# Measurement Bounds for Observability of Linear Dynamical Systems under Sparsity Constraints

Geethu Joseph and Chandra R. Murthy *Senior Member, IEEE*

*Abstract*—In this paper, we address the problem of observability of a linear dynamical system from compressive measurements and the knowledge of its external inputs. Observability of a high dimensional system state in general requires a correspondingly large number of measurements. We show that, if the initial state vector admits a sparse representation, the number of measurements can be significantly reduced by using random projections for obtaining the measurements. Our analysis gives sufficient conditions for the restricted isometry property of the observability matrix to hold, which leads to guarantees for the observability of the system. Our results depend only on the properties of system transfer and observation matrices, and are derived using tools from probability theory and compressed sensing. Unlike the prior work in this direction, our results are applicable to systems with an arbitrary nonzero system transfer matrix. Moreover, our results are stronger than the existing results in the regime where they are comparable.

*Index Terms*—Linear dynamical systems, observability, compressed sensing, sparse signal recovery

## I. INTRODUCTION

Observability is an important notion in control theory. It is concerned with the question of how well the state of a linear dynamical system can be inferred from its observations and inputs [2]. The classical observability problem involves solving a linear system of equations:

$$\widetilde{\boldsymbol{y}}_{(K)} = \widetilde{\boldsymbol{A}}_{(K)}\boldsymbol{x}_0, \qquad (1)$$

where the measurement vector $\widetilde{\boldsymbol{y}}_{(K)}$ and the observability matrix $\widetilde{\boldsymbol{A}}_{(K)}$ are known, and we need to estimate $\boldsymbol{x}_0$ exactly.[1] The standard results from linear algebra state that a discrete time system is observable if the rank of the observability matrix $\widetilde{\boldsymbol{A}}_{(K)}$ equals the system dimension [3]. This result applies to the general formulation of the problem, and hence, a large number of measurements are required to recover the initial state for systems with a high dimensional state [4]–[6]. However, if the initial state of the system is known to admit a sparse representation in a suitable basis, the number of measurements required can be potentially reduced by exploiting this additional information. For example, diffusion processes in complex networks that model phenomena like disease or epidemic spreading in the human society [7], [8], air or water pollution [9], [10], virus spreading in computer and mobile phone networks [11], [12], information propagation in online

social networks [13], [14], etc., are known to have a sparse initialization. Identifying the initial state of these processes accurately is a critical first step towards their control [15]. Thus, a key problem in this context is the recoverability of the sparse system state using as few measurements as possible. Further, in some cases, the measurements are obtained as random linear projections of the system state. For example, in the problem of finding the source of pollution in a water body or in the atmosphere, measurements collected from sensors placed at spatially random locations can be mathematically modeled as random linear projections of the system state [16]–[19]. Hence, in this paper, we provide guarantees on the observability of a system when the observability matrix is random and possibly rank deficient, and the initial state admits a sparse representation. Our work is motivated by the results from the area of sparse signal recovery or compressive sensing which studies the theory and algorithmic development for finding sparse solutions to underdetermined linear systems of equations [20]–[22]. As these results serve as the point of departure for our work, we discuss them in the next subsection.

### A. Recovery Guarantees in Compressed Sensing

The most popular approach to establishing guarantees for the exact recovery of an unknown sparse vector from underdetermined linear measurements is through the so-called restricted isometry property (RIP) [23], defined as follows. An observability matrix $\widetilde{\boldsymbol{A}}_{(K)}$ is said to satisfy the $s$-RIP with restricted isometry constant (RIC) $\delta_s$ if $\delta_s \in (0,1)$, where

$$\delta_s \triangleq \inf \left\{ \delta : 1 - \delta \leq \left\| \widetilde{\boldsymbol{A}}_{(K)}\boldsymbol{z} \right\|^2 \leq 1 + \delta, \right.$$

$$\left. \forall \left\| \boldsymbol{z} \right\| = 1, \text{ and } \left\| \boldsymbol{z} \right\|_0 \leq s \right\}. \quad (2)$$

Some examples of RIP based guarantees for exact recovery of sparse vectors with $\ell_0$ norm at most $s$ are as follows:

- $\delta_s < 1/3$ and $\delta_{2s} < \sqrt{1/2}$, (more generally, $\delta_{ts} < \sqrt{(t-1)/t}$ for $t \geq 4/3$ and $\delta_{ts} < \sqrt{t/(4-t)}$ for $0 < t < 4/3$) are sharp for recovery using basis pursuit [24]–[26]
- $\delta_{2s} < \delta(p)$ for $0 < p \leq 1$ is sharp for recovery using $\ell_p$ minimization[2] [27]
- $\delta_{3s} < 1/8$ is sufficient for recovery via the iterative hard thresholding algorithm [28]
- $\delta_{s+1} < \frac{1}{\sqrt{s+1}}$ is sufficient for recovery via the orthogonal matching pursuit (OMP) algorithm [29]

The authors are with the Dept. of ECE at the Indian Institute of Science, Bengaluru, India, Emails:{geethu, cmurthy}@iisc.ac.in.

[1]We discuss the system model in detail in Section II.

[2]We omit the definition of $\delta(p)$ specified in [27].

The RIP also ensures that the recovery process is robust to noise and is stable when the unknown vector is not precisely sparse. Thus, in order to establish bounds on the number of measurements, it is necessary to derive conditions under which the observability matrix satisfies the RIP. In this work, we analyze the RIP of the *structured random observability matrix* arising out of a linear dynamical system.

### B. Related Work

Our work focuses on two aspects: first, we explore the connection between compressed sensing and observability of the state of a linear dynamical system, and second, we derive sufficient conditions for state recovery by analyzing the RIP of the observability matrix. In our case, the observability matrix is a random matrix with a special structure. Hence, the existing results from the compressed sensing literature cannot be directly applied to our problem. In this subsection, we provide review the past literature in this direction.

*1) Compressed sensing and observability:* The connection between the compressed sensing and linear dynamical systems is a nascent topic, and has only recently been studied in the literature. The design of control algorithms based on sparsity in the state using tools from compressed sensing is presented in [30]. However, this paper does not discuss guarantees for recoverability of the system state in the proposed framework. On the other hand, [31], [32] assert that a linear dynamical system is observable if the observability matrix satisfies the RIP. However, conditions under which the observability matrix satisfies RIP are not discussed.

The results in [16]–[18] characterize the number of measurements required for the exact recovery of the initial state in a stochastic setting. However, the results are useful only under somewhat overly restrictive conditions such as the system transfer matrix being unitary, the observation matrices being i.i.d. Gaussian, and the initial state being sparse in the canonical basis. Moreover, those results depend on the smallest singular value of the transfer matrix. As a consequence, they are not independent of scaling of the transfer matrix. In our recent work [1], we derived an initial set of results on the observability of the system considered in this paper. That work is limited to the case when the observation matrices are random and mutually independent, and the initial state is sparse in the canonical basis. However, in many applications, due to hardware constraints, the measurement process could involve linear projection using a single, randomly selected matrix, rather than an independent matrix for each measurement instant. Hence, it is more pertinent to derive recovery guarantees for the case when the observation matrix is fixed, but equal to an instantiation of a random matrix. Direct extension of the techniques in [1] yields a weak result, where the sufficient condition for observability does not improve with the number of measurements available. In this paper, we present a different, new analysis to obtain guarantees for uniform recovery of the state for the identical observation matrices cases. We also study the problem of joint recovery of the initial state and sparse input vectors, which was not considered in [1].

*2) RIP of structured random matrices:* We list a few types of structured random matrices which have been shown to satisfy the RIP in the literature:

- Subsampled bounded orthonormal systems [33], [34]
- Partial random circulant matrices and partial random Toeplitz matrices [35]–[37]
- Block diagonal measurement matrices where each block on the main diagonal is a subgaussian random matrix [38]
- The column-wise Khatri-Rao product of two matrices [39].

As we will see, the RIP of the structured random observability matrix that arises in our problem has not been studied in the past. Hence, it requires new analysis using tools from non-asymptotic random matrix theory. This leads us to our contributions in this paper, which we summarize next.

### C. Our Contributions

In this work, we derive guarantees on recoverability of the sparse initial state of a linear dynamical system under a stochastic setting for two cases: (i) the observation matrices at different time instants are independent and identically distributed (i.i.d.) subgaussian random matrices; (ii) the observation matrices at all time instants are identical, and equal to a subgaussian random matrix. Our contributions are as follows:

- *Independent observation matrices:* We analyze the RIP of the observability matrix when the mutually independent observation matrices have i.i.d. subgaussian entries in Section III. Our results show that $Km = \mathcal{O}(s \ln N)$ to ensure observability with high probability, where $m$ is the number of observations per time instant and $K$ is the number of time steps over which observations are collected. Also, $s$ and $N$ denote sparsity and length of the initial state vector, respectively. We provide a detailed discussion on the implications of the result. We also show that our results are more general and stronger than existing results in the literature.
- *Identical observation matrices:* We study the conditions on $Km$ for the system to observable when the observation matrices are identical in Section IV. We show that $Km = \mathcal{O}(s \ln^2 s \ln^2 N)$ is sufficient for exact recovery. This result is on par with recovery with an unstructured subgaussian random matrix of the same size.
- *Joint recovery of sparse input vectors and initial state:* We generalize the above results to the problem of recovering both the sparse initial state as well as the set of input vectors, which are also assumed to be sparse. We give lower bounds on the number of measurements required to jointly recover both the sparse initial state and the sparse input vectors in Section V, under both independent and identical observation matrices.

In summary, we show that systems that are unobservable using classical control theory can be observable when the underlying sparsity is exploited. Also, our intermediate results could be of independent interest as we analyze the RIP of independent or identical subgaussian matrices with rows right-multiplied by different, arbitrary matrices.

**Notation:** In the sequel, boldface small letters denote vectors, boldface capital letters denote matrices, and calligraphic letters denote sets. The symbol $\mathbb{R}$ denotes the set of real numbers. Also, $\boldsymbol{A}_i$ and $\boldsymbol{A}_{i,j}$ denote the $i^{\text{th}}$ column and $(i,j)^{\text{th}}$ entry of the matrix $\boldsymbol{A}$, respectively. The symbols $\|\cdot\|$, $\|\cdot\|_1$ and $(\cdot)^{\text{T}}$ denote the $\ell_2$ norm, the $\ell_1$ norm, and the transpose of a vector (or a matrix), respectively. We use $\|\cdot\|_2$ and $\|\cdot\|_F$ to denote spectral norm and Forbenius norm of a matrix, respectively. The notation $\boldsymbol{I}$ represents the identity matrix and $\boldsymbol{0}$ represents the all zero matrix (or vector). We use $\mathbb{P}\{\cdot\}$ to denote the probability of an event, and $\mathbb{E}\{\cdot\}$ to denote the expectation operator.

## II. System Model

We consider the following discrete-time linear system:

$$\boldsymbol{x}_{k+1}= \boldsymbol{D}\boldsymbol{x}_k, \tag{3}$$

$$\boldsymbol{y}_k = \boldsymbol{A}_{(k)}\boldsymbol{x}_k, \tag{4}$$

for discrete time instants $k = 0, 1, \ldots, K - 1$. Here, $\boldsymbol{D} \in \mathbb{R}^{N \times N}$ is a nonzero system transfer matrix and $\boldsymbol{A}_{(k)} \in \mathbb{R}^{m \times N}, m \ll N$ is the observation matrix of the system at time instant $k$. We are interested in the observability of the system when the initial state is sparse. We make the following points before proceeding further:

(a) Observability of the initial sparse state $\boldsymbol{x}_0$ implies the observability of $\boldsymbol{x}_k$ for all $k$.

(b) In (3), we do not include an innovation term. Since we are considering the problem of system observability, the system input is assumed to be known. We can therefore simply subtract its effect from the system evolution as well as observation equations, resulting in the system model given by (3) and (4). We consider the joint recovery of the initial state and sparse innovation terms in Section V.

(c) The system equations do not consider measurement noise or model mismatch. However, in the presence of these impairments, our results can be extended to robust recovery of the initial state; we discuss this in Section IV-C.

In view of the above, we formally define the notion of observability as follows:

**Definition 1** (Observability). *A system is said to be observable if any unknown $s-$sparse initial state $\boldsymbol{x}_0$ can be determined uniquely from the outputs $\{\boldsymbol{y}_k\}_{k=0}^{K-1}$, the transfer matrix $\boldsymbol{D}$, and the observation matrices $\{\boldsymbol{A}_{(k)}\}_{k=0}^{K-1}$.*

To recover the sparse initial vector, we consider the following equivalent linear system at time $K$:

$$\widetilde{\boldsymbol{y}}_{(K)} = \widetilde{\boldsymbol{A}}_{(K)}\boldsymbol{x}_0, \tag{5}$$

where the measurement vector $\widetilde{\boldsymbol{y}}_{(K)} \in \mathbb{R}^{Km}$ and the observability matrix $\widetilde{\boldsymbol{A}}_{(K)} \in \mathbb{R}^{Km \times N}$ are defined as

$$\widetilde{\boldsymbol{y}}_{(K)} = \begin{bmatrix} \boldsymbol{y}_0 \\ \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_{K-1} \end{bmatrix}, \quad \widetilde{\boldsymbol{A}}_{(K)} = \begin{bmatrix} \boldsymbol{A}_{(0)} \\ \boldsymbol{A}_{(1)}\boldsymbol{D} \\ \vdots \\ \boldsymbol{A}_{(K-1)}\boldsymbol{D}^{K-1} \end{bmatrix}. \tag{6}$$

In order to ensure the recovery of $\boldsymbol{x}_0$ from (5) using sparse signal recovery techniques, we need to analyze the RIP of the observability matrix $\widetilde{\boldsymbol{A}}_{(K)}$. This, in turn, yields bounds on the number of measurement vectors required to recover any sparse initial state.

Before launching into the RIP analysis, we note that an overall scaling does not affect the RIP of a matrix. Now, let $\lambda_{\max} \neq 0$ be the largest singular value of $\boldsymbol{D}$. We can rewrite (5) as

$$\widetilde{\boldsymbol{L}}(\boldsymbol{\lambda}_{\max})\widetilde{\boldsymbol{y}}_{(K)} = \widetilde{\boldsymbol{L}}(\boldsymbol{\lambda}_{\max})\widetilde{\boldsymbol{A}}_{(K)}\boldsymbol{x}_0, \tag{7}$$

where $\boldsymbol{\lambda}_{\max} = \begin{bmatrix} 1 & \lambda_{\max} & \ldots & \lambda_{\max}^{K-1} \end{bmatrix}^{\text{T}} \in \mathbb{R}^K$ and the matrix function $\widetilde{\boldsymbol{L}}(\boldsymbol{\lambda}) : \mathbb{R}^K \to \mathbb{R}^{Km \times Km}$ is defined as

$$\widetilde{\boldsymbol{L}}(\boldsymbol{\lambda}) = \frac{1}{\sqrt{Km}} \begin{bmatrix} \lambda_1 \boldsymbol{I} & & & \\ & \lambda_2 \boldsymbol{I} & & \\ & & \ddots & \\ & & & \lambda_K \boldsymbol{I} \end{bmatrix}^{-1}, \tag{8}$$

where $\lambda_k$ denotes the $k^{\text{th}}$ element of the vector $\boldsymbol{\lambda}$, and all the identity matrices are of size $m \times m$. Therefore, we get the following relation:

$$\widetilde{\boldsymbol{L}}(\boldsymbol{\lambda}_{\max})\widetilde{\boldsymbol{A}}_{(K)} = \begin{bmatrix} \boldsymbol{A}_{(0)} \\ \boldsymbol{A}_{(1)}\bar{\boldsymbol{D}} \\ \cdots \\ \boldsymbol{A}_{(K-1)}\bar{\boldsymbol{D}}^{K-1} \end{bmatrix} \tag{9}$$

where $\bar{\boldsymbol{D}} = \boldsymbol{D}/\lambda_{\max}$ has the largest and the smallest singular values as $1$ and $\lambda$, respectively. Here, $\lambda$ is the ratio of the smallest to the largest singular value of $\boldsymbol{D}$. Analyzing the recoverability of $\boldsymbol{x}_0$ from (7), which is equivalent to (5), requires one to study the RIP of the matrix $\widetilde{\boldsymbol{L}}(\boldsymbol{\lambda}_{\max})\widetilde{\boldsymbol{A}}_{(K)}$. Therefore, in this paper, we focus on the RIP of such a scaled version of $\widetilde{\boldsymbol{A}}_{(K)}$.

## III. RIP For Independent Observation Matrices

In this section, we present a result on the RIP of the observability matrix when the observation matrices $\boldsymbol{A}_{(k)}$ are independent random matrices, and discuss its implications. We need the following definition of a subgaussian random variable and matrix to state the main result of the section.

**Definition 2** (Subgaussian random variable and matrix). *A random variable $A$ is said to be subgaussian with parameter $c$ if, for any $\theta \in \mathbb{R}$, $\mathbb{E}\{\exp(\theta A)\} \leq \exp(c\theta^2)$. Also, if the entries of the matrix $\boldsymbol{A} \in \mathbb{R}^{m \times N}$ are independent, zero mean, unit variance subgaussian random variables with the same parameter $c$, then $\boldsymbol{A}$ is said to be a subgaussian random matrix.*

The subgaussian random matrix includes a large class of random matrices including i.i.d. Gaussian random matrices, and i.i.d. Bernoulli random matrices, etc.

**Theorem 1** (Independent random observation matrices). *Suppose $\boldsymbol{A}_{(k)}, k = 0, 1, \ldots, K - 1$ are independent subgaussian*

*random matrices with parameter c. Then, if*

$$Km \left(\delta - 1 + \lambda^{2(K-1)}\right)^2 \geq \widetilde{c} \left[9s \ln\left(\frac{eN}{s}\right) + 2 \ln\left(\frac{2}{\epsilon}\right)\right],$$
(10)

*the RIC $\delta_s$ of the scaled version $\widetilde{\boldsymbol{L}}(\boldsymbol{\lambda}_{\max})\widetilde{\boldsymbol{A}}_{(K)}$ of $\widetilde{\boldsymbol{A}}_{(K)}$ satisfies $\delta_s < \delta$ for all $1 - \lambda^{2(K-1)} < \delta < 1$ with probability at least $1 - \epsilon$. Here, $\widetilde{c}$ is a constant dependent only on c, and $\lambda \leq 1$ is the ratio of the smallest to the largest singular values of $\boldsymbol{D}$. When (10) holds, the system is observable for sufficiently large $\lambda$ with high probability.*

Here, we note that the phrase "sufficiently large" $\lambda$ refers to the fact that it must be large enough so as to be able to meet the upper bound on the RIC set by the RIP based guarantees of different algorithms, as discussed in Section I-A. We discuss this point in detail in Section III-B.

Theorem 1 is proved in [1]. In the following subsections, we discuss some implications of the above results.

### A. Special Cases

1) Suppose $\boldsymbol{D}$ is a scaled unitary matrix. Then, $\lambda = 1$, and Theorem 1 simplifies to the recovery condition for the standard compressed sensing problem with $Km$ measurements. Since the RIP of a matrix is invariant to multiplication by a unitary matrix, each new observation vector adds $m$ new measurements to (5) as $K$ increases.
2) Suppose $\boldsymbol{D}$ is rank-deficient. Then, $\lambda = 0$, and (10) does not hold for any $\delta < 1$, unless

$$m \geq \frac{\widetilde{c}}{\delta^2} \left[9s \ln\left(\frac{eN}{s}\right) + 2 \ln(2\epsilon^{-1})\right].$$
(11)

This is intuitive, because when $\boldsymbol{x}_0$ lies in the null space of $\boldsymbol{D}$, $\boldsymbol{y}_k = 0$ for $k \geq 1$. Hence, the system is observable if it is observable from $\boldsymbol{y}_0$. Thus, the uniform recovery guarantee does not hold for a rank deficient $\boldsymbol{D}$.
3) Suppose that $\boldsymbol{D}$ is an ill-conditioned matrix, i.e., $\lambda$ is close to zero. Then, the upper bound on $\delta$ required to guarantee observability may not hold [24], [40], [41]. This is because right multiplication of a matrix by another ill-conditioned matrix may severely degrade its RIP.
4) For $K = 1$, Theorem 1 reduces to the recovery condition of the standard compressed sensing problem [28]. Also, if the system is observable with $m$ measurements (for example, when (11) is satisfied), the conditions in Theorem 1 hold for $K = 1$, as expected.

### B. Number of Measurements

Theorem 1 shows that $Km = \mathcal{O}(s \ln(N/s))$ is sufficient for observability. Note that the number of measurements are independent of the scaling of $\boldsymbol{D}$. Thus, the number of measurements can be greatly reduced for large dimensional systems. In contrast, $Km = \mathcal{O}(N)$ measurements are necessary for observability of a general non-sparse initial state vector. We also recall from Section I-A that the initial state can be recovered using any of the compressed sensing techniques like basis pursuit, thresholding algorithms, or greedy algorithms.

The RIP based recovery guarantees available in the literature set an upper bound on the RIC. For example, using the necessary and sufficient condition for $\ell_1$ based recovery: $\delta_s \leq 1/3$ [24], (10) reduces to

$$K \left(\lambda^{2(K-1)} - 2/3\right)^2$$
$$\geq \frac{\widetilde{c}}{m} \left[9s \ln\left(\frac{eN}{s}\right) + 2 \ln(2\epsilon^{-1})\right], \quad (12)$$

for $\lambda^{2(K-1)} \geq 2/3$. In other words, if (12) is satisfied for some $K \leq \lfloor (\ln(2/3))/(2 \ln(\lambda)) \rfloor + 1$, then the system is observable. However, note that, if the system is observable for $K_1$ measurements, it remains observable for $K > K_1$.

We note that $K \left(\lambda^{2(K-1)} - 2/3\right)^2$ is an increasing function of $K$, which gives a lower bound $m$ from (12). Therefore, for $\lambda < 1$,

$$m = \mathcal{O}\left(\frac{\ln(N/s)}{K \left(\lambda^{2(K-1)} - 2/3\right)^2}\right).$$
(13)

We also note that value of $m$ required decreases with $\lambda$ and $K$. This is in agreement with the fact that as $K$ increases, we get more measurements and a smaller $m$ suffices for ensuring successful recovery of the initial state. Also, as $\lambda$ increases, the matrix $\boldsymbol{D}$ becomes better conditioned, and, consequently, a smaller value of $m$ is sufficient for exact recovery.

### C. Comparison With Prior Work

In [17], [18], the authors address the same problem as ours and give a sufficient condition on number of measurements $Km$ for successful recovery. In this subsection, we compare and contrast the two results. We begin with the result from [17], [18], stated in our notation.

**Theorem 2** (Prior work [17], [18])**.** *Suppose that $\boldsymbol{D} = a\boldsymbol{U}$ where $a \neq 0$ and $\boldsymbol{U} \in \mathbb{R}^{N \times N}$ is unitary. Define $b \triangleq \sum_{k=1}^{K} a^{2(k-1)}$. Assume $\boldsymbol{A}_{(k)}, k = 0, 1, \ldots, K-1$ are independent Gaussian random matrices with mean zero and variance $1/m$. Then, if*

$$Km\delta^2 \geq 512 \left[s \ln\left(\frac{42}{\delta}\right) + 1 + \ln\left(\frac{N}{s}\right) + \ln\left(\frac{2}{\epsilon}\right)\right]$$
$$\left[\frac{\left|1 - a^2\right| K + \min\left\{1, a^2\right\}}{\max\left\{1, a^2\right\}}\right] \quad (14)$$

*the RIC $\delta_s$ of $\frac{1}{\sqrt{b}}\tilde{\boldsymbol{A}}_{(K)}$ satisfies $\delta_s < \delta < 1$ with probability at least $1 - \epsilon$.*

We make the following observations:

- *Restriction on $\boldsymbol{D}$*: Theorem 2 is applicable only when $\boldsymbol{D}$ is a scaled unitary matrix. Reference [18] extends the result to a certain type of positive definite matrices. Our results are more general, and hold true for any arbitrary matrix $\boldsymbol{D} \neq \boldsymbol{0}$.
- *Bound for scaled unitary matrices*: For the special case of $\boldsymbol{D} = a\boldsymbol{U}$, (10) reduces to the following:

$$Km\delta^2 \geq \widetilde{c} \left[9s \ln\left(\frac{eN}{s}\right) + 2 \ln\left(\frac{2}{\epsilon}\right)\right], \quad (15)$$

for $0 < \delta < 1$. We see that there is an extra term on the right hand side of (14) of Theorem 2. We can bound this term as follows:

$$\frac{\left|1-a^2\right|K + \min\left\{1,a^2\right\}}{\max\left\{1,a^2\right\}}$$
$$\geq \frac{\left|1-a^2\right| + \min\left\{1,a^2\right\}}{\max\left\{1,a^2\right\}} = 1, \quad (16)$$

for all $a \neq 0$. Hence, our results are stronger than Theorem 2 for the scaled unitary matrix case.

- *Dependency on the eigenvalue*: The condition (14) heavily depends on the eigenvalue $a$ of $\boldsymbol{D}$. The least number of measurements $Km$ are required for $|a| = 1$, and as $|a|$ moves away from unity, the lower bound on $Km$ increases. However, our results depend only on the ratio of the smallest to the largest singular value of $\boldsymbol{D}$, and therefore gives the best bound for all values of $a$. This is because our results make use of the fact that the recovery properties are independent of scaling due to the equivalence of (5) and (7). This critical observation allowed us to get stronger results compared to Theorem 2.

### D. Extension to Identical Observation Matrices Case

Suppose we carry out a similar analysis for the case when all observation matrices are identical $\boldsymbol{A}_{(k)} = \boldsymbol{A}$ for $k = 0, 1, \ldots, K-1$, where $\boldsymbol{A}$ is a subgaussian random matrix with parameter $c$. The sufficient condition then obtained shows that the system is recoverable if (11) is satisfied. However, this condition ensures that the system is observable with $K = 1$. This is a weak result, because it implies that the availability of additional measurements does not improve the sufficient condition for observability. This is indeed true when $\boldsymbol{D} = \alpha \boldsymbol{I}$, for some $\alpha \in \mathbb{R}$, because we are only adding scaled versions of the rows of $\boldsymbol{A}$ to $\widetilde{\boldsymbol{A}}_{(K)}$ as $K$ increases. For general $\boldsymbol{D}$, a different proof technique has to be used, which is discussed in the next section.

### IV. RIP FOR IDENTICAL OBSERVATION MATRICES CASE

In this section, we present a result on the RIP of the observability matrix when the observation matrices are identical random matrices. First, we define the following quantities:

$$\widetilde{\boldsymbol{D}}_{(K,i)} \triangleq \begin{bmatrix} \boldsymbol{I}_i & \boldsymbol{D}_i & \ldots & \boldsymbol{D}_i^{K-1} \end{bmatrix}, \quad (17)$$

$$L(\boldsymbol{D}, K) \triangleq \max_i \left\| \widetilde{\boldsymbol{D}}_{(K,i)} \right\|_2, \quad (18)$$

where $\boldsymbol{I}_i$ is the $i^{\text{th}}$ column of identity matrix of size $N \times N$ and $\boldsymbol{D}_i^k$ is the $i^{\text{th}}$ column of $\boldsymbol{D}^k$.

**Theorem 3** (Identical random observation matrices)**.** *Suppose all the observation matrices are identical, i.e., $\boldsymbol{A}_{(k)} = \boldsymbol{A}$ for $k = 0, 1, \ldots, K-1$, where $\boldsymbol{A}$ is a subgaussian random matrix with parameter $c$. Then, if*

$$Km\frac{\left(\delta - 1 + \lambda^{2(K-1)}\right)^2}{L^2(\boldsymbol{D}, K)}$$
$$\geq \widetilde{c}s \max\left\{\ln^2 s \ln^2 N, \ln(2\epsilon^{-1})\right\}, \quad (19)$$

*then the RIC $\delta_s$ of the scaled version $\widetilde{\boldsymbol{L}}(\lambda_{\max})\widetilde{\boldsymbol{A}}_{(K)}$ of $\widetilde{\boldsymbol{A}}_{(K)}$ satisfies $\delta_s < \delta$ for all $1 - \lambda^{2(K-1)} < \delta < 1$ with probability at least $1 - \epsilon$. Here, $\widetilde{c}$ is a constant dependent only on $c$, and $\lambda \leq 1$ is the ratio of the smallest to the largest singular values of $\boldsymbol{D}$. Hence, when (19) holds, the system is observable for sufficiently large $\lambda$ with high probability.*

*Proof:* See Appendix A. ∎

Next, using the proof technique of the above theorem, we can show the following interesting corollaries. We omit their proofs as they are straightforward. The first corollary extends Theorem 3 to the case when the measurements are not necessarily taken over a contiguous set of time instants.

**Corollary 1** (Non-consecutive measurements)**.** *Suppose the available measurements are $\{\boldsymbol{y}_k\}_{k\in\mathcal{K}}$, where $\mathcal{K}$ is an index set of cardinality $K$, and $\boldsymbol{A}_{(k)} = \boldsymbol{A}, k \in \mathcal{K}$ where $\boldsymbol{A}$ is a subgaussian random matrix with parameter $c$. Then, if*

$$Km\frac{\left(\delta - 1 + \lambda^{2\max\{\mathcal{K}\}}\right)^2}{L^2(\boldsymbol{D}, \mathcal{K})}$$
$$\geq \widetilde{c}s \max\left\{\ln^2 s \ln^2 N, \ln(2\epsilon^{-1})\right\}, \quad (20)$$

*then the RIC $\delta_s$ of the scaled version $\widetilde{\boldsymbol{L}}(\lambda_{\max,\mathcal{K}})\widetilde{\boldsymbol{A}}_{(K)}$ of $\widetilde{\boldsymbol{A}}_{(K)}$ satisfies $\delta_s < \delta$ for all $1 - \lambda^{2\max\{\mathcal{K}\}} < \delta < 1$ with probability at least $1 - \epsilon$. Here, $\widetilde{c}$ is a constant dependent only on $c$, and $\lambda \leq 1$ is the ratio of the smallest to the largest singular values of $\boldsymbol{D}$. $L^2(\boldsymbol{D}, \mathcal{K})$ has same definition as that of $L^2(\boldsymbol{D}, K)$ with $j^{\text{th}}$ column of $\widetilde{\boldsymbol{D}}_{(K,i)}$ as $\boldsymbol{D}_i^{\widetilde{j}}$ where $\widetilde{j}$ is the $j^{\text{th}}$ smallest element of $\mathcal{K}$. Also, $\boldsymbol{\lambda}_{\max,\mathcal{K}} \in \mathbb{R}^K$ has $j^{\text{th}}$ entry as $\lambda_{\max}^{\widetilde{j}}$. Hence, when (20) holds, the system is observable for sufficiently large $\lambda$ with high probability.*

The next corollary extends Theorem 3 to the case when $\boldsymbol{x}_0$ is sparse under an arbitrary basis $\Psi \in \mathbb{R}^{N \times N}$ rather than the canonical basis.

**Corollary 2** (Sparsifying basis other than the canonical basis)**.** *Suppose $\boldsymbol{A}_{(k)} = \boldsymbol{A}, k = 0, 1, \ldots, K-1$, where $\boldsymbol{A}$ is a subgaussian random matrix with parameter $c$, and the initial state is sparse under the basis $\Psi \in \mathbb{R}^{N \times N}$, which need not be the canonical basis. Then, if*

$$Km\frac{\left(\delta - 1 + \widetilde{\lambda}^2\right)^2}{\widetilde{L}^2(\boldsymbol{D}, K)}$$
$$\geq \widetilde{c}s \max\left\{\ln^2 s \ln^2 N, \ln(2\epsilon^{-1})\right\}, \quad (21)$$

*the RIC $\delta_s$ of the scaled version $\widetilde{\boldsymbol{L}}\left(\widetilde{\boldsymbol{\lambda}}_{\max}\right)\widetilde{\boldsymbol{A}}_{(K)}$ of $\widetilde{\boldsymbol{A}}_{(K)}$ satisfies $\delta_s < \delta$, for all $1 - \widetilde{\lambda}^2 < \delta < 1$, with probability at least $1 - \epsilon$. Here, $\widetilde{c}$ is a constant dependent only on $c$, and $\widetilde{\lambda} \leq 1$ is the ratio of the smallest to the largest singular value of $\boldsymbol{D}^{(K-1)}\Psi$. $\widetilde{L}(\boldsymbol{D}, K)$ has same definition as that of $L(\boldsymbol{D}, K)$, but with $j^{\text{th}}$ column of $\widetilde{\boldsymbol{D}}_{(K,i)}$ as $\boldsymbol{D}^{j-1}\Psi_i$. Also, $\widetilde{\boldsymbol{\lambda}}_{\max} \in \mathbb{R}^K$ has $j^{\text{th}}$ entry as the largest singular value of $\boldsymbol{D}^{j-1}\Psi$. Hence, when (21) holds, the system is observable for sufficiently large $\widetilde{\lambda}$ with high probability.*

It is also interesting to consider guarantees for the case where the matrix $\boldsymbol{D}$ is an RIP-compliant matrix. The following

corollary gives guarantees similar to Theorem 3 based on the RIC of an RIP-compliant $N \times N$ matrix $\boldsymbol{D}$.

**Corollary 3** (Relaxation based on the RIP of the transfer matrix). *Suppose $\boldsymbol{A}_{(k)} = \boldsymbol{A}, k = 0, 1, \ldots, K-1$, where $\boldsymbol{A}$ is a subgaussian random matrix with parameter c. Then, if*

$$Km \frac{\left(\delta - 1 + \bar{\lambda}^{2(K-2)}(1 - \tilde{\delta}_s)^2\right)^2}{L^2(\boldsymbol{D}, K)}$$
$$\geq \tilde{c}s \max\left\{\ln^2 s \ln^2 N, \ln(2\epsilon^{-1})\right\}, \quad (22)$$

*then the RIC $\delta_s$ of the scaled version $\widetilde{\boldsymbol{L}}(\boldsymbol{\lambda}_{\max})\tilde{\boldsymbol{A}}_{(K)}$ of $\widetilde{\boldsymbol{A}}_{(K)}$ satisfies $\delta_s < \delta$ for all $1 - \bar{\lambda}^{2(K-2)}(1 - \tilde{\delta}_s)^2 < \delta < 1$ with probability at least $1 - \epsilon$. Here, $\tilde{c}$ is a constant dependent only on c, and $\bar{\lambda} < 1$ and $\tilde{\delta}_s \leq 1$ are the smallest nonzero singular value and the RIC of $\boldsymbol{D}$ normalized to unit spectral norm. Hence, when (22) holds, the system is observable for sufficiently small $\tilde{\delta}_s$ with high probability.*

*Proof:* When the matrix $\boldsymbol{D}$ is normalized to unit spectral norm, for any unit norm $s-$sparse vector $\boldsymbol{z} \in \mathbb{R}^N$, we have

$$\left\|\boldsymbol{D}^k \boldsymbol{z}\right\| \geq \bar{\lambda}^{k-1} \left\|\boldsymbol{D}\boldsymbol{z}\right\| \geq \bar{\lambda}^{k-1}(1 - \tilde{\delta}_s), \quad (23)$$

since the vector $\boldsymbol{D}\boldsymbol{z}$ belongs to the column space of $\boldsymbol{D}$. Thus, we can replace $\lambda^{(K-1)}$ with $\bar{\lambda}^{(K-2)}(1 - \tilde{\delta}_s)$ in Theorem 3 to obtain the desired result. ∎

We note that $\bar{\lambda} \geq \lambda$ and $1 - \tilde{\delta}_s \geq \lambda$, and thus the above corollary is a stronger result than Theorem 3. However, $\lambda$ is easier to compute than the RIC constant of $\boldsymbol{D}$.

In the following subsections, we provide more insights into the above results.

*A. Special Cases*

1) Suppose $\boldsymbol{D}$ is a scaled identity matrix. Then, $\lambda = 1$, and $L^2(\boldsymbol{D}, K) = K$, and hence from Theorem 3, we retrieve the recovery condition for a standard compressed sensing problem with $m$ measurements, and the guarantee does not improve with increasing $K$. This is intuitive, because we are only adding scaled versions of the rows of $\boldsymbol{A}$ to $\widetilde{\boldsymbol{A}}_{(K)}$ as $K$ increases.
2) Suppose $\boldsymbol{D}$ is rank-deficient. Then, $\lambda = 0$, and (19) does not hold for any $\delta < 1$, unless

$$m \geq \tilde{c}s \max\left\{\ln^2 s \ln^2 N, \ln(2\epsilon^{-1})\right\}, \quad (24)$$

   as expected. However, if $\tilde{\delta}_s \neq 0$, Corollary 3 guarantees that it is possible to recover $\boldsymbol{x}_0$ even if (11) is not satisfied.
3) Suppose that $\boldsymbol{D}$ is ill-conditioned, i.e., $\lambda$ is close to zero. Then, the upper bound on $\delta$ required to guarantee observability may not hold [24], [40], [41], which is in similar vein as explained in the case of Theorem 1.
4) For $K = 1$, Theorem 3 reduces to the recovery condition of the standard compressed sensing problem [28]. Also, if the system is observable with $m$ measurements (for example, when (24) is satisfied), the conditions in Theorem 3 hold for $K = 1$, as expected.
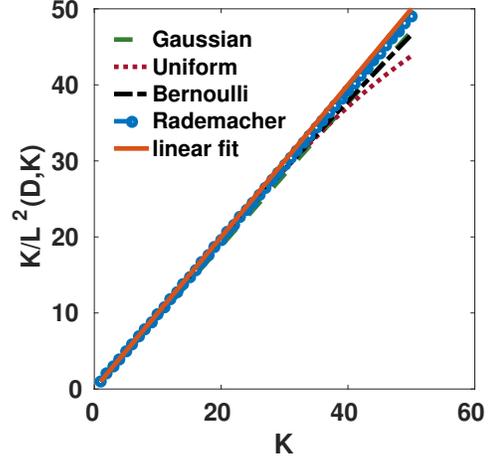


Figure 1. Variation of $K/L^2(\boldsymbol{D}, K)$ with $K$ when $\boldsymbol{D}$ generated using different distributions. We see that $K/L^2(\boldsymbol{D}, K)$ is a (linearly) non-decreasing function of $K$, and it meets the upper bound in Proposition 1. Thus, random matrices are good choices for $\boldsymbol{D}$.
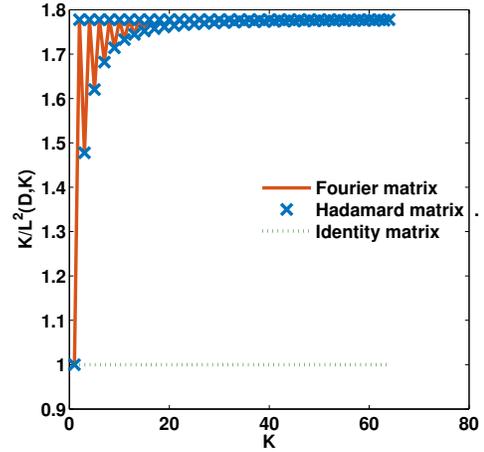


Figure 2. Variation of $K/L^2(\boldsymbol{D}, K)$ with $K$ for Fourier, Hadamard and identity constructions of $\boldsymbol{D}$. We see that $K/L^2(\boldsymbol{D}, K)$ is not an increasing function of $K$.

*B. Number of Measurements*

Theorem 3 shows that $Km = \mathcal{O}(s \ln^2 s \ln^2 N)$ is sufficient for observability, whereas $\mathcal{O}(N)$ measurements are necessary for observability of a non-sparse initial state vector. Also, as mentioned in Section I-A, the initial state can be recovered using any of the compressed sensing techniques like basis pursuit, thresholding algorithms, or greedy algorithms. As in the case of Theorem 1, the RIP based guarantees fix an upper bound on $K$, and hence a lower bound on $m$. However, note that, if the system is observable for $K_1$ measurements, it remains observable for $K > K_1$.

The main difference between the results in Theorem 1 and Theorem 3 is in the $L^2(\boldsymbol{D}, K)$ term. Hence, in order to gain intuition on the number of measurements required in the identical observation matrices case, we study the behavior of the $L(\boldsymbol{D}, K)$ term in the following proposition.

**Proposition 1.** *The term $K/L^2(\boldsymbol{D}, K)$ satisfies*

$$1 \le K/L^2(\boldsymbol{D}, K) \le K. \quad (25)$$

*Proof:* See Appendix F. ∎

We note that the upper and the lower bounds are achieved by $\boldsymbol{D} = \boldsymbol{0}$, and $\boldsymbol{D} = \boldsymbol{I}$, respectively. Further, as discussed in Section IV-A, both cases are not favorable from the observability point of view. Although the $\boldsymbol{D} = \boldsymbol{0}$ case achieves the upper bound of the term $K/L^2(\boldsymbol{D}, K)$, this choice is not desirable since $\lambda = 0$.

In Figure 1, we empirically illustrate that if $\boldsymbol{D}$ is randomly chosen, the upper bound can be nearly achieved. Also, when $\boldsymbol{D}$ is a random matrix, $\lambda$ is non-zero with high probability [42], which makes this choice desirable. Random transfer matrices occur in some linear dynamical systems with sparse initial state, which models homogeneous spreading of epidemic or information or computer virus. For example, a dynamical system corresponding to a epidemic spread can be modeled using a Erdos-Renyi model in which case the transfer matrix has independent Bernoulli distributed entries [43]–[46]. In Figure 1, we use $N = 50$ and vary $K$ from 1 to $N$ [43]. The entries of $\boldsymbol{D}$ are drawn from the following four distributions:

1) Standard Gaussian distribution
2) Uniform distribution on $[0, 1]$
3) Bernoulli distribution with parameter 0.5
4) Rademacher distribution.

The curve $f(K) = K$ is labeled as *linear fit*. The value of $K/L^2(\boldsymbol{D}, K)$ is averaged over 100 trials and plotted along the Y-axis as a function of $K$. In all the four cases, the behavior of the curves is nearly linear, and $K/L^2(\boldsymbol{D}, K) \approx K$.

Before we discuss the implications of the result, we first give some intuition on reason behind this linear behavior. Since $L^2(\boldsymbol{D}, K)$ is a complicated function of $\boldsymbol{D}$, we focus on the regime where $N$ is large and the distribution of the entries of $\boldsymbol{D}$ is Gaussian. We note that, from the Marchenko-Pastur theorem [47], the spectral norm of a Gaussian matrix with zero mean and unit variance is close to $2\sqrt{N}$. Thus, after normalization, as $K$ increases, $\boldsymbol{D}^K$ goes to $\boldsymbol{0}$, and the extra column that gets added to the matrix $\tilde{\boldsymbol{D}}_{K,i}$ as $K$ increases is approximately $\boldsymbol{0}$. Therefore, we have $\left\|\tilde{\boldsymbol{D}}_{(K,i)}\right\|_2 \approx \left\|\tilde{\boldsymbol{D}}_{(2,i)}\right\|_2$. Further, $\left\|\tilde{\boldsymbol{D}}_{(2,i)}\right\|_2$ is the same as the largest eigenvalue of the matrix: $\begin{bmatrix} 1 & \boldsymbol{D}_{ii} \\ \boldsymbol{D}_{ii} & \|\boldsymbol{D}_i\|^2 \end{bmatrix}$. Also, for large $N$, $\|\boldsymbol{D}_i\|^2 \approx 1$ and $\boldsymbol{D}_{ii}^2 \approx 0$, which gives $\left\|\tilde{\boldsymbol{D}}_{(K,2)}\right\|_2 \approx 1$. Hence, we have the following relation:

$$L(\boldsymbol{D}, K) = \arg\max_i \left\|\tilde{\boldsymbol{D}}_{(K,i)}\right\|_2 \approx \left\|\tilde{\boldsymbol{D}}_{(K,2)}\right\|_2 \approx 1. \quad (26)$$

Thus, intuitively, $K/L^2(\boldsymbol{D}, K) \approx K$ for sufficiently large $N$. Also, from Figure 1, we see that $N = 50$ is large enough for the argument to hold.

Next, we discuss some implications of Figure 1. The RIP of the observability matrix $\tilde{\boldsymbol{A}}_{(K)}$ is on par with an unstructured random Gaussian matrix. In turn, this suggests that it is not necessary to use independent observation matrices to ensure observability; identical observation matrices result in a penalty

of only $O(\ln^2 s \ln N)$ in terms of the number of measurements required, provided $\boldsymbol{D}$ is a random matrix. Hence, we conclude that even though $\boldsymbol{I}$ and $\boldsymbol{0}$ are poor choices for $\boldsymbol{D}$, matrices with good recovery properties are plenty in number.

Another important observation from the plot is that $K/L^2(\boldsymbol{D}, K)$ is a non-decreasing function of $K$. Thus, as $K$ increases, we need a smaller value of $m$ for exact recovery of the sparse initial state. That is, one can trade-off $K$ and $m$ while ensuring observability. The following result shows that the function is increasing for the special case when $\boldsymbol{D}$ is a positive semi-definite (psd) matrix.

**Proposition 2.** *When $\boldsymbol{D}$ is a psd matrix, the function $K/L^2(\boldsymbol{D}, K)$ is a non-decreasing function of $K$.*

*Proof:* See Appendix G. ∎

*Remark:* The above result does not, in general, imply that $K/L^2(\boldsymbol{D}, K)$ increases with $K$. For example, from Figure 2, we see that $K/L^2(\boldsymbol{D}, K)$ for standard dictionaries like Fourier and Hadamard matrices is not strictly increasing with $K$.

*C. Extension to Robust Recovery*

Theorem 3 can be extended to bound the error in recovery of the initial state under bounded noise and model mismatch. In this case, the system model modifies as

$$\boldsymbol{x}_k = \boldsymbol{D}^k(\boldsymbol{x}_0 + \check{\boldsymbol{x}}_0), k \ge 1 \quad (27)$$
$$\boldsymbol{y}_k = \boldsymbol{A}_{(k)}\boldsymbol{x}_k + \boldsymbol{w}_k, \quad (28)$$

for discrete time instants $k = 0, 1, \dots, K - 1$. Here, $\boldsymbol{w}_k \in \mathbb{R}^m$ denotes the bounded measurement noise: $\|\boldsymbol{w}_k\| \le W$; while $\check{\boldsymbol{x}}_0 \in \mathbb{R}^N$ represents the error in approximating the initial state by an $s-$sparse vector. That is, $\boldsymbol{x}_0 = \arg\min_{\boldsymbol{v} \in \mathbb{R}^N : \|\boldsymbol{v}\|_0 \le s} \|\boldsymbol{x}_0 + \check{\boldsymbol{x}}_0 - \boldsymbol{v}\|$. Therefore, the overall set of equations can be written as

$$\tilde{\boldsymbol{y}}_{(K)} = \tilde{\boldsymbol{A}}_{(K)}(\boldsymbol{x}_0 + \check{\boldsymbol{x}}_0) + \tilde{\boldsymbol{w}}, \quad (29)$$

where the bounded noise $\tilde{\boldsymbol{w}} \in \mathbb{R}^{Km}$ satisfies $\|\tilde{\boldsymbol{w}}\| \le \sqrt{K}W$.

**Corollary 4** (Robust recovery). *Suppose $\boldsymbol{A}_{(k)} = \boldsymbol{A}$ for $k = 0, 1, \dots, K-1$, where $\boldsymbol{A}$ is a subgaussian random matrix with parameter $c$. Also, suppose that, for some integer $p > 0$ and positive real number $c_{th}$,*

$$Km \frac{\left(c_{th} - 1 + \lambda^{2(K-1)}\right)^2}{L^2(\boldsymbol{D}, K)}$$
$$\ge \tilde{c}ps \max\left\{\ln^2(ps)\ln^2 N, \ln(2\epsilon^{-1})\right\}, \quad (30)$$

*and $\lambda^{2(K-1)} > 1 - c_{th}$. Here, $\tilde{c}$ is a constant dependent only on $c$, and $\lambda \le 1$ is the ratio of the smallest to the largest singular values of $\boldsymbol{D}$. Then, with probability at least $1 - \epsilon$, the initial vector $\boldsymbol{x}_0 + \check{\boldsymbol{x}}_0$ can be recovered with errors as follows:*

$$\|\boldsymbol{x}_0 + \check{\boldsymbol{x}}_0 - \hat{\boldsymbol{x}}_0\|_1 \le c_1 \|\check{\boldsymbol{x}}_0\|_1 + c_2 \sqrt{\frac{s(1 - \lambda_{\max}^{-2K})}{Km(1 - \lambda_{\max}^{-2})}} W$$
$$(31)$$

$$\|\boldsymbol{x}_0 + \check{\boldsymbol{x}}_0 - \hat{\boldsymbol{x}}_0\| \le \frac{c_1}{\sqrt{s}} \|\check{\boldsymbol{x}}_0\|_1 + c_2 \sqrt{\frac{(1 - \lambda_{\max}^{-2K})}{Km(1 - \lambda_{\max}^{-2})}} W,$$
$$(32)$$

where $\hat{\boldsymbol{x}}_0$ is the estimate of the initial vector, and $c_1, c_2 > 0$ are universal constants. The estimate $\hat{\boldsymbol{x}}_0$ can be computed by pre-multiplying (29) by the matrix $\widetilde{\boldsymbol{L}}\left(\widetilde{\boldsymbol{\lambda}}_{\max}\right)$, and then using one of the algorithms mentioned below.

Further, the constants $p$ and $c_{th}$ depend on the recovery algorithms as follows:

- *Basis pursuit:* $p = 2$ and $c_{th} = \frac{4}{\sqrt{41}}$.
- *Iterative hard-thresholding:* $p = 6$ and $c_{th} = \frac{1}{\sqrt{3}}$.
- *Compressive Sampling Matching Pursuit:* $p = 8$ and $c_{th} = \frac{\sqrt{\sqrt{11/3}-1}}{2}$.

*Proof:* From Theorem 3, the matrix $\widetilde{\boldsymbol{L}}\left(\widetilde{\boldsymbol{\lambda}}_{\max}\right)\widetilde{\boldsymbol{A}}_{(K)}$ satisfies RIP. Then, the result follows from the upper bound on the RIC required by the different algorithms to ensure robust recovery [28, Theorem 6.12, 6.21, 6.28]. ∎

We note the dependence on $\lambda_{\max}$ in the above expressions is not unexpected: it arises because of the scaling of the measurement matrix. The scaling operation is reasonable due to the following reasons:

- One can always scale the linear equations with no information loss. The scaling operation neither changes the problem nor affects any intuitive notion of SNR.
- The scaling matrix is diagonal, and therefore does not introduce any correlation between the noise terms which might affect the recovery. Moreover, the recovery guarantees of the algorithms listed in Corollary 4 depend only on the $\ell_2$ norm of the noise vector, and are independent of the individual variances of the noise terms.
- Note that $\lambda_{\max}$ determines the effective SNR of the system, and hence it plays an important role in recoverability of the initial state. The effect of $\lambda_{\max}$ appears as the factor $\sqrt{s}\sqrt{\frac{(1-\lambda_{\max}^{-2K})}{K(1-\lambda_{\max}^{-2})}}\frac{W}{\sqrt{m}}$ in (31). Here, $\sqrt{s}$ and $W/\sqrt{m}$ capture the same effect as those of the sparsity $s$ and the average noise power per measurement $W/\sqrt{m}$, respectively, in the standard compressed sensing results. Further, we intuitively examine the term $\sqrt{\frac{(1-\lambda_{\max}^{-2K})}{K(1-\lambda_{\max}^{-2})}}$ via three special cases of $\lambda_{\max}$ below:

  (i) $\lambda_{\max} \gg 1$: When $\lambda_{\max}$ is large, this term reduces to $1/\sqrt{K}$, which has no dependence on $\lambda_{\max}$. This is because the effective SNR is large, and hence the noise term is negligible, for all measurements except for the first measurement vector, $\boldsymbol{y}_0$. Thus, we have one noisy and $K-1$ noiseless measurements, which leads to an error bound that decreases with $K$.
  (ii) $\lambda_{\max} \approx 1$: When $\lambda_{\max}$ is close to 1, this term reduces to 1. This is equivalent to having $K$ noisy measurements with equal scaling factor and thus the error bound per measurement is independent of $K$. In this case, the advantage of having multiple observations comes in terms of the $Km$ dependence of the number of measurements in (30).
  (iii) $\lambda_{\max} \ll 1$: When $\lambda_{\max}$ is small, this term reduces to $\lambda_{\max}^{-(K-1)}/\sqrt{K}$, which is a new dependence. In this case, the noise in the later measurements gets amplified

by the scaling factor. Hence, the noise term in the last measurement dominates the average noise power. However, in practice, one would consider the smallest value of $K$ for which (30) is satisfied, and substitute that value of $K$ in (31) and (32) to get the bound on robust recovery of the initial state.

*Remark:* Theorem 1 also has corollaries similar to Corollary 1-Corollary 4. We omit those results to avoid repetition.

## V. Joint Recovery of Sparse Initial State and Sparse Inputs

We now discuss the extension of the results presented thus far to the problem of jointly estimating the initial state as well as the input sequence, under sparsity constraints [32]. Sparse inputs are relevant in the networked control of systems, where the controller and plant communicate over a network [48], [49]. In these applications, using sparse inputs helps to handle rate-limited channels over which the control signals are exchanged, because they admit compact representations. The system model in this case is as follows:

$$\boldsymbol{x}_{k+1} = \boldsymbol{D}\boldsymbol{x}_k + \boldsymbol{H}\boldsymbol{h}_{k+1}, \tag{33}$$
$$\boldsymbol{y}_k = \boldsymbol{A}_{(k)}\boldsymbol{x}_k, \tag{34}$$

where $\boldsymbol{H} \in \mathbb{R}^{n \times L}$ is the input matrix and $\boldsymbol{h}_k \in \mathbb{R}^L$ is the input vector such that $\|\boldsymbol{h}_k\|_0 \leq s_{in}$. Therefore, the sparse recovery problem is given by the following equation:

$$\widetilde{\boldsymbol{y}}_{(K)} = \tilde{\boldsymbol{A}}_{(K)}\boldsymbol{x}_0 + \tilde{\boldsymbol{J}}_{(K)}\tilde{\boldsymbol{h}}_{(K)}, \tag{35}$$

where the measurement vector $\tilde{\boldsymbol{y}}_{(K)} \in \mathbb{R}^{Km}$ (as defined in (6)), the unknown sparse vector $\widetilde{\boldsymbol{h}}_{(K)} \triangleq \begin{bmatrix} \boldsymbol{h}_1^{\mathrm{T}} & \cdots & \boldsymbol{h}_{K-1}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{(K-1)L}$ which is at most $\tilde{s} = s + (K-1)s_{in}$ sparse, and the matrix $\tilde{\boldsymbol{J}}_{(K)} \in \mathbb{R}^{Km \times (K-1)L}$ is defined follows:

$$\tilde{\boldsymbol{J}}_{(K)} = \begin{bmatrix} \boldsymbol{0} \in \mathbb{R}^{m \times (K-1)L} \\ \boldsymbol{A}_{(1)}\tilde{\boldsymbol{H}}_{(1)} \in \mathbb{R}^{m \times L} & \boldsymbol{0} \in \mathbb{R}^{m \times (K-2)L} \\ \boldsymbol{A}_{(2)}\tilde{\boldsymbol{H}}_{(2)} \in \mathbb{R}^{m \times 2L} & \boldsymbol{0} \in \mathbb{R}^{m \times (K-3)L} \\ \vdots \\ \boldsymbol{A}_{(K-1)}\tilde{\boldsymbol{H}}_{(K-1)} \in \mathbb{R}^{m \times (K-1)L} \end{bmatrix} \tag{36}$$

$$\tilde{\boldsymbol{H}}_{(k)} = \begin{bmatrix} \boldsymbol{D}^{k-1}\boldsymbol{H} & \boldsymbol{D}^{k-2}\boldsymbol{H} & \cdots & \boldsymbol{H} \end{bmatrix} \in \mathbb{R}^{N \times kL}. \tag{37}$$

Comparing (35) with (5), the effective measurement matrix of the recovery problem takes the form

$$\begin{bmatrix} \boldsymbol{A}_{(0)}\boldsymbol{U}_{(0)} \\ \boldsymbol{A}_{(1)}\boldsymbol{U}_{(1)} \\ \vdots \\ \boldsymbol{A}_{(K-1)}\boldsymbol{U}_{(K-1)}, \end{bmatrix},$$

where

$$\boldsymbol{U}_{(k)} \triangleq \begin{bmatrix} \boldsymbol{D}^k & \tilde{\boldsymbol{H}}_{(k)} & \boldsymbol{0}_{N \times (K-1-k)L} \end{bmatrix} \in \mathbb{R}^{N \times (N+(K-1)L)}. \tag{38}$$

To state results similar to Theorem 1 and Theorem 3, we define $\tilde{\delta}_{s,\max}$ as the largest of the RICs among the matrices $\left\{\boldsymbol{U}_{(k)}\right\}_{k=0}^{K-1}$. The proofs of the two theorems below are similar to that of the earlier results, and hence are omitted.

**Theorem 4** (Independent random observation matrices). *Suppose* $\boldsymbol{A}_{(k)}, k = 0, 1, \ldots, K-1$ *are independent subgaussian random matrices with parameter $c$. Then, if*

$$Km \left(\delta - 1 + (1 - \tilde{\delta}_{s,\max})^2\right)^2$$
$$\geq \tilde{c} \left[9s \ln\left(\frac{eN}{s}\right) + 2\ln(2\epsilon^{-1})\right], \quad (39)$$

*the RIC $\delta_s$ of a suitably scaled version $\tilde{\boldsymbol{L}}(\boldsymbol{\delta}) \begin{bmatrix} \tilde{\boldsymbol{A}}_{(K)} & \tilde{\boldsymbol{J}}_{(K)} \end{bmatrix}$ of $\begin{bmatrix} \tilde{\boldsymbol{A}}_{(K)} & \tilde{\boldsymbol{J}}_{(K)} \end{bmatrix}$ satisfies $\delta_s < \delta$ for all $1 - (1 - \tilde{\delta}_{s,\max})^2 < \delta < 1$ with probability at least $1 - \epsilon$. Here, $\tilde{c}$ is a constant dependent only on $c$, and $\boldsymbol{\delta}_j = 1 - \tilde{\delta}_{s,j}$ where $\tilde{\delta}_{s,j}$ is the RIC of matrix $\boldsymbol{U}_{(j)}$. Hence, when (39) holds for $s = \tilde{s}$, the system is observable for sufficiently small $\tilde{\delta}_{s,\max}$ with high probability.*

Next, to state the corresponding result for the identical observation matrices case, we define quantities similar to (17) and (18) as follows:

$$\widetilde{\boldsymbol{U}}_{(K,i)} \triangleq \begin{bmatrix} \boldsymbol{U}_{(0)i} & \boldsymbol{U}_{(1)i} & \ldots & \boldsymbol{U}_{(K-1)i} \end{bmatrix}, \quad (40)$$

$$L_{\boldsymbol{U}}\left(\boldsymbol{D}, \boldsymbol{H}, K\right) \triangleq \max_i \left\|\widetilde{\boldsymbol{U}}_{(K,i)}\right\|_2, \quad (41)$$

where $\boldsymbol{U}_{(k)i}$ denotes the $i^{\text{th}}$ column of $\boldsymbol{U}_{(k)}$.

**Theorem 5** (Identical random observation matrices). *Suppose all observation matrices are identical, i.e., $\boldsymbol{A}_{(k)} = \boldsymbol{A}$ for $k = 0, 1, \ldots, K-1$, where $\boldsymbol{A}$ is a subgaussian random matrix with parameter $c$. Then, if*

$$Km \frac{\left(\delta - 1 + (1 - \tilde{\delta}_{s,\max})^{2(K-1)}\right)^2}{L_{\boldsymbol{U}}^2\left(\boldsymbol{D}, \boldsymbol{H}, K\right)}$$
$$\geq \tilde{c}s \max\left\{\ln^2 s \ln^2 N, \ln(2\epsilon^{-1})\right\}, \quad (42)$$

*then the RIC $\delta_s$ of a suitably scaled version $\tilde{\boldsymbol{L}}(\boldsymbol{\delta}) \begin{bmatrix} \tilde{\boldsymbol{A}}_{(K)} & \tilde{\boldsymbol{J}}_{(K)} \end{bmatrix}$ of $\begin{bmatrix} \tilde{\boldsymbol{A}}_{(K)} & \tilde{\boldsymbol{J}}_{(K)} \end{bmatrix}$ satisfies $\delta_s < \delta$ for all $1 - (1 - \tilde{\delta}_{s,\max})^2 < \delta < 1$ with probability at least $1 - \epsilon$. Here, $\tilde{c}$ is a constant dependent only on $c$, and $\boldsymbol{\delta}_j = 1 - \tilde{\delta}_{s,j}$ where $\tilde{\delta}_{s,j}$ is the RIC of matrix $\boldsymbol{U}_{(j)}$. Hence, when (42) holds for $s = \tilde{s}$, the system is observable for sufficiently small $\tilde{\delta}_{s,\max}$ with high probability.*

*Remark 1:* As before, we can extend the above results to the nonconsecutive measurements, noncanonical basis and robust recovery cases. Also, conditions in Theorem 4 and Theorem 5 can be made less stringent using the RIC of $\boldsymbol{D}$. We omit explicitly stating the results to avoid repetition.

*Remark 2:* The above three theorems show how to extend three main results of the paper (Theorem 1 and Theorem 3) to derive a sufficient condition for the structured random matrix in (36) to satisfy the RIP. These results could be of independent interest: they provide insight to the RIP of two special types of structured random matrices (resulting from independent and identical $\boldsymbol{A}_{(k)}$).

## VI. Conclusions

We derived the conditions for a linear dynamical system to be observable using the knowledge of its noiseless observations and inputs, when the initial state is sparse. We derived the results in the stochastic setting, both when the observation matrices are independent random matrices and when they are identical to a single random matrix. We characterized the number of measurements that are sufficient to observe the state of the linear dynamical system, using tools from compressed sensing. We also extended the results to the joint sparse input and initial state recovery problem. The development of a similar theory on the *controllability* of a linear system under sparsity constraints is an interesting direction for future work.

## Appendix A
## Proof of Theorem 3

Before we prove the theorem, we present a set of mathematical tools used in the proof.

### A. Toolbox

Let $\mathcal{Z} \subset \mathbb{R}^{m \times N}$ be a set of matrices, and the set $\mathcal{T}_s$ denote the set of $s$-sparse vectors in $\mathbb{R}^N$: $\mathcal{T}_s = \left\{\boldsymbol{z} \in \mathbb{R}^N : \|\boldsymbol{z}\| = 1 \text{ and } \|\boldsymbol{z}\|_0 \leq s\right\}$. We need the following two definitions to state the results in this subsection.

**Definition 3** (Admissible sequence). *An admissible sequence $\mathcal{U} = \{\mathcal{U}_i\}_{i=0}^{\infty}$ on $\mathcal{Z}$ is an increasing sequence of partitions of $\mathcal{Z}$ such that $|\mathcal{U}_i| = 2^{2^i}$ and $|\mathcal{U}_0| = 1$. Here, increasing sequence of partitions implies that every set of $\mathcal{U}_i$ is contained in one of the sets of $\mathcal{U}_{i-1}$ for all $i$. Also, $\mathcal{U}_0 = \mathcal{Z}$, and every set of $\mathcal{U}_i$ is a subset of $\mathcal{Z}$. Given a matrix $\boldsymbol{Z} \in \mathcal{Z}$, we denote the unique set of $\mathcal{U}_i$ that contains $\boldsymbol{Z}$ by $\mathcal{U}_i(\boldsymbol{Z})$.*

**Definition 4.** *[Functionals on a set of matrices] We define three functionals on $\mathcal{Z}$ as follows:*

$$d_F\left(\mathcal{Z}\right) \triangleq \sup_{\boldsymbol{Z} \in \mathcal{Z}} \|\boldsymbol{Z}\|_F \quad (43)$$

$$d_2\left(\mathcal{Z}\right) \triangleq \sup_{\boldsymbol{Z} \in \mathcal{Z}} \|\boldsymbol{Z}\|_2 \quad (44)$$

$$\zeta\left(\mathcal{Z}\right) \triangleq \inf_{\mathcal{U} = \{\mathcal{U}_i\}_{i=0}^{\infty}} \sup_{\boldsymbol{Z} \in \mathcal{Z}} \sum_{i=0}^{\infty} 2^{i/2} \mathcal{D}(\mathcal{U}_i(\boldsymbol{Z})), \quad (45)$$

*where the $\inf$ is over all possible admissible sequences, and the term $\mathcal{D}(\mathcal{U}_i(\boldsymbol{Z})) \triangleq \max_{\boldsymbol{U}_{(1)}, \boldsymbol{U}_{(2)} \in \mathcal{U}_i(\boldsymbol{Z})} \left\|\boldsymbol{U}_{(1)} - \boldsymbol{U}_{(2)}\right\|$ represents the diameter the set $\mathcal{U}_i(\boldsymbol{Z})$, which is a decreasing function of $i$.*

Next, we state a result which is the main ingredient of our proof. It bounds the suprema of a chaos process indexed by the set $\mathcal{Z}$.

**Theorem 6.** *[36, Theorem 3.1] Let $\boldsymbol{u}$ be a random vector whose entries are independent zero-mean, unit-variance subgaussian random variables with common parameter $c$. Let*

$$F_1 \triangleq \zeta\left(\mathcal{Z}\right)\left[\zeta\left(\mathcal{Z}\right) + d_F\left(\mathcal{Z}\right)\right] + d_F\left(\mathcal{Z}\right) d_2\left(\mathcal{Z}\right) \quad (46)$$

$$F_2 \triangleq d_2^2\left(\mathcal{Z}\right)\left[\zeta\left(\mathcal{Z}\right) + d_F\left(\mathcal{Z}\right)\right]^2 \quad (47)$$

$$F_3 \triangleq d_2^2\left(\mathcal{Z}\right). \quad (48)$$

*Then, for $t > 0$, it holds that*

$$\mathbb{P}\left\{\sup_{\boldsymbol{Z} \in \mathcal{Z}} \left| \|\boldsymbol{Z}\boldsymbol{u}\|^2 - \mathbb{E}\left\{ \|\boldsymbol{Z}\boldsymbol{u}\|^2 \right\} \right| > c_1 F_1 + t \right\}$$
$$\leq 2 \exp\left( -c_2 \min\left\{ \frac{t^2}{F_2}, \frac{t}{F_3} \right\} \right), \quad (49)$$

*where $c_1, c_2 > 0$ are universal positive constants which depend only on $c$.*

It is difficult to directly apply the above theorem due to the complicated form of the functional $\zeta(\mathcal{Z})$. We need a result that bounds the function $\zeta(\mathcal{Z})$ using the notion of the covering number. The covering number is defined as follows.

**Definition 5** (Covering number). *Given $u > 0$, the covering number $CN\{\mathcal{Z}, u\}$ is defined as the smallest integer $p$ such that one can find a subset $\mathcal{Z}' \subset \mathcal{Z}$ satisfying $|\mathcal{Z}'| \leq p$ and*

$$\min_{\boldsymbol{Z}' \in \mathcal{Z}'} \|\boldsymbol{Z} - \boldsymbol{Z}'\|_2 \leq u, \forall \boldsymbol{Z} \in \mathcal{Z}. \quad (50)$$

**Lemma 1.** *[50]. For the functional $\zeta(\mathcal{Z})$, it holds that*

$$\zeta(\mathcal{Z}) \leq \int_0^\infty \ln^{1/2}\left( CN\{\mathcal{Z}, u\} \right) du. \quad (51)$$

The covering number is hard to compute in closed form. Hence, we use the following lemma to further bound the covering number of the set of interest to us.

**Lemma 2.** *[38, Lemma 6] Let a linear map $\boldsymbol{A} : \mathbb{R}^N \to \mathbb{R}^m$ be such that*

$$\left\| \frac{1}{\sqrt{s}} \boldsymbol{A}(\boldsymbol{z}) \right\| \leq \kappa \|\boldsymbol{z}\|_1, \forall \boldsymbol{z} \in \mathbb{R}^N, \kappa > 0. \quad (52)$$

*Then, for the set $\mathcal{T}_s$, it holds that*

$$\ln\left( CN\{\boldsymbol{A}(\mathcal{T}_s), u\} \right)$$
$$\begin{cases} \leq Cs \min\left\{ \frac{\kappa^2}{u^2} \ln^2 N, s \ln N + s \ln\left(1 + \frac{2\kappa}{u}\right) \right\} \\ \qquad\qquad\qquad\qquad for\ 0 < u < \kappa \quad (53) \\ = 0 \qquad\qquad\qquad\qquad for\ u > \kappa, \end{cases}$$

*where $C$ is a universal positive constant.*

We will also need the following result from elementary calculus in the proof to tackle the integral in Lemma 1.

**Lemma 3.** *For $a \leq b$, it holds that*

$$\int_0^a \sqrt{\ln\left(1 + \frac{b}{u}\right)} du \leq \frac{3}{2} a \sqrt{\ln\left(1 + \frac{b}{a}\right)}. \quad (54)$$

*Proof:* See Appendix B. ∎

We have now presented all the mathematical tools that are required to prove the theorem. In the next subsection, we formally prove the desired result.

### B. Proof of Theorem 3

As mentioned in Section II, (7) is equivalent to (5). Therefore, without loss of generality, we assume that the largest and the smallest singular values of $\boldsymbol{D}$ are 1 and $\lambda$, respectively. We recall that our goal is to obtain a probabilistic bound on $\left\| \widetilde{\boldsymbol{A}}_{(K)} \boldsymbol{z} \right\|$ for $\boldsymbol{z} \in \mathcal{T}_s$, using Theorem 6. At a high level, there are four main steps to the proof:

- First, we convert $\widetilde{\boldsymbol{A}}_{(K)} \boldsymbol{z}$ to the form given in Theorem 6, i.e., the product of a matrix and a subgaussian vector.
- Second, we bound the three functionals $d_2$, $d_F$ and $\zeta$ in Definition 4.
- Third, using the bounds in the previous step, we bound $F_1, F_2$ and $F_3$ in Theorem 6, since the three quantities are functions of $d_2$, $d_F$ and $\zeta$.
- Fourth, we apply Theorem 6 with the upper bounds on $F_1, F_2$ and $F_3$. Here, we note that Theorem 6 holds for upper bounds on $F_1, F_2$ and $F_3$. This yields a concentration inequality bounding the deviation of the random variable $\left\| \frac{1}{\sqrt{Km}} \widetilde{\boldsymbol{A}}_{(K)} \boldsymbol{z} \right\|^2$ from its mean $\frac{1}{K} \sum_{k=0}^{K-1} \left\| \boldsymbol{D}^k \boldsymbol{z} \right\|^2$. Finally, we establish the desired result by suitable algebraic manipulation of the concentration inequality.

In the remainder of this section, we provide the details of each of these steps.

For the first step, we consider the following:

$$\left\| \frac{1}{\sqrt{Km}} \widetilde{\boldsymbol{A}}_{(K)} \boldsymbol{z} \right\|^2 = \sum_{k=0}^{K-1} \left\| \frac{1}{\sqrt{Km}} \boldsymbol{A} \boldsymbol{D}^k \boldsymbol{z} \right\|^2 \quad (55)$$

$$= \left\| \frac{1}{\sqrt{Km}} \boldsymbol{A} \boldsymbol{Z}_{(K)}(\boldsymbol{z}) \right\|_F^2 \quad (56)$$

where the matrix function $\boldsymbol{Z}_{(K)} : \mathcal{T}_s \to \mathbb{R}^{N \times K}$ is defined as follows:

$$\boldsymbol{Z}_{(K)}(\boldsymbol{z}) \triangleq \begin{bmatrix} \boldsymbol{z} & \boldsymbol{D}\boldsymbol{z} & \dots \boldsymbol{D}^{K-1}\boldsymbol{z} \end{bmatrix}, \boldsymbol{z} \in \mathcal{T}_s. \quad (57)$$

Further, we have

$$\left\| \frac{1}{\sqrt{Km}} \widetilde{\boldsymbol{A}}_{(K)} \boldsymbol{z} \right\|^2 = \left\| \frac{1}{\sqrt{Km}} \boldsymbol{Z}_{(K)}^{\mathrm{T}}(\boldsymbol{z}) \boldsymbol{A}^{\mathrm{T}} \right\|_F^2 \quad (58)$$

$$= \left\| \widetilde{\boldsymbol{Z}}_{(K)}(\boldsymbol{z}) \mathrm{vec}\left\{ \boldsymbol{A}^{\mathrm{T}} \right\} \right\|^2, \quad (59)$$

where $\mathrm{vec}\left\{ \boldsymbol{A}^{\mathrm{T}} \right\} \in \mathbb{R}^{mN}$ is the vectorized version of the matrix $\boldsymbol{A}^{\mathrm{T}}$, which has subgaussian entries with common parameter $c$. The matrix function $\widetilde{\boldsymbol{Z}}_{(K)} : \mathcal{T}_s \to \mathbb{R}^{Km \times mN}$ is a block diagonal matrix with $\frac{1}{\sqrt{Km}} \boldsymbol{Z}_{(K)}^{\mathrm{T}}(\boldsymbol{z})$ as the $m^{\mathrm{th}}$ block diagonal entries, for all $\boldsymbol{z} \in \mathcal{T}_s$. Thus, the first step is complete.

The next step is bound the three terms $d_2$, $d_F$ and $\zeta$ using the following lemmas.

**Lemma 4.** *For the set $\widetilde{\mathcal{Z}}_{(K)}$,*

$$d_F\left( \widetilde{\mathcal{Z}}_{(K)} \right) \leq 1 \quad (60)$$

$$d_2\left( \widetilde{\mathcal{Z}}_{(K)} \right) \leq \sqrt{\frac{s}{Km}} L(\boldsymbol{D}, K), \quad (61)$$

*where $L(\boldsymbol{D}, K)$ is as defined in (18).*

*Proof:* See Appendix C. ∎

**Lemma 5.** *The functional $\zeta\left(\widetilde{\mathcal{Z}}_{(K)}\right)$ can be bounded as follows:*

$$\zeta\left(\widetilde{\mathcal{Z}}_{(K)}\right) \leq C\sqrt{\frac{s}{Km}} L\left(\boldsymbol{D}, K\right) \ln N \ln s, \qquad (62)$$

*for some $C$ which is universal positive constant that depends only on the subgaussian parameter c.*

*Proof:* See Appendix D. ∎

Now, we combine the results in the second step to obtain bounds on $F_1, F_2$, and $F_3$. Further, we note that we need to bound $\mathbb{P}\left\{\left|\left\|\frac{1}{\sqrt{Km}}\widetilde{\boldsymbol{A}}_{(K)}\boldsymbol{z}\right\|^2 - \frac{1}{K}\sum_{k=0}^{K-1}\left\|\boldsymbol{D}^k\boldsymbol{z}\right\|^2\right| > \widetilde{\delta}\right\}$, for some $0 \leq \widetilde{\delta} < 1$ using Theorem 6. To this end, we use the assumptions of Theorem 3 to further upper bound $F_1$ to make it a multiple of $\widetilde{\delta}$. We summarize the third step in the following lemma.

**Lemma 6.** *Using the definitions in Theorem 6 and under the assumptions of Theorem 3, there exists a constant $C_1 > 0$ such that*

$$F_1 \leq \widetilde{\delta}/2c_1 \qquad (63)$$

$$F_2 \leq C_1 \frac{s}{Km} L^2\left(\boldsymbol{D}, K\right) \qquad (64)$$

$$F_3 \leq \frac{s}{Km} L^2\left(\boldsymbol{D}, K\right), \qquad (65)$$

*where $c_1$ is the same constant as in Theorem 6 and*

$$\widetilde{\delta} \triangleq \delta - 1 + \lambda^{2(K-1)}. \qquad (66)$$

*Proof:* See Appendix E. ∎

Now, we are ready to go the final step of the proof. We apply Theorem 6 to (59) with $t = \widetilde{\delta}/2$ to get

$$\mathbb{P}\left\{\left|\left\|\frac{1}{\sqrt{Km}}\widetilde{\boldsymbol{A}}_{(K)}\boldsymbol{z}\right\|^2 - \frac{1}{K}\sum_{k=0}^{K-1}\left\|\boldsymbol{D}^k\boldsymbol{z}\right\|^2\right| > \widetilde{\delta}\right\}$$
$$\leq 2\exp\left(-C\min\left\{\widetilde{\delta}^2, \widetilde{\delta}\right\}\frac{Km}{sL^2\left(\boldsymbol{D}, K\right)}\right) \qquad (67)$$

$$\leq 2\exp\left(-C\frac{Km\widetilde{\delta}^2}{sL^2\left(\boldsymbol{D}, K\right)}\right) \qquad (68)$$

$$\leq \epsilon, \qquad (69)$$

where the universal positive constant $C$ depends on the subgaussian parameter $c$, and we use (19) of Theorem 3 to bound using $\epsilon$ in the last step.

Thus, for all $\boldsymbol{z} \in \mathbb{R}^N$ such that $\|\boldsymbol{z}\| = 1$ and $\|\boldsymbol{z}\|_0 \leq s$, with probability at least $1 - \epsilon$,

$$\left|\left\|\frac{1}{\sqrt{Km}}\widetilde{\boldsymbol{A}}_{(K)}\boldsymbol{z}\right\|^2 - \frac{1}{K}\sum_{k=0}^{K-1}\left\|\boldsymbol{D}^k\boldsymbol{z}\right\|^2\right| \leq \widetilde{\delta}. \qquad (70)$$

Therefore, for $0 \leq \widetilde{\delta} < \lambda^{2(K-1)}$,

$$\lambda^{2(K-1)} - \widetilde{\delta} < \frac{1}{Km}\left\|\widetilde{\boldsymbol{A}}_{(K)}\boldsymbol{z}\right\|^2 < 1 + \widetilde{\delta}, \qquad (71)$$

since $\lambda^{K-1} \leq \lambda^k \leq \left\|\boldsymbol{D}^k\boldsymbol{z}\right\| \leq 1$. We also use (66) to relate $\delta$ and $\widetilde{\delta}$ as follows:

$$1 - \delta < \frac{1}{Km}\left\|\widetilde{\boldsymbol{A}}_{(K)}\boldsymbol{z}\right\|^2 < 1 + \delta, \qquad (72)$$

for $\delta > 1 - \lambda^{2(K-1)}$, with probability at least $1 - \epsilon$. Hence, $\frac{1}{Km}\widetilde{\boldsymbol{A}}_{(K)}$ satisfies RIP of order $s$ with RIC as $\delta$, with probability at least $1 - \epsilon$. Thus, the proof is complete.

# APPENDIX B
## PROOF OF LEMMA 3

We have

$$\int_0^a \sqrt{\ln\left(1 + \frac{b}{u}\right)}\, du$$
$$= b\int_{\sqrt{\ln(1+b/a)}}^{\infty} t\, d\left(\frac{1}{\exp(t^2) - 1}\right) \qquad (73)$$
$$= a\sqrt{\ln\left(1 + \frac{b}{a}\right)} + b\int_{\sqrt{\ln(1+b/a)}}^{\infty} \frac{1}{\exp(t^2) - 1}\, dt, \qquad (74)$$

where we use the substitution $t = \sqrt{\ln\left(1 + \frac{b}{u}\right)}$ in (73) and integration by parts to get (74). Now, the second term simplifies as follows:

$$\int_{\sqrt{\ln(1+b/a)}}^{\infty} \frac{1}{\exp(t^2) - 1}\, dt$$
$$\leq \int_{\sqrt{\ln(1+b/a)}}^{\infty} \frac{t}{\sqrt{\ln(1+b/a)}} \frac{e^{-t^2}}{1 - e^{-t^2}}\, dt \qquad (75)$$
$$= \frac{1}{2\sqrt{\ln(1+b/a)}} \ln\left(1 + \frac{a}{b}\right). \qquad (76)$$

Therefore, we get

$$\int_0^a \sqrt{\ln\left(1 + \frac{b}{u}\right)}\, du$$
$$\leq a\sqrt{\ln\left(1 + \frac{b}{a}\right)} + \frac{b}{2\sqrt{\ln(1+b/a)}} \ln\left(1 + \frac{a}{b}\right) \qquad (77)$$
$$= a\sqrt{\ln\left(1 + \frac{b}{a}\right)} \left[1 + \frac{b}{2a}\left(1 - \frac{\ln\left(\frac{b}{a}\right)}{\ln(1+b/a)}\right)\right]. \qquad (78)$$

Now, we need to show that $\frac{b}{a}\left(1 - \frac{\ln\left(\frac{b}{a}\right)}{\ln(1+b/a)}\right) \leq 1$ to complete the proof. So, we consider the function $h(u) \triangleq u\left(1 - \frac{\ln u}{\ln(u+1)}\right)$, by replacing $b/a = u \geq 1$. Further, we note that $h(1) = 1$, and therefore it suffices to show that $\frac{d}{du}h(u) \leq 0$, which then implies that $h(u) \leq h(1) = 1$, for all $u \geq 1$. We have

$$\frac{d}{du}h(u) = 1 - \frac{\ln u}{\ln(u+1)} - \frac{(u+1)\ln(u+1) - u\ln u}{(u+1)\ln^2(u+1)} \qquad (79)$$

$$= \frac{\widetilde{h}(u)}{(u+1)\ln^2(u+1)}, \qquad (80)$$

where we define

$$\widetilde{h}(u) \triangleq (u+1)\ln^2(u+1) - (u+1)\ln(u+1)\ln u$$
$$- (u+1)\ln(u+1) + u\ln u. \qquad (81)$$

Now, $\frac{d}{du}h(u) \leq 0$ if $\tilde{h}(u) \leq 0$. Therefore, we show that $\frac{d}{du}\tilde{h}(u) \leq 0$, which implies that $\tilde{h}(u) \leq \tilde{h}(1) = 2\ln^2 2 - 2\ln 2 < 0$. Then, we get

$$\frac{d}{du}\tilde{h}(u) = \ln(u+1)\left(\ln(u+1) - \ln u - \frac{u+1}{u}\right) \quad (82)$$

$$= -\ln(u+1)\left(\ln u + 1 - \ln(u+1) + \frac{1}{u}\right). \quad (83)$$

Using the same technique again, we now consider the function $\ln u + 1 - \ln(u+1)$. Since derivative of $\ln u + 1 - \ln(u+1)$ is $\frac{1}{u(u+1)} > 0$, for $u \geq 1$, $\ln u + 1 - \ln(u+1) \geq 1 - \ln 2 > 0$. Therefore, $\frac{d}{du}\tilde{h}(u) \leq 0$ because $\ln u + 1 - \ln(u+1) \geq 0$ and $\frac{1}{u} \geq 0$, for $u \geq 1$. Hence, we get $\tilde{h}(u) < 0$. This implies that $\frac{d}{du}h(u) < 0$, and thus, $h(u) \leq h(1) = 1$, for $u \geq 1$. Substituting this in (78) completes the proof.

## APPENDIX C
## PROOF OF LEMMA 4

To show the first part of the lemma, we have

$$d_F\left(\widetilde{\mathcal{Z}}_{(K)}\right) = \sup_{z \in \mathcal{T}_s}\left\|\widetilde{\boldsymbol{Z}}_{(K)}(z)\right\|_F \quad (84)$$

$$= \frac{1}{\sqrt{K}}\sup_{z \in \mathcal{T}_s}\left\|\boldsymbol{Z}_{(K)}(z)\right\|_F \quad (85)$$

$$= \sup_{z \in \mathcal{T}_s}\frac{1}{\sqrt{K}}\sqrt{\sum_{k=0}^{K-1}\left\|\boldsymbol{D}^k z\right\|^2} \leq 1, \quad (86)$$

where the last step follows from the definition of $\mathcal{T}_s$, and the fact that the largest singular value of $\boldsymbol{D}$ is unity.

To show the second part of the lemma, we have

$$d_2\left(\widetilde{\mathcal{Z}}_{(K)}\right) = \sup_{z \in \mathcal{T}_s}\left\|\widetilde{\boldsymbol{Z}}_{(K)}(z)\right\|_2 \quad (87)$$

$$= \frac{1}{\sqrt{Km}}\sup_{z \in \mathcal{T}_s}\left\|\boldsymbol{Z}_{(K)}(z)\right\|_2 \quad (88)$$

$$= \frac{1}{\sqrt{Km}}\sup_{z \in \mathcal{T}_s}\left\|\sum_{i=1}^N \widetilde{\boldsymbol{D}}_{(K,i)}z_i\right\|_2 \quad (89)$$

$$\leq \frac{1}{\sqrt{Km}}\sup_{z \in \mathcal{T}_s}\sum_{i=1}^N |z_i|\left\|\widetilde{\boldsymbol{D}}_{(K,i)}\right\|_2 \quad (90)$$

$$\leq \frac{L(\boldsymbol{D}, K)}{\sqrt{Km}}\sup_{z \in \mathcal{T}_s}\|z\|_1 \quad (91)$$

$$\leq L(\boldsymbol{D}, K)\sqrt{\frac{s}{Km}}\|z\|, \quad (92)$$

$$= \sqrt{\frac{s}{Km}}L(\boldsymbol{D}, K). \quad (93)$$

where (89) and (91) follow from the definitions of $\widetilde{\boldsymbol{D}}_{(K,i)}$ and $L(\boldsymbol{D}, K)$ in (17) and (18), respectively. Also, (92) is because $z$ is at most $s-$sparse. Hence, the proof is complete.

## APPENDIX D
## PROOF OF LEMMA 5

From Lemma 4, for all $\boldsymbol{Z} \in \widetilde{\mathcal{Z}}_{(K)}$ and any $z \in \mathbb{R}^{mN}$,

$$\left\|\frac{1}{\sqrt{s}}\boldsymbol{Z}z\right\| \leq \sqrt{\frac{1}{Km}}L(\boldsymbol{D}, K)\|z\| \leq \sqrt{\frac{1}{Km}}L(\boldsymbol{D}, K)\|z\|_1. \quad (94)$$

Then, for some positive constant $C'$, we have

$$\frac{1}{\sqrt{s}}\zeta\left(\widetilde{\mathcal{Z}}_{(K)}\right)$$

$$\leq \frac{1}{\sqrt{s}}\int_0^\infty \ln^{1/2}\left[\text{CN}\left\{\widetilde{\mathcal{Z}}_{(K)}, u\right\}\right]du \quad (95)$$

$$= C'\frac{1}{\sqrt{s}}\int_0^{\frac{L(\boldsymbol{D}, K)}{\sqrt{sKm}}}\ln^{1/2}\left[\text{CN}\left\{\widetilde{\mathcal{Z}}_{(K)}, u\right\}\right]du$$

$$+ C'\frac{1}{\sqrt{s}}\int_{\frac{L(\boldsymbol{D}, K)}{\sqrt{sKm}}}^{\frac{L(\boldsymbol{D}, K)}{\sqrt{Km}}}\ln^{1/2}\left[\text{CN}\left\{\widetilde{\mathcal{Z}}_{(K),u}\right\}\right]du \quad (96)$$

$$\leq C'\int_0^{\frac{L(\boldsymbol{D}, K)}{\sqrt{sKm}}}\sqrt{s\ln N + s\ln\left(1 + \frac{2L(\boldsymbol{D}, K)}{u\sqrt{Km}}\right)}du$$

$$+ C'\int_{\frac{L(\boldsymbol{D}, K)}{\sqrt{sKm}}}^{\frac{L(\boldsymbol{D}, K)}{\sqrt{Km}}}\frac{L(\boldsymbol{D}, K)}{u\sqrt{Km}}\ln N du, \quad (97)$$

where (95) and (97) follow from Lemma 1 and Lemma 2 with $\kappa = \sqrt{\frac{1}{Km}}L(\boldsymbol{D}, K)$, respectively. Further, we have

$$\frac{1}{\sqrt{s}}\zeta\left(\widetilde{\mathcal{Z}}_{(K)}\right)$$

$$\leq C'\int_0^{\frac{L(\boldsymbol{D}, K)}{\sqrt{sKm}}}\sqrt{s\ln N} + \sqrt{s\ln\left(1 + \frac{2L(\boldsymbol{D}, K)}{u\sqrt{Km}}\right)}du$$

$$+ C'\frac{L(\boldsymbol{D}, K)}{\sqrt{Km}}\ln N \ln\sqrt{s} \quad (98)$$

$$\leq C'\frac{L(\boldsymbol{D}, K)}{\sqrt{Km}}\left(\sqrt{\ln N} + 3/2\sqrt{\ln(1 + 2\sqrt{s})}\right.$$

$$\left. + \ln N \ln\sqrt{s}\right) \quad (99)$$

$$\leq C\frac{L(\boldsymbol{D}, K)}{\sqrt{Km}}\ln N \ln s \quad (100)$$

where $C = 3C'$. Also, (98) uses the fact that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, for any $a, b > 0$, and (99) uses Lemma 3. Thus, the proof is complete.

## APPENDIX E
## PROOF OF LEMMA 6

From Lemma 4 and Lemma 5, we get

$$F_1 = \zeta\left(\widetilde{\mathcal{Z}}_{(K)}\right)\left[\zeta\left(\widetilde{\mathcal{Z}}_{(K)}\right) + d_F\left(\widetilde{\mathcal{Z}}_{(K)}\right)\right]$$

$$+ d_F\left(\widetilde{\mathcal{Z}}_{(K)}\right)d_2\left(\widetilde{\mathcal{Z}}_{(K)}\right) \quad (101)$$

$$\leq CL(\boldsymbol{D}, K)\sqrt{\frac{s}{Km}}\ln N \ln s$$

$$\left(CL(\boldsymbol{D}, K)\sqrt{\frac{s}{Km}}\ln N \ln s + 1\right)$$

$$+ L(\boldsymbol{D}, K)\sqrt{\frac{s}{Km}}\frac{\ln N \ln s}{\ln^2 2}, \quad (102)$$

where we use the bound $\frac{\ln N \ln s}{\ln^2 2} > 1$ when $N \geq s > 1$ to get (102). Next, we use assumption (19) in Theorem 3, i.e.,

$$L(\boldsymbol{D}, K)\sqrt{\frac{s}{Km}}\ln N \ln s \leq \sqrt{\tilde{c}\delta}, \quad (103)$$

to get

$$F_1 \leq C\sqrt{\widetilde{c}}\widetilde{\delta}\left(C\sqrt{\widetilde{c}}\widetilde{\delta} + 1\right) + \frac{\sqrt{\widetilde{c}}\widetilde{\delta}}{\ln^2 2} \qquad (104)$$

$$\leq \sqrt{\widetilde{c}}\widetilde{\delta}\left(C^2\sqrt{\widetilde{c}} + C + \frac{1}{\ln^2 2}\right). \qquad (105)$$

The last step above follows because of the bound $\widetilde{\delta} \leq 1$. Finally, we choose $\widetilde{c}$ in (19) of Theorem 3 such that

$$\sqrt{\widetilde{c}}\left(C^2\sqrt{\widetilde{c}} + C + \frac{1}{\ln^2 2}\right) \leq \frac{1}{2c_1}, \qquad (106)$$

where $c_1$ and $C$ are the same constants as in Theorem 6 and Lemma 5, respectively. We note that $\sqrt{\widetilde{c}}\left(C^2\sqrt{\widetilde{c}} + C + \frac{1}{\ln^2 2}\right)$ is strictly increasing with $\widetilde{c}$, for $\widetilde{c} \geq 0$, and the left hand side equals zero when $\widetilde{c} = 0$. Thus, there exists a $\widetilde{c} > 0$ that satisfies (106), for any $c_1$ and $C$. Thus, from (105) and (106) we get

$$F_1 \leq \frac{\widetilde{\delta}}{2c_1}, \qquad (107)$$

and thus, we complete of the first part of the proof.

Similarly, we simplify $F_2$ using Lemma 4, Lemma 5 and (103) as follows:

$$F_2 = d_2^2\left(\widetilde{\mathcal{Z}}_{(K)}\right)\left[\zeta\left(\widetilde{\mathcal{Z}}_{(K)}\right) + d_F\left(\widetilde{\mathcal{Z}}_{(K)}\right)\right]^2 \qquad (108)$$

$$\leq \frac{s}{Km}L^2\left(\boldsymbol{D}, K\right)\left(CL\left(\boldsymbol{D}, K\right)\sqrt{\frac{s}{Km}}\ln N\ln s + 1\right)^2 \qquad (109)$$

$$\leq \frac{s}{Km}L^2\left(\boldsymbol{D}, K\right)\left(C\sqrt{\widetilde{c}}\widetilde{\delta} + 1\right)^2 \qquad (110)$$

$$\leq C_1\frac{s}{Km}L^2\left(\boldsymbol{D}, K\right) \qquad (111)$$

where we use the fact that $\widetilde{\delta} < 1$ and define $C_1 \triangleq (C\sqrt{\widetilde{c}}+1)^2$. Finally, we have,

$$F_3 = d_2^2\left(\widetilde{\mathcal{Z}}_{(K)}\right) \leq \frac{s}{Km}L^2\left(\boldsymbol{D}, K\right), \qquad (112)$$

which completes the proof.

## APPENDIX F
## PROOF OF PROPOSITION 1

To prove the result, we first upper and lower bound the term $L(\boldsymbol{D}, K)$. We have,

$$L(\boldsymbol{D}, K) \leq \max_i\left\|\widetilde{\boldsymbol{D}}_{(K,i)}\right\|_F \qquad (113)$$

$$= \max_i\sqrt{\sum_{k=0}^{K-1}\left\|\boldsymbol{D}_i^k\right\|^2} \leq \sqrt{K}, \qquad (114)$$

where we obtain the last step from the fact that the largest singular value of $\boldsymbol{D}^k$ is at most unity, and Rayleigh-Ritz theorem [51, Theorem 4.2.2] which gives

$$1 = \sup_{\boldsymbol{z}\in\mathbb{R}^N, \boldsymbol{z}\neq 0} \frac{\left\|\boldsymbol{z}^{\mathrm{T}}\boldsymbol{D}^{k\mathrm{T}}\boldsymbol{D}^k\boldsymbol{z}\right\|}{\left\|\boldsymbol{z}\right\|^2}$$

$$\geq \max_i\left(\boldsymbol{D}^{k\mathrm{T}}\boldsymbol{D}^k\right)_{i,i} = \max_i\left\|\boldsymbol{D}_i^k\right\|^2. \qquad (115)$$

Similarly, we also have,

$$L(\boldsymbol{D}, K)^2 = \max_{i=1,2,\ldots,N}\left[\sup_{\boldsymbol{z}\in\mathbb{R}^K, \boldsymbol{z}\neq 0}\frac{\left\|\boldsymbol{z}^{\mathrm{T}}\widetilde{\boldsymbol{D}}_{(K,i)}^{\mathrm{T}}\widetilde{\boldsymbol{D}}_{(K,i)}\boldsymbol{z}\right\|}{\left\|\boldsymbol{z}\right\|^2}\right] \qquad (116)$$

$$\geq \max_{\substack{i=1,2,\ldots,N \\ k=0,1,\ldots,K-1}}\left(\widetilde{\boldsymbol{D}}_{(K,i)}^{\mathrm{T}}\widetilde{\boldsymbol{D}}_{(K,i)}\right)_{k,k} \qquad (117)$$

$$= \max_i\left[\max_{k=0,1,\ldots,K-1}\left\|\boldsymbol{D}_i^k\right\|^2\right] = 1, \qquad (118)$$

where (118) uses the fact that $\left\|\boldsymbol{D}_i^0\right\|^2 = \left\|\boldsymbol{I}_i\right\|^2 = 1$ and $\left\|\boldsymbol{D}_i^k\right\|^2 \leq 1$, for $k = 1, 2, \ldots, K-1$ from (115). Combining (114) and (118), we obtain the desired result.

## APPENDIX G
## PROOF OF PROPOSITION 2

To prove the monotonicity of the function, we need to show that

$$L^2\left(\boldsymbol{D}, K+1\right) \leq L^2\left(\boldsymbol{D}, K\right)\left(1 + \frac{1}{K}\right). \qquad (119)$$

Therefore, we consider the following relation connecting $L^2\left(\boldsymbol{D}, K+1\right)$ and $L^2\left(\boldsymbol{D}, K\right)$:

$$L^2\left(\boldsymbol{D}, K+1\right) = \max_i\left\|\widetilde{\boldsymbol{D}}_{(K+1,i)}\right\|_2 \qquad (120)$$

$$= \max_i\sup_{\substack{\boldsymbol{z}\in\mathbb{R}^N \\ \|\boldsymbol{z}\|=1}}\boldsymbol{z}^{\mathrm{T}}\widetilde{\boldsymbol{D}}_{(K+1,i)}\widetilde{\boldsymbol{D}}_{(K+1,i)}^{\mathrm{T}}\boldsymbol{z} \qquad (121)$$

$$= \max_i\sup_{\substack{\boldsymbol{z}\in\mathbb{R}^N \\ \|\boldsymbol{z}\|=1}}\sum_{j=0}^{K}\left|\boldsymbol{z}^{\mathrm{T}}\boldsymbol{D}^j\boldsymbol{I}_i\right|^2 \qquad (122)$$

$$\leq \max_i\sup_{\substack{\boldsymbol{z}\in\mathbb{R}^N \\ \|\boldsymbol{z}\|=1}}\sum_{j=0}^{K-1}\left|\boldsymbol{z}^{\mathrm{T}}\boldsymbol{D}^j\boldsymbol{I}_i\right|^2$$

$$+ \max_i\sup_{\substack{\boldsymbol{z}\in\mathbb{R}^N \\ \|\boldsymbol{z}\|=1}}\left|\boldsymbol{z}^{\mathrm{T}}\boldsymbol{D}^K\boldsymbol{I}_i\right|^2 \qquad (123)$$

$$\leq L^2\left(\boldsymbol{D}, K\right) + \max_i\left\|\boldsymbol{D}^K\boldsymbol{I}_i\right\|^2, \qquad (124)$$

where we use the definition of $L^2\left(\boldsymbol{D}, K\right)$ and Cauchy-Schwarz inequality to get the last step.

Now, to complete the proof, it suffices to show that

$$\max_i\left\|\boldsymbol{D}^K\boldsymbol{I}_i\right\|^2 \leq L^2\left(\boldsymbol{D}, K\right)/K. \qquad (125)$$

Since $\boldsymbol{D}$ is a psd matrix with largest singular value as unity, $\boldsymbol{D}^j - \boldsymbol{D}^k$ is a psd matrix, for any pair of integers $j \leq k$. Then,

we have

$$K^2 \left\| \boldsymbol{D}^K \boldsymbol{I}_i \right\|^2 = K^2 \boldsymbol{I}_i^{\mathrm{T}} \boldsymbol{D}^{2K} \boldsymbol{I}_i \leq \sum_{j=0}^{K-1} \sum_{k=0}^{K-1} \boldsymbol{I}_i^{\mathrm{T}} \boldsymbol{D}^{j+k} \boldsymbol{I}_i \tag{126}$$

$$= \sum_{j=0}^{K-1} \sum_{k=0}^{K-1} \boldsymbol{I}_i^{\mathrm{T}} \boldsymbol{D}^{j\mathrm{T}} \boldsymbol{D}^k \boldsymbol{I}_i \tag{127}$$

$$= K \left( \boldsymbol{1}/\sqrt{K} \right)^{\mathrm{T}} \widetilde{\boldsymbol{D}}_{(K+1,i)}^{\mathrm{T}} \widetilde{\boldsymbol{D}}_{(K+1,i)} \left( \boldsymbol{1}/\sqrt{K} \right) \tag{128}$$

$$\leq K L^2 \left( \boldsymbol{D}, K \right). \tag{129}$$

Hence, (125) holds, which in turn shows (119). Thus, the proof is complete.

## References

[1] G. Joseph and C. R. Murthy, "On the observability of a linear system with a sparse initial state," *IEEE Signal Process. Lett.*, vol. 25, no. 7, pp. 994 – 998, Jul. 2018.

[2] R. Kalman, "On the general theory of control systems," *IRE Trans. Autom. Control*, vol. 4, no. 3, pp. 110–110, Dec 1959.

[3] C. T. Chen, *Linear System Theory and Design*. Oxford University Press, 1999.

[4] C. Böß, "Using model reduction techniques within the incremental 4D-Var method," Ph.D. dissertation, Universität Bremen, 2008.

[5] M. Verlaan, "Efficient Kalman filtering algorithms for hydrodynamic models," Ph.D. dissertation, Delft University of Technology, 1998.

[6] A. Lawless, N. Nichols, C. Boess, and A. Bunse-Gerstner, "Using model reduction methods within incremental four-dimensional variational data assimilation," *Monthly Weather Review*, vol. 136, no. 4, pp. 1511–1522, Apr. 2008.

[7] G. Neumann, T. Noda, and Y. Kawaoka, "Emergence and pandemic potential of swine-origin H1N1 influenza virus," *Nature*, vol. 459, no. 7249, pp. 931–939, Jun. 2009.

[8] M. Hvistendahl, D. Normile, and J. Cohen, "Despite large research effort, H7N9 continues to baffle," *Science*, vol. 340, no. 6131, pp. 414–415, Apr. 2013.

[9] C. A. Pope III, R. T. Burnett, M. J. Thun, E. E. Calle, D. Krewski, K. Ito, and G. D. Thurston, "Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution," *Jama*, vol. 287, no. 9, pp. 1132–1141, Mar. 2002.

[10] M. Shao, X. Tang, Y. Zhang, and W. Li, "City clusters in China: air and surface water pollution," *Frontiers in Ecology and the Environment*, vol. 4, no. 7, pp. 353–361, Sep. 2006.

[11] A. L. Lloyd and R. M. May, "How viruses spread among computers and people," *Science*, vol. 292, no. 5520, pp. 1316–1317, May 2001.

[12] P. Wang, M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding the spreading patterns of mobile phone viruses," *Science*, vol. 324, no. 5930, pp. 1071–1076, May 2009.

[13] M. Cha, A. Mislove, and K. P. Gummadi, "A measurement-driven analysis of information propagation in the flickr social network," in *Proc. WWW*, Apr. 2009, pp. 721–730.

[14] I. Miholic, "Information propagation in online social networks," Bachelor's thesis, University of Zagreb, 2014.

[15] Z.-L. Hu, X. Han, Y.-C. Lai, and W.-X. Wang, "Optimal localization of diffusion sources in complex networks," *R. Soc. Open Sci.*, vol. 4, no. 4, p. 170091, Apr. 2017.

[16] M. B. Wakin, B. M. Sanandaji, and T. L. Vincent, "On the observability of linear systems from random, compressive measurements," in *Proc. CDC*, Dec. 2010, pp. 4447 – 4454.

[17] B. M. Sanandaji, M. B. Wakin, and T. L. Vincent, "Technical report: Observability with random observations," *CoRR*, vol. abs/1211.4077, 2013. [Online]. Available: http://arxiv.org/abs/1211.4077

[18] ——, "Observability with random observations," *IEEE Trans. Autom. Control*, vol. 59, no. 11, pp. 3002–3007, Nov. 2014.

[19] K. P. Singh, A. Malik, and S. Sinha, "Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques-a case study," *Analytica Chimica Acta*, vol. 538, no. 1, pp. 355 – 374, May 2005.

[20] L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[21] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[22] R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[23] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[24] T. T. Cai and A. Zhang, "Sharp RIP bound for sparse signal and low-rank matrix recovery," *CoRR*, vol. abs/1302.1236, 2013. [Online]. Available: http://arxiv.org/abs/1302.1236

[25] ——, "Sparse representation of a polytope and recovery of sparse signals and low-rank matrices," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 122–132, Jan. 2014.

[26] R. Zhang and S. Li, "A proof of conjecture on restricted isometry property constants $\delta_{tk}$ $\left( 0 < t < \frac{4}{3} \right)$," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1699–1705, Mar. 2018.

[27] ——, "Optimal rip bounds for sparse signals recovery via $\ell_p$ minimization," *Appl. Comput. Harmon. Anal.*, Oct. 2017.

[28] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Birkhäuser, 2013.

[29] J. Wen, Z. Zhou, J. Wang, X. Tang, and Q. Mo, "A sharp condition for exact support recovery of sparse signals with orthogonal matching pursuit," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2016, pp. 2364 – 2368.

[30] S. Bhattacharya and T. Basar, "Sparsity based feedback design: A new paradigm in opportunistic sensing," in *Proc. ACC*, Jun. 2011, pp. 3704 – 3709.

[31] W. Dai and S. Yüksel, "Observability of a linear system under sparsity constraints," *IEEE Trans. Autom. Control*, vol. 58, no. 9, pp. 2372–2376, Sep. 2013.

[32] S. Sefati, N. J. Cowan, and R. Vidal, "Linear systems with sparse inputs: Observability and input recovery," in *Proc. ACC*, Jul. 2015, pp. 5251 – 5257.

[33] H. Rauhut, "Compressive sensing and structured random matrices," in *Theoretical foundations and numerical methods for sparse recovery*, M. Fornasier, Ed. Walter de Gruyter, 2011, ch. 1–10, pp. 1–92.

[34] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Comm. Pure Appl. Math.*, vol. 61, no. 8, pp. 1025–1045, Aug. 2008.

[35] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5862–5875, Nov 2010.

[36] F. Krahmer, S. Mendelson, and H. Rauhut, "Suprema of chaos processes and the restricted isometry property," *Comm. Pure Appl. Math.*, vol. 67, no. 11, pp. 1877–1904, Nov. 2014.

[37] H. Rauhut, J. K. Romberg, and J. A. Tropp, "Restricted isometries for partial random circulant matrices," *Appl. Comput. Harmon. Anal.*, vol. 32, no. 2, pp. 242–254, Mar. 2012.

[38] A. Eftekhari, H. Yap, C. Rozell, and M. Wakin, "The restricted isometry property for random block diagonal matrices," *Appl. Comput. Harmon. Anal.*, vol. 38, no. 1, pp. 1–31, Jan. 2015.

[39] S. Khanna and C. R. Murthy, "On the restricted isometry of the columnwise Khatri-Rao product," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1170 – 1183, Dec. 2017.

[40] Q. Mo and S. Li, "New bounds on the restricted isometry constant $\delta_{2k}$," *Appl. Comput. Harmon. Anal.*, vol. 31, no. 3, pp. 460 – 468, 2011.

[41] T. T. Cai, L. Wang, and G. Xu, "New bounds for restricted isometry constants," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4388–4394, Sep. 2010.

[42] T. Tao and V. Vu, "Random matrices: The distribution of the smallest singular values," *Geom. Funct. Anal.*, vol. 20, no. 1, pp. 260–297, Jun. 2010.

[43] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, "Epidemic thresholds in real networks," *ACM Trans. Inf. Syst. Security*, vol. 10, no. 4, p. 1, Jan 200-8.

[44] R. Durrett, "Some features of the spread of epidemics and information on a random graph," *Proc. of the National Acad. of Sciences*, Feb. 2010.

[45] ——, *Random graph dynamics*. Cambridge university press Cambridge, 2007.

[46] M. E. Newman, S. Forrest, and J. Balthrop, "Email networks and the spread of computer viruses," *Physical Review E*, vol. 66, no. 3, p. 035101, Sep. 2002.

[47] L. A. Pastur and M. Shcherbina, *Eigenvalue distribution of large random matrices*. American Math. Soc., 2011.

[48] M. Nagahara and D. E. Quevedo, "Sparse representations for packetized predictive networked control," *IFAC Proc. Vol.*, vol. 44, no. 1, pp. 84–89, Jan. 2011.

[49] Z. Li, Y. Xu, H. Huang, and S. Misra, "Sparse control and compressed sensing in networked switched systems," *IET Control Theory & Appl.*, vol. 10, no. 9, pp. 1078–1087, Jun. 2016.

[50] R. M. Dudley, "The sizes of compact subsets of Hilbert space and continuity of Gaussian processes," *Journal of Functional Analysis*, vol. 1, no. 3, pp. 290–330, Oct. 1967.

[51] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

**Geethu Joseph** received the B. Tech. degree in Electronics and Communication Engineering from National Institute of Technology, Calicut, India, in 2011, and the M. E. degree in Signal Processing from Indian Institute of Science (IISc), Bangalore, India, in 2014. She was awarded the Prof. I. S. N. Murthy medal in 2014 for being the best M. E. (signal processing) student in the Dept. of Electrical Communication Engineering (ECE), IISc. She is currently working towards the Ph. D. degree at the Dept. of ECE, IISc, Bangalore, India. Her research interests include statistical signal processing, adaptive filter theory, sparse Bayesian learning and compressive sensing.

**Chandra R. Murthy** received the B. Tech. degree in Electrical Engineering from the Indian Institute of Technology Madras, Chennai, India, in 1998, the M.S. and Ph.D. degrees in Electrical and Computer Engineering from Purdue University, West Lafayette, IN and the University of California, San Diego, CA, in 2000 and 2006, respectively. From 2000 to 2002, he worked as an engineer for Qualcomm Inc., San Jose, USA, where he worked on WCDMA baseband transceiver design and 802.11b baseband receivers. From 2006 to 2007, he worked as a staff engineer at Beceem Communications Inc., Bangalore, India on advanced receiver architectures for the 802.16e Mobile WiMAX standard. Currently, he is working as a Professor in the department of Electrical Communication Engineering at the Indian Institute of Science, Bangalore, India.

His research interests are in the areas of energy harvesting communications, multiuser MIMO systems, and sparse signal recovery techniques applied to wireless communications. His paper won the best paper award in the Communications Track at NCC 2014 and a paper co-authored with his student won the student best paper award at the IEEE ICASSP 2018. He has 50+ journal papers and 80+ conference papers to his credit. He was an associate editor for the IEEE Signal Processing Letters during 2012-16. He is an elected member of the IEEE SPCOM Technical Committee for the years 2014-16, and has been re-elected for the 2017-19 term. He is a past Chair of the IEEE Signal Processing Society, Bangalore Chapter. He is currently serving as an associate editor for the IEEE Transactions on Signal Processing and IEEE Transactions on Information Theory, and as an editor for the IEEE Transactions on Communications.