On the Observability of a Linear System with a Sparse Initial State

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

Abstract—In this paper, we address the problem of observability of a linear dynamic system from compressive measurements and the knowledge of its external inputs. Observability of a high dimensional system state may require a large number of measurements in general, but 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. We derive guarantees for the observability of the system using tools from probability theory and compressed sensing. Our analysis uses properties of the transfer matrix and random measurement matrices to derive concentration of measure bounds, which lead to sufficient conditions for the restricted isometry property of the observability matrix to hold. Hence, under the derived conditions, the initial state can be recovered by solving a computationally tractable convex optimization problem.

I. INTRODUCTION

Observability is a major notion in control theory which is concerned with the question of how well the state of a linear dynamic system can be inferred from its observations and inputs [1]. The classical observability problem involves solving a linear system of equations of the form: $\widetilde{\boldsymbol{y}}_{(K)} = \widetilde{\boldsymbol{A}}_{(K)}\boldsymbol{x}_0$, where the measurement vector $\boldsymbol{y}_{(K)}$ and the observability matrix $\widetilde{\boldsymbol{A}}_{(K)}$ are known, and the state $\boldsymbol{x}_0$ needs to be estimated.[1] 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 [2]. Hence, in the general formulation of the problem, a large number of measurements are required to recover the initial state for systems with a high dimensional state [3]–[5]. However, if the system is known to admit only initial state with a specific structure, the number of measurements required can be significantly brought down by exploiting this additional information. For example, diffusion processes in complex networks that model phenomena like disease or epidemic spreading in the human society [6], [7], air or water pollution [8], [9], virus spreading in computer and mobile phone networks [10], [11], information propagation in online social networks [12], [13], etc, are known to have sparse initialization. Identifying the initial state of these processes accurately is critical to control or eliminate the spreading process [14]. Thus, an important problem in this context is the recoverability of a spare initial state using as few measurements as possible. Further, in some cases, the measurements are obtained as random linear projections of the system

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

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 [15], [16]. The focus of this paper is on providing guarantees on the observability of a system when the observability matrix is possibly rank deficient and random, and the initial state admits a sparse representation in some suitable basis. Our work is motivated by the results from the area of sparse signal recovery or compressive sensing which studies the theory and algorithms for the reconstruction of sparse solutions to linear underdetermined systems.

The tools from compressive sensing provide conditions under which an underdetermined system of linear equations admits a unique sparse solution that can be efficiently computed [17]–[19]. One of the properties of the observability or sensing matrix for guaranteeing the recovery is its *restricted isometric property* (RIP) [20]. Some examples of RIP based conditions that guarantee exact recovery of vectors with ℓ_0 norm at most s are: $\delta_{2s} < 0.493$ is sufficient for basis pursuit (BP) [21]; $\delta_s < 0.307$ is sufficient for BP [22]; $\delta_s < 1/3$ is sharp for BP [23]; $\delta_{2s} < \sqrt{1/2}$, (or generally $\delta_{ts} < \sqrt{(t-1)/t}$ for $t > 4/3$) is sharp for BP [24]; $\delta_{3s} < 1/8$ is sufficient for iterative hard thresholding (IHT) [25]; and $\delta_{s+1} < \frac{1}{\sqrt{s+1}}$ is sufficient for orthogonal matching pursuit (OMP) [26], where δ_s denotes the restricted isometry constant (RIC) of the measurement matrix of order s. The RIP also ensures that the recovery process is robust to noise and is stable when the unknown vector is not precisely sparse. Hence, we analyze the RIP of the observability matrix of a linear dynamical system.

The connection between the compressed sensing and the observability of a linear system has received little attention in the literature. The design of control *algorithms* based on sparsity in the state using tools from compressive sensing is presented in [27]. However, this paper does not discuss *guarantees* for recoverability of the system state in the proposed framework. Guarantees on the observability of the linear system based on the RIP of the observability matrix are derived in [28], [29]. A drawback of these approaches is that one has to keep collecting observations until the observability matrix satisfies the RIP, which in turn is hard to verify in practice. To overcome this difficulty, another paper characterizes the number of measurements required for the exact recovery of the initial state in a stochastic setting [15]. 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.

In this work, we derive guarantees on the system observability under a stochastic setting when the observation matrices are i.i.d. subgaussian random matrices and the system transfer matrix is nonzero. The key novelty in the results is the derivation of a concentration of measure bound for the norm of a sparse vector transformed using the observability matrix, which in turn allows us to characterize the number of measurements required for sparse state recovery. We show that Km should scale as $s \ln \left(\frac{eN}{s} \right)$ to ensure exact recoverability of an s-sparse state 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.

II. SYSTEM MODEL

We consider the following discrete-time linear system:

$$\boldsymbol{x}_{k+1} = \boldsymbol{D}\boldsymbol{x}_k \qquad \boldsymbol{y}_k = \boldsymbol{A}_{(k)}\boldsymbol{x}_k, \qquad (1)$$

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 $\boldsymbol{x}_0$ is sparse, i.e., $\|\boldsymbol{x}_0\|_0 \leq s, s \ll N$. We make the following points before proceeding further:

1) Observability of the initial state $\boldsymbol{x}_0$ implies the observability of $\boldsymbol{x}_k$ for all k.
2) In (1), 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 (1).
3) 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.

Our starting point is the following equivalent linear system:

$$\widetilde{\boldsymbol{y}}_{(K)} = \widetilde{\boldsymbol{A}}_{(K)}\boldsymbol{x}_0, \qquad (2)$$

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 follows:

$$\widetilde{\boldsymbol{y}}_{(K)} = \begin{bmatrix} \boldsymbol{y}_0 \\ \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_{K-1} \end{bmatrix}, \qquad \widetilde{\boldsymbol{A}}_{(K)} = \begin{bmatrix} \boldsymbol{A}_{(0)} \\ \boldsymbol{A}_{(1)}\boldsymbol{D} \\ \vdots \\ \boldsymbol{A}_{(K-1)}\boldsymbol{D}^{K-1} \end{bmatrix}. \qquad (3)$$

In order to ensure the recovery of $\boldsymbol{x}_0$ from (2) using sparse signal recovery techniques, we need to analyze the RIP of the structured observability matrix $\tilde{\boldsymbol{A}}_{(K)}$, which is not available in the literature. Using our analysis of the RIP of $\tilde{\boldsymbol{A}}_{(K)}$, we present bounds on the number of measurement vectors and observations required to recover any sparse initial state.

III. PRELIMINARIES

In this section, we define a subgaussian random matrix and summarize some of its properties.

Definition 1 (Subgaussian random variable and matrix). *A random variable A is said to be subgaussian with parameter c if, for any $\theta \in \mathbb{R}$, $\mathbb{E}\{\exp(\theta A)\} \leq \exp(c\theta^2)$. A random matrix $\boldsymbol{A} \in \mathbb{R}^{m \times N}$ is said to be a subgaussian random matrix if its entries are independent zero mean and unit variance subgaussian random variables with common parameter c.*

The subgaussian random matrix includes a large class of random matrices including independent and identically distributed (i.i.d.) Gaussian random matrices, and i.i.d. Bernoulli random matrices, etc. Next, we present two results that are necessary for the derivation of the main results in the paper.

Lemma 1. *If A is a subgaussian random variable with parameter c, then $A^2 - \mathbb{E}\{A^2\}$ is a subexponential random variable with parameter $16c$, i.e., for $|\theta| \leq \frac{1}{16c}$, we have*

$$\mathbb{E}\left\{\exp\left[\theta\left(A^2 - \mathbb{E}\{A^2\}\right)\right]\right\} \leq \exp\left(128\theta^2 c^2\right). \qquad (4)$$

Proof: See [30, Lemma 1.12]. ∎

Proposition 1 (Bernstein-type inequality). *Let $\{A_l\}_{l=1,2,\ldots,m}$ be independent subexponential random variables such that $a_{\min} \leq \mathbb{E}\{A_l\} \leq a_{\max}$. That is, for all $t \geq 0$,*

$$\mathbb{P}\{|A_l - \mathbb{E}\{A_l\}| \geq t\} \leq c_1 \exp(-c_2 t), \qquad (5)$$

for $l = 1, 2, \ldots, m$, and some constants $c_1, c_2 > 0$. Then, for any $t > m \max\{a_{\max}, -a_{\min}\}$,

$$\mathbb{P}\left\{\left|\sum_{l=1}^{m} A_l\right| \geq t\right\} \leq \exp\left(-\frac{c_2^2(t - ma_{\max})^2/2}{m(2c_1 + c_2 a_{\max}) + c_2 t}\right)$$
$$+ \exp\left(-\frac{c_2^2(t + ma_{\min})^2/2}{m(2c_1 + c_2 a_{\min}) + c_2 t}\right). \qquad (6)$$

Proof: See Appendix A. ∎

IV. MAIN RESULTS AND DISCUSSION

In this section, we present the main results of the paper and discuss their implications.

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

$$Km\left(\delta - 1 + \lambda^{2(K-1)}\right)^2 \geq \widetilde{c}\left[9s \ln\left(\frac{eN}{s}\right) + 2\ln(2\epsilon^{-1})\right], \qquad (7)$$

the RIC δ_s of $\widetilde{\boldsymbol{A}}_{(K)}$ satisfies $\delta_s < \delta$ for all $\delta > 1 - \lambda^{2(K-1)}$ 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 (7) holds, the system is observable for sufficiently large λ with high probability.

Proof: See Appendix B. ∎

In the above, the phrase *sufficiently large λ* refers to the λ required to meet the upper bound on the RIC set by the RIP based guarantees of different algorithms, as discussed in Section I. We discuss this further in the next subsection.

A. Discussion

- Theorem 1 shows that $Km \geq \mathcal{O}(s \ln(N/s))$ is sufficient for observability. Further, ℓ_1 minimization can be posed as a linear program and solved in polynomial time complexity in N [25]. In contrast, $\mathcal{O}(N)$ measurements are necessary for observability of a non-sparse initial state vector.

- 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 (2) as K increases.

- Suppose $\boldsymbol{D}$ is rank-deficient. Then, $\lambda = 0$, and (7) does not hold for any $\delta < 1$, unless

$$m \geq \frac{2}{3\tilde{c}\delta^2}\left[9s \ln\left(\frac{eN}{s}\right) + 2\ln(2\epsilon^{-1})\right]. \quad (8)$$

Intuitively, if $\boldsymbol{x}_0$ lies in the null space of $\boldsymbol{D}$, $\boldsymbol{y}_k = 0$ for $k \geq 1$. Hence, the system is observable for all sparse $\boldsymbol{x}_0$ if it is observable from $\boldsymbol{y}_0$. By following a different approach based on the RIP properties of $\boldsymbol{D}$, one can derive sufficient conditions that improve with K even when $\boldsymbol{D}$ is rank-deficient, but a proof of this result is beyond the scope of this paper.

- Suppose that $\boldsymbol{D}$ is an ill-conditioned matrix, i.e., λ is close to zero. Then, the upper bound on δ required to guarantee observability may not hold [21]–[23]. For example, using the necessary and sufficient condition for ℓ_1 based recovery: $\delta_s \leq 1/3$ [23], (7) reduces to

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

for $\lambda^{2(K-1)} \geq 2/3$. In other words, if (9) is satisfied for some $K \leq \lfloor (\ln(2/3))/(2\ln(\lambda)) \rfloor + 1$, then the system is observable. This is intuitive, because right multiplication by an ill-conditioned matrix may severely degrade its RIP. We also note that the ratio of the smallest to the largest singular values of $\boldsymbol{D}^K$ decreases, i.e., $\boldsymbol{D}^K$ becomes ill-conditioned as K increases. This results in an upper bound on K as mentioned above. However, note that, if the system is observable for K_1 measurements, it remains observable for $K > K_1$.

- For $K = 1$, Theorem 1 reduces to the recovery condition of the standard compressed sensing problem [25]. Also, if the system is observable with m measurements (for example, when (8) is satisfied), the conditions in Theorem 1 hold for $K = 1$, as expected.

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 (8) is satisfied. However, this condition implies that the system is observable with $K = 1$. This is a weak result, as the sufficient condition for

observability does not improve when additional measurements are available. This is indeed true when $\boldsymbol{D} = \alpha\boldsymbol{I}$, for some $\alpha \in \mathbb{R}$, as we are only adding scaled versions of the rows of $\boldsymbol{A}$ to $\tilde{\boldsymbol{A}}_{(K)}$, as K increases. Unfortunately, the current tools from concentration of measure used for characterizing the RIP of a random matrix are inadequate for bounding the number of measurements required for the general $\boldsymbol{D}$ case. This an interesting direction for future work.

B. RIP of the product of matrices

We can derive a sufficient condition for the product of a subgaussian matrix and a deterministic matrix to satisfy the RIP property as follows:

Corollary 1. *Suppose $\boldsymbol{A} \in \mathbb{R}^{m \times N}$ is subgaussian random matrix with parameter c. Then, if*

$$m\left(\delta - 1 + \lambda^2\right)^2 \geq \frac{2}{3\tilde{c}}\left[9s \ln\left(\frac{eN}{s}\right) + 2\ln(2\epsilon^{-1})\right], \quad (10)$$

the RIC δ_s of $\boldsymbol{AD}$ satisfies $\delta_s < \delta$, for all $\delta > 1 - \lambda^2$, with probability at least $1 - \epsilon$. Here, $\tilde{c}$ is a constant dependent only on c, and $\lambda \leq 1$ is the ratio of the smallest and the largest singular values of $\boldsymbol{D}$.

Corollary 1 is an immediate by-product of the proof of Theorem 1, but is an interesting result in its own right, as it provides conditions under which right-multiplication of a subgaussian random matrix by a deterministic matrix $\boldsymbol{D}$ preserves its RIP.

C. Extension to Robust Recovery

The RIP based analysis allows us to extend Theorem 1 to bound error in recovering $\boldsymbol{x}_0$ under bounded noise and model mismatch, that is, when the measurements are noisy and the initial state is not exactly sparse, respectively. In this case, the system model modifies as follows:

$$\boldsymbol{x}_{k+1} = \boldsymbol{D}^{k+1}(\boldsymbol{x}_0 + \check{\boldsymbol{x}}_0) \quad (11)$$
$$\boldsymbol{y}_k = \boldsymbol{A}_{(k)}\boldsymbol{x}_k + \boldsymbol{w}_k, \quad (12)$$

for discrete time instants $k = 0, 1, \ldots, K-1$. Here, $\boldsymbol{w}_k \in \mathbb{R}^m$ denotes the bounded measurement noise: $\|\boldsymbol{w}_k\| \leq W$. $\check{\boldsymbol{x}}_0 \in \mathbb{R}^N$ represents the bounded noise in the initial state. Here, $\boldsymbol{x}_0$ denote s the $s-$sparse approximation of the initial state: $\boldsymbol{x}_0 = \underset{\boldsymbol{v} \in \mathbb{R}^N : \|\boldsymbol{v}\|_0 \leq s}{\arg\min} \|\boldsymbol{x}_0 + \check{\boldsymbol{x}}_0 - \boldsymbol{v}\|$. Therefore, the overall set of equations can be written as

$$\tilde{\boldsymbol{y}}_{(K)} = \tilde{\boldsymbol{A}}_{(K)}(\boldsymbol{x}_0 + \check{\boldsymbol{x}}_0) + \tilde{\boldsymbol{w}}, \quad (13)$$

where bounded noise $\tilde{\boldsymbol{w}} \in \mathbb{R}^{Km}$ satisfies $\|\tilde{\boldsymbol{w}}\| \leq KW$.

Corollary 2. *Suppose $\boldsymbol{A}_{(k)}, k = 0, 1, \ldots, K-1$ are independent subgaussian random matrices with parameter c. Then, for some integer $p > 0$ and positive real number C_{th}, suppose*

$$Km\left(C_{th} - 1 + \lambda^{2(K-1)}\right)^2 \geq \tilde{c}\left[9ps \ln\left(\frac{eN}{ls}\right) + 2\ln(2\epsilon^{-1})\right], \quad (14)$$

and $\lambda^{2(K-1)} > 1 - C_{th}$. Here, $\tilde{c}$ 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 (14) *holds, with probability at least $1 - \epsilon$, the initial vector $\boldsymbol{x}_0 + \check{\boldsymbol{x}}_0$ can be recovered with errors bounded as follows:*

$$\|\boldsymbol{x}_0 + \check{\boldsymbol{x}}_0 - \hat{\boldsymbol{x}}_0\|_1 \leq c_1 \|\check{\boldsymbol{x}}_0\|_1 + c_2 \sqrt{s} KW \quad (15)$$

$$\|\boldsymbol{x}_0 + \check{\boldsymbol{x}}_0 - \hat{\boldsymbol{x}}_0\| \leq c_1 \|\check{\boldsymbol{x}}_0\|_1 / \sqrt{s} + c_2 KW, \quad (16)$$

where $\hat{\boldsymbol{x}}_0$ is the estimate of the initial vector obtained from a sparse recovery algorithm, and $c_1, c_2 > 0$ are universal constants. The constants p and C_{th} are dependent on the recovery algorithms as follows: for BP, $p = 2$ and $C_{th} = \frac{4}{\sqrt{41}}$; for IHT, $p = 6$ and $C_{th} = \frac{1}{\sqrt{3}}$; and for compressive sampling matched pursuit (CoSAMP), $p = 8$ and $C_{th} = 0.478$.

Proof: Follows from the upper bound on the RIC required by the different algorithms to ensure robust recovery [25, Theorem 6.12, 6.21, 6.28]. ∎

V. CONCLUSIONS

We derived sufficient conditions for a linear dynamical system to be observable when its initial state is sparse. Our results are applicable under the stochastic setting where the observation matrices are independent and subgaussian random matrices. We showed that systems that are unobservable using classical control theory can be observable under the sparsity constraints. We also presented an extension of the results to the robust recovery of the system state under model mismatch and noisy observations. Deriving similar results for the case when all observation matrices are identical is an interesting direction for future work.

.

APPENDIX A
PROOF OF PROPOSITION 1

Proof: Using [25, Corollary 7.32] and [25, Theorem 7.30], we can show that, for $t > 0$,

$$\mathbb{P} \left\{ \sum_{l=1}^{m} (A_l - \mathbb{E}\{A_l\}) \geq t \right\} \leq \exp \left(-\frac{c_2^2 t^2 / 2}{2 c_1 m + c_2 t} \right),$$

$$\mathbb{P} \left\{ \sum_{l=1}^{m} (-A_l + \mathbb{E}\{A_l\}) \geq t \right\} \leq \exp \left(-\frac{c_2^2 t^2 / 2}{2 c_1 m + c_2 t} \right).$$

Therefore, for $t > m \max\{a_{\max}, -a_{\min}\}$,

$$\mathbb{P} \left\{ \sum_{l=1}^{m} A_l \geq t \right\} \leq \exp \left(-\frac{c_2^2 (t - m a_{\max})^2 / 2}{2 c_1 m + c_2 (t - m a_{\max})} \right),$$

$$\mathbb{P} \left\{ \sum_{l=1}^{m} -A_l \geq t \right\} \leq \exp \left(-\frac{c_2^2 (t + m a_{\min})^2 / 2}{2 c_1 m + c_2 (t + m a_{\min})} \right).$$

We get the desired result by combining the above inequalities using the union bound. ∎

APPENDIX B
PROOF OF THEOREM 1

Proof: First, we note that an overall scaling does not affect the RIP of a matrix. Hence, without loss of generality, we assume that the largest and the smallest singular values of

$\boldsymbol{D} \neq \boldsymbol{0}$ are 1 and λ, respectively. For any $\boldsymbol{z} \in \mathbb{R}^N$ such that $\|\boldsymbol{z}\|^2 = 1$ and $t \in (0, 1)$, we have

$$\mathbb{P} \left\{ \left| \frac{1}{Km} \left\| \tilde{\boldsymbol{A}}_{(K)} \boldsymbol{z} \right\|^2 - \|\boldsymbol{z}\|^2 \right| \geq t \right\}$$

$$= \mathbb{P} \left\{ \left| \sum_{k=0}^{K-1} \sum_{l=1}^{m} \left(a_{k,l} + \left\| \boldsymbol{D}^k \boldsymbol{z} \right\|^2 - \|\boldsymbol{z}\|^2 \right) \right| \geq Kmt \right\}, \quad (17)$$

where $a_{k,l} \triangleq \left| (\boldsymbol{A}_{(k)}^{\mathrm{T}})_l^{\mathrm{T}} \boldsymbol{D}^k \boldsymbol{z} \right|^2 - \left\| \boldsymbol{D}^k \boldsymbol{z} \right\|^2$, where $(\boldsymbol{A}_{(k)}^{\mathrm{T}})_l^{\mathrm{T}}$ is the lth row of the matrix $\boldsymbol{A}_{(k)}$. Here, the term $(\boldsymbol{A}_{(k)}^{\mathrm{T}})_l^{\mathrm{T}} \boldsymbol{D}^k \boldsymbol{z}$ is the inner product between a row of $\tilde{\boldsymbol{A}}_{(K)}$ and $\boldsymbol{z}$. It is easy to see that $(\boldsymbol{A}_{(k)}^{\mathrm{T}})_l^{\mathrm{T}} \boldsymbol{D}^k \boldsymbol{z}$ is a subgaussian random variable with parameter $c \left\| \boldsymbol{D}^k \boldsymbol{z} \right\|^2$. Also, using the independence and unit variance property of the entries of $(\boldsymbol{A}_{(k)}^{\mathrm{T}})_l$, we have $\mathbb{E}\{a_{k,l}\} = 0$. Thus, from Lemma 1, for $|\theta| \leq \frac{1}{16c \|\boldsymbol{D}^k \boldsymbol{z}\|^2}$ and hence for $|\theta| \leq \frac{1}{16c}$, we have

$$\mathbb{E}\{\exp(\theta a_{k,l})\} \leq \exp \left(128\theta^2 c^2 \left\| \boldsymbol{D}^k \boldsymbol{z} \right\|^4 \right) \leq \exp \left(128\theta^2 c^2 \right),$$

which follows since the largest singular value of $\boldsymbol{D}$ is 1. Note that this holds true even if $\boldsymbol{D}$ is not invertible. Hence, using the Chernoff bound, for all $t > 0$,

$$\mathbb{P}\{|a_{k,l}| \geq t\} \leq 2 \min_{0 < \theta \leq \frac{1}{16c}} \exp \left(128\theta^2 c^2 \right) \exp(-\theta t) \quad (18)$$

$$\leq 2 \exp(1/8) e^{-t/(32c)} \quad (19)$$

where (19) is obtained by setting $\theta = 1/(32c)$. Further, independence of the rows of $\boldsymbol{A}_{(k)}$ for $k = 1, 2, \ldots, K$ implies that $a_{k,l}$ are independent. Therefore, $a_{k,l} + \left\| \boldsymbol{D}^k \boldsymbol{z} \right\|^2 - \|\boldsymbol{z}\|^2$ satisfies the conditions required to apply Proposition 1. Thus, (17), along with the fact $\lambda^{2(K-1)} - 1 \leq \left\| \boldsymbol{D}^k \boldsymbol{z} \right\|^2 - \|\boldsymbol{z}\|^2 \leq 0$ yields, for $t \in (1 - \lambda^{2(K-1)}, 1)$

$$\mathbb{P} \left\{ \left| \frac{1}{Km} \left\| \tilde{\boldsymbol{A}}_{(K)} \boldsymbol{z} \right\|^2 - \|\boldsymbol{z}\|^2 \right| \geq t \right\}$$

$$\leq \exp \left(-\frac{c_2^2 (Kmt)^2 / 2}{2 c_1 Km + c_2 Kmt} \right)$$

$$+ \exp \left(-\frac{c_2^2 \left[Kmt + Km \left(\lambda^{2(K-1)} - 1 \right) \right]^2 / 2}{2 c_1 Km + c_2 Km \left(t + \lambda^{2(K-1)} - 1 \right)} \right) \quad (20)$$

$$\leq \exp \left(-\frac{c_2^2 Kmt^2}{2 (2c_1 + c_2 t)} \right)$$

$$+ \exp \left(-\frac{c_2^2 Km \left(t + \lambda^{2(K-1)} - 1 \right)^2}{2 (2c_1 + c_2 t)} \right) \quad (21)$$

$$\leq 2 \exp \left(-\tilde{c} Km \left(t - 1 + \lambda^{2(K-1)} \right)^2 \right), \quad (22)$$

where $c_1 = 2 \exp(1/8)$, $c_2 = 1/(32c)$ and $\tilde{c} = \frac{c_2^2}{2(2c_1 + c_2)}$. Also, the last step follows because when $t \in (1 - \lambda^{2(K-1)}, 1)$, we have $t^2 \geq \left(t + \lambda^{2(K-1)} - 1 \right)^2$. Now, using the proof technique in [25, Theorem 9.11], we get that if (7) holds, the RIC δ_s of A satisfies $\delta_s < \delta$, for all $\delta > 1 - \lambda^{2(K-1)}$, with probability at least $1 - \epsilon$. This completes the proof. ∎

REFERENCES

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

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

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

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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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, 2006.

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

[11] 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.

[12] 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.

[13] I. Miholic, "Information propagation in online social networks," 2014.

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

[15] 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.

[16] 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.

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

[18] 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.

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

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

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

[22] 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.

[23] 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

[24] ——, "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.

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

[26] 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.

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

[28] 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.

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

[30] P. Rigollet, "18.S997: High-Dimensional Statistics, Cambridge, MA, USA:MIT OpenCourseWare," Jul. 2015. [Online]. Available: http://www-math.mit.edu/ rigollet/PDFs/RigNotes15.pdf