Space–Time Sampling for Network Observability

Arash Amini, Hossein K. Mousavi, Student Member, IEEE, Qiyu Sun, and Nader Motee, Senior Member, IEEE

Abstract—Designing sparse sampling strategies is crucial in having resilient estimation and control in networked systems as they make sampling less expensive and networks more resilient w.r.t where and when samples are collected. It is shown under what conditions taking coarse samples from a network will contain the same amount of information as a finer set of samples. Our goal is to estimate the initial condition of linear time-invariant networks using a set of noisy measurements. The observability condition is reformulated as the frame condition, where one can easily trace the location and time stamps of each sample. We compare the estimation quality of various sampling strategies using estimation measures, which depend on the spectrum of the corresponding frame operators. Using properties of the minimal polynomial of the state matrix, deterministic and randomized methods are suggested to construct observability frames. Intrinsic tradeoffs assert that collecting samples from fewer subsystems dictates taking more samples (in average) per subsystem. Three scalable algorithms are developed to generate sparse space–time sampling strategies with explicit error bounds.

Index Terms—Networked control systems, state estimation, signal sampling, sparse sampling.

I. INTRODUCTION

A COMMON assumption in estimation theory is that samples are collected periodically from some prescribed output sensors [1], [2]. In practice, sampling strategies are designed subject to given performance criteria and hardware/software constraints, e.g., achieving certain estimation quality, data processing power, and battery life of the sensors and processors. Although oversampling may result in superior estimation quality, it is usually undesirable in networked systems that are equipped with spatially distributed sensors. The design of sparse sampling strategies, which allows collecting samples aperiodically from only a fraction of subsystems, will reduce sensing costs due to the existing algorithmic, physical, hardware, and software constraints. Our goal in this article is to propose a formal methodology to study properties and performance of various sampling strategies and devise scalable algorithms for sparse sampling strategies in space and time with provable performance bounds.

There have been recent interests in revisiting the notion of observability in the context of networked control systems. Bianchin et al. [3] revisited the notion of observability radius for a class of linear networks, whose state matrices are adjacency matrices of some weighted graphs. In [4] and [5], the problem of minimum constraint input selection is considered, where the objective is to find the smallest subset of inputs to ensure controllability. While it is shown in general that this problem is NP-hard, a subclass of such problems can be solved efficiently with the aid of network graph algorithms. Olshovsky [6] shows that the problem of approximating the minimum number of input (output) variables to guarantee controllability (and, equivalently, observability) is NP-hard. Tzoumas et al. [7] proposed the problem of sensing-constrained LQG control, where contrary to the classical LQG, they look at the minimal sensing requirements for the desired control objective and tackle the problem by solving for suboptimal sensing strategies by focusing only on finitely many sensing schedules. Our work is close in spirit to [8] and [9], where the authors investigate the controllability of linear-time invariant (LTI) networks and utilize randomized algorithms for sparsification [10] and (greedy) deterministic algorithms to obtain a sparse actuator scheduling. Prior to their work, several authors had also considered problems related to the sensor or actuator scheduling for control and estimation, for instance, see [11], [12], [13], [14], and the references therein. In related work, Ranieri et al. [15] consider the sensor placement problem with only a few sensors and employ ideas from frame theory to propose a near-optimal (in mean-square sense) greedy algorithm with guaranteed performance bounds. Our proposed methodology in this article is close in spirit to that of [15] as both works use frame theoretic ideas to obtain efficient approximate solutions for combinatorial sensor selection problems.

In this article, our focus is on estimating the initial condition of an LTI network from a set of state samples that are collected sparsely from a subset of subsystems aperiodically over some time interval. In Section IV, we apply tools from (finite) frame theory to reformulate the network observability problem and show that one can extract an observability frame from any given set of samples that solves the observability problem. This key idea allows us to cast observability conditions as whether a set of vectors forms a frame for the Euclidean space. This is particularly useful as every frame element is labeled by where and when that sample was taken. In Section VI, we propose deterministic and randomized methods to generate observability frames for
a given LTI network. It is shown that the minimum required number of samples from each subsystem (location) depends on the degree of the minimal polynomial of the state matrix. We show in Section VIII that there are inherent fundamental limits on the best achievable levels of estimation quality, and intrinsic tradeoffs reveal an interplay between space–time samples: Taking fewer samples (on average) per subsystem mandates collecting samples from more subsystems. In Section VII, we discuss the following three methods for frame sparsification:

1) sparsification by leverage scores, which is developed based on notions of spectral graph sparsification [10];
2) random partitioning using the Kadison–Singer paving solution [16];
3) greedy elimination using the Sherman–Morrison rank-one update rule [17].

In all these algorithms, we obtain explicit error bounds for the estimation-quality loss. We assert that our bounds are rather conservative. The reason is that, contrary to the results in [8], [9], [10], and [18], elements of a sparsified observability frame cannot be rescaled to compensate for the estimation-quality loss. This article is an outgrowth of its conference version [19] and contains several new technical results, proofs, and simulation results. All technical proofs of our results can be found in [20].

II. MATHEMATICAL NOTATIONS

The set of complex numbers, real numbers, nonnegative numbers, integers, and nonnegative integers are shown by C, R, R+ , Z, and Z+ , respectively, and the imaginary number √−1 by j. For a given number γ ∈ C, we define γZ := {γk | k ∈ Z}. For the n-dimensional Euclidean space Rn, we denote its standard basis by {e1, . . . , en} and the inner product of x, y ∈ Rn by ⟨x, y⟩. For a vector x ∈ Rn, ||x|| stands for its Euclidean 2-norm. For two families of vectors Φ1 and Φ2, Φ1 ⊂ Φ2 implies that φ ∈ Φ2 for all φ ∈ Φ1. We use the block capital letters to denote a matrix or a linear operator, e.g., X. The transpose of a matrix X is denoted by XT , the matrix exponential of a square matrix X by eX, and the identity matrix of appropriate size by I. Eigenvalues of a positive semidefinite matrix X ∈ Rn×n are indexed in ascending order, i.e., 0 ≤ λ1(X) ≤ · · · ≤ λn(X); similarly, singular values of a square matrix X are indexed from the smallest to the largest as 0 ≤ σ1(X) ≤ · · · ≤ σn(X); and the induced 2-norm is denoted by ∥X∥ = σn(X). Given two positive semidefinite matrices X and Y, we say that X ≤ Y if Y − X is positive semidefinite, and that X ≪ Y if Y − X is positive definite. A normal random variable with mean μ and covariance matrix Σ is denoted by N(μ, Σ). The expected value of a random variable is shown by E{·} and the probability of an event is denoted by P{·}. The cumulative distribution function (cdf) of a scalar normal variable N(μ, σ) is denoted by F(x; μ, σ). For sequences {an}n≥1 and {bn}n≥1 with positive elements, notation an = O(bn) implies that an/bn is bounded.

III. PROBLEM STATEMENT

We consider linear dynamical networks that consist of multiple subsystems with state vector

\[ x := [x_1, \ldots, x_n]^T \]  

where x ∈ R is the state variable of subsystem i ∈ {1, 2, . . . , n}. These subsystems are interconnected and their collective dynamics is governed by

\[ \dot{x} = Ax \]  

in which A is time-invariant. It is assumed that the initial state \( x_0 \in \mathbb{R}^n \) of the network is unknown. In order to recover the initial state, suppose that samples can only be collected from a subset of subsystems \( \Omega = \{i_1, i_2, \ldots, i_p\} \subset \{1, \ldots, n\} \), where \( \Omega \) is called the set of sampling locations. At every spatial location \( i \in \Omega \), sensors are allowed to take a finite number of samples with different time stamps; the set of such sampling times is denoted by \( \Theta_i \). A sampling strategy for subsystem \( i \in \Omega \) is given by the set of ordered pairs \( \mathcal{S}_i = \{(i, t) | t \in \Theta_i\} \).

A sampling strategy for the entire network can be obtained by

\[ \mathcal{S} = \bigcup_{i \in \Omega} \mathcal{S}_i. \]

For a given sampling strategy \( \mathcal{S} \), we stack all output samples (observations) in a column vector of dimension \( m \times 1 \) and denote it by

\[ y = [x_1(t) + \xi_1(t)]_{(i, t) \in \mathcal{S}} \]

where measurement noises \( \xi_i(t) \) in all samples are assumed to be independent from each other and have normal (Gaussian) distributions with zero mean and \( \sigma^2 \) variance. We highlight that \( y \) is not the output vector; it is the collection of all sampled outputs using \( \mathcal{S} \). For a given set of sampling locations \( \Omega \), the corresponding output matrix is defined by

\[ C_\Omega = [e_{i_1} | \ldots | e_{i_p}]^T. \]

Assumption 1: The set of sampling locations \( \Omega \) is chosen such that the pair (A, C_\Omega) is observable.

Verifying the observability of a network with respect to a given set of sampling locations is an important and active field of research on its own, and it is different from what we investigate here. For instance, Karger [24] adapts a graphical approach to identify those sensors that are necessary for the reconstruction of the initial state. Such results may offer options for A and \( \Omega \) that satisfy Assumption 1.

The research problem of this article is to characterize properties of sampling strategies that allow us to estimate the initial state of the linear network (2) using sparse sets of samples in space and time via scalable and efficient algorithms.

IV. CHARACTERIZATION OF SAMPLING STRATEGIES

We apply tools from finite frame theory to reformulate the observability problem and characterize its feasible sampling strategies.

A. Reconstruction in Frame Theory

The contents of this subsection are based on adjusted materials from [21].

Definition 1: For a given family of vectors \( \{\phi_i\}_{i=1,\ldots,m} \) in \( \mathbb{R}^n \), the corresponding analysis operator \( T : \mathbb{R}^n \to \mathbb{R}^m \) is defined

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
by
\[ T(x) := [\langle x, \phi_i \rangle]_{i=1, \ldots, m} \]
and its frame operator \( S : \mathbb{R}^n \to \mathbb{R}^n \) is defined by
\[ S(x) := \sum_{i=1}^{m} \langle x, \phi_i \rangle \phi_i. \]

It is straightforward to verify that operator \( T \) admits the following canonical matrix representation:
\[ T = [\phi_1 \ldots \phi_m]^T \in \mathbb{R}^{m \times n}. \]

Thus, the canonical matrix representation of the frame operator is
\[ S = T^T T \in \mathbb{R}^{n \times n}. \]

**Definition 2:** A family of vectors \( \langle \phi_i \rangle_{i=1, \ldots, m} \) in \( \mathbb{R}^n \) is a frame for \( \mathbb{R}^n \) if there exists constants \( 0 < \alpha \leq \beta \) such that
\[ \alpha \| x \|^2 \leq \sum_{i=1}^{m} | \langle x, \phi_i \rangle |^2 \leq \beta \| x \|^2 \]
for all \( x \in \mathbb{R}^n \).

The largest lower frame bound and smallest upper frame bound are called the optimal frame bounds.

**Proposition 1:** Let us consider a family of vectors \( \Phi = \langle \phi_i \rangle_{i=1, \ldots, m} \) in \( \mathbb{R}^n \). The following statements are equivalent.

i) The family of vectors \( \Phi \) forms a frame for \( \mathbb{R}^n \).

ii) The set of vectors \( \Phi \) spans \( \mathbb{R}^n \). Thus, \( m = |\Phi| \geq n \).

iii) The corresponding frame operator is positive definite, i.e., \( S \geq 0 \), with optimal frame bounds \( \alpha = \lambda_1(S) \) and \( \beta = \lambda_n(S) \).

iv) The corresponding analysis operator \( T \) is injective\(^1\) with a pseudoinverse
\[ T^\dagger := (T^T T)^{-1} T^T = S^{-1} T^T \]
which is a left-inverse of \( T \) that satisfies \( T^\dagger T = I \).

One of the well-studied problems in frame theory is to reconstruct an unknown vector \( x \in \mathbb{R}^n \) from the following vector of observations:
\[ y = T x = [\langle x, \phi_i \rangle]_{i=1, \ldots, m} \in \mathbb{R}^m. \]

The following known result highlights the role of \( T^\dagger \) in the reconstruction process from these observations.

**Proposition 2:** If the family of vectors \( \Phi \) forms a frame for \( \mathbb{R}^n \), then any vector \( x \in \mathbb{R}^n \), with a corresponding vector of observations \( y \in \mathbb{R}^m \) as in (10), can be reconstructed via
\[ x = T^\dagger y \]
where \( T \) is the analysis operator of \( \Phi \) and \( T^\dagger \) is given by (9).

**B. Initial State Reconstruction**

The solution of the linear network (2) is given by
\[ x(t) = e^{At} x_0 \]
where its \( i \)-th component is
\[ x_i(t) = e_{ii}^T e^{At} x_0 = \langle x_0, e^{At} e_i \rangle. \]

Based on the definition of a frame, (12) reveals that the following families of vectors are the only candidates for building constructors to recover the initial state of the network.

**Theorem 1:** Suppose that \( \Omega \) is the set of sampling locations and \( \Theta_i \) is the set of sampling times for each location \( i \in \Omega \). Every initial state of linear network (2) can be reconstructed from the set of samples that are collected according to sampling strategy \( \Theta = \{(i, t) | i \in \Omega, t \in \Theta_i \} \) if and only if the family of vectors
\[ \Phi(A, \Theta) = \left( e^{At} e_i \mid (i, t) \in \Theta \right) \]
is a frame for \( \mathbb{R}^n \).

The conclusion in Theorem 1 asserts that the initial state of the network can be recovered from the vector of observations
\[ y = T x_0 = [x_i(t)]_{(i, t) \in \Theta} \]
using the following equation:
\[ x_0 = T^\dagger y \]
where \( T \) is the analysis matrix of frame (13).

**Remark 1:** A frame for \( \mathbb{R}^n \) must contain at least \( n \) vectors. Hence, the number of components in frame (13) satisfies
\[ \sum_{i \in \Omega} |\Theta_i| \geq n. \]

This inequality implies that taking fewer spatial samples should be compensated by taking more temporal samples. This hints at an inherent tradeoff between the minimum number of samples in space and time required for successful initial state reconstruction.

It turns out that observability at the sampling locations is a necessary condition for the reconstruction problem.

**Lemma 1:** Suppose that the family of vectors (13) forms a frame for \( \mathbb{R}^n \). Then, the pair \((A, C_{\Omega})\) is observable.

This can be interpreted as follows: If the sampling locations \( \Omega \) create an unobservable output matrix \( C_{\Omega} \), then the initial state reconstruction will be always infeasible independent of the number of time samples.

In the rest of this article, whenever it is not ambiguous, we drop the argument of \( \Phi(A, \Theta) \) in (13) and simply write \( \Phi \). Whenever (13) forms a frame, it will be referred to as an observability frame. The space of all observability frames in \( \mathbb{R}^n \) is denoted by \( \mathcal{F} \).

**V. Estimation Measures**

In the previous section, the reconstruction problem was formulated in noise absence. One needs to solve an estimation problem when measurement noise is presented, which requires some appropriate mechanism to measure the quality of the resulting estimations. We start this section by showing that some useful estimation measures can be quantified in terms of the frame eigenvalues (i.e., eigenvalues of the frame matrix).
A. Estimation Measures

Instead of pure measurements (10), suppose that a noisy observation vector is collected
\[ \hat{y} = y + \xi \in \mathbb{R}^m \]  
(14)
in which \( \xi \in \mathbb{R}^m \) is a zero-mean Gaussian measurement noise with independent components and covariance \( \mathbb{E}\{\xi \xi^T\} = \sigma^2 I \).

For the linear network (2), the (14) can be rewritten in the following form:
\[ \hat{y} = \mathbf{T} x_0 + \xi \]  
(15)
where \( \mathbf{T} \) is the analysis matrix associated with the observability frame \( \Phi \) in (13). Let us denote an estimation of \( x_0 \) by \( \hat{x}_0 \) and define the corresponding estimation error as
\[ \eta := \hat{x}_0 - x_0. \]  
(16)

**Definition 3:** An operator \( \rho : \mathfrak{G} \to \mathbb{R} \) is called (decreasingly) monotone if \( \rho(\Phi_2) \leq \rho(\Phi_1) \) for all \( \Phi_1 \subseteq \Phi_2 \).

In the following, we discuss two common estimation measures to compare different observability frames.

i) **Standard Deviation of the Estimation Error:** For a given noisy observation vector (15) with underlying observability frame \( \Phi \), this estimation measure \( \rho_d(\Phi) := \sqrt{\mathbb{E}\{\|\eta\|^2\}} \).

This measure has been widely used to compute an optimal estimation via least-squares approximation \([22]\).

**Proposition 3:** Suppose that a noisy observation vector \( \hat{y} \) as in (15) is given. Then
\[ \hat{x}_0 = \mathbf{T}^\dagger \hat{y} \]  
(17)
is an unbiased estimator for \( x_0 \) with \( \mathbb{E}\{\hat{x}_0\} = x_0 \) that minimizes \( \|\mathbf{T} \hat{x}_0 - \hat{y}\|_2 \). Moreover, the (least-squares) estimation measure \( \rho_d(\Phi) := \sqrt{\mathbb{E}\{\|\eta\|^2\}} \).

Proposition 4: Under the Gaussian measurement noise assumption, the value of differential entropy of the estimation error is given by
\[ h(\eta) = \frac{1}{2} \rho_e(\Phi) + \frac{n}{2} \left(1 + \log(2\pi\sigma^2)\right) \]
with
\[ \rho_e(\Phi) = -\sum_{i=1}^n \log \left(\lambda_i(S) \right). \]  
(20)

Moreover, the above operator \( \rho_e : \mathfrak{G} \to \mathbb{R} \) is monotone.

B. Effects of Dwell-Time on Quality of Estimation

Suppose that sensors are scheduled to take samples according to a sampling strategy \( \mathcal{S} \), but actual measurements are taken with a uniform dwell time \( \delta \in \mathbb{R} \). Let us represent the resulting family of vectors by
\[ \Phi_\delta = \left( e^{A_T(t + \delta)} e_i \right)_{(i, t) \in \mathcal{S}}. \]  
(21)

One can equivalently represent this set using (13) as
\[ \Phi_\delta = \left( B_\delta \phi \mid \phi \in \Phi \right) \]
where \( B_\delta := e^{A_T\delta} \) is full rank for all \( \delta \in \mathbb{R} \). It is straightforward to verify that elements of \( \Phi_\delta \) span \( \mathbb{R}^n \) if and only if the elements of \( \Phi \) span \( \mathbb{R}^n \). Thus, the family of vectors \( \Phi_\delta \) is a frame for \( \mathbb{R}^n \) if and only if \( \Phi \) forms a frame for \( \mathbb{R}^n \). The next result shows that the estimation quality is not shift-invariant.

**Proposition 5:** Suppose that measurement noise (14) has normal distribution \( \mathcal{N}(0, \sigma^2 I) \). Then
\[ \rho_d(\Phi_\delta) \leq \sigma \left( \sum_{i=1}^n \sigma_i^2 \left( e^{-A_\delta} \right) \lambda_i(S)^{-1} \right)^{1/2} \]
and
\[ \rho_e(\Phi_\delta) = \rho_e(\Phi) - \sum_{i=1}^n \log \left( \sigma_i^2 \left( e^{-A_\delta} \right) \right) \]
in which \( \sigma_i(e^{-A_\delta}) \) is the ith singular value of \( e^{-A_\delta} \).

The upper bound (22) becomes tight for \( \delta = 0 \) because \( \sigma_i(e^{-A_\delta}) = \sigma_i(I) = 1 \) for all \( i = 1, \ldots, n \). When \( A \) is Hurwitz, according to inequality (22), the estimation quality deteriorates as \( \delta \) gets larger. The reason is that magnitude of samples decreases and the measurement noise (with constant intensity) becomes more dominant as time goes by. In fact, (22) implies that antistable state matrices neutralize negative effects of dwell time on the quality of estimation.

VI. CONSTRUCTION OF OBSERVABILITY FRAMES

Let us represent distinct eigenvalues of state matrix \( A \) by distinct eigenvalues \( \lambda_1(A), \ldots, \lambda_q(A) \) for some \( q \leq n \) and its corresponding minimal polynomial\(^2\) by
\[ p_A(\lambda) = \prod_{m=1}^q \left(\lambda - \lambda_m(A)\right)^{p_m} \]
(24)
\(^2\)The minimal polynomial of matrix \( A \) is the monic polynomial in \( A \) of the smallest degree such that \( p_A(A) = 0 \). This should not be confused with the characteristic polynomial of a matrix, which is always of degree \( n \) and only in certain cases coincides with the minimal polynomial \([23]\).

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
for some positive integers $p_m$, whose degree is denoted by $\delta(A)$, which is less than or equal to $n$. To state our next result, we need to define the row vector map

$$E(t) := \left[ E^{(1)}(t), \ldots, E^{(q)}(t) \right] \in \mathbb{R}^{1 \times \delta(A)}$$

(25)

where for $m = 1, \ldots, q$, we have

$$E^{(m)} = e^{i \Delta m(A)^T t} = \left[ 1, t, \ldots, t^{p_m-1} \right] \in \mathbb{R}^{1 \times p_m}.$$  

**Theorem 2:** Suppose that a sampling strategy $\mathcal{S} = \{ \{ i, t \} \mid i \in \Omega, \ t \in \Theta_i \}$ is adopted such that the following statements hold:

1) At every $i \in \Omega$, $M_i := |\Theta_i| \geq \delta(A)$ samples are collected.

2) $E_i$ has full column rank, where

$$E_i := \left[ E(t) \right]_{t \in \Theta_i} \in \mathbb{R}^{M \times \delta(A)}.$$  

Then, under Assumption 1, the family of vectors

$$\Phi = \left( e^{A^T t} e_i \mid (i, t) \in \mathcal{S} \right)$$  

(27)

forms a frame for $\mathbb{R}^n$.

Any frame for $\mathbb{R}^n$ must have at least $n$ vectors. Hence, prior to the application of Theorem 2, a necessary condition for the total number of sampling times is

$$|\mathcal{S}| = \sum_{i \in \Omega} |\Theta_i| \geq n.$$  

(28)

On the other hand, Theorem 2 requires $|\Theta_i| \geq \delta(A)$. Comparing these two arguments implies that the resulting frame $\Phi$ from Theorem 2 will have many redundant elements as the number of locations $|\Omega|$ increases. This motivates our investigation in the next section to seek scalable algorithms to construct sparse frames (in space and time) out of highly redundant observability frames.

According to Theorem 2, the sufficient number of samples at each location is $n$. This condition is rather conservative as it takes into account situations where samples are taken only from a very small (compared to $n$) subset of spatial locations. In Theorem 4, it is shown that if $|\Omega| = n$, then we may collect as few as one sample from each spatial location.

The set of all time instances for which $E_i$ is not full-column rank has zero Lebesgue measure in the corresponding design space. In fact, Theorem 2 suggests that one can comfortably skip the rank verification step.

**Corollary 1:** For a given $\tau > 0$, suppose that the sampling times in Theorem 2 are drawn randomly and independently from the uniform distribution over $[0, \tau]$. Then, with probability 1, the family of vectors in (27) is a frame for $\mathbb{R}^n$.

**Remark 2:** Theorem 2 does not directly advise us to choose certain locations and times for an optimal estimation quality. Nevertheless, the latter corollary motivates an approach for finding a sparse sampling strategy with acceptable quality. First, we can randomly construct a rich and dense set of space–time sampling indices. Then, we can use sparsification to discover the pivotal components of the sampling strategy (see Section VII).

**Example 1:** Let us consider the 2-D system

$$\dot{x} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} x.$$  

(29)

We choose to sample only from the first subsystem; i.e., $\Omega = \{1\}$, $C_\Omega = \begin{bmatrix} 1 & 0 \end{bmatrix}$. Thus, $(A, C_\Omega)$ is observable and

$$e^{A t} = \begin{bmatrix} \cos(t) & - \sin(t) \\ \sin(t) & \cos(t) \end{bmatrix}.$$  

Let us pick sampling times $t_1, t_2 \in [0, \tau]$, i.e., $M_1 = 2$. The corresponding family of vectors is

$$\Phi = \left( e^{A^T t_1} e_1, e^{A^T t_2} e_1 \right) = \begin{bmatrix} \cos(t_1) \\ \sin(t_1) \end{bmatrix}, \begin{bmatrix} \cos(t_2) \\ \sin(t_2) \end{bmatrix}.$$  

In this case, matrix (26) is

$$E_1 = \begin{bmatrix} e^{j t_1} & e^{-j t_1} \\ e^{j t_2} & e^{-j t_2} \end{bmatrix} \Rightarrow \det(E_1) = 2 \sin(t_1 - t_2).$$

Hence, according to Theorem 2, if $t_1 - t_2 \neq k \pi$ for $k \in \mathbb{Z}$, then $\Phi$ is a frame. Alternatively, if we compute the frame matrix $S$, using trigonometric identities, we get

$$\det(S) = \sin^2(t_1 - t_2)$$

which gives us the same constraints on the sampling times. Since the Lebesgue measure of the points for which $\sin^2(t_1 - t_2) = 0$ is indeed zero in $[0, \tau] \times [0, \tau]$, any random choices for $t_1$ and $t_2$ will result into a frame with probability 1. The latter observation agrees with Corollary 1. Next, we consider sampling $M_1 = M$ samples at location 1 for $M > 2$. By induction on sampling times $\Theta = \{ t_1, \ldots, t_M \}$, we have

$$\det(S) = \sum_{i = 1}^{t_M} \sin^2(t_i - t_j).$$

(30)

Again, if $t_i - t_j \neq k \pi$ for $k \in \mathbb{Z}$, we get a frame out of these observations. Random sampling also results in a frame with probability 1. Moreover, $\rho_2(\Phi) = \sqrt{2} \sigma / \sqrt{\det(S)}$ and it is shift-invariant.

Now, we briefly look at periodic sampling strategy, i.e., $\Theta_i = \{0, \delta, \ldots, (M_i - 1) \delta\}$ for every $i \in \Omega$, where $\delta > 0$ is a known sampling step-size.

**Theorem 3:** Suppose that the sampling step-size satisfies

$$\lambda_m(A) - \lambda_{m'}(A) \delta \notin 2\pi j \mathbb{Z}$$

(31)

for all distinct eigenvalues $\lambda_m(A)$ and $\lambda_{m'}(A)$ of the state matrix $A$. If $M_i \geq \delta(A)$, then the family of vectors

$$\Phi = \left( B_\delta^i e_i \mid i \in \Omega, k = 0, \ldots, M_i - 1 \right)$$

(32)

forms a frame for $\mathbb{R}^n$, where $B_\delta := e^{A^T \delta}$.

**Example 2 (Example 1 continued):** A sufficient condition for the sampling step-size for linear system (29) is

$$j \delta - (-j) \delta = 2j \delta \notin 2\pi \mathbb{Z} \Rightarrow \delta \notin \pi \mathbb{Z}.$$
Alternatively, because $t_i - t_j = (i - j)\delta$, using the expression for $\det(S)$ in (30), $\Phi$ is a frame if $\delta \notin \pi \mathbb{Z}$.

For a state matrix $A$ whose all eigenvalues are real, one may verify that the requirement (32) for any positive step-size $\delta$. Therefore we have the following corollary by Theorem 3.

**Corollary 2:** If all eigenvalues of $A$ are real, then for every step-size $\delta > 0$ and sampling horizon $M_i \geq \delta(A), i \in \Omega$, the family of vectors (32) is a frame for $\mathbb{R}^n$.

Now, we consider the case where collecting samples from all locations is possible and samples are taken in a small time range.

**Theorem 4:** Suppose that $\Omega = \{1, \ldots, n\}$ is the set of sampling locations and the set of sampling times $\Theta$, for each sampling location $i \in \Omega$, is chosen such that $|\Theta_i| = M_i \geq 1$ and $|t^*, t^* + \delta^*) \cap \Theta_i \neq \emptyset$ for some $t^* \in \mathbb{R}$, where step-size $\delta^* > 0$ is given by

$$\delta^* < (\ln 2) \|A\|^{-1}. \quad (34)$$

Then, the sampling strategy

$$\mathcal{S} = \{ (i, t) \mid i = 1, \ldots, n \text{ and } t \in \Theta_i \}$$

results in a family of vectors

$$\Phi = \left( e^{A^T t \epsilon_i} \mid (i, t) \in \mathcal{S} \right) \quad (35)$$

that forms a frame for $\mathbb{R}^n$.

The time range $\delta^*$ in Theorem 4 only depends on the state matrix $A$ and is strictly positive. Next, we consider the case where collecting samples from all locations is possible and samples are taken randomly in a time range $[0, \tau]$, which is not necessarily in a small time range.

**Corollary 3:** Suppose that samples are collected from all subsystems, i.e., $\Omega = \{1, \ldots, n\}$, at least once, i.e., $|\Theta_i| \geq 1$. Sampling times are drawn randomly and independently from the uniform distribution over interval $[0, \tau]$. Then, the resulting family of vectors (35) is a frame for $\mathbb{R}^n$ with probability 1.

**Example 3 (Example 1 continued):** For the linear system (29), let us consider a full state sampling with strategy $\mathcal{S} = \{ (1, t_1), (2, t_2) \}$ that results in vectors

$$\Phi = \left( e^{A^T t \epsilon_1}, e^{A^T t \epsilon_2} \right) = \left( \begin{bmatrix} \cos(t_1) \\ -\sin(t_1) \end{bmatrix}, \begin{bmatrix} \cos(t_2) \\ \sin(t_2) \end{bmatrix} \right). \quad (36)$$

For the corresponding frame matrix, we have

$$\det(S) = 1 - \sin^2(t_1 - t_2).$$

Thus, $\Phi$ is a frame for $\mathbb{R}^2$ if and only if

$$t_1 - t_2 \neq \left( k + \frac{1}{2} \right) \pi \text{ for all } k \in \mathbb{Z}. \quad (38)$$

Alternatively, $\delta^*$ in Theorem 4 satisfies

$$\delta^* < \ln 2. \quad (39)$$

According to Theorem 4, if sampling times $t_1$ and $t_2$ satisfy

$$|t_1 - t_2| < \ln 2 \quad (39)$$

then the family of vectors (36) is a frame for $\mathbb{R}^n$. Comparing the two constraints characterized by (39) and (38) reveals that the resulting condition for sampling times from Theorem 4 is more conservative. To verify the effectiveness of Corollary 3, one can verify that the Lebesgue measure of all pairs of points $\{t_1, t_2\}$ in $[0, \tau]^2 \subset \mathbb{R}^2$ for which $\det(S) = 0$. As a result, one can randomly select sampling times $\{t_1, t_2\}$ to construct a frame with probability 1.

**VII. FRAME SPARSIFICATION**

Suppose that a highly redundant set of samples from network (2) is provided for the estimation problem. This usually happens when a conservative sampling strategy $\mathcal{G}$ is used and all subsystems are allowed to collect numerous samples over time. Even if the corresponding family of vectors

$$\Phi = \left( e^{A^T t \epsilon_i} \mid (i, t) \in \mathcal{S} \right) \quad (40)$$

forms a frame, i.e., the resulting estimation problem is feasible, there are still unnecessary and undesired degrees of redundancy that should be trimmed away in order to enhance scalability properties of the estimation algorithms.

**Definition 4:** For a given design parameter $\theta > 0$, a family of vectors $\Phi_s$ in $\mathbb{R}^n$ is called a $\theta$-approximation of frame $\Phi$ if the following statements hold.

i) $\Phi_s \subset \Phi$ and it has at most $\theta|\Omega|$ elements.

ii) $\Phi_s$ is a frame for $\mathbb{R}^n$.

For a given (highly redundant) frame (40) and some parameter $\theta > 0$, our goal is to find a sampling strategy $\mathcal{S}_s$ whose corresponding family of vectors

$$\Phi_s = \left( e^{A^T t \epsilon_i} \mid (i, t) \in \mathcal{S}_s \right)$$

is a $\theta$-approximation of (40). Condition (i) mandates the sampling strategy $\mathcal{S}_s$ to collect at most $\theta$ samples in average from all subsystems, i.e.,

$$\Theta(\mathcal{S}_s) = \frac{1}{|\Omega|} \sum_{i \in \Omega} |\Theta_i(\mathcal{S}_s)| = \frac{|\mathcal{S}_s|}{|\Omega|} \leq \frac{\theta|\Omega|}{|\Omega|} = \theta.$$

Condition (ii) ensures the feasibility of the resulting estimation problem. In the following, we will discuss three methods to achieve our goal.

**Remark 3:** Sparsity is a relative notion. In this article, it is reasonable to consider a sampling strategy to be sparse if it takes an almost linear number of samples (in terms of network size) in a given time window. This is the level of sparsity that the main result of this section, Theorem 5, achieves; we refer to the following section.

**A. Sparsification by Leverage Scores**

Our first approach is based on sparsification via the notion of effective resistances [10], which depends on the concentration properties of the sums of random outer-products and was originally developed for sparsification of weighted graph Laplacians. Similar to graph Laplacian, the frame matrix $S$ is also a sum of rank-one matrices

$$S = \sum_{\phi \in \Phi} \phi \phi^T.$$
Algorithm 1: Randomized Frame Sparsification.

**input:** frame $\Phi = \{\phi_1, \ldots, \phi_{|\Omega|}\}$ and design parameters $q, \epsilon > 0$

**output:** set of vectors $\Phi_s$ and weight function $w_s$

**initialize:** $\Phi_s = \emptyset, w_s(.) = 0, S_s = 0$

for $k = 1$ to $q$ do

sample an element from $\Phi$ with probability distribution $\pi \rightarrow \phi$

update weight function: $w_s(\phi) \leftarrow w_s(\phi) + (q\pi(\phi))^{-1}$

if $\phi \notin \Phi_s$, then

add $\phi$ to $\Phi_s$

update the frame matrix: $S_s \leftarrow S_s + \phi\phi^T$

end if

end for

**Definition 5:** For a given finite frame $\Phi$, the leverage scores are positive numbers that are defined by

$$r_\phi(S) := \phi^T S^{-1} \phi$$  \hspace{1cm} (41)

for every $\phi \in \Phi$.

One can associate a probability mass function $\pi : \Phi \rightarrow [0, 1]$ to a given frame $\Phi$ using its leverage scores as follows:

$$\pi(\phi) = \frac{r_\phi(S)}{n}.$$ \hspace{1cm} (42)

This gives a well-defined mass function as

$$\sum_{\phi \in \Phi} \pi(\phi) = \frac{1}{n} \sum_{\phi \in \Phi} \text{Tr}(S^{-1} \phi\phi^T) = 1.$$ \hspace{1cm} (43)

As it is summarized in Algorithm 1, elements of $\Phi$ are sampled iteratively and independently, with replacement, according to the probability mass function (42). A sampled element will be added to $\Phi_s$ if it is not already in $\Phi_s$. The resulting sparsified frame $\Phi_s$ will have at most $q$ elements. One may estimate $|\Omega|$, i.e., the number of sampling locations after sparsification, and obtain a reasonable estimate for $\theta \leq q/|\Omega|$. Algorithm 1 also assigns a weight to every element of $\Phi_s$ via weight function $w_s : \Phi \rightarrow \mathbb{R}$. Each execution of Algorithm 1 returns a different realization of $w_s$, where $w_s(\phi) = 0$ for $\phi \notin \Phi_s$. These weights are useful in quantifying the estimation-quality loss due to sparsification. In fact, the weight function $w_s$ is a bounded random variable, where $w_s(\phi)$ may assume different realizations drawn from $\{p(q\pi(\phi))^{-1} | p = 0, 1, \ldots, q\}$.

**Theorem 5:** For a given frame $\Phi$ in $\mathbb{R}^n$, let us fix parameter $\epsilon \in (1/\sqrt{n}, 1]$ and the number of samples $q = O(n \log n/\epsilon^2)$. Then, the resulting set of elements $\Phi_s$ from Algorithm 1 is also a frame for $\mathbb{R}^n$ with a probability of at least $1/2$. Furthermore, with probability at least $1/4$, the estimation-quality losses satisfy

$$\frac{\rho_d(\Phi_s) - \rho_d(\Phi)}{\rho_d(\Phi)} \leq -1 + \sqrt{\frac{4\bar{\chi}}{1 - \epsilon}}$$ \hspace{1cm} (43)

$$\rho_e(\Phi_s) - \rho_e(\Phi) \leq n \log \left(\frac{4\bar{\chi}}{1 - \epsilon}\right)$$ \hspace{1cm} (44)

Using monotonicity property of the estimation measures, one can show that upper bounds in (43) and (44) are nonnegative.

where $\bar{\chi} := \mathbb{E}\{\chi\}$ and $\chi$ is a random variable given by

$$\chi := \inf \left\{ \gamma > 0 \mid \sum_{\phi \in \Phi} w_s(\phi) (\gamma - w_s(\phi)) \phi\phi^T \geq 0 \right\}.$$ \hspace{1cm} (45)

The backbone of this result is based on [10, Th. 1] and asserts that Algorithm 1 trims off a given (highly redundant) frame and returns, with probability of more than 0.5, a new frame whose size is almost linear in network size. Moreover, it is shown that, with probability at least 0.25, the estimation measures of the new (sparsified) frame stay within constant multiples/difference of the estimation measure of the original (redundant) frame.

**Corollary 4:** The random variable $\chi$, which is defined by (45), satisfies

$$\chi \leq \max_{\phi \in \Phi} w_s(\phi)$$

almost surely. Moreover, under the settings of Theorem 5, inequalities

$$\frac{\rho_d(\Phi_s) - \rho_d(\Phi)}{\rho_d(\Phi)} \leq -1 + \sqrt{\frac{4w_{\text{max}}}{1 - \epsilon}}$$ \hspace{1cm} (46)

$$\rho_e(\Phi_s) - \rho_e(\Phi) \leq n \log \left(\frac{4w_{\text{max}}}{1 - \epsilon}\right)$$ \hspace{1cm} (47)

holds with probability at least $1/4$, where

$$w_{\text{max}} := \mathbb{E}\left\{\max_{\phi \in \Phi} w_s(\phi)\right\}.$$ \hspace{1cm}

The random variable (45) depends on random variable $w_s$, which itself depends on the leverage scores. Thus, depending on what type of frame elements are fed into Algorithm 1, the performance bounds in this corollary may become tight and show a meaningful relationship between the actual performance loss and the magnitude of parameter $w_{\text{max}}$.

The leverage scores disclose the importance of every component with respect to the entire frame for the sake of estimation. For instance, if the network is asymptotically stable, a component with a relatively large time label is expected to have a relatively small leverage score. Thus, such insignificant components are less likely to be sampled by Algorithm 1 and can be trimmed off to achieve a comparable estimation quality.

**Running Time Analysis:** Computing the inverse of $S$ can be done in $O(n^3)$ operations while computing the leverage scores using this matrix requires $O(|\Phi|n^2)$. We need to check for repeated samples, which does not increase the running time of the algorithm. Computing the frame matrix $S_s$ can be done in $O(n \log n/\epsilon^2 \times n^2) = O(n^3 \log n/\epsilon^2)$ operations. Hence, the total running time of Algorithm 1 is $O(n^3 \log n/\epsilon^2 + |\Phi|n^2)$.

**B. Random Partitioning and Kadison–Singer Paving Solution**

If the leverage scores (41) are uniformly bounded by a small enough number, then components of a frame can be partitioned in a balanced manner in order to obtain two separate subframes with explicit bounds on their spectra. Such spectral bounds are useful to find bounds on the estimation quality of the resulting subframes. The next theorem is based on [16, Corollary 1.3]
that gives us a paving solution to the famous Kadison–Singer problem.

**Proposition 6:** Suppose that the leverage scores in (41) satisfy
\[
\rho_d(\Phi_j) - \rho_d(\Phi) \leq \kappa(r^*)
\]
for all \( \phi \in \Phi \) and some positive number \( r^* < 1.5 - \sqrt{2} \). Let us randomly partition \( \Phi \) into two subfamilies \( \Phi_1 \) and \( \Phi_2 \) such that every element of \( \Phi \), independent of others, belongs to either of the partitions with probability 1/2. Then, with a positive probability, the resulting partition will satisfy
\[
\left(1 - \frac{1 + \sqrt{2r^*}}{2}\right) S \leq \sum_{\phi \in \Phi_j} \phi \phi^T \leq \left(1 + \frac{1 + \sqrt{2r^*}}{2}\right) S
\]
for \( j = 1 \) and 2.

For a given (highly redundant) observability frame, the result of Proposition 6 allows us to compute the relative/absolute estimation-quality degradation of the resulting partitions.

**Theorem 6:** Suppose that the leverage scores (41) satisfy (48). Then, the randomly partitioned subfamilies \( \Phi_1 \) and \( \Phi_2 \) from Proposition 6 are both frames for \( \mathbb{R}^n \) with a positive probability and the estimation-quality can be bounded as follows:
\[
\rho_d(\Phi_j) - \rho_d(\Phi) \leq \kappa(r^*)
\]
for \( j = 1, 2 \), where
\[
\kappa(r^*) := \left(1 - \frac{\sqrt{2r^* + 1}}{2}\right)^{-1/2} - 1.
\]
The quantity \( \kappa(r^*) \) is a worst case bound on the relative performance degradation of the randomly partitioned subframes \( \Phi_1 \) and \( \Phi_2 \). This function has been illustrated in Fig. 1. In simulations, we observe that comparably better bounds are achievable.

**Applicability of Random Partitioning:** When we deal with massive incoming samples (data) from the sensors, we expect the leverage scores to be small. In fact, we observe that the leverage scores satisfy
\[
\sum_{\phi \in \Phi} r_\phi(S) = n \Rightarrow \tilde{r}(S) = \frac{n}{|\Phi|}
\]
where \( \tilde{r}(\cdot) \) stands for the average of the leverage scores. Hence, as a rule of thumb, one should expect that for fairly balanced observability frames with size
\[
|\Phi| > \left[\frac{6 + 4\sqrt{2}}{n}\right]^2
\]
the leverage score of a typical component in \( \Phi \), in average, is less than \((6 + 4\sqrt{2})^{-1} = 1.5 - \sqrt{2} \approx 0.0858\).

**Running Time Analysis:** Computing the frame matrix for each partition can be done in \( O(|\Phi|n^2) \). Hence, the random partitioning can be done in \( O(|\Phi|n^2) \) operations.

**C. Greedy Sparsification**

In order to maintain a predetermined level of estimation quality and sparsity, one may consider using greedy algorithms that have been demonstrated to be useful in practice with satisfactory performance for a broad range of combinatorial problems [24], [25]. In greedy frame sparsification, the core idea at every iteration is to eliminate that component of the frame that will increase the value of a given estimation measure less than the others. At iteration \( k \), let us denote the remaining frame and its frame matrix by \( \Phi_k \) and \( S_k \), respectively. Eliminating a component \( \phi \) from \( \Phi_k \) corresponds to the following rank-one update
\[
S_{k+1} = S_k - \phi \phi^T
\]
with \( S_0 = S \). According to the Sherman–Morrison formula [26], one gets update rule
\[
S_{k+1}^{-1} = S_k^{-1} + S_k^{-1} \phi \phi^T S_k^{-1} \frac{1}{1 - \phi^T S_k^{-1} \phi}.
\]

**Proposition 7:** Upon eliminating a component \( \phi \) from an observability frame \( \Phi_k \), the estimation measures are updated according to
\[
\rho_d(\Phi_{k+1}) = \sqrt{\rho_d^2(\Phi_k) + \frac{\sigma^2 r_\phi(S_k^2)}{1 - r_\phi(S_k)}}
\]
\[
\rho_e(\Phi_{k+1}) = \rho_e(\Phi_k) - \log \left(1 - \rho_d(S_k)\right).
\]
At every iteration, the optimizer of \( \rho_d(\Phi_{k+1}) \) can be determined by solving the optimization problem
\[
\phi^*_d = \arg \min_{\phi \in \Phi_k} \frac{||S_k^{-1} \phi||^2}{1 - \phi^T S_k^{-1} \phi}
\]
and for \( \rho_e(\Phi_{k+1}) \) by solving
\[
\phi^*_e = \arg \min_{\phi \in \Phi_k} \phi^T S_k^{-1} \phi.
\]

Algorithm 2 details all necessary steps to compute a sparsification of a (redundant) observability frame, where we use notation \( \square \in \{d, e\} \). The algorithm stops whenever either the desired sparsity level \( s \in (0, 1) \) or a maximum allowable relative estimation error \( \epsilon > 0 \) has been achieved. This algorithm resembles the procedure of updating a performance measure of a linear consensus network when a new coupling link is added to the network [27]. The performance guarantees of the greedy methods in this context are a well-studied subject. In general, derivation of performance bounds heavily depends on...
Algorithm 2: Greedy Frame Sparsification.

**input:** frame $\Phi$ and its frame matrix $S > 0$, $s \in (0, 1)$, $\epsilon > 0$

**output:** frame $\Phi_s$

**initialize:** $S_s = S$, $\Phi_s = \Phi$

while $\frac{\rho_{\Theta}(\Phi_s)}{\rho_{\Theta}(\Phi)} \leq \epsilon$ and $|\Phi_s| \geq s$ do

find minimizer $\phi_{1s}^\ast$ via solving (56) or (57)

update the frame matrix

$$S_s^{-1} \leftarrow S_s^{-1} \frac{S_s^{-1} \phi_{1s}^T \phi_{1s}^{\ast T} S_s^{-1} \phi_{1s}^{\ast}}{1 - \phi_{1s}^T S_s^{-1} \phi_{1s}^{\ast}}.$$

update $\Phi_s \leftarrow \Phi_s \backslash \phi_{1s}^\ast$

end while

on the curvature conditions of the specific class of objective functions, e.g., submodularity, supermodularity, or weak forms of these properties; please see [28] and references in there.

**Running Time Analysis:** Since the frame matrix $S$ is positive definite, the Cholesky factorization is usually utilized to compute its inverse $S^{-1}$ with computational cost $n^3/3 + O(n^2)$; we refer to [29] and references therein for various algorithms to compute the inverse of a positive definite matrix. Then, at each iteration, one needs to compute and update the value of the estimation measure for every vector in $\Phi$, which takes $O(|\Phi| n^3)$. Thus, in order to achieve sparsity level $s$, one needs $O((1 - s)|\Phi| \times |\Phi| n^3) = O((1 - s)|\Phi|^2 n^3)$ operations. Since $|\Phi| \geq n$, Algorithm 2 can be implemented in $O((1 - s)|\Phi|^2 n^3)$ operations. Compared to the running time of the randomized sparsification $O(n^3 \log n) = |\Phi|^2 n^3$, the running time of the greedy method can be higher by an order of $|\Phi|^2$.

**Remark 4:** Our proposed algorithms in this section employ some results from [10] and [16]. The idea of sparsification via effective resistances are recently applied in the controls community to obtain network abstraction as well as actuator scheduling in large-scale network control systems [8] and [18]. The Kadison–Singer paving solution in [16] is also utilized in [8] as a method of randomized actuator scheduling.

**VIII. FUNDAMENTAL LIMITS AND TRADEOFFS**

For a given linear network (2) whose state vector is sampled based on an arbitrary sampling strategy, we show that there are fundamental limits and tradeoffs on the best achievable values for the estimation measures and space–time sparsity.

**Theorem 7:** Suppose that state of the $n$-dimensional linear network (2) is sampled under a sampling strategy $\Omega$ with a total number of samples $|\Omega|$. Then, the best achievable estimation measures are bounded from below by constants that are quantified by

$$\rho_d(\Phi) \geq \frac{\sigma_n}{\nu \sqrt{|\Omega|}}$$

(58)

and

$$\rho_e(\Phi) \geq n \log \left( \frac{n}{\nu^2 |\Omega|} \right)$$

(59)

where $\Phi$ is the observability frame corresponding to $\Omega$ and $\nu := \nu(A, \Omega)$ is defined by

$$\nu(A, \Omega) = \max_{t \in \Theta_1 \cup \cdots \cup \Theta_r} \| e^{A_1 T} \|.$$  

(60)

The inequalities (58) and (59) give us some convenient rules of thumb about the scaling properties of the estimation measures. For instance, if $|\Omega| = O(n^2)$, then it can be deduced from (58) that

$$\rho_d(\Phi) \geq \frac{cn}{\nu}$$

for some constant $c$. This implies that the estimation quality cannot be enhanced beyond a hard limit. Such limitations are important in network design as they are independent of the network size and sampling strategy. For more discussions on the significant role of fundamental limits in control, we refer to [30].

**Theorem 8:** For a given linear network (2), let us assume that there exists $\delta > 0$ such that all distinct eigenvalues of its state matrix satisfy

$$\lambda_1(A) - \lambda_k(A) \geq 2\pi |Z$$

(61)

and it is sampled according to a sampling strategy with property $\Theta_i \subset \delta Z_+$ for all $i \in \Omega$. If $A$ is Hurwitz, then universal (i.e., independent of the number of samples) fundamental limits on the best achievable estimation measures emerge as follows:

$$\rho_d(\Phi) \geq \sigma \sqrt{\text{Tr}(Q^{-1})}$$

(62)

$$\rho_e(\Phi) \geq \sqrt{- \text{Tr}(\log(Q))}$$

(63)

where $Q > 0$ is the observability Gramian, i.e., the unique solution of the Lyapunov equation

$$e^{A_1 T} Q e^{A_1} - Q + C_1^T C_1 \geq 0.$$  

(64)

Also, the lower bounds can be achieved if and only if

$$\hat{\Theta}(\Omega) := \frac{1}{|\Omega|} \sum_{i \in \Omega} |\Theta_i(\Omega)|.$$

(65)

**Theorem 9:** For given desired levels of estimation qualities $\rho_d^*, \rho_e^* > 0$ and parameter $\nu^* > 0$, let us consider all
Fig. 3. Space–time representation of the sampling strategies corresponding to frame Φ and sparsified frame Φs, with |Φ| = 1760 and |Φs| = 503. The space-axis shows the label of subsystems and time-axis represents the sampling time. Every black or blue dot depicts whether a sample is collected from the corresponding subsystem at that time. It is evident that there are significantly fewer black dots than the blue dots.

The result of Theorem 9 asserts that intrinsic tradeoffs emerge among all linear networks with similar estimation quality and parameter (60): Reducing the number of sampling locations must be compensated by increasing the average number of samples per subsystem and vice versa.

**IX. NUMERICAL SIMULATIONS**

Let us consider a linear dynamical network (2) that consists of n subsystems, which are randomly and uniformly distributed over a square-shape spatial domain [0, 1] × [0, 1]. The Euclidean (spatial) distance between the subsystems i and j is denoted by dis(i, j). If two subsystems lie in each other’s connectivity range, then there will be a coupling between the two subsystems and the corresponding entries in the state space will be nonzero numbers. More precisely, the state matrix \( A = [a_{ij}] \) is defined by

\[
a_{ij} = \begin{cases} 
\zeta_{ij} e^{-a \text{dis}(i,j)^b}, & \text{dis}(i, j) \leq d \\
0, & \text{dis}(i, j) > d
\end{cases}
\]

for some \( d > 0 \). The parameters \( a > 0 \) and \( 0 < b < 1 \) determine the decay rate of the couplings and spatial localization properties of the network. For instance, larger values of a and b result in more localized networks with short-range couplings. To make our study generic, the coefficients \( \zeta_{ij} \) are independently and randomly chosen from \( N(0, 1) \). In our simulations, we set \( n = 40, \ a = 1, \ b = 0.5, \) and \( d = 0.3 \). We generate and save one state matrix \( A \) that has both stable and unstable modes and use it in the following simulation studies. The spatial locations of subsystems and their coupling topology are illustrated in Fig. 2.

Constructing Observability Frame: We set \( \Omega = \{1, \ldots, n\} \), \( \tau = 0.12 \), and \( M_i = M = 44 \) and utilize Corollary 1 to construct an observability frame \( \Phi \) with format (27). The resulting frame contains \( nM = 1760 \) vectors with a space–time representation illustrated by blue dots in Fig. 3. For a noise intensity of \( \sigma = 0.1 \), the value of the (least-squares) estimation measure is \( \rho_d(\Phi) \approx 0.0967 \).

Randomized Frame Sparsification: For design parameters \( \epsilon = 0.5 \) and \( q = 590 \), we apply Theorem 5 and find a sparsified frame with \( |\Phi_s| = 503 \). The spatial locations and time stamps of the sampling strategy corresponding to the sparsified frame are

\[
\begin{align*}
\hat{\Theta}(\mathcal{S}) \cdot |\Omega| & \geq \left( \frac{\sigma \nu}{\nu' \rho_s} \right)^2 \quad (66) \\
\hat{\Theta}(\mathcal{S}) \cdot |\Omega| & \geq \frac{n}{\nu'^2} \exp \left( -\frac{\rho^*}{n} \right) \quad (67)
\end{align*}
\]
AMINI et al.: SPACE–TIME SAMPLING FOR NETWORK OBSERVABILITY

Fig. 6. Space–time representation of the sampling points of the sequence of frames \( \Phi_1, \ldots, \Phi_{12} \) that are separated by the dashed lines.

Fig. 7. Estimation measure of the shifted frames is compared to our theoretical upper bound (22).

Illustrated by black circles in Fig. 3. The value of the estimation measure is

\[ \rho_d(\Phi_s) \approx 0.1883. \]

The number of space–time samples has been reduced by 71% for the price of 95% relative estimation-quality loss.

Sparsification via Random Partitioning: Let us reconsider the frame \( \Phi \) shown in Fig. 3. By applying Proposition 6, random partitioning of \( \Phi \) leads into two subsets \( \Phi_1 \) and \( \Phi_2 \). In our simulations, we repeat the random partitioning procedure \( 5 \times 10^5 \) times. In each case, the minimum and maximum relative estimation-quality losses for \( \Phi_1 \) and \( \Phi_2 \), i.e.,

\[ \min_{j=1,2} \frac{\rho_d(\Phi_j) - \rho_d(\Phi)}{\rho_d(\Phi)} \quad \text{and} \quad \max_{j=1,2} \frac{\rho_d(\Phi_j) - \rho_d(\Phi)}{\rho_d(\Phi)} \]

is computed and saved. The histogram of these data is depicted in Fig. 4. In this simulation, the minimum and maximum degradations are less than 0.55 with a high probability. The theoretical estimate from Theorem 6 is

\[ r^* = \max_{\phi \in \Phi} r_\phi(S) \approx 0.0315 \Rightarrow \kappa(r^*) \approx 1.1441. \]

Our extensive simulations reveal that, in practice, one typically achieves comparably better estimation quality than our theoretical bounds (49). In these simulations, the number of space–time samples are reduced by almost 50% for the price of 55% estimation-quality loss (in most outcomes of the simulations).

Performance of Randomized Versus Greedy Algorithms: In this simulation, we compare the estimation quality of the resulting sparsified frames from Algorithms 1 and 2. Using Algorithm 1, we construct 25 different sparsified frames by selecting 25 different values for \( \epsilon \) in \( \{1/\sqrt{n}, 1\} \). We treat \( q \) as a control parameter and vary its value between 5902 and 153. For a fixed \( q \), we compute the value of the estimation measure for all 25 frames and save the one with the minimum value. When applying Algorithm 2, we change the desired sparsity level \( s \) to get a sequence of sparsified frames. The outcome of our simulations is depicted in Fig. 5, where one can observe that both methods result in almost similar estimation qualities. The only difference we can report is their running time (on a personal computer with an Intel processor using MATLAB): The randomized method (including 25 experiments per \( q \)) took about 1.68 s while the greedy method took 26.71 s. This is consistent with our running time analysis for both algorithms.

Sequential Frame Construction: We compute the value of \( \delta^* \), which is defined in Theorem 4, using the saved state matrix \( A \) and get \( \delta^* \approx 0.0434 \). We merge \( N = 12 \) subframes in the time horizon, where \( c_j = 0.25 \) is for every subframe \( \Phi_j \) for \( j = 1, \ldots, N \). Sampling times \( t_{ij} \) for each frame is chosen randomly and uniformly from time interval \( [t_j, t_{j+1}] \). The space–time representation of these frames is illustrated in Fig. 6. The resulting concatenated frame has \( |\Phi| = Nn = 480 \) components with estimation quality

\[ \rho_d(\Phi) \approx 0.1862. \]

Estimation Quality Deterioration With Time Shifts: Let us consider the first observability frame \( \Phi_1 \) in Fig. 6 with \( n \) components. First, we construct a family of observability frame \( \Phi_\delta \), which is defined in (21), by increasing \( \delta \) from 0 to 0.5. We compute the exact value of the (least-squares) estimation measure for every \( \Phi_\delta \). The result of our simulations is depicted in Fig. 7 along with our theoretical upper bound (22). One observes that our proposed upper bound is rather tight for all values of \( \delta \) in \([0,0.5]\).
X. CONCLUSION

In this article, we assume that measurement noises in (14) are Gaussian and independent of each other. When the measurement noises are dependent, the covariance of the estimation error (16) will be

\[ \Sigma_\eta = \left( T^T \Sigma_\xi^{-1} T \right)^{-1} \]  

(69)

where \( \Sigma_\xi \) is the covariance of the measurement noise. In general, we may not be able to expand (69) as a sum of rank-one matrices made of frame components. This was a useful property for our developments in Section VII. This case needs a thorough analysis, which is beyond the scope of this article. However, in the case that \( \Sigma_\xi \) is diagonal and with entries

\[ \Sigma_\xi = \text{diag}(\sigma_{i,t}^2)_{(i,t)\in \mathcal{E}} \]

i.e., the observations are spatially and temporally independent, one can conduct a similar analysis by defining the frame components as

\[ \Phi(A, \mathcal{E}) = \sigma_{i,t}^{-1} e^{A^T i} e_i \quad (i,t) \in \mathcal{E} \]

where \( \sigma_{i,t} \) is the variance of the observation error at \((i,t)\). Our results can be extended to include linear dynamical networks with arbitrary output matrices. In such networks, the sampling locations will be different from subsystem locations and the components of the observability frame (13) will take form \( e^{A^T i} e_i \), where \( c \) is a row of the output matrix. The main reason for working with output matrix (4) is to highlight inherent tradeoffs between the minimum required number of samples in space and time in order to achieve a certain estimation quality in linear dynamical networks; see the results of Section VIII.

The significance of fundamental limits and tradeoffs in Section VIII is that they reveal what is achievable and what is not. This is practically plausible as it prevents us from searching for sampling strategies with unachievable estimation qualities. The results of Section VII provide three methods to sparsify a given observability frame. Our theoretical error bounds are rather conservative. However, our extensive simulations assert that our proposed algorithms can achieve comparably better error bounds in practice.

REFERENCES

[1] B. D. Anderson and J. B. Moore, Optimal Filtering, vol. 21, Englewood Cliffs, NJ, USA: Prentice-Hall, pp. 29–95, 1979.

[2] S. J. Julier and J. K. Uhlmann, “Unscented filtering and nonlinear estimation,” Proc. IEEE, vol. 92, no. 3, pp. 401–422, Mar. 2004.

[3] G. Bianchin, P. Frasca, A. Gasparri, and F. Pasqualetti, “The observability radius of networks,” IEEE Trans. Autom. Control, vol. 62, no. 6, pp. 3006–3013, Jun. 2017.

[4] S. Pequito, S. Kar, and A. P. Aguiar, “On the complexity of the constrained input selection problem for structural linear systems,” Automatica, vol. 62, pp. 193–199, 2015.

[5] S. D. Pequito, S. Kar, and A. P. Aguiar, “A framework for structural input/output and control configuration selection in large-scale systems,” IEEE Trans. Autom. Control, vol. 61, no. 2, pp. 303–318, Feb. 2016.

[6] A. Olshesky, “Minimal controllability problems,” IEEE Trans. Control Netw. Syst., vol. 1, no. 3, pp. 249–258, Sep. 2014.

[7] A. Olshesky, “(Non) supermodularity of average control energy,” IEEE Trans. Control Netw. Syst., vol. 5, no. 3, pp. 1711–1718, Sep. 2018.

[8] P. Bientinesi, B. Gunter, and R. A. V. de Geijn, “Families of algorithms related to the inversion of a symmetric positive definite matrix,” ACM Trans. Math. Softw., vol. 35, pp. 1–22, 2009.

[9] R. H. Middleton, “Trade-offs in linear control system design,” Automatica, vol. 27, no. 2, pp. 281–292, 1991.

M. Siami, A. Olshesky, and A. Jadabaei, “Deterministic and randomized actuator scheduling with guaranteed performance bounds,” IEEE Trans. Autom. Control, vol. 66, no. 4, pp. 1686–1701, 2020.

A. Jadabaei, A. Olshesky, and M. Siami, “Limitations and tradeoffs in minimum input selection problems,” in Proc. Annu. Amer. Control Conf., 2018, pp. 185–190.

D. A. Spielman and N. Srivastava, “Graph sparsification by effective resistances,” SIAM J. Comput., vol. 40, no. 6, pp. 1913–1926, 2011.

V. Gupta, T. H. Chung, B. Hassibi, and R. M. Murray, “On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage,” in Proc. Annu. Amer. Control Conf., 2002, pp. 251–260, 2006.

J. Wu, Q.-S. Jia, K. H. Johansson, and L. Shi, “Event-based sensor data scheduling: Trade-off between communication rate and estimation quality,” IEEE Trans. Autom. Control, vol. 58, no. 4, pp. 1041–1046, Apr. 2013.

L. Shi, P. Cheng, and J. Chen, “Optimal periodic sensor scheduling with limited resources,” IEEE Trans. Autom. Control, vol. 56, no. 9, pp. 2190–2195, Sep. 2011.

T. Tanaka and H. Sandberg, “SDP-based joint sensor and controller design for information-regularized optimal LQG control,” in Proc. IEEE 54th Annu. Conf. Decis. Control, 2015, pp. 4486–4491.

J. Ranieri, A. Chebira, and M. Vetterli, “Near-optimal sensor placement for linear inverse problems,” IEEE Trans. Signal Process., vol. 62, no. 5, pp. 1135–1146, Mar. 2014.

A. W. Marcus, D. A. Spielman, and N. Srivastava, “Interlacing families II: Mixed characteristic polynomials and the Kadison-Singer problem,” Ann. Math., pp. 327–350, 2015.

M. Brookes, The Matrix Reference Manual, vol. 3, London, U.K.: Imperial College London, 2005.

M. Siami and N. Motiee, “Network abstraction with guaranteed performance bounds,” IEEE Trans. Autom. Control, vol. 63, no. 10, pp. 3301–3316, Oct. 2018.

H. K. Mousavi, Q. Sun, and N. Motiee, “Space-time sampling for network observability,” IFAC-PapersOnLine, vol. 51, no. 23, pp. 408–413, 2018.

H. K. Mousavi, Q. Sun, and N. Motiee, “Space-time sampling for network observability,” 2018, arXiv:1811.01303.

P. G. Casazza, G. Kutyniok, and F. Philipp, “Introduction to finite frame theory,” in Finite Frames. Berlin, Germany: Springer, 2013, pp. 1–53.

V. K. Goyal, J. Kovačević, and J. A. Kelner, “Quantized frame expansions with erasures,” Appl. Comput. Harmon. Anal., vol. 10, no. 3, pp. 203–233, 2001.

M. Burrow, “The minimal polynomial of a linear transformation,” Amer. Math. Monthly, vol. 80, no. 10, pp. 1129–1131, 1973.

D. R. Karger, “Random sampling and greedy sparsification for matroid optimization problems,” Math. Program., vol. 82, no. 1/2, pp. 41–81, 1998.

M. Siami and N. Motiee, “Growing linear dynamical networks endowed by spectral systemic performance measures,” IEEE Trans. Autom. Control, vol. 63, no. 7, pp. 2091–2106, Jul. 2018.

M. S. Bartlett, “An inverse matrix adjustment arising in discriminant analysis,” Ann. Math. Statist., vol. 22, no. 1, pp. 107–111, 1951.

M. Siami and N. Motiee, “Growing linear consensus networks endowed by spectral systemic performance measures,” IEEE Trans. Autom. Control, vol. 63, no. 7, pp. 2091–2106, Jul. 2018.

A. Olshesky, “(Non) supermodularity of average control energy,” IEEE Trans. Control Netw. Syst., vol. 5, no. 3, pp. 1177–1181, Sep. 2018.

P. Bientinesi, B. Gunter, and R. A. V. de Geijn, “Families of algorithms related to the inversion of a symmetric positive definite matrix,” ACM Trans. Math. Softw., vol. 35, pp. 1–22, 2009.

R. H. Middleton, “Trade-offs in linear control system design,” Automatica, vol. 27, no. 2, pp. 281–292, 1991.
Hossein K. Mousavi (Student Member, IEEE) received the B.Sc. degree in mechanical engineering from the Sharif University of Technology, Tehran, Iran, in 2014. Since 2015, he has been working toward the Ph.D. degree in mechanical engineering with Lehigh University, Bethlehem, PA, USA.

His research interests include optimal control and estimation in complex systems, decentralized decision-making, and machine learning.

Qiyu Sun received the Ph.D. degree in mathematics from Hangzhou University, Hangzhou, China, in 1990.

He is currently a Professor of Mathematics with the University of Central Florida, Orlando, FL, USA. He has authored or coauthored more than 100 papers and written a book An Introduction to Multiband Wavelets with N. Bi and D. Huang. His research interests include applied and computational harmonic analysis, sampling theory, phase retrieval, and graph signal processing.

Nader Motee (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2000, and the M.Sc. and Ph.D. degrees in electrical and systems engineering from the University of Pennsylvania, Philadelphia, PA, USA, in 2006 and 2007, respectively.

From 2008 to 2011, he was a Postdoctoral Scholar with the Control and Dynamical Systems Department, California Institute of Technology, Pasadena, CA, USA. He is currently a Professor with the Department of Mechanical Engineering and Mechanics, Lehigh University, Bethlehem, PA, USA. His research interests include distributed control systems, risk analysis, and robot perception.

Dr. Motee was the recipient of several awards including the 2019 Best SIAM Journal of Control and Optimization Paper Prize, the 2008 AACC Hugo Schuck Best Paper Award, the 2007 ACC Best Student Paper Award, the 2008 Joseph and Rosaline Wolf Best Thesis Award, the 2013 Air Force Office of Scientific Research Young Investigator Program Award, 2015 NSF Faculty Early Career Development Award, and a 2016 Office of Naval Research Young Investigator Program Award.