

Observability of a Linear System Under Sparsity Constraints

Wei Dai and Serdar Yüksel

Abstract—Consider an n -dimensional linear system where it is known that there are at most $k < n$ nonzero components in the initial state. The observability problem, that is the recovery of the initial state, for such a system is considered. We obtain sufficient conditions on the number of available observations to be able to recover the initial state exactly for such a system. Both deterministic and stochastic setups are considered for system dynamics. In the former setting, the system matrices are known deterministically, whereas in the latter setting, all of the matrices are picked from a randomized class of matrices. The main message is that one does not need to obtain full n observations to be able to uniquely identify the initial state of the linear system, even when the observations are picked randomly, when the initial condition is known to be sparse.

Index Terms—Linear systems, observability, stochastic systems.

I. INTRODUCTION

A discrete-time linear system of dimension n is said to be observable if an ensemble of at most n successive observations guarantees the recovery of the initial state. Observability is an essential notion in control theory as, with the sister notion of controllability, these form the essence of modern linear control theory.

In this technical note, we consider the observability problem when the number of nonzeros in the initial state in a linear system is strictly less than the dimension of the system. This might arise in systems where natural or external forces give rise to a certain subset of components of a linear system to be activated or excited, for example an external force may give rise to a subset of locally unstable states while keeping certain other states intact.

Furthermore, with the increasing emphasis on networked control systems, it has been realized that the controllability and observability concepts for linear systems with controllers having full access to sensory information is not practical. Many research efforts have focused on both stochastic settings, as well as information theoretic settings to adapt the observability notion to control of linear systems with limited information. One direction in this general field is the case when the observations available at a controller comes at random intervals. In this context, in both the information theory literature as well as automatic control literature, a rich collection of papers have studied the recursive estimation problem and its applications in remote control.

Before proceeding further, we introduce the notation adopted in the technical note. In this note, bold-face, capital letters refer to matrices, and bold-face, lower-case letters denote vectors. Calligraphic letters, e.g., \mathcal{I} , denote sets. The symbols \mathbb{Z} , \mathbb{Z}^+ , and \mathbb{R} represent the sets of integers, positive integers, and real numbers, respectively. The set $[n]$ is defined as $[n] = \{1, 2, \dots, n\} \subset \mathbb{Z}^+$. For a given set $\mathcal{I} \subset \mathbb{Z}^+$, $\mathbf{v}_{\mathcal{I}}$ denotes the sub-vector formed by the entries indexed by \mathcal{I} , and

$\mathbf{M}_{\mathcal{I}}$ refers to the sub-matrix formed by the columns indexed by \mathcal{I} . The superscript T denotes the transpose operation. For a given vector \mathbf{v} , $\|\mathbf{v}\|_0$ gives the number of nonzero components in \mathbf{v} ($\|\cdot\|_0$ is often referred to as the ℓ_0 -norm [1, eq. (1.1)], even though it is not a well-defined norm), and $\|\mathbf{v}\|_1 = \sum_i |v_i|$ denotes the ℓ_1 -norm of \mathbf{v} .

In the following, we describe the system model. Preliminaries on compressive sensing are presented in Section III. A formal discussion of observability of linear systems follows: since the analytical tools and results are significantly different for different cases, we study first a deterministic setup in Section IV and then a stochastic setup in Section V. Detailed proofs are given in Section VI. Concluding remarks are discussed in Section VII.

II. PROBLEM FORMULATION

For the purpose of observability analysis, we consider the following discrete-time linear time-invariant system (with zero control input): $\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t$, $\mathbf{y}_t = \eta_t \mathbf{C}\mathbf{x}_t$, where $t \in \mathbb{Z}_+$ denotes the discrete time instant, $\mathbf{x}_t \in \mathbb{R}^n$ and $\mathbf{y}_t \in \mathbb{R}^{d_y}$ are the state of the system and the observation of the system respectively, the matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{C} \in \mathbb{R}^{d_y \times n}$ denote the state transfer matrix and the observation matrix respectively, and η_t takes value either 0 or 1 ($\eta_t = 1$ means an observation at time t is available, and $\eta_t = 0$ otherwise).

The problem we are interested in is the observability of a system with a sparse initial state. Given $m < n$ observations (m instances where $\eta_t = 1$), can we reconstruct the initial state $\mathbf{x}_0 \in \mathbb{R}^n$ exactly? Suppose that the receiver observes the output of the system \mathbf{y}_t at the (stopping) time instances t_1, t_2, \dots, t_m . Let the overall observation matrix be the stacked observation matrices $\mathbf{O}_{\mathcal{T}_m} = [(\mathbf{C}\mathbf{A}^{t_1})^T, (\mathbf{C}\mathbf{A}^{t_2})^T, \dots, (\mathbf{C}\mathbf{A}^{t_m})^T]^T$ and the overall observation be $\mathbf{y}_{\mathcal{T}_m} = [\mathbf{y}_{t_1}^T, \mathbf{y}_{t_2}^T, \dots, \mathbf{y}_{t_m}^T]^T$, where the subscript \mathcal{T}_m emphasizes that only the observations at time instants $\mathcal{T}_m := [t_1, t_2, \dots, t_m]^T$ are available. Then $\mathbf{y}_{\mathcal{T}_m} = \mathbf{O}_{\mathcal{T}_m} \mathbf{x}_0$. In order to infer the initial state \mathbf{x}_0 from $\mathbf{y}_{\mathcal{T}_m}$, the columns of $\mathbf{O}_{\mathcal{T}_m}$ have to be linearly independent, or equivalently, the null-space of the matrix $\mathbf{O}_{\mathcal{T}_m}$ must be trivial.

While the general setup has been well understood, the problem of our particular interest is the observability when the initial state \mathbf{x}_0 is sparse.

Definition 1: Let $\mathbf{B} \in \mathbb{R}^{n \times n}$ be an orthonormal basis, i.e., \mathbf{B} contains n orthonormal columns. A vector $\mathbf{x} \in \mathbb{R}^n$ is K -sparse under $\mathbf{B} \in \mathbb{R}^{n \times n}$ if $\mathbf{x} = \mathbf{B}\mathbf{s}$ for some $\mathbf{s} \in \mathbb{R}^n$ with $\|\mathbf{s}\|_0 \leq K$.

We recall that $\|\mathbf{s}\|_0$ gives the number of nonzero components in the vector \mathbf{s} . Our formulation appears to be new in the control theory literature, except for a paper [2] which considers a similar setting for observability properties of a stochastic model to be considered later in the technical note. The differences between the approaches in the stochastic setup are presented in Section V. Another related work is [3] which designs control algorithms based on sparsity in the state, where compressive sensing tools are used to reconstruct the state for control purposes.

III. PRELIMINARIES AND COMPRESSIVE SENSING

Compressive sensing is a signal processing technique that encodes a signal \mathbf{x} of dimension n by computing a measurement vector \mathbf{y} of dimension $m \ll n$ via linear projections, i.e., $\mathbf{y} = \Phi \mathbf{x}$, where $\Phi \in \mathbb{R}^{m \times n}$ is referred to as the *measurement matrix*. In general, it is not possible to uniquely recover the unknown signal \mathbf{x} using measurements \mathbf{y} with reduced-dimensionality. Nevertheless, if the input signal is sufficiently sparse, exact reconstruction is possible. In this context, suppose that the unknown signal $\mathbf{x} \in \mathbb{R}^n$ is at most K -sparse,

Manuscript received July 14, 2011; revised April 11, 2012; accepted October 25, 2012. Date of publication March 19, 2013; date of current version August 15, 2013. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of a Discovery Grant. Recommended by Associate Editor P. Pepe.

W. Dai is with Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: wei.dai@imperial.ac.uk).

S. Yüksel is with the Department of Mathematics and Statistics, Queen's University, Kingston, ON Canada K7L 3N6 (e-mail: yuksel@mast.queensu.ca).

Digital Object Identifier 10.1109/TAC.2013.2253272

i.e., there are at most K nonzero entries in \mathbf{x} . A naive reconstruction method is to search among all possible signals and find the sparsest one which is consistent with the linear measurements. This method requires only $m = 2K$ random linear measurements, but finding the sparsest signal representation is an NP-hard problem. On the other hand, Donoho and Candés *et al.* [1], [4] demonstrated that reconstruction of \mathbf{x} from \mathbf{y} is a *polynomial time* problem if more measurements are taken. This is achieved by casting the reconstruction problem as an ℓ_1 -minimization problem, i.e., $\min \|\mathbf{x}\|_1$ subject to $\mathbf{y} = \Phi \mathbf{x}$. It is a convex optimization problem and can be solved efficiently by linear programming (LP) techniques. The reconstruction complexity equals $O(m^2 n^{3/2})$ if the convex optimization problem is solved using interior point methods [5]. More recently, an iterative algorithm, termed *subspace pursuit (SP)*, was proposed independently in [6] and [7]. The corresponding computational complexity is $O(Km(n + K^2))$, which is significantly smaller than that of ℓ_1 -minimization when $K \ll n$.

A sufficient and necessary condition for ℓ_1 -minimization to perform exact reconstruction is the so called *the null-space condition* [8].

Theorem 1: The ℓ_1 -minimization method reconstructs \mathbf{x} exactly if and only if for all $\mathbf{w} \in \mathbb{R}^n$ such that $\Phi \mathbf{w} = \mathbf{0}$, the property $c \|\mathbf{w}_{\mathcal{I}}\|_1 \leq \|\mathbf{w}_{\mathcal{I}^c}\|_1$ holds for some constant $c > 1$ and for all sets $\mathcal{I} \subset [n]$ such that $|\mathcal{I}| \leq K$. Here, \mathcal{I}^c denotes the set $[n] - \mathcal{I}$.

A sufficient condition for both the ℓ_1 -minimization and SP algorithms to perform exact reconstruction is based on the so called *restricted isometry property (RIP)* [1]. A matrix $\Phi \in \mathbb{R}^{m \times n}$ is said to satisfy the Restricted Isometry Property (RIP) with coefficients (k, δ) for $k \leq m$, $0 \leq \delta \leq 1$, if for all index sets $I \subset \{1, \dots, n\}$ such that $|I| \leq k$ and for all $\mathbf{q} \in \mathbb{R}^{|I|}$, one has

$$(1 - \delta) \|\mathbf{q}\|_2^2 \leq \|\Phi_I \mathbf{q}\|_2^2 \leq (1 + \delta) \|\mathbf{q}\|_2^2$$

where Φ_I denotes the matrix formed by the columns of Φ with indices in I . The *RIP parameter* δ_k of a given matrix Φ is defined as $\inf \{\delta : \Phi \text{ satisfies RIP with } (k, \delta)\}$. It was shown in [1], [6], [9] that both ℓ_1 -minimization and SP algorithms lead to exact reconstructions of K -sparse signals if the matrix Φ satisfies the RIP with a constant parameter, i.e., $\delta_k \leq c_0$ where $c_0 \in (0, 1)$ and k is linear in the sparsity K .

IV. THE DETERMINISTIC MODEL

This section characterizes the number of measurements needed for observability for different scenarios. We follow Definition 1 to assume that \mathbf{x}_0 is K -sparse under a given basis $\mathbf{B} \in \mathcal{S}_{n,n}(\mathbb{R})$. Recall that observability generally requires that the observability matrix \mathbf{O}_{T_m} has full rank, i.e., at least n measurements should be collected. When \mathbf{x}_0 is sparse, the number of observations required for observability can be significantly reduced. In the following, the cases where the state transfer matrix \mathbf{A} is of a diagonal form, a Jordan canonical form (block diagonal form), and the general form will be discussed.

We start with a special case where the number of required observations is $2K + 1$.

Proposition 1: Suppose that \mathbf{x}_0 is K -sparse under the natural basis $\mathbf{B} = \mathbf{I}$. Assume that $\mathbf{A} \in \mathbb{R}^{n \times n}$ is diagonal, and that all diagonal entries are nonzero and distinct. Let all of the entries of $\mathbf{C} \in \mathbb{R}^{1 \times n}$ ($d_y = 1$) be nonzero. Then \mathbf{x}_0 can be exactly reconstructed after exactly $2K + 1$ measurements by algorithms with polynomial complexity in n .

See Section VI-A for the proof. Note that the reconstruction relies on the Reed-Solomon decoding method [10], which is not robust to noise. The following proposition considers the case where ℓ_1 -minimization is used for reconstruction. We have further restrictions on the initial state and observation time.

Proposition 2: Suppose that \mathbf{x}_0 is K -sparse under $\mathbf{B} = \mathbf{I}$. Let all of the entries of $\mathbf{C} \in \mathbb{R}^{1 \times n}$ ($d_y = 1$) be nonzero. Suppose $c_i x_{0,i} \geq 0$ for all i , where $\mathbf{C} = [c_1, \dots, c_n]$.¹ Further assume that $\mathbf{A} \in \mathbb{R}^{n \times n}$ is diagonal, and that the diagonal entries are nonzero and distinct. If the decoder receives $2K + 1$ successive observations at times $t = 0, \dots, 2K$, the decoder can reconstruct the initial state perfectly and the unique solution can be obtained by the solution of the linear program $\min \|\mathbf{x}\|_1$ s.t. $\mathbf{O}_t \mathbf{x} = \mathbf{y}$, where $\mathbf{O}_t = [\mathbf{C}^T, (\mathbf{C}\mathbf{A})^T, \dots, (\mathbf{C}\mathbf{A}^{2K})^T]^T$.

The proof is presented in Section VI-B. We note that, one can relax the above to the case when the observations are periodic such that $t_2 - t_1 = t_3 - t_2 = \dots = t_m - t_{m-1}$, where $1, 2, \dots, m$ are the observation times.

In the following, we consider the case where \mathbf{A} is of a Jordan canonical form.

Proposition 3: Suppose that \mathbf{x}_0 is K -sparse under $\mathbf{B} = \mathbf{I}$. Suppose that $\mathbf{A} \in \mathbb{R}^{n \times n}$ is of Jordan canonical form, all diagonal entries are nonzero, and the eigenvalues corresponding to different Jordan blocks are distinct. Let the entries of $\mathbf{C} \in \mathbb{R}^{1 \times n}$ ($d_y = 1$) be nonzero for all the leading components of Jordan blocks (that is, for the first entry corresponding to a Jordan block). If the decoder receives m random observations, at random times $T_m = \{t_1, t_2, \dots, t_m\}$, let $\mathbf{O}_{T_m} = [(\mathbf{C}\mathbf{A}^{t_1})^T, (\mathbf{C}\mathbf{A}^{t_2})^T, \dots, (\mathbf{C}\mathbf{A}^{t_m})^T]^T$. Let $\mathbf{O}_{T_m}(i)$ denote the i^{th} column of \mathbf{O}_{T_m} for $1 \leq i \leq n$. Define

$$M(\mathbf{T}_m) = \max_{i \neq j} \left\langle \frac{1}{\|\mathbf{O}_{T_m}(i)\|_2} \mathbf{O}_{T_m}(i), \frac{1}{\|\mathbf{O}_{T_m}(j)\|_2} \mathbf{O}_{T_m}(j) \right\rangle.$$

Suppose that $M(\mathbf{T}_m) < 1$.² Then \mathbf{x}_0 can be exactly reconstructed after m measurements if $\|\mathbf{x}_0\|_0 \leq 1/2(1 + 1/M(\mathbf{T}_m))$ by algorithms with polynomial complexity in n . In particular, a linear program (LP) can be used to recover the initial state.

Proof: See Section VI-C. ■

Remark 1: We recall that the observability of a linear system described by the pair (\mathbf{A}, \mathbf{C}) can be verified by the following criterion, known as the Hautus-Rosenbrock test: The pair is observable if and only if for all $\lambda \in \mathbb{C}$, the matrix $[(\lambda \mathbf{I} - \mathbf{A})^T, \mathbf{C}^T]^T$ is full rank. Clearly, one needs to check the rank condition only for the eigenvalues of \mathbf{A} . It is a consequence of the above that, if the component of \mathbf{C} corresponding to the first entry of a Jordan block is zero, then the corresponding component cannot be recovered even with n successive observations, since this is a necessary condition for observability.

The next proposition considers the general setting.

Proposition 4: Given $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{C} \in \mathbb{R}^{d_y \times n}$ and $\mathbf{T}_m = [t_1, \dots, t_m]^T$, if $\Phi = \mathbf{O}_{T_m} \mathbf{B}$ satisfies the null-space condition specified in Theorem 1, then ℓ_1 -minimization $\min \|\mathbf{s}\|_1$ s.t. $\mathbf{y}_t = \mathbf{O}_{T_m} \mathbf{B} \mathbf{s}$ reconstructs \mathbf{s} and $\mathbf{x}_0 = \mathbf{B} \mathbf{s}$ exactly. Suppose that Φ satisfies the RIP with proper parameters, both ℓ_1 -minimization and SP algorithm lead to exact reconstruction of the initial state \mathbf{x}_0 .

This proposition is a direct application of the results presented in Section III. This result implies a protocol in which one keeps collecting available observations $\mathbf{y}_{t_1}, \mathbf{y}_{t_2}, \dots$ until the null-space or RIP condition is satisfied. However, the computation complexity of verifying either of them generally increases exponentially with n . There are two approaches to avoid this extremely expensive computational cost. The first approach is reconstruction on the fly by trying to reconstruct the

¹In some control systems, the initial state is known to belong to a particular polytope (see for example [11] and the references therein). Hence, there may be some side information about where the system starts itself.

²By the Cauchy-Schwarz inequality, $M(\mathbf{T}_m) \leq 1$. Note that $M(\mathbf{T}) = 1$ implies that there exists two rows in \mathbf{O}_{T_m} that are linearly dependent. We assume $M(\mathbf{T}_m) < 1$ to exclude any "repeated" observations.

unknown initial state \mathbf{x}_0 every time when certain number of new observations are received; and continue this process until the reconstruction is good enough. In the second approach, certain suboptimal but computationally more efficient conditions, for example, the incoherence condition, are employed to judge whether current observations are sufficient for reconstruction.

V. THE STOCHASTIC MODEL

In this section, we discuss a stochastic model for the system matrices. We note that, even though the general case under the deterministic model has been discussed in Proposition 4, there is no explicit quantification on the required number of observations. However, under the stochastic model, such a quantification is possible, see Corollary 1. Our analysis is based on the concept of rotational invariance, defined in Section V-A. The intuition is that rotational invariance provides a rich structure to “mix” the nonzeros in the initial state and this “mixing” ensures an observability with significantly reduced number of observations.

During the preparation of this technical note, we noticed that the stochastic model was also discussed in an independent work [2]. The major differences between our approach and that in [2] are as follows. First, in [2], the observation matrix \mathbf{C}_k 's are assumed to be random Gaussian matrices. In contrast, our model relies on rotationally invariant random matrices, which are much more general. Second, though the work [2] is targeted for general state transition matrix \mathbf{A} , the analysis and results best suit for the \mathbf{A} matrices with concentrated spectrum, for example, unitary matrices. As a comparison, in our stochastic model, we separate the rotational invariance and the spectral property and hence the spectral property can be very much relaxed.

A. The Isotropy of Random Matrices

To define rotational invariance, we need to define the set of rotational matrices, often referred to as the Stiefel manifold. Formally, the *Stiefel manifold* $S_{n,k}(\mathbb{R})$ is defined as $S_{n,k}(\mathbb{R}) = \{\mathbf{U} \in \mathbb{R}^{n \times k} : \mathbf{U}^T \mathbf{U} = \mathbf{I}_k\}$, where \mathbf{I}_k is the $k \times k$ identity matrix. When $n = k$, a matrix in $S_{n,n}(\mathbb{R})$ is an orthonormal matrix and represents a rotation. A left rotation of a measurable set $\mathcal{H} \subset \mathbb{R}^{m \times n}$ under a given rotation represented by $\mathbf{A} \in S_{m,m}$ is given by the set $\mathbf{A}\mathcal{H} = \{\mathbf{A}\mathbf{H} : \mathbf{H} \in \mathcal{H}\} \subset \mathbb{R}^{m \times n}$. Similarly define the right rotation of \mathcal{H} given by $\mathcal{H}\mathbf{B}$ for a given $\mathbf{B} \in S_{n,n}$. An *invariant/isotropic* probability measure μ_I [13, Sections 2 and 3] is defined by the property that for any measurable set $\mathcal{M} \subset \mathbb{R}^{m \times n}$ and rotation matrices $\mathbf{A} \in S_{m,m}(\mathbb{R})$ and $\mathbf{B} \in S_{n,n}(\mathbb{R})$, $\mu_I(\mathcal{M}) = \mu_I(\mathbf{A}\mathcal{M}) = \mu_I(\mathcal{M}\mathbf{B})$. The invariant probability on the Stiefel manifold is essentially the uniform probability measure, i.e., $\mu_I(\{\mathbf{A} \in S_{n,k}(\mathbb{R}) : \|\mathbf{A} - \mathbf{U}\|_F \leq \epsilon\})$ is independent of the choice of $\mathbf{U} \in S_{n,k}(\mathbb{R})$.

The main results in this subsection are Lemmas 1 and 2, which show that an rotationally invariant random matrix admits rotationally invariant matrix products and decompositions. These results are key for proving results regarding observability in Section V-B. Due to the space constraints, the proofs are omitted here. The reader can find detailed proofs in the corresponding technical report [14].

Lemma 1: Let $\mathbf{A} \in S_{n,k}(\mathbb{R})$ be isotropically distributed. Let $\mathbf{B} \in S_{n,n}(\mathbb{R})$ be random. Let $\mathbf{C} = \mathbf{B}\mathbf{A}$. Then $\mathbf{C} \in S_{n,k}(\mathbb{R})$ is isotropically distributed and independent of \mathbf{B} .

Let $\mathbf{H} \in \mathbb{R}^{n \times n}$ be a standard Gaussian random matrix, i.e., the entries of \mathbf{H} are independent and identically distributed Gaussian random variables with zero mean and unit variance. Consider the Jordan matrix decomposition $\mathbf{H} = \mathbf{P}\mathbf{J}\mathbf{P}^{-1}$, where \mathbf{J} is often referred to as the Jordan normal form of \mathbf{H} . Let $\mathbf{P} = \mathbf{U}_P \mathbf{\Lambda}_P \mathbf{V}_P^T$ be the singular value decomposition of \mathbf{P} , where $\mathbf{\Lambda}_P$ is the diagonal matrix composed of singular values of \mathbf{P} . Then $\mathbf{P}^{-1} = \mathbf{V}_P \mathbf{\Lambda}_P^{-1} \mathbf{U}_P^T$. The following lemma states that the orthogonal matrix \mathbf{U}_P is isotropically distributed.

Let $\mathbf{H} \in \mathbb{R}^{n \times n}$ be a standard Gaussian random matrix, let $\mathbf{H} = \mathbf{P}\mathbf{J}\mathbf{P}^{-1}$ be the corresponding Jordan matrix decomposition, and let $\mathbf{P} = \mathbf{U}_P \mathbf{\Lambda}_P \mathbf{V}_P^T$ be the singular value decomposition of \mathbf{P} . Then $\mathbf{U}_P \in S_{n,n}(\mathbb{R})$ is isotropically distributed and independent of \mathbf{J} , $\mathbf{\Lambda}_P$ and \mathbf{V}_P .

Remark 2: Although Lemma 2 only treats standard Gaussian random matrices, the same result holds for general random matrix ensembles whose distributions are left and right rotationally invariant: the proof of Lemma 2 can be carried over.

B. Results for Stochastic Models

Recall that a general linear system is observable if and only if the observability matrix \mathbf{O}_{T_m} has full row rank. One may expect that the row rank of \mathbf{O}_{T_m} still indicates the observability of a linear system with sparse initial state and partial observations. The next theorem confirms the intimate relation between the row rank and the observability. The difference between our results and the standard results is that the required minimum rank is much smaller than the signal dimension n in our setting.

Theorem 2: Suppose that $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{C} \in \mathbb{R}^{d_y \times n}$ are independent drawn from random matrix ensembles whose distributions are left and right rotationally invariant. Let r be the row rank of the overall observation matrix \mathbf{O}_{T_m} . If $r \geq O(K \log n / K)$, then the ℓ_1 -minimization method perfectly reconstructs \mathbf{x}_0 from $\mathbf{y}_t = \mathbf{O}_t \mathbf{x}_0$ (where we write $\mathbf{t} = T_m$ for notational convenience) with high probability (at least $1 - e^{-nc}$ for some positive constant c independent of n and r).

Note that the positive constant c appearing in the probability $1 - e^{-nc}$ depends on the specific underlying probability distribution and is difficult to quantify [15]. The common experience is that when n is in hundreds, the probability of perfect reconstruction is practically one [1], [4], [6]–[8]. The proof of Theorem 2 rests on the following Lemma.

Lemma 3: Assume the same set-ups as in Theorem 2 and let $\mathbf{t} = T_m$ for notational convenience. Let $\mathbf{O}_t = \mathbf{U}_t \mathbf{\Lambda}_t \mathbf{V}_t^T$ be the corresponding singular value decomposition, where $\mathbf{U}_t \in S_{m d_y, m d_y}(\mathbb{R})$, $\mathbf{V}_t \in S_{n,n}(\mathbb{R})$ are the left and right singular vector matrices respectively. Then \mathbf{V}_t is isotropically distributed and independent of \mathbf{U}_t and $\mathbf{\Lambda}_t$.

While Lemma 3 is proved in Section VI-D, the detailed proof of Theorem 2 is presented in Section VI-E. The detailed reconstruction procedure using ℓ_1 -minimization is explicitly presented in the proof.

The next corollary presents a special case where the diagonal form is involved.

Corollary 1: Suppose that $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{C} \in \mathbb{R}^{1 \times n}$ ($d_y = 1$) are independent drawn from random matrix ensembles whose distribution is left and right rotationally invariant. Suppose that the Jordan normal form $\mathbf{J} = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}$ is diagonal with distinct diagonal entries with probability one. Then after $m \geq O(K \log n / K)$ measurements, the ℓ_1 -minimization method perfectly reconstructs \mathbf{x}_0 with high probability (at least $1 - e^{-nc}$ for some positive constant c).

The proof is presented in Section VI-F. Note that if $\mathbf{A} = \mathbf{H}\mathbf{H}^T$ where $\mathbf{H} \in \mathbb{R}^{n \times n}$ is a standard Gaussian random matrix, and if $\mathbf{C} \in \mathbb{R}^{1 \times n}$ is also drawn from the standard Gaussian random matrix ensemble, then all the assumptions in this corollary hold. Hence, blindly collecting $m \geq O(K \log n / K)$ observations is sufficient for perfect reconstruction with high probability.

VI. PROOFS

A. Proof of Proposition

Let $\mathbf{A} = \text{diag}(\boldsymbol{\lambda})$ where $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_n]^T$ is the vector containing the diagonal entries of \mathbf{A} . Let c_i denote the i^{th} entry of the row vector \mathbf{C} . Then, $\mathbf{C}\mathbf{A}^{t_i} = [c_1 \lambda_1^{t_i}, c_2 \lambda_2^{t_i}, \dots, c_n \lambda_n^{t_i}] = [\lambda_1^{t_i}, \lambda_2^{t_i}, \dots, \lambda_n^{t_i}] \text{diag}(\mathbf{C})$,

where $\text{diag}(\mathbf{C})$ is the diagonal matrix whose i^{th} diagonal entry is c_i . Hence

$$\mathbf{y}_{T_m} = \mathbf{O}_{T_m} \mathbf{x}_0 = \underbrace{\begin{bmatrix} \lambda_1^{t_1} & \lambda_2^{t_1} & \cdots & \lambda_n^{t_1} \\ \lambda_1^{t_2} & \lambda_2^{t_2} & \cdots & \lambda_n^{t_2} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{t_m} & \lambda_2^{t_m} & \cdots & \lambda_n^{t_m} \end{bmatrix}}_{\mathbf{A}_t} \text{diag}(\mathbf{C}) \mathbf{x}_0.$$

Since all the entries of \mathbf{C} are nonzero, $\text{diag}(\mathbf{C}) \mathbf{x}_0$ is K -sparse under the natural basis. On the other hand, since $\lambda_1, \lambda_2, \dots, \lambda_n$ are all distinct, the matrix \mathbf{A}_t is a truncation of the full rank Vandermonde matrix [17]. Now according to the Reed-Solomon decoding method presented in [10] and the corresponding proof, as long as $m \geq 2K + 1$, one can exactly reconstruct $\text{diag}(\mathbf{C}) \mathbf{x}_0$ and therefore \mathbf{x}_0 from \mathbf{y}_t with the number of algebraic operations polynomial in n . This proposition is therefore proved. \diamond

B. Proof of Proposition 2

Since \mathbf{A} is diagonal, it is of the form $\mathbf{A} = \text{diag}([\lambda_1, \dots, \lambda_n])$. Furthermore, assume that $\mathbf{C} = [c_1, \dots, c_n]$ is a row vector. With m many successive observations, we have a linear system described by

$$\mathbf{y}_t = \underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{m-1} & \lambda_2^{m-1} & \cdots & \lambda_n^{m-1} \end{bmatrix}}_{\mathbf{M}} \text{diag}([c_1, \dots, c_n]) \mathbf{x}_0.$$

Define $\mathbf{z} \in \mathbb{R}^n$ such that $z_i = c_i x_{0,i} \geq 0$. Then the corresponding ℓ_1 -minimization problem becomes

$$\min_{\mathbf{z}} \|\mathbf{z}\|_1 \text{ subject to } \mathbf{y}_t = \mathbf{M}\mathbf{z}. \quad (1)$$

Once we solve the above optimization problem, it is clear that $x_{0,i} = z_i / (\lambda_i^{t_1} c_i)$ where $t_1 = 0$.

For this case, we first show that the ℓ_1 -minimization has a unique solution. Via duality theory, for a constrained minimization problem of a convex function with an equality constraint, the minimization has a unique solution if one can find a Lagrange multiplier (in the dual space) for which the Lagrangian at the solution is locally stationary. More specifically, let $\mathbf{M}_{:,i}$ be the i^{th} column of the matrix \mathbf{M} . Let i_1, \dots, i_K be the indices of the nonzero entries of \mathbf{x}_0 . Clearly, i_1, \dots, i_K are also the indices of the nonzero entries of the corresponding $\mathbf{z} = \text{diag}([\lambda_1^{t_1} c_1, \dots, \lambda_n^{t_1} c_n]) \mathbf{x}_0$. If there exists a vector $\mathbf{g} \in \mathbb{R}^m$ so that

$$\begin{cases} \langle \mathbf{g}, \mathbf{M}_{:,i} \rangle = 1 & \forall i \in \{i_1, i_2, \dots, i_K\} \\ \langle \mathbf{g}, \mathbf{M}_{:,i} \rangle < 1 & \forall i \notin \{i_1, i_2, \dots, i_K\} \end{cases}$$

then the duality theory implies that the optimization problem in (1) has a unique minimizer that is K -sparse and has nonzero entries at indices i_1, \dots, i_K .

In the following we construct a subdifferential which is essentially what Fuchs constructed in [18]. Consider a polynomial in λ of the form $P(\lambda) = \prod_{k=1}^K (\lambda_{i_k} - \lambda)^2 = \alpha_0 \lambda^{2K} + \alpha_1 \lambda^{2K-1} + \dots + \alpha_{2K}$. It is clear that

$$\begin{cases} P(\lambda_i) = 0 & \forall i \in \{i_1, i_2, \dots, i_K\} \\ P(\lambda_i) > 0 & \forall i \notin \{i_1, i_2, \dots, i_K\} \end{cases}$$

where the inequality holds since λ_i 's are distinct. Let $\mathbf{f} \in \mathbb{R}^m$, $\mathbf{f} := [\alpha_{2K}, \alpha_{2K-1}, \dots, \alpha_1, \alpha_0, 0, 0, \dots, 0]^T$. It can be verified that the inner product $\langle \mathbf{f}, [1, \lambda_i, \dots, \lambda_i^{m-1}]^T \rangle = \prod_{k=1}^K (\lambda_{i_k} - \lambda_i)^2 = P(\lambda_i)$. Now, define a vector $\mathbf{g} \in \mathbb{R}^m$ as $\mathbf{g} = [1, 0, 0, \dots, 0]^T - \mathbf{f}$. Then

$\langle \mathbf{g}, [1, \lambda_i, \dots, \lambda_i^{m-1}]^T \rangle = 1 - P(\lambda_i)$. The vector \mathbf{g} is the desired Lagrange vector. Hence, the optimization problem (1) has a unique minimizer.

What now needs to be shown is that there is a unique solution to the original problem under the ℓ_0 constraint. In other words, we wish to show that there is a unique K -sparse \mathbf{z} such that $\mathbf{y}_t = \mathbf{M}\mathbf{z}$. Now, let there be another K -sparse solution \mathbf{z}' . Then, $\mathbf{M}(\mathbf{z} - \mathbf{z}') = \mathbf{0}$. But, since any $2K$ columns of the Vandermonde matrix \mathbf{M} are linearly independent, $\mathbf{z} - \mathbf{z}'$ has to be the zero vector. Hence, this ensures the the found ℓ_1 solution is the sought ℓ_0 solution. \diamond

C. Proof of Proposition 3

We now consider a Jordan matrix \mathbf{A} . Observe that

$$\mathbf{J} = \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{bmatrix} \Rightarrow \mathbf{J}^n = \begin{bmatrix} \lambda_1^n & \binom{n}{1} \lambda_1^{n-1} & \binom{n}{2} \lambda_1^{n-2} \\ 0 & \lambda_1^n & \binom{n}{1} \lambda_1^{n-1} \\ 0 & 0 & \lambda_1^n \end{bmatrix}.$$

Thus, it follows that if \mathbf{A} is of Jordan canonical form, then the observation matrix writes as

$$\Phi = \begin{bmatrix} c_1 \lambda_1^{t_1} & c_1 t_1 \lambda_1^{t_1-1} + c_2 \lambda_1^{t_1} & \cdots & c_n \lambda_n^{t_1} \\ \vdots & \vdots & \ddots & \vdots \\ c_1 \lambda_1^{t_m} & c_1 t_m \lambda_1^{t_m-1} + c_2 \lambda_1^{t_m} & \cdots & c_n \lambda_n^{t_m} \end{bmatrix}.$$

If c_1 is nonzero, and the entries corresponding to leading entries of Jordan blocks are nonzero, the columns of the matrix become linearly independent. By multiplying the initial condition with a diagonal matrix, we can normalize the columns such that the ℓ_2 norm of each column is equal to 1. The rest of the proof follows from Theorem 3 of [19]. \diamond

D. Proof of Lemma 3

Consider the Jordan decomposition $\mathbf{A} = \mathbf{P}\mathbf{J}\mathbf{P}^{-1}$ and the singular value decomposition $\mathbf{P} = \mathbf{U}_P \mathbf{\Lambda}_P \mathbf{V}_P^T$. It is clear that $\mathbf{P}^{-1} = \mathbf{V}_P \mathbf{\Lambda}_P^{-1} \mathbf{U}_P^T$. For notational convenience, let $\tilde{\mathbf{A}} = \mathbf{\Lambda}_P \mathbf{V}_P^T \mathbf{J} \mathbf{V}_P \mathbf{\Lambda}_P^{-1}$ so that $\mathbf{A} = \mathbf{U}_P \tilde{\mathbf{A}} \mathbf{U}_P^T$. It is elementary to verify that $\mathbf{A}^{t_i} = \mathbf{U}_P \tilde{\mathbf{A}}^{t_i} \mathbf{U}_P^T$. Hence

$$\mathbf{O}_t = \begin{bmatrix} \mathbf{C} \mathbf{A}^{t_1} \\ \vdots \\ \mathbf{C} \mathbf{A}^{t_m} \end{bmatrix} = \begin{bmatrix} \mathbf{C} \mathbf{U}_P \tilde{\mathbf{A}}^{t_1} \mathbf{U}_P^T \\ \vdots \\ \mathbf{C} \mathbf{U}_P \tilde{\mathbf{A}}^{t_m} \mathbf{U}_P^T \end{bmatrix}.$$

We shall show that \mathbf{U}_P is independent of both $\tilde{\mathbf{A}}$ and $\mathbf{C} \mathbf{U}_P$. Since \mathbf{A} is left and right rotation-invariantly distributed, according to Remark 2, \mathbf{U}_P is isotropically distributed and independent of $\tilde{\mathbf{A}}$. In order to show that \mathbf{U}_P is independent of $\mathbf{C} \mathbf{U}_P$, we resort to the singular value decomposition $\mathbf{C} = \mathbf{U}_C \mathbf{\Lambda}_C \mathbf{V}_C^T$. Since \mathbf{C} is right rotation-invariantly distributed, \mathbf{V}_C is isotropically distributed. Thus $\tilde{\mathbf{V}}_C := \mathbf{V}_C^T \mathbf{U}_P$ is isotropically distributed and independent of \mathbf{U}_P according to Lemma 1. As a result, $\mathbf{C} \mathbf{U}_P = \mathbf{U}_C \mathbf{\Lambda}_C \tilde{\mathbf{V}}_C$ is independent of \mathbf{U}_P . Write $\mathbf{O}_t = \tilde{\mathbf{O}}_t \mathbf{U}_P^T$, where $\tilde{\mathbf{O}}_t = [(\mathbf{C} \mathbf{U}_P \tilde{\mathbf{A}}^{t_1})^T, \dots, (\mathbf{C} \mathbf{U}_P \tilde{\mathbf{A}}^{t_m})^T]^T$. Since \mathbf{U}_P is independent of both $\tilde{\mathbf{A}}$ and $\mathbf{C} \mathbf{U}_P$, \mathbf{U}_P is independent of $\tilde{\mathbf{O}}_t$. Write the singular value decompositions of \mathbf{O}_t and $\tilde{\mathbf{O}}_t$ as $\mathbf{O}_t = \mathbf{U}_t \mathbf{\Lambda}_t \mathbf{V}_t^T$ and $\tilde{\mathbf{O}}_t = \mathbf{U}_t \mathbf{\Lambda}_t \tilde{\mathbf{V}}_t^T$. Clearly $\mathbf{V}_t = \mathbf{U}_P \tilde{\mathbf{V}}_t$. Since \mathbf{U}_P is isotropically distributed and independent of $\tilde{\mathbf{O}}_t$, $\mathbf{V}_t = \mathbf{U}_P \tilde{\mathbf{V}}_t$ is isotropically distributed and independent of both $\mathbf{\Lambda}_t$ and \mathbf{U}_t according to Lemma 1. \diamond

E. Proof of Theorem 2

We transfer the considered reconstruction problem to the standard compressive sensing reconstruction. Let $\lambda_1, \lambda_2, \dots, \lambda_r$ be the r

nonzero singular values of \mathbf{O}_t and $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_r]^T$. The singular value decomposition of \mathbf{O}_t can be written in the form

$$\mathbf{O}_t = \mathbf{U}_t \begin{bmatrix} \text{diag}(\boldsymbol{\lambda}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}_t^T$$

where $\text{diag}(\boldsymbol{\lambda})$ is the diagonal matrix generated from $\boldsymbol{\lambda}$. Note that

$$\mathbf{U}_t^T \mathbf{y}_t = \begin{bmatrix} \text{diag}(\boldsymbol{\lambda}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}_t^T \mathbf{x}_0.$$

The $r+1, r+2, \dots, m$ entries of $\mathbf{U}_t^T \mathbf{y}_t$ are zeros: they do not carry any information about \mathbf{x}_0 . Define $\tilde{\mathbf{y}}_t$ be the vector containing the first r entries of $\mathbf{U}_t^T \mathbf{y}_t$. We have $\tilde{\mathbf{y}}_t = [\text{diag}(\boldsymbol{\lambda})\mathbf{0}]\mathbf{V}_t^T \mathbf{x}_0$ and therefore

$$\text{diag}(\boldsymbol{\lambda})^{-1} \tilde{\mathbf{y}}_t = [\mathbf{I}_r \quad \mathbf{0}]\mathbf{V}_t^T \mathbf{x}_0 = [\mathbf{I}_r \quad \mathbf{0}]\mathbf{V}_t^T \mathbf{B} \mathbf{s}$$

where \mathbf{I}_r is the $r \times r$ identity matrix. The unknown \mathbf{s} (K -sparse) can be reconstructed by ℓ_1 -minimization with high probability. Since \mathbf{V}_t is isotropically distributed and independent of \mathbf{B} , the matrix $\mathbf{V}_t^T \mathbf{B}$ is isotropically distributed. The matrix $([\mathbf{I}_r \mathbf{0}]\mathbf{V}_t^T \mathbf{B})^T \in \mathcal{S}_{n,r}(\mathbb{R})$, containing the first r rows of $\mathbf{V}_t^T \mathbf{B}$ as columns, is therefore isotropically distributed. Provided that $r \geq O(K \log(n/K))$, the unknown signal \mathbf{s} can be exactly reconstructed from $\text{diag}(\boldsymbol{\lambda})^{-1} \tilde{\mathbf{y}}_t$ via ℓ_1 -minimization [15]. Theorem 2 is proved. \diamond

Remark 3: The reconstruction procedure involves singular value decomposition, matrix production, and ℓ_1 -minimization. The numbers of algebraic operations required for all these steps are polynomial in n . Hence, the complexity of the whole reconstruction process is polynomial in n .

F. Proof of Corollary

Since both \mathbf{A} and \mathbf{C} are left and right rotation-invariantly distributed, Theorem 2 can be applied. Let $\mathbf{A} = \mathbf{P}\mathbf{J}\mathbf{P}^{-1}$ be a Jordan decomposition. Corollary 1 holds if

$$\mathbf{O}_t = \begin{bmatrix} \mathbf{C}\mathbf{A}^{t_1} \\ \vdots \\ \mathbf{C}\mathbf{A}^{t_m} \end{bmatrix} = \begin{bmatrix} \mathbf{C}\mathbf{P}\mathbf{J}^{t_1} \\ \vdots \\ \mathbf{C}\mathbf{P}\mathbf{J}^{t_m} \end{bmatrix} \mathbf{P}^{-1}$$

is full row ranked with probability one, i.e., $\text{rank}(\mathbf{O}_t) = m \geq O(K \log n/K)$ with probability one.

Suppose that the Jordan normal form $\mathbf{J} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ is diagonal. Denote the j^{th} diagonal entry of \mathbf{J} by J_j . Note that

$$\begin{aligned} \mathbf{C}\mathbf{P}\mathbf{J}^{t_i} &= [(\mathbf{C}\mathbf{P})_1 J_1^{t_i}, (\mathbf{C}\mathbf{P})_2 J_2^{t_i}, \dots, (\mathbf{C}\mathbf{P})_n J_n^{t_i}] \\ &= [J_1^{t_i}, J_2^{t_i}, \dots, J_n^{t_i}] \text{diag}(\mathbf{C}\mathbf{P}) \end{aligned}$$

where $\text{diag}(\mathbf{C}\mathbf{P})$ is the diagonal matrix generated from the row vector $\mathbf{C}\mathbf{P}$. Define $\mathbf{J}_{V,t}$ and \mathbf{J}_V , respectively, as

$$\begin{bmatrix} J_1^{t_1} & J_2^{t_1} & \dots & J_n^{t_1} \\ J_1^{t_2} & J_2^{t_2} & \dots & J_n^{t_2} \\ \vdots & \vdots & \ddots & \vdots \\ J_1^{t_m} & J_2^{t_m} & \dots & J_n^{t_m} \end{bmatrix}, \begin{bmatrix} 1 & 1 & \dots & 1 \\ J_1 & J_2 & \dots & J_n \\ \vdots & \vdots & \ddots & \vdots \\ J_1^{n-1} & J_2^{n-1} & \dots & J_n^{n-1} \end{bmatrix}.$$

Then $\mathbf{O}_t = \mathbf{J}_{V,t} \text{diag}(\mathbf{C}\mathbf{P})\mathbf{P}^{-1}$. Note that $\mathbf{J}_{V,t}$ is a sub-matrix composed of m rows of the Vandemonde matrix \mathbf{J}_V , which has full rank. Hence, the matrix $\mathbf{J}_{V,t}$ has full row rank. By definition of \mathbf{P} , \mathbf{P}^{-1} has full rank as well. Therefore, \mathbf{O}_t has full row rank if and only if $\mathbf{C}\mathbf{P}$ does not contain any zero entries.

The fact that the row vector $\mathbf{C}\mathbf{P}$ does not contain any zero entries holds with probability one. This fact will be established by the isotropy

of \mathbf{C} . Let $\mathbf{P}_{\cdot,j}$ denote the j^{th} column of \mathbf{P} . Since \mathbf{P} is full rank, $\mathbf{P}_{\cdot,j} \neq \mathbf{0}$ for all $j = 1, 2, \dots, n$. By assumption, \mathbf{C} is isotropically distributed. This implies that $\mathbf{C}\mathbf{P}_{\cdot,j} \neq 0$ with probability one [13]. $\mathbf{C}\mathbf{P}$ is composed of finite columns. It follows that with probability one, no entry of $\mathbf{C}\mathbf{P}$ is zero.

So far, we have proved that \mathbf{O}_t has full row rank with probability one if the Jordan normal form $\mathbf{J} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ is diagonal. Note that by assumption, the Jordan normal form is diagonal with probability one. We have $\text{rank}(\mathbf{O}_t) = m \geq O(K \log n/K)$ with probability one. \diamond

VII. CONCLUDING REMARKS

In this technical note, we discussed the observability of a linear system where the number of nonzeros in the initial state is smaller than the dimensionality of the system. We observed that a much smaller number of observations (even when the observations are randomly picked) can be used to recover the initial condition.

REFERENCES

- [1] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [2] M. Wakin, B. Sanandaji, and T. Vincent, "On the observability of linear systems from random, compressive measurements," in *Proc. IEEE Conf. Decision and Control (CDC)*, 2010, pp. 4447–4454.
- [3] S. Bhattacharya and T. Başar, "Sparsity based feedback design: A new paradigm in opportunistic sensing," in *Proc. IEEE American Control Conf.*, 2011.
- [4] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [5] I. E. Nesterov, A. Nemirovskii, and Y. Nesterov, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA: SIAM, 1994.
- [6] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inform. Theory*, vol. 55, pp. 2230–2249, 2009.
- [7] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comp. Harmonic Anal.*, vol. 26, pp. 301–321, May 2009.
- [8] W. Xu and B. Hassibi, "Compressive Sensing Over the Grassmann Manifold: A Unified Geometric Framework arXiv:1005.3729v1, 2010.
- [9] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [10] M. Akcakaya and V. Tarokh, "On sparsity, redundancy and quality of frame representations," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jun. 24–29, 2007, pp. 951–955.
- [11] M. Broucke, "Reach control on simplices by continuous state feedback," *SIAM J. Control and Optimiz.*, vol. 48, pp. 3482–3500, Feb. 2010.
- [12] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.
- [13] A. T. James, "Normal multivariate analysis and the orthogonal group," *Ann. Math. Statist.*, vol. 25, no. 1, pp. 40–75, 1954.
- [14] W. Dai and S. Yüksel, "Observability of a Linear System Under Sparsity Constraints arXiv:1204.3097, Tech. Rep., 2012.
- [15] M. Rudelson and R. Vershynin, "Geometric approach to error-correcting codes and reconstruction of signals," *Int. Math. Res. Notices*, vol. 2005, pp. 4019–4041, 2005.
- [16] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [17] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. New York: Cambridge University Press, 1991.
- [18] J.-J. Fuchs, "Sparsity and uniqueness for some specific under-determined linear systems," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Mar. 2005, vol. 5, pp. v/729–v/732, Vol. 5.
- [19] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1341–1344, Jun. 2004.