditions for the correctness of the solution of problem (14).

**Proposition 1.** *Let  $\Pi$  be the projection onto the orthogonal complement of the column space of the observability matrix  $\mathcal{O}_N$ . If  $\mathcal{O}_N$  is full rank and the projected matrix  $\Pi\Gamma_N$  is incoherent,  $x_0$  and  $U_N$  can be uniquely recovered from  $Y_N$  as the solution to (14).*

*Proof.* Since  $\mathcal{O}_N$  is full rank, it follows from (10) that we can solve for the initial condition  $x_0$  as a function of the unknown input sequence  $U_N$  as

$$x_0 = (\mathcal{O}_N^T \mathcal{O}_N)^{-1} \mathcal{O}_N^T (Y_N - \Gamma_N U_N). \quad (16)$$

Substituting this expression for  $x_0$  back into the optimization problem in (14) we obtain

$$\min_{U_N} \|U_N\|_1 \quad \text{subject to} \quad Y_{\Pi} = \Gamma_{\Pi} U_N, \quad (17)$$

where  $Y_{\Pi} = \Pi Y_N$ ,  $\Gamma_{\Pi} = \Pi \Gamma_N$ , and  $\Pi$  is the projection matrix onto the orthogonal complement of the column space of the observability matrix  $\mathcal{O}_N$ , which is given by:

$$\Pi = I - \mathcal{O}_N (\mathcal{O}_N^T \mathcal{O}_N)^{-1} \mathcal{O}_N^T, \quad (18)$$

where  $I$  is the identity matrix of appropriate size.

The formulation in (17) is a standard  $l_1$  minimization problem, where  $Y_{\Pi}$  are the measurements,  $\Gamma_{\Pi}$  is the dictionary, and  $U_N$  is the sparse vector to be recovered. As a consequence, sufficient conditions for the correctness of the recovery of  $U_N$  follow directly from the incoherence conditions applied to  $\Gamma_{\Pi}$  (see Equation (5)). Once  $U_N$  is recovered from (17),  $x_0$  can be recovered from (16).  $\square$

Note that Proposition 1 assumes that  $U_N$ , the vector of all inputs, is sparse. This assumption does not necessarily require the input signal to be sparse at *each* time step. Therefore, Proposition 1 applies to systems that are more general than (13) in which the input at each time step is assumed to be  $S$ -sparse. As a result, the perfect joint recovery condition in Proposition 1 may be too strong for (13), and weaker conditions could be obtained by using a different sparsity inducing regularizer in (14).

### C. Sequential Input Recovery with Zero Initial Conditions

An alternative to batch recovery is to estimate each input sequentially as each new output becomes available. Since the input signal at each time step,  $u_i$ , is sparse with respect to the dictionary  $\Psi$ , step-by-step sparse recovery of the unknown input signal can be formulated as follows:

$$\hat{u}_k = \arg \min_{u_k} \|u_k\|_1 \quad \text{subject to} \quad \tilde{y}_{k+1} = C\Psi u_k \quad (19)$$

where

$$\tilde{y}_1 = y_1 \quad \text{and} \quad \tilde{y}_{k+1} = y_{k+1} - \sum_{j=0}^{k-1} C A^{k-j} \Psi \hat{u}_j \quad (20)$$

for  $k = \{1, 2, \dots, N-1\}$ . This procedure is potentially more efficient in terms of computational cost, because each step solves an optimization problem with far fewer variables and constraints compared to the batch recovery.

Notice that a sufficient condition for this convex program to recover a sparse solution is that the matrix  $C\Psi$  be incoherent. However, this sparse solution need not coincide with the true sparse solution, as the step-by-step recovery procedure disregards some of the linear constraints from the equation (10), hence it is not equivalent to the batch procedure in (14). In fact, the overall performance of the step-by-step recovery hinges on stable and exact recovery of the input signals at each time step. Failure in recovering the input signal in any one step results in the propagation of the error for the rest of the steps. The challenge of capitalizing on the efficiency of step-by-step recovery, while not being

hampered by a failure along the way, is an important problem but is beyond the scope of the present paper.

## IV. SIMULATION RESULTS

In this section, we evaluate the proposed convex optimization algorithms on synthetically generated time-series data.

### A. Experimental Setup

The data was synthesized by simulating (13) to generate times series data for a wide range of parameters of the LDS-SI. The entries of the dynamic matrix,  $A \in \mathbb{R}^{n \times n}$ , were generated as i.i.d. samples from a Gaussian random variable with mean zero and standard deviation  $1/\sqrt{n}$ . It has been shown that the distribution of eigenvalues of these random matrices obey the circular law as  $n \rightarrow \infty$  [37]. To ensure the stability of the  $A$  matrices generated in our experiments, sampled matrices with eigenvalues of maximum modulus greater than 0.9 were discarded. The dictionary matrix  $\Psi$  was generated in the same way as  $A$ , but without enforcing stability. In all experiments, the entries of the measurement matrix,  $C \in \mathbb{R}^{p \times n}$ , were generated as i.i.d. samples from a standard normal distribution. The value of each entry of the non-sparse initial condition,  $x_0$ , was sampled uniformly on  $[-5, 5]$ . For the input signal with sparsity level  $S$  at time step  $k$ ,  $\text{supp}(u_k)$  is a set of  $S$  integers sampled uniformly on  $[1, m]$  ( $|\text{supp}(u_k)| = S$ ). The values of the non-zero entries of  $u_k$  were also sampled uniformly on  $[-5, 5]$ .

Values for the number of steps,  $N = 50$ , and the dimension of the inputs at each time step,  $m = 50$ , were kept fixed in all our experiments. The supports and values of the input were allowed to vary from step to step. The dimension of the dynamical system was varied from 10 to 50 with increments of 10, i.e.,  $n \in \{10, 20, 30, 40, 50\}$ . For each choice of  $n$ , the dimension of the measurements was varied from 5 to  $n$  with increments of 5, i.e.  $p \in \{5, 10, \dots, n\}$ . For each combination of  $n$  and  $p$ , the sparsity level of the inputs at each time step was varied from 1 to 20, i.e.  $S \in \{1, 2, 3, \dots, 20\}$ . This range of sparsity allows for up to 40% of the entries of  $u_k$  to be nonzero at each time step ( $S_{\max}/m = 20/50 = 0.4$ ).

Each replicate of the experiment consisted of 600 sets of simulations. Ten replicates of the experiment were carried out. In all simulations we used the CVX software package [38], [39] to solve the optimization problems.

### B. Joint Recovery of the Initial Condition and Sparse Unknown Input

The probability of perfect joint recovery as a function of the sparsity level of the inputs at each time step,  $S$ , and the dimension of measurements at each time step,  $p$ , is shown in Fig. 1. The probability of perfect joint recovery is estimated using 10 simulated replicates. For a given number of measurements,  $p$ , and for the range parameters explored in this study, the ratio of the dimension of the dynamical system to the number of inputs,  $n/m$ , does not significantly change the level of sparsity that admits perfect recovery. This can be seen by comparing the results in Fig. 1(A) and (B)

for  $p \leq 30$  (where both systems were simulated). Note that this ratio,  $n/m$ , is the ratio of the number of rows to the number of columns of the dictionary matrix,  $\Psi$ . Of course, as  $n$  gets very small compared to the number of inputs, the overcomplete dictionary becomes highly coherent (as an extreme example when  $n = 1$ ,  $\mu(\Psi) = 1$ ) and it becomes impossible to exactly recover the sparse input. Notice that

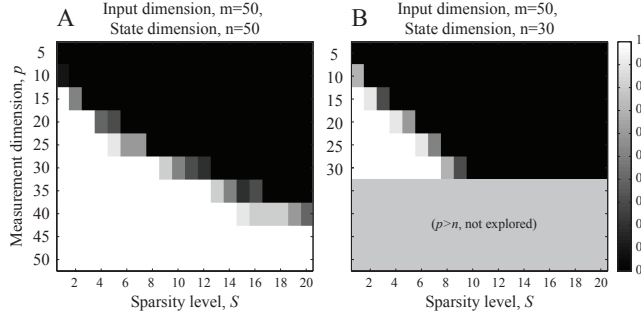


Fig. 1. Probability of perfect joint recovery estimated from 10 simulated replicates. Probability (scale on right) is given as a function of the sparsity level of the input at each time step, and the measurement dimension. (A)  $n = m = 50$ . Dimension of the dynamical model equals to the number of inputs at each time step. (B)  $30 = n < m = 50$ .  $\Psi$  is an overcomplete dictionary.

the initial state may be recovered independently from the inputs and directly from the measurements. For example, when  $p \geq n$  joint recovery becomes trivial in the sense that the initial condition can be recovered independently from the inputs, directly from the first measurement,  $y_0 = Cx_0$ , provided the observation matrix,  $C$ , has rank  $n$ .

### C. Mutual Coherence

The mutual coherence of the matrices,  $\Psi$ ,  $C\Psi$ ,  $\Gamma_N$ , and  $\Gamma_{\Pi}$  as a function of the dimension of the measurements at each time step is shown in Fig. 2. Error bars show one standard deviation variation across 10 replicates of simulated data. For a fixed choice of  $n$  and  $m$ , the coherence of the randomly sampled dictionary matrix,  $\mu(\Psi)$ , does not statistically change as a function of  $p$  (shown in green). For a fixed choice of  $m$  (number of inputs), as  $n$  decreases, the dictionary,  $\Psi$ , becomes more coherent. As a result, for a given dimension of the measurements, the coherence of  $C\Psi$  also increases as  $n$  decreases. From the structure of

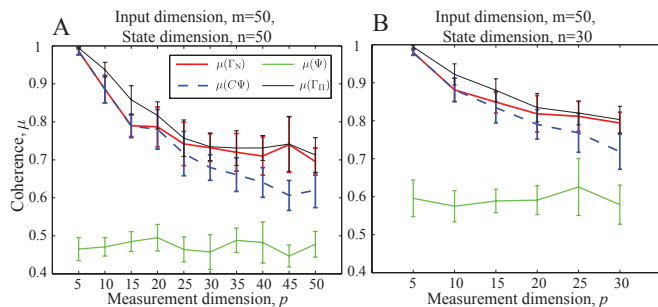


Fig. 2. Mutual coherence of several matrices (see legend) as a function of measurement dimension,  $p$ . The coherence of the  $\Gamma_N$  matrix is always bounded from below by the coherence of  $C\Psi$ . (A)  $n = 50$ . (B)  $n = 30$ .

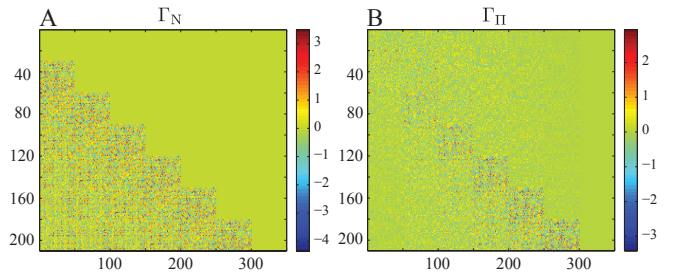


Fig. 3. Visualization of the entries of  $\Gamma_N$  and  $\Gamma_{\Pi}$  of a representative simulation set for  $m = 50, N = 7, n = 50, p = 30$ . Note that in this example, after projection onto the orthogonal complement of the column space of  $\mathcal{O}_N$ , the colors in the first few blocks (50 columns per block) are more “muted” than before projection, suggesting that coherence is compromised as verified numerically (see Fig. 2).

the matrix  $\Gamma_N$  (see (12)), it can be easily verified that its mutual coherence,  $\mu(\Gamma_N)$  (shown in red), is always bounded from below by  $\mu(C\Psi)$  (shown in blue). Note that for almost all simulations, the coherence of  $\Gamma$  is larger than about 0.6. Based on the sufficient recovery condition in (6), this implies that the maximum sparsity level we can tolerate to guarantee recovery, independent of the number of measurements, is about 1.3. Experimentally, and for the joint recovery of the initial condition and sparse inputs, this sufficient condition turns out to be quite conservative as shown in Fig. 1.

### D. Discussion on Optimization Formulation

One can jointly recover the initial condition and input by first projecting both sides of the equation onto the orthogonal complement of the column space of the observability matrix ( $\mathcal{O}_N$ ) as described in Section III-B. This approach appears promising because one can first solve (17) to recover the unknown inputs,  $U_N$ , and then solve (10) to recover the initial condition. Simulation results reveals that when (14) fails to jointly recover the initial condition and unknown sparse input, (17) also fails to recover the unknown input perfectly. This result can be explained by comparing the coherence of  $\Gamma_N$  (before projection) and  $\Gamma_{\Pi}$  (after projection). As shown in Fig. 2, the coherence of  $\Gamma_{\Pi}$  (black) is always greater than the coherence of  $\Gamma_N$  (see Fig. 3).

## V. CONCLUSION AND FUTURE WORK

In this paper, we introduced a new class of linear time-invariant dynamical systems with sparse inputs (LDS-SI). In the proposed framework, at each time instant, the input to the system is sparse with respect to an overcomplete dictionary of inputs. We proposed a convex optimization formulation to jointly recover the initial condition and unknown but sparse inputs of a linear dynamical system. While this formulation allows the sparsity of the initial condition, and state-transition matrices that preserve sparsity, it does not require these potentially limiting assumptions. Simulation results show that recovery of sparse inputs are achievable even for signals that are sparse with respect to an overcomplete dictionary.

In this study, we did not assume any structured sparsity pattern in the input signals. One possible example is the case in which non-zero entries of control inputs are clustered. In

such scenarios, the performance of the optimization solver may be improved by the use of norms that induce sparsity at the group level (e.g., the  $l_1/l_2$  norm). Structured sparsity-inducing norms other than  $l_1$  can be formulated for the specific problem at hand [40]. Moreover, sparse representations are shown to be useful for classification of unlabeled data [12], [13]. In future work, we aim to extend the results of the current study to the classification problem of segmented data such as surgical gesture classification [4]. This will require solving not only the state estimation and sparse input recovery problem, but also the system identification and dictionary learning problems.

#### ACKNOWLEDGMENT

Authors would like to thank Bijan Afsari and Narges Ahmidi for insightful discussions and comments.

#### REFERENCES

- [1] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto, "Dynamic textures," *International Journal of Computer Vision*, vol. 51, no. 2, pp. 91–109, 2003.
- [2] A. Ravichandran and R. Vidal, "Video registration using dynamic textures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 1, pp. 158–171, 2011.
- [3] B. Béjar, L. Zappella, and R. Vidal, "Surgical gesture classification from video data," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, vol. 7510, 2012, pp. 34–41.
- [4] L. Zappella, B. Béjar, G. Hager, and R. Vidal, "Surgical gesture classification from video and kinematic data," *Medical image analysis*, vol. 17, no. 7, pp. 732–745, 2013.
- [5] A. Bissacco, A. Chiuso, Y. Ma, and S. Soatto, "Recognition of human gaits," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2001, pp. II–52.
- [6] A. Veeraraghavan, A. K. Roy-Chowdhury, and R. Chellappa, "Matching shape sequences in video with applications in human movement analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1896–1909, 2005.
- [7] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal, "Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1932–1939.
- [8] R. J. Martin, "A metric for ARMA processes," *IEEE Transactions on Signal Processing*, vol. 48, no. 4, pp. 1164–1170, 2000.
- [9] S. Vishwanathan, A. J. Smola, and R. Vidal, "Binet-cauchy kernels on dynamical systems and its application to the analysis of dynamic scenes," *International Journal of Computer Vision*, vol. 73, no. 1, pp. 95–119, 2007.
- [10] B. Afsari, R. Chaudhry, A. Ravichandran, and R. Vidal, "Group action induced distances for averaging and clustering linear dynamical systems with applications to the analysis of dynamic scenes," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2208–2215.
- [11] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [12] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 2790–2797.
- [13] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [14] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vision research*, vol. 37, no. 23, pp. 3311–3325, 1997.
- [15] N. Vaswani, "Kalman filtered compressed sensing," in *IEEE International Conference on Image Processing*, 2008, pp. 893–896.
- [16] M. B. Wakin, B. M. Sanandaji, and T. L. Vincent, "On the observability of linear systems from random, compressive measurements," in *IEEE Conference on Decision and Control*, 2010, pp. 4447–4454.
- [17] A. Charles, M. S. Asif, J. Romberg, and C. Rozell, "Sparsity penalties in dynamical system estimation," in *Annual Conference on Information Sciences and Systems*, 2011, pp. 1–6.
- [18] M. Hou and P. Muller, "Design of observers for linear systems with unknown inputs," *IEEE Transactions on Automatic Control*, vol. 37, no. 6, pp. 871–875, 1992.
- [19] T. Floquet, C. Edwards, and S. K. Spurgeon, "On sliding mode observers for systems with unknown inputs," *International Journal of Adaptive Control and Signal Processing*, vol. 21, no. 8-9, pp. 638–656, 2007.
- [20] S. Sundaram and C. N. Hadjicostis, "Delayed observers for linear systems with unknown inputs," *IEEE Transactions on Automatic Control*, vol. 52, no. 2, pp. 334–339, 2007.
- [21] Y. C. Pati, R. Rezaifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Asilomar Conference on Signals, Systems and Computers*, 1993, pp. 40–44.
- [22] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM journal on scientific computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [23] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.
- [24] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathématique*, vol. 346, no. 9, pp. 589–592, 2008.
- [25] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, 2006.
- [26] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2210–2219, 2008.
- [27] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2011.
- [28] Y. Pati, R. Rezaifar, P. S. Krishnaprasad, and W. P. Dayawansa, "A fast recursive algorithm for system identification and model reduction using rational wavelets," in *Asilomar Conference on Signals, Systems and Computers*, 1993, pp. 35–39.
- [29] A. Chiuso and G. Pillonetto, "Learning sparse dynamic linear systems using stable spline kernels and exponential hyperpriors," in *Neural Information Processing Systems NIPS*, 2010, pp. 397–405.
- [30] B. M. Sanandaji, T. L. Vincent, M. B. Wakin, R. Tóth, and K. Poolla, "Compressive system identification of lti and ltv arx models," in *Conference on Decision and Control and European Control Conference (CDC-ECC)*, 2011, pp. 791–798.
- [31] D. Napolitano, T. D. Sauer, et al., "Reconstructing the topology of sparsely connected dynamical networks," *Physical Review E-Statistical Nonlinear and Soft Matter Physics*, vol. 77, no. 2, p. 26103, 2008.
- [32] B. M. Sanandaji, T. L. Vincent, and M. B. Wakin, "Exact topology identification of large-scale interconnected dynamical systems from compressive observations," in *American Control Conference*, 2011, pp. 649–656.
- [33] S. Bhattacharya and T. Basar, "Sparsity based feedback design: a new paradigm in opportunistic sensing," in *American Control Conference*, 2011, pp. 3704–3709.
- [34] H. L. Taylor, S. C. Banks, and J. F. McCoy, "Deconvolution with the  $l_1$  norm," *Geophysics*, vol. 44, no. 1, pp. 39–52, 1979.
- [35] M. S. O'Brien, A. N. Sinclair, and S. M. Kramer, "Recovery of a sparse spike time series by  $l_1$  norm deconvolution," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3353–3365, 1994.
- [36] M. Raptis, K. Wnuk, and S. Soatto, "Spike train driven dynamical models for human actions," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2077–2084.
- [37] A. Edelman, "The probability that a random real gaussian matrix has  $k$  real eigenvalues, related distributions, and the circular law," *Journal of Multivariate Analysis*, vol. 60, no. 2, pp. 203–232, 1997.
- [38] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab software for disciplined convex programming," 2008.
- [39] M. C. Grant and S. P. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent advances in learning and control*. Springer, 2008, pp. 95–110.
- [40] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Structured sparsity through convex optimization," *Statistical Science*, vol. 27, no. 4, pp. 450–468, 2012.