Observability with Random Observations

Borhan M. Sanandaji,* Michael B. Wakin,◇ and Tyrone L. Vincent◇

Abstract

Recovery of the initial state of a high-dimensional system can require a large number of measurements. In this paper, we explain how this burden can be significantly reduced when randomized measurement operators are employed. Our work builds upon recent results from Compressive Sensing (CS). In particular, we make the connection to CS analysis for random block diagonal matrices. By deriving Concentration of Measure (CoM) inequalities, we show that the observability matrix satisfies the Restricted Isometry Property (RIP) (a sufficient condition for stable recovery of sparse vectors) under certain conditions on the state transition matrix. For example, we show that if the state transition matrix is unitary, and if independent, randomly-populated measurement matrices are employed, then it is possible to uniquely recover a sparse high-dimensional initial state when the total number of measurements scales *linearly* in the sparsity level (the number of non-zero entries) of the initial state and logarithmically in the state dimension. We further extend our RIP analysis for scaled unitary and symmetric state transition matrices. We support our analysis with a case study of a two-dimensional diffusion process.

Index Terms

Observability, Restricted Isometry Property, Concentration of Measure Inequalities, Block Diagonal Matrices, Compressive Sensing

I. Introduction

In this paper, we consider the problem of recovering the initial state of a high-dimensional system from compressive measurements (i.e., we take fewer measurements than the system dimension).

*B. M. Sanandaji is with the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley, CA 94720, USA. Email: sanandaji@eecs.berkeley.edu. ◇M. B. Wakin and T. L. Vincent are with the Department of Electrical Engineering and Computer Science at the Colorado School of Mines, Golden, CO 80401, USA. Email: {mwakin, tvincent}@mines.edu. Preliminary versions of portions of this work appeared in [1]. This work was partially supported by AFOSR Grant FA9550-09-1-0465, NSF Grant CCF-0830320, DARPA Grant HR0011-08-1-0078, and NSF Grant CNS-0931748.

A. Measurement Burdens in Observability Theory

Consider an N-dimensional discrete-time linear dynamical system described by the state equation[1]

$$\begin{aligned} \boldsymbol{x}_k &= A\boldsymbol{x}_{k-1} \\ \boldsymbol{y}_k &= C_k\boldsymbol{x}_k, \end{aligned} \tag{1}$$

where $\boldsymbol{x}_k \in \mathbb{R}^N$ represents the state vector at time $k \in \{0, 1, 2, \dots\}$, $A \in \mathbb{R}^{N\times N}$ represents the state transition matrix, $\boldsymbol{y}_k \in \mathbb{R}^M$ represents a set of measurements (or "observations") of the state at time k, and $C_k \in \mathbb{R}^{M\times N}$ represents the measurement matrix at time k. (Observe that the number of measurements at each sample time is M.) For any finite set $\Omega \subset \{0, 1, 2, 3, \dots\}$, define the *generalized observability matrix* as

$$\mathcal{O}_\Omega := \begin{bmatrix} C_{k_0}A^{k_0} \\ C_{k_1}A^{k_1} \\ \vdots \\ C_{k_{K-1}}A^{k_{K-1}} \end{bmatrix} \in \mathbb{R}^{MK\times N}, \tag{2}$$

where $\Omega = \{k_0, k_1, \dots, k_{K-1}\}$ contains K observation times. Note that this definition extends the traditional definition of the observability matrix by allowing arbitrary time samples in (2) and matches the traditional definition when $\Omega = \{0, 1, \dots, K-1\}$. The primary use of observability theory is in ensuring that a state (say, an initial state $\boldsymbol{x}_0$) can be recovered from a collection of measurements $\{\boldsymbol{y}_{k_0}, \boldsymbol{y}_{k_1}, \dots, \boldsymbol{y}_{k_{K-1}}\}$. In particular, defining $\boldsymbol{y}_\Omega := \begin{bmatrix} \boldsymbol{y}_{k_0}^T & \boldsymbol{y}_{k_1}^T & \cdots & \boldsymbol{y}_{k_{K-1}}^T \end{bmatrix}^T \in \mathbb{R}^{MK}$, we have

$$\boldsymbol{y}_\Omega = \mathcal{O}_\Omega\boldsymbol{x}_0. \tag{3}$$

Although we will consider situations where C_k changes with each k, we first discuss the classical case where $C_k = C$ (C is assumed to have full row rank) for all k and $\Omega = \{0, 1, \dots, K-1\}$ (the observation times are consecutive). In this setting, an important and classical result [2] states that a system described by the state equation (1) is observable if and only if $\mathcal{O}_\Omega$ has rank N (full column rank) where $\Omega = \{0, 1, \dots, N-1\}$. One challenge in exploiting this fact is that for some systems, N can be quite large. For example, distributed systems evolving on a spatial domain can have a large state space

[1]The results of this paper also apply directly to systems described by a state equation of the form

$$\begin{aligned} \boldsymbol{x}_k &= A\boldsymbol{x}_{k-1} + B\boldsymbol{u}_k \\ \boldsymbol{y}_k &= C_k\boldsymbol{x}_k + D\boldsymbol{u}_k, \end{aligned}$$

where $\boldsymbol{u}_k \in \mathbb{R}^P$ is the input vector at sample time k and $B \in \mathbb{R}^{N\times P}$ and $D \in \mathbb{R}^{M\times P}$ are constant matrices. Indeed, initial state recovery is independent of B and D when it is assumed that the input vector $\boldsymbol{u}_k$ is known for all sample times k.

even after taking a spatially-discretized approximation. In such settings, we might therefore require a very large total number of measurements (MK with $K = N$) to identify an initial state, and moreover, inverting the matrix $\mathcal{O}_\Omega$ could be very computationally demanding.

This raises an interesting question: under what circumstances might we be able to infer the initial state of a system when $MK < N$? We might imagine, for example, that the measurement burden could be alleviated in cases when there is a model for the state $\boldsymbol{x}_0$ that we wish to recover. Alternatively, we may have cases where, rather than needing to recover $\boldsymbol{x}_0$ from $\boldsymbol{y}_\Omega$, we desire only to solve a much simpler inference problem such as a binary detection or a classification problem. In this paper, inspired by the emerging theory of Compressive Sensing (CS) [3]–[5], we explain how such assumptions can indeed reduce the measurement burden and, in some cases, even allow recovery of the initial state when $MK < N$ and the system of equations (3) is guaranteed to be underdetermined. More broadly, exploiting CS concepts in the analysis of sparse dynamical systems from limited information has gained much attention over the last few years in applications such as system identification [6]–[9], control feedback design [10], and interconnected networks [11], [12], among others.

B. Compressive Sensing and Randomized Measurements

The CS theory states that it is possible to solve certain rank-deficient sets of linear equations by imposing a model assumption on the signal to be recovered. In particular, suppose $\boldsymbol{y} = \Phi\boldsymbol{x}$ where Φ is an $\widetilde{M} \times N$ matrix with $\widetilde{M} < N$. Suppose also that $\boldsymbol{x} \in \mathbb{R}^N$ is S-sparse, meaning that only S out of its N entries are non-zero.[2] If Φ satisfies a condition called the Restricted Isometry Property (RIP) of order $2S$ for a sufficiently small isometry constant δ_{2S}, then it is possible to *uniquely* recover any S-sparse signal $\boldsymbol{x}$ from the measurements $\boldsymbol{y} = \Phi\boldsymbol{x}$ using a tractable convex optimization program known as ℓ_1-minimization [3], [4]. The RIP also ensures that the recovery process is robust to noise and stable in cases where $\boldsymbol{x}$ is not precisely sparse [13]. In the following, we provide the definition of the RIP.

Definition 1: A matrix $\Phi \in \mathbb{R}^{\widetilde{M}\times N}$ ($\widetilde{M} < N$) is said to satisfy the RIP of order S with isometry constant $\delta_S \in (0, 1)$ if

$$(1 - \delta_S)\|\boldsymbol{x}\|_2^2 \leq \|\Phi\boldsymbol{x}\|_2^2 \leq (1 + \delta_S)\|\boldsymbol{x}\|_2^2 \tag{4}$$

holds for all S-sparse vectors $\boldsymbol{x} \in \mathbb{R}^N$.

Observe that the RIP is a property of a matrix and has a deterministic definition. However, checking whether the RIP holds for a given matrix Φ is computationally expensive and is almost impossible when

[2]This is easily extended to the case where $\boldsymbol{x}$ is sparse in some transform domain.

N is large. A common way to establish the RIP for Φ is to populate Φ with random entries. If Φ is populated with independent and identically distributed (i.i.d.) Gaussian random variables having zero mean and variance $\frac{1}{\widetilde{M}}$, for example, then Φ will satisfy the RIP of order S with isometry constant δ_S with very high probability when $\widetilde{M}$ is proportional to $\delta_S^{-2} S \log \frac{N}{S}$ [5], [14], [15]. This result is significant because it indicates that the number of measurements sufficient for correct recovery scales *linearly* in the sparsity level S and only *logarithmically* in the ambient dimension N. Other random distributions may also be considered, including matrices with uniform entries of random signs. Consequently, a number of new sensing hardware architectures, from analog-to-digital converters to digital cameras, are being developed to take advantage of the benefits of random measurements [16]–[19].

A simple way [14], [20] of proving the RIP for a randomized construction of Φ involves first showing that the matrix satisfies a Concentration of Measure (CoM) inequality akin to the following.

Definition 2: A random matrix (a matrix whose entries are drawn from a particular probability distribution) $\Phi \in \mathbb{R}^{\widetilde{M} \times N}$ is said to satisfy the Concentration of Measure (CoM) inequality if for any fixed signal $\boldsymbol{x} \in \mathbb{R}^N$ (not necessarily sparse) and any $\epsilon \in (0, \overline{\epsilon})$,

$$\mathbf{P}\left\{\left|\|\Phi \boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_2^2\right| > \epsilon \|\boldsymbol{x}\|_2^2\right\} \leq 2\exp\left\{-\widetilde{M} f(\epsilon)\right\}, \tag{5}$$

where $f(\epsilon)$ is a positive constant that depends on the isometry constant ϵ, and $\overline{\epsilon} \leq 1$ is some maximum value of the isometry constant for which the CoM inequality holds.

Note that the failure probability in (5) decays exponentially fast in the number of measurements $\widetilde{M}$ times some constant $f(\epsilon)$ that depends on the isometry constant ϵ. For most interesting random matrices, including matrices populated with i.i.d. Gaussian random variables, $f(\epsilon)$ is quadratic in ϵ as $\epsilon \to 0$.

Baraniuk et al. [14] and Mendelson et al. [21] showed that a CoM inequality of the form (5) can be used to prove the RIP for random compressive matrices. This result is rephrased by Davenport [15].

Lemma 1: [15] Let $\mathcal{X}$ denote an S-dimensional subspace in $\mathbb{R}^N$. Let $\delta_S \in (0,1)$ denote a distortion factor and $\nu \in (0,1)$ denote a failure probability, and suppose Φ is an $\widetilde{M} \times N$ random matrix that satisfies the CoM inequality (5) with $\widetilde{M} \geq \frac{S\log(\frac{42}{\delta_S}) + \log(\frac{2}{\nu})}{f(\frac{\delta_S}{\sqrt{2}})}$. Then with probability at least $1-\nu$, for all $\boldsymbol{x} \in \mathcal{X}$,

$$(1-\delta_S)\|\boldsymbol{x}\|_2^2 \leq \|\Phi \boldsymbol{x}\|_2^2 \leq (1+\delta_S)\|\boldsymbol{x}\|_2^2.$$

Through a union bound argument (see, for example, Theorem 5.2 in [14]) and by applying Lemma 1 for all $\binom{N}{S}$ S-dimensional subspaces that define the space of S-sparse signals in $\mathbb{R}^N$, one can show that Φ satisfies the RIP (of order S and with isometry constant δ_S) with high probability when $\widetilde{M}$ scales *linearly* in S and *logarithmically* in N.

C. Observability from Random, Compressive Measurements

In order to exploit CS concepts in observability analysis, we consider scenarios where the measurement matrices C_k are populated with random entries. Physically, such randomized measurements may be taken using the types of CS protocols and hardware mentioned above. Our analysis is therefore appropriate in cases where one has some control over the sensing process.

As is apparent from (2), even with randomness in matrices C_k, the observability matrix $\mathcal{O}_\Omega$ contains some structure and cannot be simply modeled as a matrix populated with i.i.d. Gaussian random variables and thus, existing results cannot be directly applied. Our work builds on a recent paper by Park et al. in which CoM inequalities are derived for random block diagonal matrices [22]. Our concentration results cover a large class of systems (not necessarily unitary) and initial states (not necessarily sparse), and apart from guaranteeing recovery of sparse initial states, other inference problems concerning $\boldsymbol{x}_0$, such as detection or classification of more general initial states and systems, can also be solved using random, compressive measurements, and the performance of such techniques can be studied using the CoM bounds that we provide [23].

Our CoM results have important implications for establishing the RIP. Such RIP results provide a sufficient number of measurements for exact initial state recovery when the initial state is known to be sparse a priori. The results of this paper show that under certain conditions on A (e.g., for unitary, scaled unitary, and certain symmetric matrices A), the observability matrix $\mathcal{O}_\Omega$ will satisfy the RIP with high probability when the total number of measurements MK scales linearly in S and logarithmically in N. Before going into the details of the derivations and proofs, we first state in Section II our main results on establishing the RIP for the observability matrix. We then present in Section III the CoM results upon which the conclusions for establishing the RIP are based. Finally, in Section IV we support our results with a case study involving a diffusion process starting from a sparse initial state.

D. Related Work

Questions involving observability in compressive measurement settings have also been raised in a recent paper [24] concerned with tracking the state of a system from nonlinear observations. Due to the intrinsic nature of the problems in that paper, however, the observability issues raised are quite different. For example, one argument appears to assume that $M \geq S$, a requirement that we do not have. In a recent technical report [25], Dai et al. have also considered a similar sparse initial state recovery problem. However, their approach is different and the results are only applicable in noiseless and perfectly sparse initial state recovery problems. In this paper, we establish the RIP for the observability matrix, which

implies not only that perfectly sparse initial states can be recovered exactly when the measurements are noiseless but also that the recovery process is robust with respect to noise and that nearly-sparse initial states can be recovered with high accuracy [13]. Finally, we note that a matrix vaguely similar to the observability matrix has been studied by Yap et al. in the context of quantifying the memory capacity of echo state networks [26]. The recovery results and guarantees presented in this paper are substantially different from the above mentioned papers as we derive CoM inequalities for the observability matrix and then establish the RIP based on these CoM inequalities.

II. Restricted Isometry Property and the Observability Matrix

When the observability matrix $\mathcal{O}_\Omega$ satisfies the RIP of order $2S$ with isometry constant $\delta_{2S} < \sqrt{2}-1$, an initial state with S or fewer non-zero elements can be stably recovered by solving an ℓ_1-minimization problem [13]. In this section, we present cases where the total number of measurements sufficient for establishing the RIP scales linearly in S and only logarithmically in the state dimension N. As in standard observability theory, the state transition matrix A plays a crucial role in the analysis. Because the analysis is somewhat complex, results for completely general A are difficult to obtain. However, we present results here for unitary, scaled unitary, and certain symmetric matrices, and we believe that these can give interesting insight into the essential issues driving the initial state recovery problem.

To assist in interpreting the RIP results, let us point out that we actually state our RIP bounds in terms of the scaled observability matrix $\frac{1}{\sqrt{b}}\mathcal{O}_\Omega$ where b is defined below and is chosen to ensure that the measurements are properly normalized to be compatible with (4). In noiseless recovery, this scaling is unimportant. When noise occurs, however, the scaling enters into the effective signal to noise ratio for the problem, as the measurements will be more sensitive to noise when b is small.

Our first result, stated in Theorem 1, applies to a system with dynamics represented by a scaled unitary matrix when measurements are taken at the first K sample times, starting at zero.

Theorem 1: Assume $\Omega = \{0, 1, \ldots, K-1\}$. Suppose that $A \in \mathbb{R}^{N\times N}$ can be represented as $A = aU$ where $a \in \mathbb{R}$ ($a \neq 0$) and $U \in \mathbb{R}^{N\times N}$ is unitary. Define $b := 1 + a^2 + a^4 + \cdots + a^{2(K-1)}$. Assume each of the measurement matrices $C_k \in \mathbb{R}^{M\times N}$ is populated with i.i.d. Gaussian random entries with mean zero and variance $\frac{1}{M}$. Assume all matrices C_k are generated independently of each other. Suppose that N, S, and $\delta_S \in (0, 1)$ are given. Then with probability exceeding $1-\nu$, $\frac{1}{\sqrt{b}}\mathcal{O}_\Omega$ satisfies the RIP of order

S with isometry constant δ_S whenever

$$MK \geq \begin{cases} \dfrac{512\left((1-a^2)K+a^2\right)\left(S(\log(\frac{42}{\delta_S})+1+\log(\frac{N}{S}))+\log(\frac{2}{\nu})\right)}{\delta_S^2}, & |a|<1 \qquad (6) \\ \dfrac{512\left((1-a^{-2})K+a^{-2}\right)\left(S(\log(\frac{42}{\delta_S})+1+\log(\frac{N}{S}))+\log(\frac{2}{\nu})\right)}{\delta_S^2}, & |a|>1. \qquad (7) \end{cases}$$

Proof See Section III-B. ■

One should note that when $A = aU$ ($a \neq 1$), the results of Theorem 1 have a dependency on K (the number of sampling times). This dependency is not desired in general. When $a = 1$ (i.e., A is unitary), a result (Theorem 2) can be obtained in which the total number of measurements MK scales linearly in S and with no dependency on K. Our general results for $A = aU$ also indicate that when $|a|$ is close to the origin (i.e., $|a| \ll 1$), and by symmetry when $|a| \gg 1$, worse recovery performance is expected as compared to the case when $a = 1$. When $|a| \ll 1$, as an example, the effect of the initial state will be highly attenuated as we take measurements at later times. A similar intuition holds when $|a| \gg 1$. When $a = 1$ (i.e., unitary A), we can relax the consecutive sample times assumption in Theorem 1 (i.e., $\Omega = \{0, 1, \ldots, K-1\}$). We have the following RIP result when K arbitrarily-chosen samples are taken.

Theorem 2: Assume $\Omega = \{k_0, k_1, \ldots, k_{K-1}\}$. Suppose that $A \in \mathbb{R}^{N\times N}$ is unitary. Assume each of the measurement matrices $C_k \in \mathbb{R}^{M\times N}$ is populated with i.i.d. Gaussian random entries with mean zero and variance $\frac{1}{M}$. Assume all matrices C_k are generated independently of each other. Suppose that N, S, and $\delta_S \in (0, 1)$ are given. Then with probability exceeding $1 - \nu$, $\frac{1}{\sqrt{K}}\mathcal{O}_\Omega$ satisfies the RIP of order S with isometry constant δ_S whenever

$$MK \geq \frac{512\left(S(\log(\frac{42}{\delta_S})+1+\log(\frac{N}{S}))+\log(\frac{2}{\nu})\right)}{\delta_S^2}. \qquad (8)$$

Proof See Section III-B. ■

Theorem 2 states that under the assumed conditions, $\frac{1}{\sqrt{K}}\mathcal{O}_\Omega$ satisfies the RIP of order S with isometry constant δ_S with high probability when the total number of measurements MK scales linearly in the sparsity level S and logarithmically in the state ambient dimension N. Consequently under these assumptions, *unique* recovery of any S-sparse initial state $\boldsymbol{x}_0$ is possible from $\boldsymbol{y}_\Omega = \mathcal{O}_\Omega \boldsymbol{x}_0$ by solving the ℓ_1-minimization problem whenever MK is proportional to $S\log(\frac{N}{S})$. This is in fact a significant reduction in the sufficient total number of measurements for correct initial state recovery as compared to traditional observability theory.

We further extend our analysis and establish the RIP for certain symmetric matrices A. We believe this analysis has important consequences in analyzing problems of practical interest such as diffusion (see, for example, Section IV). Suppose $A \in \mathbb{R}^{N\times N}$ is a positive semidefinite matrix with the eigendecomposition

$$A = U\Lambda U^T = \left[U_1 \middle| U_2\right] \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \left[U_1 \middle| U_2\right]^T, \tag{9}$$

where $U \in \mathbb{R}^{N\times N}$ is unitary, $\Lambda \in \mathbb{R}^{N\times N}$ is a diagonal matrix with non-negative entries, $U_1 \in \mathbb{R}^{N\times L}$, $U_2 \in \mathbb{R}^{N\times(N-L)}$, $\Lambda_1 \in \mathbb{R}^{L\times L}$, and $\Lambda_2 \in \mathbb{R}^{(N-L)\times(N-L)}$. The submatrix Λ_1 contains the L largest eigenvalues of A. The value for L can be chosen as desired; our results below give the strongest bounds when all eigenvalues in Λ_1 are large compared to all eigenvalues in Λ_2. Let $\lambda_{1,\text{min}}$ denote the smallest entry of Λ_1, $\lambda_{1,\text{max}}$ denote the largest entry of Λ_1, and $\lambda_{2,\text{max}}$ denote the largest entry of Λ_2.

In the following, we show that in the special case where the matrix $U_1^T \in \mathbb{R}^{L\times N}$ ($L < N$) happens to itself satisfy the RIP (up to a scaling), then $\mathcal{O}_\Omega$ satisfies the RIP (up to a scaling). Although there are many state transition matrices A that do not have a collection of eigenvectors U_1 with this special property, we do note that if A is a circulant matrix, its eigenvectors will be the Discrete Fourier Transform (DFT) basis vectors, and it is known that a randomly selected set of DFT basis vectors will satisfy the RIP with high probability [27].

Theorem 3: Assume $\Omega = \{k_0, k_1, \ldots, k_{K-1}\}$. Assume A has the eigendecomposition given in (9) and $U_1^T \in \mathbb{R}^{L\times N}$ ($L < N$) satisfies a scaled version[3] of the RIP of order S with isometry constant δ_S. Formally, assume for $\delta_S \in (0,1)$ that

$$(1-\delta_S)\frac{L}{N}\|\boldsymbol{x}_0\|_2^2 \leq \|U_1^T\boldsymbol{x}_0\|_2^2 \leq (1+\delta_S)\frac{L}{N}\|\boldsymbol{x}_0\|_2^2 \tag{10}$$

holds for all S-sparse $\boldsymbol{x}_0 \in \mathbb{R}^N$. Assume each of the measurement matrices $C_k \in \mathbb{R}^{M\times N}$ is populated with i.i.d. Gaussian random entries with mean zero and variance $\frac{1}{M}$. Assume all matrices C_k are generated independently of each other. Let $\nu \in (0,1)$ denote a failure probability and $\delta \in (0, \frac{16}{\sqrt{K}})$ denote a distortion factor. Then with probability exceeding $1-\nu$,

$$(1-\delta)\left((1-\delta_S)\frac{L}{N}\sum_{i=0}^{K-1}\lambda_{1,\text{min}}^{2k_i}\right) \leq \frac{\|\mathcal{O}_\Omega\boldsymbol{x}_0\|_2^2}{\|\boldsymbol{x}_0\|_2^2} \leq (1+\delta)\left((1+\delta_S)\frac{L}{N}\sum_{i=0}^{K-1}\lambda_{1,\text{max}}^{2k_i} + \sum_{i=0}^{K-1}\lambda_{2,\text{max}}^{2k_i}\right) \tag{11}$$

for all S-sparse $\boldsymbol{x}_0 \in \mathbb{R}^N$ whenever

$$MK \geq \frac{512K\left(S\left(\log(\frac{42}{\delta})+1+\log(\frac{N}{S})\right)+\log(\frac{2}{\nu})\right)}{\rho\delta^2}, \tag{12}$$

[3]The $\frac{L}{N}$ scaling in (10) is to account for the unit-norm rows of U_1^T.

where

$$\rho := \inf_{S-\text{sparse } \boldsymbol{x}_0 \in \mathbb{R}^N} \Gamma(\mathcal{A}_\Omega \boldsymbol{x}_0) \tag{13}$$

and

$$\Gamma\left(\mathcal{A}_\Omega \boldsymbol{x}_0\right) := \frac{\left(\|A^{k_0}\boldsymbol{x}_0\|_2^2 + \|A^{k_1}\boldsymbol{x}_0\|_2^2 + \cdots + \|A^{k_{K-1}}\boldsymbol{x}_0\|_2^2\right)^2}{\|A^{k_0}\boldsymbol{x}_0\|_2^4 + \|A^{k_1}\boldsymbol{x}_0\|_2^4 + \cdots + \|A^{k_{K-1}}\boldsymbol{x}_0\|_2^4}.$$

Proof See Appendix A of a companion technical report [28] which contains additional details on several topics from our paper. ■

The result of Theorem 3 is particularly interesting in applications where the largest eigenvalues of A all cluster around each other and the rest of the eigenvalues cluster around zero. Put formally, we are interested in applications where $0 \approx \lambda_{2,\text{max}} \ll \frac{\lambda_{1,\text{min}}}{\lambda_{1,\text{max}}} \approx 1$. The following corollary of Theorem 3 considers an extreme case when $\lambda_{1,\text{max}} = \lambda_{1,\text{min}}$ and $\lambda_{2,\text{max}} = 0$.

Corollary 1: Assume $\Omega = \{k_0, k_1, \ldots, k_{K-1}\}$. Assume each of the measurement matrices $C_k \in \mathbb{R}^{M\times N}$ is populated with i.i.d. Gaussian random entries with mean zero and variance $\frac{1}{M}$. Assume all matrices C_k are generated independently of each other. Suppose A has the eigendecomposition given in (9) and $U_1^T \in \mathbb{R}^{L\times N}$ ($L < N$) satisfies a scaled version of the RIP of order S with isometry constant δ_S as given in (10). Assume $\lambda_{1,\text{max}} = \lambda_{1,\text{min}} = \lambda$ ($\lambda \neq 0$) and $\lambda_{2,\text{max}} = 0$. Let $\nu \in (0,1)$ denote a failure probability and $\delta \in (0,1)$ denote a distortion factor. Define $C := \sum_{i=0}^{K-1} \lambda^{2k_i}$ and $\delta_S' := \delta_S + \delta + \delta_S\delta$. Then with probability exceeding $1-\nu$,

$$(1-\delta_S')\|\boldsymbol{x}_0\|_2^2 \leq \|\sqrt{\frac{N}{LC}}\mathcal{O}_\Omega \boldsymbol{x}_0\|_2^2 \leq (1+\delta_S')\|\boldsymbol{x}_0\|_2^2 \tag{14}$$

for all S-sparse $\boldsymbol{x}_0 \in \mathbb{R}^N$ whenever

$$MK \geq \begin{cases} \dfrac{512(1+\delta_S)^2\lambda^{-4(k_{K-1}-k_0)}\left(S\left(\log(\frac{42}{\delta})+1+\log(\frac{N}{S})\right)+\log(\frac{2}{\nu})\right)}{(1-\delta_S)^2\delta^2}, & \lambda \leq 1 \quad (15) \\ \dfrac{512(1+\delta_S)^2\lambda^{4(k_{K-1}-k_0)}\left(S\left(\log(\frac{42}{\delta})+1+\log(\frac{N}{S})\right)+\log(\frac{2}{\nu})\right)}{(1-\delta_S)^2\delta^2}, & \lambda > 1. \quad (16) \end{cases}$$

Proof See Appendix B of our technical report [28]. ■

These results essentially indicate that the more λ deviates from one, the more total measurements MK are required to ensure unique recovery of any S-sparse initial state $\boldsymbol{x}_0$. The bounds on ρ (which we state in [28, Appendix B], to derive Corollary 1 from Theorem 3) also indicate that when $\lambda \neq 1$, the smallest number of measurements is required when the sample times are *consecutive* (i.e., when $k_{K-1} - k_0 = K$). Similar to what we mentioned earlier in our analysis for a scaled unitary A, when $\lambda \neq 1$ the effect of the

initial state will be highly attenuated as we take measurements at later times (i.e., when $k_{K-1}-k_0 > K$) which results in a larger total number of measurements MK.

III. Concentration of Measure Inequalities and the Observability Matrix

In this section, we derive CoM inequalities for the observability matrix when the measurement matrices C_k are populated with i.i.d. Gaussian random entries. These inequalities are the foundation for establishing the RIP presented in the previous section, via Lemma 1. However, they are also of independent interest for other types of problems involving the states of dynamical systems, such as detection and classification [23], [29], [30]. As mentioned earlier, we make a connection to the analysis for block diagonal matrices from Compressive Sensing (CS). To begin, note that when $\Omega = \{k_0, k_1, \ldots, k_{K-1}\}$ we can write

$$\mathcal{O}_\Omega = \underbrace{\begin{bmatrix} C_{k_0} & & & \\ & C_{k_1} & & \\ & & \ddots & \\ & & & C_{k_{K-1}} \end{bmatrix}}_{\mathcal{C}_\Omega \in \mathbb{R}^{\widetilde{M}\times\widetilde{N}}} \underbrace{\begin{bmatrix} A^{k_0} \\ A^{k_1} \\ \vdots \\ A^{k_{K-1}} \end{bmatrix}}_{\mathcal{A}_\Omega \in \mathbb{R}^{\widetilde{N}\times N}}, \tag{17}$$

where $\widetilde{M} := MK$ and $\widetilde{N} := NK$. In this decomposition, $\mathcal{C}_\Omega$ is a block diagonal matrix whose diagonal blocks are the measurements matrices, C_k. Assume all matrices C_k are generated independently of each other. Focusing just on $\mathcal{C}_\Omega$ for the moment, we have the following bound on its concentration behavior.[4]

Theorem 4: [22] Assume each of the measurement matrices $C_k \in \mathbb{R}^{M\times N}$ is populated with i.i.d. Gaussian random entries with mean zero and variance $\frac{1}{M}$. Assume all C_k are generated independently of each other. Let $\boldsymbol{v}_{k_0}, \boldsymbol{v}_{k_1}, \ldots, \boldsymbol{v}_{k_{K-1}} \in \mathbb{R}^N$ and define $\boldsymbol{v} = \begin{bmatrix} \boldsymbol{v}_{k_0}^T & \boldsymbol{v}_{k_1}^T & \cdots & \boldsymbol{v}_{k_{K-1}}^T \end{bmatrix}^T \in \mathbb{R}^{KN}$. Then

$$\mathbf{P}\left\{ \left| \|\mathcal{C}_\Omega \boldsymbol{v}\|_2^2 - \|\boldsymbol{v}\|_2^2 \right| > \epsilon \|\boldsymbol{v}\|_2^2 \right\} \leq \begin{cases} 2\exp\{-\dfrac{M\epsilon^2\|\boldsymbol{\gamma}\|_1^2}{256\|\boldsymbol{\gamma}\|_2^2}\}, & 0 \leq \epsilon \leq \frac{16\|\boldsymbol{\gamma}\|_2^2}{\|\boldsymbol{\gamma}\|_\infty \|\boldsymbol{\gamma}\|_1} \quad (18) \\ 2\exp\{-\dfrac{M\epsilon\|\boldsymbol{\gamma}\|_1}{16\|\boldsymbol{\gamma}\|_\infty}\}, & \epsilon \geq \frac{16\|\boldsymbol{\gamma}\|_2^2}{\|\boldsymbol{\gamma}\|_\infty \|\boldsymbol{\gamma}\|_1}, \quad (19) \end{cases}$$

where $\boldsymbol{\gamma} = \boldsymbol{\gamma}(\boldsymbol{v}) := \begin{bmatrix} \|\boldsymbol{v}_{k_0}\|_2^2 & \|\boldsymbol{v}_{k_1}\|_2^2 & \ldots & \|\boldsymbol{v}_{k_{K-1}}\|_2^2 \end{bmatrix} \in \mathbb{R}^K$.

As we will be frequently concerned with applications where ϵ is small, consider the first of the cases given in the right-hand side of the above bound. (It can be shown [22] that this case always permits any

[4]All results in this section may be extended to the case where all C_k are populated with sub-Gaussian random variables, as in [22].

value of ϵ between 0 and $\frac{16}{\sqrt{K}}$.) Define

$$\Gamma = \Gamma(\boldsymbol{v}) := \frac{\|\boldsymbol{\gamma}(\boldsymbol{v})\|_1^2}{\|\boldsymbol{\gamma}(\boldsymbol{v})\|_2^2} = \frac{\left(\|\boldsymbol{v}_{k_0}\|_2^2 + \|\boldsymbol{v}_{k_1}\|_2^2 + \cdots + \|\boldsymbol{v}_{k_{K-1}}\|_2^2\right)^2}{\|\boldsymbol{v}_{k_0}\|_2^4 + \|\boldsymbol{v}_{k_1}\|_2^4 + \cdots + \|\boldsymbol{v}_{k_{K-1}}\|_2^4} \tag{20}$$

and note that for any $\boldsymbol{v} \in \mathbb{R}^{KN}$, $1 \leq \Gamma(\boldsymbol{v}) \leq K$. This simply follows from the standard relation that $\|\boldsymbol{z}\|_2 \leq \|\boldsymbol{z}\|_1 \leq \sqrt{K}\|\boldsymbol{z}\|_2$ for all $\boldsymbol{z} \in \mathbb{R}^K$. The case $\Gamma(\boldsymbol{v}) = K$ is quite favorable because the failure probability will decay exponentially fast in the total number of measurements MK. A simple comparison between this result and the CoM inequality for a *dense* Gaussian matrix stated in Definition 2 reveals that we get the same degree of concentration from the $MK \times NK$ block diagonal matrix $\mathcal{C}_\Omega$ as we would get from a dense $MK \times NK$ matrix populated with i.i.d. Gaussian random variables. This event happens if and only if the components $\boldsymbol{v}_{k_i}$ have equal energy, i.e., if and only if $\|\boldsymbol{v}_{k_0}\|_2 = \|\boldsymbol{v}_{k_1}\|_2 = \cdots = \|\boldsymbol{v}_{k_{K-1}}\|_2$. On the other hand, the case $\Gamma(\boldsymbol{v}) = 1$ is quite unfavorable and implies that we get the same degree of concentration from the $MK \times NK$ block diagonal matrix $\mathcal{C}_\Omega$ as we would get from a dense Gaussian matrix having size only $M \times NK$. This event happens if and only if $\|\boldsymbol{v}_{k_i}\|_2 = 0$ for all $i \in \{0, 1, \ldots, K-1\}$ but one i. Thus, more uniformity in the values of the $\|\boldsymbol{v}_{k_i}\|_2$ ensures a higher probability of concentration. We now note that, when applying the observability matrix to an initial state, we will have $\mathcal{O}_\Omega \boldsymbol{x}_0 = \mathcal{C}_\Omega \mathcal{A}_\Omega \boldsymbol{x}_0$. This leads us to the following corollary of Theorem 4.

Corollary 2: Suppose the same notation and assumptions as in Theorem 4. Then for any fixed initial state $\boldsymbol{x}_0 \in \mathbb{R}^N$ and for any $\epsilon \in (0, \frac{16}{\sqrt{K}})$,

$$\mathbf{P}\left\{\left|\|\mathcal{O}_\Omega \boldsymbol{x}_0\|_2^2 - \|\mathcal{A}_\Omega \boldsymbol{x}_0\|_2^2\right| > \epsilon \|\mathcal{A}_\Omega \boldsymbol{x}_0\|_2^2\right\} \leq 2\exp\left\{-\frac{M\Gamma(\mathcal{A}_\Omega \boldsymbol{x}_0)\,\epsilon^2}{256}\right\}. \tag{21}$$

There are two important phenomena to consider in this result, and both are impacted by the interaction of A with $\boldsymbol{x}_0$. First, on the left-hand side of (21), we see that the point of concentration of $\|\mathcal{O}_\Omega \boldsymbol{x}_0\|_2^2$ is around $\|\mathcal{A}_\Omega \boldsymbol{x}_0\|_2^2$, where

$$\|\mathcal{A}_\Omega \boldsymbol{x}_0\|_2^2 = \|A^{k_0}\boldsymbol{x}_0\|_2^2 + \|A^{k_1}\boldsymbol{x}_0\|_2^2 + \cdots + \|A^{k_{K-1}}\boldsymbol{x}_0\|_2^2. \tag{22}$$

For a concentration bound of the same form as Definition 2, however, $\|\mathcal{O}_\Omega \boldsymbol{x}_0\|_2^2$ should concentrate around some constant multiple of $\|\boldsymbol{x}_0\|_2^2$. In general, for different initial states $\boldsymbol{x}_0$ and transition matrices A, we may see widely varying ratios $\frac{\|\mathcal{A}_\Omega \boldsymbol{x}_0\|_2^2}{\|\boldsymbol{x}_0\|_2^2}$. However, further analysis is possible in scenarios where this ratio is predictable and fixed. Second, on the right-hand side of (21), we see that the exponent of the concentration failure probability scales with

$$\Gamma(\mathcal{A}_\Omega \boldsymbol{x}_0) = \frac{\left(\|A^{k_0}\boldsymbol{x}_0\|_2^2 + \|A^{k_1}\boldsymbol{x}_0\|_2^2 + \cdots + \|A^{k_{K-1}}\boldsymbol{x}_0\|_2^2\right)^2}{\|A^{k_0}\boldsymbol{x}_0\|_2^4 + \|A^{k_1}\boldsymbol{x}_0\|_2^4 + \cdots + \|A^{k_{K-1}}\boldsymbol{x}_0\|_2^4}. \tag{23}$$

As mentioned earlier, $1 \leq \Gamma\left(\mathcal{A}_\Omega \boldsymbol{x}_0\right) \leq K$. The case $\Gamma\left(\mathcal{A}_\Omega \boldsymbol{x}_0\right) = K$ is quite favorable and happens when $\|A^{k_0}\boldsymbol{x}_0\|_2 = \|A^{k_1}\boldsymbol{x}_0\|_2 = \cdots = \|A^{k_{K-1}}\boldsymbol{x}_0\|_2$; this occurs when the state "energy" is preserved over time. The case $\Gamma\left(\mathcal{A}_\Omega \boldsymbol{x}_0\right) = 1$ is quite unfavorable and happens when $k_0 = 0$ and $\boldsymbol{x}_0 \neq 0 \in \mathrm{null}(A)$.

A. Unitary and Scaled Unitary System Matrices

In the special case of unitary A (i.e., $\|A^{k_i}\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|_2^2$ for all $\boldsymbol{x} \in \mathbb{R}^N$ and for any power k_i), we can draw a particularly strong conclusion. Because a unitary A guarantees both that $\|\mathcal{A}_\Omega \boldsymbol{x}_0\|_2^2 = K\|\boldsymbol{x}_0\|_2^2$ and that $\Gamma\left(\mathcal{A}_\Omega \boldsymbol{x}_0\right) = K$, we have the following result.

Corollary 3: Assume $\Omega = \{k_0, k_1, \ldots, k_{K-1}\}$. Suppose the same notation and assumptions as in Theorem 4. Suppose that A is a unitary operator. Then for any fixed initial state $\boldsymbol{x}_0 \in \mathbb{R}^N$,

$$\mathbf{P}\left\{\left|\|\frac{1}{\sqrt{K}}\mathcal{O}_\Omega \boldsymbol{x}_0\|_2^2 - \|\boldsymbol{x}_0\|_2^2\right| > \epsilon\|\boldsymbol{x}_0\|_2^2\right\} \leq 2\exp\left\{-\frac{MK\epsilon^2}{256}\right\}, \qquad \epsilon \in (0,1).^5 \tag{24}$$

This means that we get the same degree of concentration from the $MK \times N$ matrix $\frac{1}{\sqrt{K}}\mathcal{O}_\Omega$ as we would get from a fully dense $MK \times N$ matrix populated with i.i.d. Gaussian random variables. Observe that this concentration result is valid for any $\boldsymbol{x}_0 \in \mathbb{R}^N$ (not necessarily sparse) and can be used, for example, to prove that finite point clouds [31] and low-dimensional manifolds [32] in $\mathbb{R}^N$ can have stable, approximate distance-preserving embeddings under the matrix $\frac{1}{\sqrt{K}}\mathcal{O}_\Omega$. In each of these cases we may be able to solve very powerful signal inference and recovery problems with $MK \ll N$. When $\Omega = \{0, 1, \ldots, K-1\}$, one can further derive CoM inequalities for a scaled unitary matrix A.

Corollary 4: Assume $\Omega = \{0, 1, \ldots, K-1\}$. Suppose the same notation and assumptions as in Theorem 4. Suppose that $A = aU$ ($a \in \mathbb{R}, a \neq 0$) and $U \in \mathbb{R}^{N\times N}$ is unitary. Define $b := \Sigma_{k=0}^{K-1} a^{2k}$. Then for any fixed initial state $\boldsymbol{x}_0 \in \mathbb{R}^N$ and for any $\epsilon \in (0,1)$,

$$\mathbf{P}\left\{\left|\|\frac{1}{\sqrt{b}}\mathcal{O}_\Omega \boldsymbol{x}_0\|_2^2 - \|\boldsymbol{x}_0\|_2^2\right| > \epsilon\|\boldsymbol{x}_0\|_2^2\right\} \leq \begin{cases} 2\exp\left\{-\dfrac{MK\epsilon^2}{256\left(\left(1-a^2\right)K + a^2\right)}\right\}, & |a| < 1 \quad (25) \\ 2\exp\left\{-\dfrac{MK\epsilon^2}{256\left(\left(1-a^{-2}\right)K + a^{-2}\right)}\right\}, & |a| > 1. \quad (26) \end{cases}$$

Proof of Corollary 4 First note that when $A = aU$ (U is unitary) and $\Omega = \{0, 1, \ldots, K-1\}$ then $\|\mathcal{A}_\Omega \boldsymbol{x}_0\|_2^2 = (1 + a^2 + \cdots + a^{2(K-1)})\|\boldsymbol{x}_0\|_2^2 = b\|\boldsymbol{x}_0\|_2^2$. From (23) when $|a| < 1$,

$$\Gamma\left(\mathcal{A}_\Omega \boldsymbol{x}_0\right) = \frac{\left(1 + a^2 + \cdots + a^{2(K-1)}\right)^2}{1 + a^4 + \cdots + a^{4(K-1)}} = \frac{\left(1-a^{2K}\right)\left(1+a^2\right)}{\left(1+a^{2K}\right)\left(1-a^2\right)} = \frac{\frac{1+a^2}{1+a^{2K}}}{\frac{1-a^2}{1-a^{2K}}}. \tag{27}$$

[5]Please refer to our technical report [28] for more details on ϵ.

Also observe[6] that when $|a| < 1$,

$$\frac{1-a^2}{1-a^{2K}} \leq (1-a^2) + \frac{a^2}{K}. \tag{28}$$

Thus, from (27) and (28) and noting that $1 + a^2 \geq 1 + a^{2K}$ when $|a| < 1$, $\Gamma\left(\mathcal{A}_\Omega \boldsymbol{x}_0\right) \geq \frac{K}{(1-a^2)K+a^2}$. Similarly, one can show that when $|a| > 1$, $\frac{1-a^{-2}}{1-a^{-2K}} \leq (1-a^{-2}) + \frac{a^{-2}}{K}$ and consequently, $\Gamma\left(\mathcal{A}_\Omega \boldsymbol{x}_0\right) \geq \frac{K}{(1-a^{-2})K+a^{-2}}$. With appropriate scaling of $\mathcal{O}_\Omega$ by $\frac{1}{\sqrt{b}}$, the CoM inequalities follow from Corollary 2. ■

B. Implications for the RIP

Our CoM inequalities have immediate implications in establishing the RIP for the observability matrix. Based on Definition 2 and Lemma 1, we prove Theorems 1 and 2.

Proof of Theorem 1 In order to establish the RIP based on Lemma 1, we simply need to evaluate $f(\epsilon)$ in our CoM result derived in Corollary 4. One can easily verify that

$$f(\epsilon) = \begin{cases} \dfrac{\epsilon^2}{256\left((1-a^2)K + a^2\right)}, & |a| < 1 \quad (29) \\ \dfrac{\epsilon^2}{256\left((1-a^{-2})K + a^{-2}\right)}, & |a| > 1. \quad (30) \end{cases}$$

Through a union bound argument and by applying Lemma 1 for all $\binom{N}{S}$ S-dimensional subspaces in $\mathbb{R}^N$, the RIP result follows. ■

Proof of Theorem 2 In order to establish the RIP based on Lemma 1, we simply need to evaluate $f(\epsilon)$ in our CoM result derived in Corollary 3. In this case, $f(\epsilon) = \frac{\epsilon^2}{256}$. Through a union bound argument and by applying Lemma 1 for all $\binom{N}{S}$ S-dimensional subspaces in $\mathbb{R}^N$, the RIP result follows. ■

IV. Case Study: Estimating the Initial State in a Diffusion Process

So far we have provided theorems that provide a sufficient number of measurements for stable recovery of a sparse initial state under certain conditions on the state transition matrix and under the assumption that the measurement matrices are independent and populated with random entries. In this section, we use a case study to illustrate some of the phenomena implied by these bounds.

[6]In order to prove (28), for a given $|a| < 1$, let $C\left(a\right)$ be a constant such that for all K (K only takes positive integer values), $\frac{1}{1-a^{2K}} \leq 1 + \frac{C(a)}{K}$. By this assumption, $C\left(a\right) \geq \frac{Ka^{2K}}{1-a^{2K}} =: g\left(a, K\right)$. Observe that for a given $|a| < 1$, $g\left(a, K\right)$ is a decreasing function of K and its maximum is achieved when $K = 1$. Choosing $C\left(a\right) = g\left(a, 1\right) = \frac{a^2}{1-a^2}$ completes the proof of (28).

A. System Model

We consider the problem of estimating the initial state of a diffusion system. For detailed description and formulation of the considered diffusion process, please refer to our technical report [28]. Using a first difference approximation in space and time, a diffusion process can be approximated by $\boldsymbol{x}_k = A\boldsymbol{x}_{k-1}$, where $A = I_N + GT_s$, G represents the discrete Laplacian, and T_s is the discretization sampling time.

B. Diffusion and its Connections to Theorem 3

Before presenting the recovery results from compressive measurements, we would like to mention that our analysis in Theorem 3 gives some insight into (but is not precisely applicable to) the diffusion problem. In particular, the discrete Laplacian matrix G and the corresponding state transition matrix A (see below) are almost circulant, and so their eigenvectors will closely resemble the DFT basis vectors. The largest eigenvalues correspond to the lowest frequencies, and so the U_1 matrix corresponding to G or A will resemble a basis of the lowest frequency DFT vectors. While such a matrix does not technically satisfy the RIP, matrices formed from random sets of DFT vectors do satisfy the RIP with high probability [27]. Thus, even though we cannot apply Theorem 3 directly to the diffusion problem, it does provide some intuition that sparse recovery should be possible in the diffusion setting.

C. State Recovery From Compressive Observations

In this section, we consider a two-dimensional diffusion process with two types of measuring processes. The diffusion process is modeled on a two-dimensional grid with a total of 100 nodes, so that $N = 100$. We first look at random measurement matrices $C_k \in \mathbb{R}^{M\times N}$ where the entries of each matrix are i.i.d. Gaussian random variables with mean zero and variance $\frac{1}{M}$. Note that this type of measurement matrix falls within the assumptions of our theorems in Sections II and III. In the following, we refer to such measurements as "Dense Measurements." We also consider a more practical measuring process in which at each sample time the operator measures the nodes of the grid occurring along a line with random slope and random intercept. We refer to such measurements as "Line Measurements." For detailed descriptions of the considered measurement scenarios, please refer to our technical report [28].

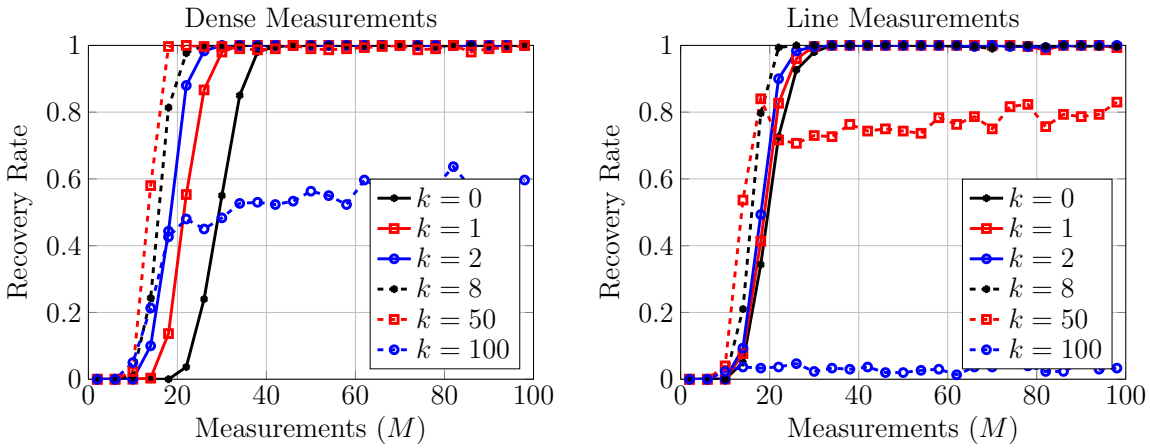
We first consider the situation where we collect measurements only at a single time k. We fix the sparsity level of $\boldsymbol{x}_0$ to $S = 9$. For various values of M, we construct measurement matrices C_k according to the two models explained above. At each trial, we collect the measurements $\boldsymbol{y}_k = C_k\boldsymbol{x}_k$ and attempt to recover $\boldsymbol{x}_0$ using the canonical ℓ_1-minimization problem from CS:

$$\widehat{\boldsymbol{x}}_0 = \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \|\boldsymbol{x}\|_1 \ \text{ subject to } \ \boldsymbol{y}_k = C_k A^k \boldsymbol{x}. \tag{31}$$

In order to imitate what might happen in reality (e.g., a drop of poison being introduced to a lake of water at $k = 0$), we assume the initial contaminant appears in a cluster of nodes on the associated diffusion grid. In our simulations, we assume the $S = 9$ non-zero entries of the initial state correspond to a 3×3 square-neighborhood of nodes on the grid. For each M, we repeat the recovery problem for 300 trials; in each trial we generate a random sparse initial state $\boldsymbol{x}_0$ (an initial state with a random location of the 3×3 square and random values of the 9 non-zero entries) and a measurement matrix C_0 as explained above. In order to see how the recovery performance would change as we take measurements at different times, we repeat the simulations for $k = \{0, 1, 2, 8, 50, 100\}$. The results are shown in Fig. 1(a) and Fig. 1(b) for Dense and Line Measurements, respectively. In both cases, the recovery performance starts to improve as we take measurements at later times. However, in both measuring scenarios, the recovery performance tends to decrease if we wait too long to take measurements. For example, as shown in Fig. 1(a), the recovery performance is significantly decreased at time $k = 100$ when Dense Measurements are employed. A more dramatic decrease in the recovery performance can be observed when Line Measurements are employed in Fig. 1(b). This behavior is expected and can be interpreted with the diffusion phenomenon. In fact, the bounds of Corollary 1 also show that how the requirement on the sufficient number of measurements grows at later times. If we wait too long to take measurements from the field of study (e.g., the lake of water), the effect of the initial contaminant starts to disappear (due to diffusion) and consequently measurements at later times contain less information. In summary, one could conclude that taking compressive measurements of a diffusion process at times that are too early or too late might decrease the recovery performance.

In another example, we fix $M = 32$, consider the same model for the sparse initial states with $S = 9$ as in the previous examples, introduce white noise in the measurements with standard deviation 0.05, use a noise-aware version of the ℓ_1 recovery algorithm [13], and plot a histogram of the recovery errors $\|\widehat{\boldsymbol{x}}_0-\boldsymbol{x}_0\|_2$. As can be seen in Fig. 2(a), at time $k = 2$ the Dense Measurements have lower recovery errors (almost half) compared to the Line Measurements. If we take measurements at later times, however, the recovery error of both measurement processes tends to be similar. Please refer to our technical report [28] to see how the recovery errors differ at $k = 2$ and $k = 10$, as an example.

Of course, it is not necessary to take all of the measurements only at one observation time. To this end, we perform the signal recovery experiments when a total of $MK = 32$ measurements are spread over $K = 4$ observation times (at each observation time we take $M = 8$ measurements). Figure 2(b) depicts a histogram of the recovery errors $\|\widehat{\boldsymbol{x}}_0 - \boldsymbol{x}_0\|_2$ when $MK = 32$ measurements are spread over 4 sample times at $k = \{10, 20, 30, 40\}$. As can be seen, the overall recovery performance is not much different



(a) (b)

Fig. 1: Signal recovery from compressive measurements of a diffusion process which has initiated from a sparse initial state of dimension $N = 100$ and sparsity level $S = 9$. The plots show the percent of trials (out of 300 trials in total) with perfect recovery of the initial state $\boldsymbol{x}_0$ versus the number of measurements M taken at observation times $k = \{0, 1, 2, 8, 50, 100\}$. (a) Recovery from compressive Dense Measurements. (b) Recovery from compressive Line Measurements.

compared to, say, taking $M = 32$ measurements at a single instant and so there is no significant penalty that one pays by slightly spreading out the measurement collection process in time, as long as a different random measurement matrix is used at each sample time. For detailed descriptions and simulation results, please refer to our technical report [28] where we repeat the simulations for different sample sets.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Alejandro Weinstein, Armin Eftekhari, Han Lun Yap, Chris Rozell, Kevin Moore, and Kameshwar Poolla for insightful comments and valuable discussions.

REFERENCES

[1] M. B. Wakin, B. M. Sanandaji, and T. L. Vincent, "On the observability of linear systems from random, compressive measurements," *in Proc. 49th IEEE Conference on Decision and Control*, pp. 4447–4454, 2010.

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

[3] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] E. J. Candès, "Compressive sampling," *in Proc. Int. Congress Math.*, vol. 3, pp. 1433–1452, 2006.

[5] E.J. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

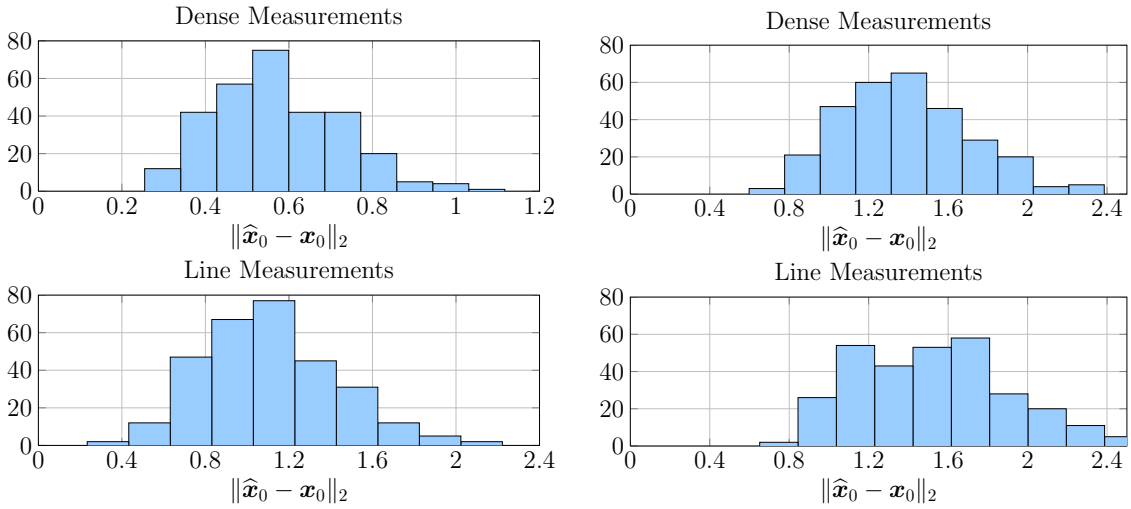


Fig. 2: Recovery error of the initial state ($N = 100$, $S = 9$) $\|\boldsymbol{e}\|_2 = \|\widehat{\boldsymbol{x}}_0 - \boldsymbol{x}_0\|_2$ over 300 trials from a total of $KM = 32$ observations: (a) All observations are taken at $k = 2$. (b) Observations are spread over $K = 4$ observation times while at each time, $M = 8$ measurements are taken.

[6] H. Ohlsson, L. Ljung, and S. Boyd, "Segmentation of ARX-models using sum-of-norms regularization," *Automatica*, vol. 46, no. 6, pp. 1107–1111, 2010.

[7] R. Tóth, B. M. Sanandaji, K. Poolla, and T. L. Vincent, "Compressive system identification in the linear time-invariant framework," *in Proc. 50th IEEE Conference on Decision and Control and European Control Conference*, pp. 783–790, 2011.

[8] B. M. Sanandaji, *Compressive System Identification (CSI): Theory and Applications of Exploiting Sparsity in the Analysis of High-Dimensional Dynamical Systems*, Ph.D. thesis, Colorado School of Mines, 2012.

[9] Parikshit Shah, Badri Narayan Bhaskar, Gongguo Tang, and Benjamin Recht, "Linear system identification via atomic norm regularization," *in Proc. 51th IEEE Conference on Decision and Control*, pp. 6265–6270, 2012.

[10] Jianguo Zhao, Ning Xi, Liang Sun, and Bo Song, "Stability analysis of non-vector space control via compressive feedbacks," *in Proc. 51th IEEE Conference on Decision and Control*, pp. 5685–5690, 2012.

[11] B. M. Sanandaji, T. L. Vincent, and M. B. Wakin, "Compressive topology identification of interconnected dynamic systems via clustered orthogonal matching pursuit," *in Proc. 50th IEEE Conference on Decision and Control*, pp. 174–180, 2011.

[12] Wei Pan, Ye Yuan, Jorge Goncalves, and Guy-Bart Stan, "Reconstruction of arbitrary biochemical reaction networks: A compressive sensing approach," *in Proc. 51th IEEE Conference on Decision and Control*, pp. 2334–2339, 2012.

[13] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *in Compte Rendus de l'Academie des Sciences, Paris, Series I*, vol. 346, pp. 589–592, 2008.

[14] R. G. Baraniuk, M. A. Davenport, R. DeVore, and M. B. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[15] M. A. Davenport, *Random Observations on Random Observations: Sparse Signal Acquisition and Processing*, Ph.D.

thesis, Rice University, 2010.

[16] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[17] D. Healy and D. J. Brady, "Compression at the physical interface," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 67–71, 2008.

[18] M. B. Wakin, S. Becker, E. Nakamura, M. Grant, E. Sovero, D. Ching, J. Yoo, J. K. Romberg, A. Emami-Neyestanak, and E. J. Candès, "A non-uniform sampler for wideband spectrally-sparse environments," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 3, pp. 516–529, 2012.

[19] J. Yoo, C. Turnes, E. Nakamura, C. Le, S. Becker, E. Sovero, M. B. Wakin, M. Grant, J. K. Romberg, A. Emami-Neyestanak, and E. J. Candès, "A compressed sensing parameter extraction platform for radar pulse signal acquisition," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 3, pp. 626–638, 2012.

[20] R. DeVore, G. Petrova, and P. Wojtaszczyk, "Instance-optimality in probability with an ℓ_1-minimization decoder," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 275–288, 2009.

[21] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Uniform uncertainty principle for Bernoulli and sub-Gaussian ensembles," *Constructive Approximation*, vol. 28, no. 3, pp. 277–289, 2008.

[22] J. Y. Park, H. L. Yap, C. J. Rozell, and M. B. Wakin, "Concentration of measure for block diagonal matrices with applications to compressive signal processing," *IEEE Trans. on Signal Processing*, vol. 59, no. 12, pp. 5859–5875, 2011.

[23] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. G. Baraniuk, "Signal processing with compressive measurements," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 445–460, 2010.

[24] E. Wang, J. Silva, and L. Carin, "Compressive particle filtering for target tracking," *in Proc. Statistical Signal Processing Workshop*, pp. 233–236, 2009.

[25] W. Dai and S. Yuksel, "Technical report: Observability of a linear system under sparsity constraints," *ArXiv preprint arXiv:1204.3097*, 2012.

[26] H. L. Yap, A. S. Charles, and C. J. Rozell, "The restricted isometry property for echo state networks with applications to sequence memory capacity," *in Proc.* 2012 *IEEE Statistical Signal Processing Workshop*, pp. 580–583, 2012.

[27] M. Rudelson and R. Vershynin, "On sparse reconstruction from fourier and gaussian measurements," *Communications on Pure and Applied Mathematics*, vol. 61, no. 8, pp. 1025–1045, 2008.

[28] B. M. Sanandaji, M. B. Wakin, and T. L. Vincent, "Technical report: Observability with random observations," *ArXiv preprint arXiv:1211.4077*, 2012.

[29] B. M. Sanandaji, T. L. Vincent, and M. B. Wakin, "Concentration of measure inequalities for compressive Toeplitz matrices with applications to detection and system identification," *in Proc.* 49*th IEEE Conference on Decision and Control*, pp. 2922–2929, 2010.

[30] B. M. Sanandaji, T. L. Vincent, and M. B. Wakin, "Concentration of measure inequalities for Toeplitz matrices with applications," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 109–117, 2013.

[31] P. Indyk and R. Motwani, "Approximate nearest neighbors: towards removing the curse of dimensionality," *in Proc. ACM Symposium on Theory of Computing*, pp. 604–613, 1998.

[32] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Foundations of Computational Mathematics*, vol. 9, no. 1, pp. 51–77, 2009.