Localization and Estimation of Unknown Forced Inputs: A Group LASSO Approach

Rajasekhar Anguluri, Member, IEEE, Oliver Kosut, Member, IEEE, and Lalitha Sankar, Senior Member, IEEE

Abstract—This article studies the problem of locating the sparse set of sources of forcing inputs driving linear systems from noisy measurements when the initial state is unknown. This problem is particularly relevant to detecting forced oscillations in electric power networks. We express measurements as an additive model comprising the initial state and inputs grouped over time, both expanded in terms of the basis functions (i.e., impulse response coefficients).

Using this model, with probabilistic guarantees, we recover the locations and simultaneously estimate the initial state and forcing inputs using a variant of the group linear absolute shrinkage and selection operator (LASSO) method. Specifically, we provide upper bounds on 1) the probability that the group LASSO estimator incorrectly identifies the locations and 2) the $\ell_2$-norm of the estimation error. Our bounds depend on the number of measurements, inputs, and sensors; the sensor noise variance; and the minimum singular value of the observability and impulse response matrices.

Our theoretical analysis is one of the first ones to provide a complete treatment for the group LASSO estimator for the left invertible linear systems with delay. Finally, we validate the performance of the estimator on synthetic models and the IEEE 68-bus, 16-machine power system.

Index Terms—Forced oscillations (FOs), group linear absolute shrinkage and selection operator (LASSO), invariant zeros, source localization, sparsity, unknown input.

I. INTRODUCTION

LOW-FREQUENCY oscillations in the electric transmission grid are indicative of the type of disturbance afflicting the system. Natural oscillations, with frequencies between 0.1 and 2 Hz, are triggered by random load fluctuations and sudden network switching. In contrast, forced oscillations (FOs), with frequencies between 0.1 and 14 Hz, result from external inputs injected by malfunctioned devices, such as power system stabilizers (PSS), generator controllers and exciters, cyclic loads, etc. [1], [2]. FOs remain undamped for longer periods of time, and if not mitigated, they pose a greater risk to the power systems operation, potentially causing blackouts. A popular and inexpensive method adopted to mitigate FOs in power systems is disconnecting the sources triggering the oscillations [2], [3], [4]. This amounts to accurately locating the sources of FOs. Because installing sensors at each potential source is an impractical solution, recent research suggests using phasor measurement unit (PMU) measurements-based localization algorithms (see [2] for a survey on many algorithms ranging from physics-based to fully data-driven approaches). However, albeit their good performance on test cases, many algorithms lack theoretical guarantees. This deficiency makes it harder to quantify the performance and limitations of measurement-based methods on what is and is not possible.

We address the lack of guarantees of existing approaches by posing the localization problem as a regularized optimization problem—referred to as the group linear absolute shrinkage and selection operator (LASSO) estimator. The regularization term imposes sparsity constraints on the number of source locations, which is often the case in many practical systems, including power systems [3], [5]. The input to our optimization problem is the noisy measurements and dynamical system matrices. It returns the source locations and estimates of the unknown initial state and inputs (oscillatory or not) injected by these sources. Formally, we consider

$$\begin{array}{l}
\mathbf{\hat{x}}_0 \\
\mathbf{\hat{u}}
\end{array} \in \arg \min_{\mathbf{x}_0, \mathbf{\hat{u}}, \{\mathbf{u}_j\}_1^{\infty}} \left\| \mathbf{y} - \mathbf{O}\mathbf{x}_0 - \sum_{j=1}^{m} \mathbf{J}_j \mathbf{u}_j \right\|^2_2 + \lambda \sum_{j=1}^{m} \left\| \mathbf{u}_j \right\|^2_2 \tag{1}
$$

where $\mathbf{u}_j = [u_j(0), \ldots, u_j[N]]^T$ is a vector of inputs injected by the $j$th source, $j \in \{1, \ldots, m\}$, over a discrete time horizon $\{0, \ldots, N\}$; $\mathbf{y}$ is the noisy batch measurements collected by $p$ sensors over $\{0, \ldots, N\}$; $\mathbf{O}$ and $\mathbf{J}_j$ are the observability and forced impulse response matrices, respectively (see Section II for the actual expressions); and $\lambda \geq 0$ is the tuning parameter.

Let $\mathbf{S} = (\mathbf{x}_0, \mathbf{u}_1, \ldots, \mathbf{u}_m)$ be the unknown ground truth and $\mathbf{\hat{S}} = \{j : \mathbf{u}_j \neq 0\} \subset \{1, \ldots, m\}$ be the set of active sources. By a sparse number of sources, we mean that $|\mathbf{S}| = m^* \ll m$.

In the context of regression models, including linear, logistic, and functional models, a vast body of literature exists on quantifying the theoretical performance of the group LASSO estimator.
and its variants; for a sample, see [6], [7], [8]. However, these studies assume either \( \mathbf{J}_j \) and \( \mathbf{O} \) to be random or satisfy rather restrictive assumptions, both of these might not hold for \( \mathbf{J}_j \) and \( \mathbf{O} \) obtained from linear dynamical systems. Furthermore, \( \mathbf{J}_j \) associated with the nonzero input \( \mathbf{u}_j^r \) could be rank deficient, especially if the underlying linear dynamical system is only \( d \)-delay left invertible\(^1 \) [9]. Consequently, the optimization in (1) is not strictly convex in the optimization variables, even when the true sources set \( \mathcal{S} \) is known. Thus, there may exist multiple optimal solutions \( \hat{\mathbf{J}} \), and it is not clear if \( \hat{\mathcal{S}} \) is common for all these solutions. In this article, we address all these issues by imposing physically meaningful assumptions on \( \mathbf{O} \) and \( \mathbf{J}_j \).

**Paper Contributions:** The problem we introduce in (1) is distinct from state-of-the-art regularized-based optimization methods in seeking to localize inputs and estimate the initial state using delayed measurements over a block of time. For the estimator in (1), our main contributions are as follows.

1) We derive sufficient conditions under which the following holds with high probability: a) The estimation error in the \( \ell_2 \)-sense is bounded, and b) the localized sources match the true sources. A key contribution is that despite the rank deficiency of model matrices, we guarantee that the group LASSO can localize the sources correctly. For rank deficient \( \mathbf{J}_j \), we provide estimation guarantees for the delayed inputs (see Section III). Our result hinges on introducing and thresholding a mutual incoherence condition (MIC) on the augmented \( \mathbf{O} \) and \( \mathbf{J}_j \) matrices.

2) The time-domain MIC condition we introduce requires computing correlations among the matrices \( \mathbf{O} \) and \( \mathbf{J}_j \). This computation can be hard, especially for estimation horizon \( N \). To tackle this hurdle, we upper bound the time-domain MIC with a frequency-domain MIC. Interestingly, the latter is a sufficient condition if we were to consider a LASSO estimator in the frequency domain. We also establish a fundamental relationship between the performance of the proposed estimator and the absence of invariant zeros for the subsystem excited by nonzero inputs and thresholding the frequency-domain MIC.

3) We validate the group LASSO estimator’s performance on synthetic data and the IEEE 68-bus, 16-machine system. We implement our estimator using the alternating direction method multipliers (ADMM) method [10].

Going beyond the motivating example of FOs in electric power systems, the problem setup in (1) is general, and the formal results in this article can be used to localize and reconstruct sparse inputs for a variety of practical engineering systems modeled as linear dynamical systems.

**Related Literature:** In power systems, energy methods based on frequency-domain data and statistical signal processing methods (e.g., auto regressive (AR) and auto regressive moving average (ARMA) models) are commonly used to localize unknown forced oscillatory inputs. In [11], a Bayesian approach was used to localize sources based on the generators’ frequency response functions. In [12], the pseudoinverse of system transfer functions was used to localize the sources. Zhou et al. [13] leveraged the magnitude and phase responses of transfer functions between different buses to localize sources. Finally, machine learning and PCA methods for localization were explored in [3], [14] All these methods primarily focus on oscillatory inputs, which might not apply to nonoscillatory inputs, including malicious attacks. We address this limitation by casting the source localization problem as an unknown input recovery in linear dynamical systems.

The problem of source identification in the context input and sensor and attacks was studied in [15], [16], [17], and [18]. But they fail to address input estimation and are applicable only for noise-free systems with more sensors than inputs. Moreover, these methods rely on banks of input observers, which might not be practical for large-scale critical infrastructures. In [19] and [20], assuming known inputs, the authors obtained sample complexity results for reconstructing the initial state with sparsity constraints from randomly sampled measurements. Instead, the authors in [21] and [22] considered sparse input and nonsparse state reconstruction using offline and sequential measurements. However, these works do not address location recovery guarantees for the sparse set of unknown sources in the presence of an unknown nonsparse initial state. In contrast to these works, we consider a unified framework, based on a LASSO method, to jointly locate the sources, and estimate the sparse inputs along with the unknown initial state. As highlighted in several other nonsparsity-based input identification methods [15], [23], [24], our results also highlight the role of invariant zeros for sparse input recovery.

**Mathematical Notation:** We denote the vectors and matrices by boldface lowercase and uppercase letters. Denote the \( d \times d \) identity matrix by \( \mathbf{I}_d \). Denote the pseudoinverse of \( \mathbf{X} \) by \( \mathbf{X}^\dagger \). The range space of \( \mathbf{X} \) is defined by \( \mathcal{R}(\mathbf{X}) = \{ \mathbf{X} \mathbf{z} \vert \mathbf{z} \in \mathbb{R}^m \} \). Given \( \mathcal{S} \subseteq \{1, \ldots, m\} \) and \( \mathbf{x} \in \mathbb{R}^m \), we write \( \mathbf{x}_\mathcal{S} \) for the subvector of \( \mathbf{x} \) formed from the entries of \( \mathbf{z} \) indexed by \( \mathcal{S} \). Similarly, we write \( \mathbf{M}_\mathcal{S} \) for the submatrix of \( \mathbf{M} \) formed from the columns of \( \mathbf{M} \) indexed by \( \mathcal{S} \). For \( 1 \leq p < \infty \) and the vector \( \mathbf{x} = [x_1, \ldots, x_m]^T \), denote \( \| \mathbf{x} \|_p = \left( \sum_{i=1}^{m} |x_i|^p \right)^{1/p} \). Instead, \( \| \mathbf{x} \|_\infty = \max_i |x_i| \). The \( \ell_{a,b} \)-mixed-norm, with \( a, b \geq 0 \), of \( \mathbf{z} = [z_1^T, \ldots, z_r^T]^T \) is given by \( \| \mathbf{z} \|_{a,b} = \sum_{j=1}^{r} \| z_j \|_a^b \). By convention, \( \| \mathbf{z} \|_{a,0} \triangleq \sum_{j=1}^{r} I(\|z_j\|_a \neq 0) \), where \( I(\cdot) \) is the indicator function, counts the number of nonzero vectors.

**II. PROBLEM SETUP AND PRELIMINARIES**

In this section, we formulate a group LASSO optimization problem to estimate the initial state and inputs and to locate unknown sources in linear dynamical systems.

**A. Linear Dynamics Under Sparse Forced Inputs**

Consider the continuous-time linear system

\[
\dot{\mathbf{x}}_c(t) = \mathbf{A}_c \mathbf{x}_c(t) + \mathbf{B}_c \mathbf{u}_c^r(t), \quad t \in \mathbb{R}
\]

where \( \mathbf{x}_c(t) \in \mathbb{R}^m \) and \( \mathbf{u}_c^r(t) \in \mathbb{R}^m \) are the state and input, respectively. We assume the input to be sparse, that is, \( \| \mathbf{u}_c^r(t) \|_0 \leq m^* \ll m \) for all \( t \in \mathbb{R} \). In power systems, \( \mathbf{x}_c(t) \) consists of states of generators and their control systems, including rotor angles, speed deviations, field excitation voltage, etc. Instead, \( \mathbf{u}_c^r(t) = [u_{r,1}^c(t), \ldots, u_{r,m}^c(t)]^T \) is the vector of inputs triggered by the sources of FOs, among which only \( m^* \) locations are active.

---

\(^1\)A dynamical system is \( d \)-delay left invertible if \( u_j[k] \), for \( j \in \{1, \ldots, m\} \), can be uniquely recovered from noise-less measurements \( \{y[k], \ldots, y[k+d]\} \).
However, our model in (2), except for sparsity constraints, is general and allows for multidimensional unmodeled exogenous disturbances, benign faults, or adversarial attacks.

We consider the discrete-time dynamics of (2) together with a measurement equation
\[ x[k+1] = Ax[k] + Bu^*[k] \quad (3) \]
\[ y[k] = Cx[k] + v[k], \quad k = 0, 1, \ldots, \quad (4) \]
where \( A = e^{A dt}, \) \( B = \left( \int_0^t e^{A \tau} d\tau \right) B_z, \) and \( \delta t \) is the sampling time period, and \( u^*[k] = [u_1[k], \ldots, u_m[k]]^T. \) Furthermore, \( y[k] = [y_1[k], \ldots, y_p[k]]^T \) is the measurement, \( v[k] \) is distributed as \( N(0, \sigma^2 I) \) noise, and \( C \in \mathbb{R}^{p \times n} \) is the sensor matrix.

In our extended paper [25, Sec. IV.A.], we consider dynamics in (3) with process noise and also relax the diagonal covariance assumption on \( v[k]. \)

Let \( S = \{ j : u_j[k] \neq 0 \} \) for at least one \( k \geq 0 \) \( \subset \{1, \ldots, m\} \) and \( S^c = \{1, \ldots, m\} \setminus S. \) We refer \( S \) and \( S^c \) as the active and inactive sets, respectively. Partition \( B \) as \( B = [B_S B_{S^c}] \) and \( u^*[k] = [u^*_S[k], u^*_{S^c}[k]]^T, \) with \( u^*_S[k] = [u_1^*[k], \ldots, u_m^*[k]] \) and \( B_{S^c} = [b_1^*, \ldots, b_n^*], \) where \( i \in S^c \) and \( r = |S^c| = m - m^*. \) Similarly, define \( u^*_{S^c}[k] \) and \( B_S. \) Then, we have
\[ Bu^*[k] = \sum_{j=1}^m b_j u^*_j[k] = \sum_{j \in S} b_j u^*_j[k] + \sum_{j \in S^c} b_j u^*_j[k] = B_S u^*_S[k] + B_{S^c} u^*_{S^c}[k]. \quad (5) \]

The above representations will play a key role in formulating our group LASSO problem in Section II-B.

Using (3)-(4), we express the batch measurements \( y \) (see the following) as a linear model with added noise. Define the vectors
\[ y = \begin{bmatrix} y[0] \\ \vdots \\ y[N] \end{bmatrix}, \quad v = \begin{bmatrix} v[0] \\ \vdots \\ v[N] \end{bmatrix}, \quad \text{and} \quad u^*_j = \begin{bmatrix} u^*_j[0] \\ \vdots \\ u^*_j[N] \end{bmatrix}, \quad (6) \]
where \( y, v \in \mathbb{R}^{p(N+1)} \) and \( u^*_j \in \mathbb{R}^{N+1}, \) for all \( j \in S \cup S^c. \) Here, \( N + 1, \) with \( N > 0 \) is the length of the estimation horizon.

We also define the observability matrix \( O \in \mathbb{R}^{p(N+1) \times n} \) and the impulse response matrix \( J_j \in \mathbb{R}^{p(N+1) \times N+1} \) as
\[ O = \begin{bmatrix} C & CA & CA^2 & \cdots & CA^{N-1} \\ \vdots \\ \vdots \\ C A^N \end{bmatrix}; \quad J_j = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & H_1^{(j)} & \cdots & H_{N-1}^{(j)} \\ H_1^{(j)} & \cdots & \vdots & \cdots & \vdots \\ H_{N}^{(j)} & H_{N-1}^{(j)} & \cdots & \cdots & \vdots \end{bmatrix}, \quad (7) \]
where \( j \in S \cup S^c, \) and, for any \( l \geq 1, \) the \( l \)th impulse response parameter \( H_l^{(j)} \in \mathbb{R}^{p \times 1} \) is defined as \( H_l^{(j)} := CA^{l-1} b_j. \)

Let \( x[0] = x_0^* \) be the unknown initial state. From (3) and (4) and the fact that \( Bu^*[k] = \sum_{j=1}^m b_j u^*_j[k], \) we observe that
\[ y = Ox_0^* + \sum_{j=1}^m J_j u^*_j + v \quad (8) \]
where \( v \sim \mathcal{N}(0, \sigma^2 I_{p(N+1)}), \) \( u^*_j \) is in (6) and \( J_j \) is in (7).

B. Initial State and Unknown Input Estimation Under Sparsity Constraints: A Group LASSO for Approach

Based on the measurement model in (8), we introduce the group LASSO estimator to estimate \( \{x_0^*, u_1^*, \ldots, u_m^*\} \) and also the active set \( S. \) Let \( J = [J_1^T, \ldots, J_m^T] \) and \( u = [u_1^T, \ldots, u_m^T], \) where \( u_j \in \mathbb{R}^{N+1}. \) Recall the definition of \( \ell_{p,0} \)-norm from the notation section, and consider
\[ \left[ \hat{x}_0^* \hat{u} \right] = \arg \min_{x_0^*, u} \left\{ \frac{1}{2T} \|y - Ox_0 - Ju\|_2^2 + \lambda_T \|u\|_{p,0} \right\} \quad (9) \]
where the regularization parameter \( \lambda_T \geq 0 \) and \( T = p(N + 1) \) is the dimension of \( y \) in (8). The above problem is called a subset or (block-column) selection problem because the optimization problem amounts to finding \( J_j \) that contributes to \( y \) in (8).

Unfortunately, (9) is a combinatorial optimization problem, and its computational complexity is exponential in \( m. \) We circumvent this difficulty by replacing the \( \|u\|_{p,0} \) with the \( \|u\|_{p,1} \) norm. This is a common relaxation technique widely used in the literature of compressed sensing and statistics; see [26], [27]. Thus, we end up with the group LASSO problem
\[ \left[ \hat{x}_0^* \hat{u} \right] \in \arg \min_{x_0^*, u} \left\{ \frac{1}{2T} \|y - Ox_0 - Ju\|_2^2 + \lambda_T \|u\|_{p,1} \right\}. \quad (10) \]

For definiteness, we set \( p = 2, \) although our analysis extends to the case \( p \neq 2. \) In the literature, \( \|u\|_{2,1} = \sum_{j=1}^m \|u_j\|_2 \) is referred to as the block or group norm. Our problem (10) differs from the traditional group LASSO [7] because we do not penalize \( x_0. \) This is a subtle yet important distinction because the initial state is rarely sparse in many applications, including power systems. In Section VI, we provide details on how to numerically solve (10). Instead, in Section III, for a specific range of \( \lambda_T, \) we show that the group-norm-based regularizer promotes group sparsity in \( \hat{u} \) and that \( \hat{S} = S \) holds with high probability, where \( \hat{S} \triangleq \{ j : \hat{u}_j \neq 0 \} \).

Due to the presence of additive noise in the measurement vector \( y \) in (8), neither the estimate \( \hat{\beta} = [\hat{x}_0^*, \hat{u}] \) in (10) need to identically match \( \beta^* = (x_0^*, u^*) \) nor does \( \hat{S} = S. \) Thus, we evaluate the quality of our estimates (i.e., the hatted quantities) in a probabilistic sense using the error metrics, which are as follows.

1) \( \hat{\beta} \) is said to be \( \ell_2\)-consistent if \( \|\hat{\beta} - \beta^*\|_2 \leq o(T) \) with the probability of at least \( 1 - c_1 \exp(-c_2 T), \) for some \( c_1, c_2 > 0. \)

2) \( \hat{u} \) is said to be location recovery consistent if \( \hat{S} = S \) with the probability of at least \( 1 - c_3 \exp(-c_4 T), \) for \( c_3, c_4 > 0. \)

Here, \( o(T) \) implies that the upper bound on the error tends to zero as \( T \to \infty. \) The \( \ell_2\)-error bound ensures that the estimate \( \hat{\beta} \approx \beta^* \) by increasing \( T = p(N + 1). \) Finally, the location selection consistency ensures that as long as \( T \) is sufficiently large, \( \hat{S} \) correctly identifies the true sources of FOs.

III. DELAYED ESTIMATION AND INvariant Zeros

In this section, we cull recent results on the initial state and delayed input recovery using a finite number of measurements [28]
is the residual. In associated with delay \( J = 0 \), for any \( \ell < - \). Does \( \beta \) say that \( \beta \) using \( (14) = (0) \in J \). Let \( J = \) be a concatenation of \( J \) in \( d \geq 1 \), Note that \( y \) in (8) and equals \( y \) in (11). Importantly, \( u^*_S \) in (12) is a concatenation of inputs \( u^*_S[k] \) associated with \( S \) from \( k = 0 \) (top) to \( N \) (bottom), but not a concatenation of \( u^*_j \) in (6), for all \( j \in S \).

To show that the group LASSO is location recovery consistent, or \( \tilde{S} = S \) holds with high probability, \( \Psi_S = [0 \ \ J_S] \) in (11) should be of full column rank. To see this, suppose that \( \sigma^2 = 0 \) and that we know \( S \). Then, by substituting \( u^*_j = 0 \), for all \( j \in S^c \), and \( v = 0 \) in \( y \) in (11), it follows that

\[
y = \Psi_S \beta^*_S.
\]

Thus, to perfectly recover \( \beta^*_S = (x^*_0, u^*_S[0], \ldots, u^*_S[N]) \), even with noise-free measurements and the knowledge of \( S \), the matrix \( \Psi_S \) should be full column rank. However, unfortunately, unlike the model matrices, such as random design and Fourier basis matrices, considered in signal processing and statistics applications, \( \Psi_S \) could be rank deficient. This is so because the system in (3)–(4) may not be initial state and input observable [9]; that is, either \( O \) or \( J_S \) is rank deficient, or both \( O \) and \( J_S \) have full ranks, but \( O \) and \( J_S \) is rank deficient.

From the foregoing discussion, it is clear that recovering \( \beta^*_S \) and the full rank of \( \Psi_S \) are intimately connected. Interestingly, for \( d \)-delay invertible linear systems, even when \( \beta^*_S \) is not recoverable, a portion of it is perfectly recoverable [9], [28]. In fact, we can recover \( \beta^*_{S,[0:N-d]} = (x^*_0, u^*_S[0], \ldots, u^*_S[N-d]) \), where \( N \geq d \), from \( y^T = [y^T[0], \ldots, y^T[N]] \). Here, \( d \geq 0 \) is called delay and we refer \( \beta^*_{S,[0:N-d]} \) as the delayed input. As a result, we show that a specific submatrix of \( \Psi_S \) has full column rank even when \( \Psi_S \) is rank deficient.

We formalize the notion of \( d \)-delay. Let \( x_0 = 0 \) to note that \( \Psi_S = J_S \) and \( \beta^*_S = u^*_S \). Substituting \( J_S \) (12) into (13), yields

\[
\begin{bmatrix}
y[0] \\
y[1] \\
\vdots \\
y[N] \\
\end{bmatrix} = \begin{bmatrix}
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{H}^S & \mathbf{0} & \cdots & \mathbf{0} \\
\vdots & \ddots & \ddots & \vdots \\
\mathbf{H}^S & \mathbf{H}^S & \cdots & \mathbf{H}^S \\
\end{bmatrix} \begin{bmatrix}
u[0] \\
u_S^*[1] \\
\vdots \\
u_S^*[N] \\
\end{bmatrix}.
\]

Notice that \( J_S = J_{S,[0:N]} \) and \( u^*_S = u^*_S,[0:N] \). Define

\[
J_{S,[0:N]} = \begin{bmatrix}
J_{S,N} & J_{S,N-1} & \cdots & J_{S,0}
\end{bmatrix}
\]

for \( J_{S,[d]} \) defined in (14) and \( m^* \) is the dimension of \( u^*_S[k] \). The smallest \( d \) that satisfies (16) is denoted as \( \eta_S \).

We assume that \( d = \eta_S < \infty \) and that \( d = \infty \) if (16) does not hold for any \( d \geq 0 \). Then, from [28], we note that the block matrix \( \mathbf{M}^d_S \) in (15) is full column rank. Thus, there exists a matrix \( \mathbf{Q} \) such that \( \mathbf{Q} \Psi_{k+d} = u^*_S[k] \), where \( y_{k+d} = [y^T[k], \ldots, y^T[d+k]]^T \). We can then recover the input vector \( u^*_S[k+1] \) using the formula \( u^*_S[k+1] = \mathbf{Q} \Psi_{k+d} \Psi_{k+d} \mathbf{K}_{k+d} \). Here, \( \mathbf{K}_{k+d} \) is the residual. In words, we first subtract the response of \( u^*_S[k] \) from the measurements collected over \( \{k+1, \ldots, k+d+1\} \). We then use the residual \( \tilde{y}_{k+d+1} \) to recover the input \( u^*_S[k+1] \). Let \( k = 0 \). Then, by iterating the above procedure, we can recover \( u^*_S[0:N-d], \ldots, u^*_S[N-d] \) from \( y^N \).

We extend the rank condition in (16) to recover \( \beta^*_{S,[0:N-d]} = (x^*_0, u^*_S[0:N-d]) \) using \( y_N \). First, we define the smallest delay for recovering \( x^*_0 \) in the presence of input

\[
\mu_S \triangleq \min \{d \geq 0 : \text{Rank}(O_d J_{S,[d]} \Psi_{S}) = \text{Rank}(J_{S,[d]} \Psi_{S}) = n\}
\]

where \( O_d = [\mathbf{C}^T (\mathbf{C}^T \mathbf{A})^T, \ldots, (\mathbf{C}^T \mathbf{A}^d)^T], \) and \( n \) is the dimension of \( A \). The rank condition in (17) says that \( O_d \) has full column rank (= \( n \)) and that the columns in \( O_d \) are linearly independent of columns in \( J_{S,[d]} \). This condition is stronger than the system in (3)–(4) being observable, as shown ahead.

\textbf{Example 1:} Let \( A = \begin{bmatrix} 1 & 0; 2 & 0 \end{bmatrix}, B_S = \begin{bmatrix} 2 \ 3 \end{bmatrix}^T \), and \( C = \begin{bmatrix} 1 & 0 \end{bmatrix} \). Then, \( \eta_S = 1 \) and \( \text{Rank}(O_1) = 2, \) for any \( \ell \geq 2; \) that is, the system is observable. However, \( \mu_S = \infty \). To see this,
note that the second column of $A$ is identical to $B_S$; thus, the matrices $O_d$ and $J_{S,[d:0]}$ have some columns in common, and consequently, (17) does not hold for any $d \geq 0$.

For $N \geq d \geq 0$, define $\Psi_{S,[N:d]} = \{O, J_{S,N} \ldots J_{S,d}\}$ and $\Psi_{S,[d-1:0]} = [J_{S,d-1} \ldots J_{S,0}]$, where $J_{S,j}$ is in (15). Then

$$
\Psi_S \triangleq \left[ O \quad J_S \right] = \left[ \Psi_{S,[N:d]} \quad \Psi_{S,[d-1:0]} \right]
$$

(18)

with an understanding that $\Psi_S = \Psi_{S,[N:d]}$ for $d = 0$. Let $\Psi_S^+$ be the pseudoinverse of $\Psi_S$. The following proposition establishes conditions under which we can recover $(\tilde{x}_0^*, u_{S,[0:N-d]}^*)$.

**Proposition 1:** Suppose that $\eta_S$ in Definition 1 and $\mu_S$ in (17) are finite. Then, for $N \geq \max\{\eta_S, \mu_S\}$ with $d \geq \eta_S$, we have the following.

1) $\Psi_{S,[N:d]}$ defined in (18) has full column rank.
2) $\mathcal{R}(\Psi_{S,[N:d]} \cap \mathcal{R}(\Psi_{S,[d-1:0]} = \{0\})$.

Moreover, for $t \equiv (N - d + 1)m^*$ and $m^* = |S|$, we have

$$
\begin{bmatrix}
\tilde{x}_0^* \\
\tilde{u}_{S,[0:N-d]}^*
\end{bmatrix} = \left[ I_{n+t} \quad 0_{(n+t) \times d m^*} \right] \Psi_S^+ y. \tag{19}
$$

The proof of this fact is given in [28, Th. 7]. Part 1) of proposition states that the submatrices $\Psi_{S,[N:d]}$ has full rank even when $\Psi_{S,[0:N]}$ is rank deficient. This fact plays a vital role in the performance analysis of the group LASSO estimate.

For Proposition 1 to hold, we require $\eta_S, \mu_S < \infty$. Using the notion of zeros and rank of the system matrix (see the following), we state verifiable conditions to check if $\eta_S, \mu_S < \infty$. For all $z \in \mathbb{C}$, define the Rosenbrock system matrix

$$
Z_S[z] \triangleq \begin{bmatrix}
I_{n} - A & -B_S \\
0 & C
\end{bmatrix}.
$$

(20)

Let $\text{nRank}Z_S \triangleq \max_{z \in \mathbb{C}} \text{Rank}Z_S[z]$ be its normal rank. A number $z_0 \in \mathbb{C}$ is called the invariant zero of $(A, B_S, C)$ if $\text{nRank}Z_S[z_0] < \text{nRank}Z_S$. If $(A, B_S, C)$ has invariant zeros, there exists $u_S^0 \neq 0$ and $x_0 \neq 0$ such that (noise-free) $y[k] = 0$, for all $k \geq 0$ [29]. Thus, we cannot distinguish between nonzero and zero inputs from $y_N$. Hence, $\eta_S, \mu_S = \infty$.

**Lemma 2:** Suppose that $(A, B_S, C)$ has no invariant zeros. Then, (i) $\mu_S < \infty$, and for $N \geq \mu_S$, system in (3)–(4) is initial state observable; and (ii) $\eta_S < \infty$ if $\text{nRank}Z_S = n + m^*$.

Proof for statement (i) can be found in [28, Proposition 5]. Instead, statement (ii) follows from [9, Th. 1, p. 227]. Thus, if $(A, B_S, C)$ satisfies conditions in Lemma 2, the assumptions in Proposition 1 hold. Hence, the submatrix $\Psi_{S,[N:d]}$ has full rank and we can recover $(\tilde{x}_0^*, u_{S,[0:N-d]}^*)$.

IV. LOCATION RECOVERY AND ESTIMATION CONSISTENCY OF THE GROUP LASSO ESTIMATOR

We theoretically investigate the performance of the group LASSO estimator in (10) using the previously stated results for the delayed input estimation. Our results generalize the existing group LASSOs guarantees for static (or nondynamical) systems [6], [30] to the dynamical systems with delay $d \geq 0$.

Recall that the estimate $\hat{\beta}$ in (10) is $(\tilde{x}_0^*, \hat{u}_1, \ldots, \hat{u}_m)$ with $\hat{u}_j = [\hat{u}_j[0], \ldots, \hat{u}_j[N]]^T$. For any $S \subset \{1, \ldots, m\}$, we define $\hat{u}_S[k] = [\hat{u}_j[k], \ldots, \hat{u}_j[k]]$, for all $k \geq 0$ and $S \in \mathbb{S}$; that is, we group the estimated inputs associated with the set $S$. Define $\hat{u}_S^j = [\hat{u}_j[0], \ldots, \hat{u}_j[N]]$ and $\beta_S = (\tilde{x}_0^*, \hat{u}_S)$. For $S = \{j : u_j^* \neq 0\}$ and $S = \{j : \hat{u}_j = 0\}$, we derive conditions under which

1) $S = S$ and 2) $\| \beta_S - \beta_S \|_2 \leq \epsilon$, for any $\epsilon > 0$, hold with high probability.

**Assumption 3 (Identifiability and MICs [27]):** Consider the following conditions.

A1) Group normalizaton: There exists a constant $C > 0$ such that $O$ and $J_4$ in (7) satisfy the normalization condition

$$
\max \{ \|O\|_2, \|J_1\|_2, \ldots, \|J_m\|_2 \} \leq C \sqrt{T} < \infty. \tag{21}
$$

A2) Model identifiability: The parameters $\eta_S$ in (16) and $\mu_S$ in (17) are finite; and $N \geq \max\{\eta_S, \mu_S\}$ with $d \geq \eta_S$. Furthermore, there exists a constant $c_{\min} > 0$ such that

$$
\left\| \left( \Psi_S^{T,[N:d]} \Theta \Psi_S^{T,[N:d]} \right)^{-1} \right\|_2 \leq \frac{1}{c_{\min}} < \infty. \tag{22}
$$

where $\Psi_S^{T,[N:d]}$ and $\Psi_S^{T,[d-1:0]}$ are given in (18), and $\Theta \triangleq [I - \Psi_S^{T,[d-1:0]} \Psi_S^{T,[N:d]}]$.

A3) Mutual incoherence: There exists some $\alpha \in \{0, 1\}$, referred to as “mutual incoherence” parameter, such that

$$
\text{MIC} \triangleq \max_{j \in \mathbb{S}} \left\| J_j^T \Psi_S \left( \Psi_S^T \Psi_S \right)^{-1} \right\|_2 \leq \alpha/m^*. \tag{23}
$$

Assumptions (A1) and the bound in (22) hold for stable and asymptotically unstable systems if $N < \infty$. However, these assumptions hold only for stable systems if $N \to \infty$. Furthermore, as discussed in Section III, we need the requirement on $N, \mu_S$, and $\eta_S$ in Assumption (A2) to ensure that the initial state and the input associated with the source set $S$ are identifiable. In light of Lemma 2, this requirement is satisfied if $(A, B_S, C)$ has no invariant zeros and that $\text{nRank}Z_S = n + m^*$. Finally, the constants $C$ and $c_{\min}$ do not depend on the horizon $N$. They capture the inherent complexity in estimating the unknown parameters and play a vital role in true support recovery.

Assumption (A3) is satisfied if $\Psi_S$ and $J_j$ are orthogonal $(J_j^T \Psi_S = 0$, for all $j \in \mathbb{S})$. Orthogonality does not hold if inputs are more than outputs. Or columns of $B_S$ in (3) are linear combinations of $b_j$, for $j \in \mathbb{S}$. Nonetheless, (A3) imposes a type of “approximate” orthogonality between $J_j$ and $\Psi_S$, mediated by the parameter $\alpha$. The $\ell_2$-norm bound in (23) could be conservative as the bound depends on $m^*$. This dependence can be avoided by working with the $\ell_1$-norm bound; that is, $\max_{j \in \mathbb{S}} \left\| J_j^T \Psi_S \left( \Psi_S^T \Psi_S \right) \right\|_1 \leq \alpha$. However, we stick with (23) as it is useful to derive an upper bound on MIC in (23) using the system transfer function. In simulations, we study the conservatism due to $\ell_2$-norm based MIC.

2At least one of the eigenvalues of $A$ lie outside the complex unit circle.
Theorem 4 (Location recovery consistency): Suppose that the linear model in (11) satisfies assumptions (A1)–(A3) with $S = \{1, \ldots, m^*\}$. For some $\delta > 0$ and $c_1 = \log(5)$, let
\[
\lambda_T = \sqrt{\frac{32C\sigma}{1-\alpha}} \left\{ \frac{\sqrt{(N+1)c_1 + \log(m-m^*)}}{T} + \frac{\delta}{2} \right\}.
\] (24)

Then, for the group LASSO estimate $\hat{\beta} = (\hat{x}_0, \hat{u})$ defined in (10), with the probability of at least $1 - 4 \exp(-T\delta^2/2)$, we have the following.

1) (Nonuniqueness): There are infinite solutions for (10).
2) (Non false inclusion): The support set of any estimate $\hat{\beta}$ lies in the true support set; that is, $\hat{S} \subseteq S$.
3) (Minimum input magnitude and no false exclusion): The delayed inputs satisfy the bound:
\[
\max_{j \in S} \|\hat{u}_{j,[0:N-d]} - u_{j,[0:N-d]}\|_2 \leq g_{\min}(\lambda_T, \Psi),
\] with
\[
g_{\min}(\lambda_T, \Psi) = \frac{\sigma}{\sqrt{c_{\min}}} \left\{ \frac{2\log((N-d+1)m^*)}{T} + \delta \right\} + \lambda_T \left\| \Pi_{S,\{0:N-d\}} \left( \Psi_{\hat{S},[0:N-d]} \right) \right\|_\infty.
\] (25)

and $\Pi_{S,\{0:N-d\}} = \{I_{\{0:N-d\}}, 0_{\{0:N-d\} \times d^*}\}$, where $t_S = (N - d + 1)m^*$, satisfies $\Pi_{S,\{0:N-d\}}\beta^*_S = u_{S,\{0:N-d\}}^*$.
4) (Minimum input magnitude and no false exclusion): If $\min_{j \in S} \|u_{j,[0:N-d]}^*\|_2 \geq g_{\min}(\lambda_T, \Psi), \Psi$), we have $\hat{S} = S$.

Corollary 5: Let $\beta^*_S,\{0:N-d\} = (\hat{x}_0, \hat{u}_{S,\{0:N-d\}})$ and similarly define $\beta^*_S,\{0:N-d\}$. Suppose that $\|\Psi_{\hat{S},\{0:T\}}\|_2 \leq 1/c_{\min}$. Under the assumptions stated in Theorem 4, with the probability of at least $1 - \exp(-\delta^2 T/2)$, we have
\[
\|\beta^*_S,\{0:N-d\} - \hat{\beta}^*_S,\{0:N-d\}\|_2 \leq \frac{2\sigma}{\sqrt{c_{\min}}} \left\{ \frac{2c_1(n + t_S)}{T} + \delta \right\} + \lambda_T \sqrt{m^*} / c_{\min}.
\] (26)

Proof: See Appendix.

We use the primal–dual witness (PDW) technique [27], [30], [31] to prove Theorem 4. The details of this technique are in Appendix. The location recovery consistency results in Theorem 4 in the literature are referred to as support recovery; here, the support means the indices of nonzero $u_j$, which is $\hat{S}$. In the following, we comment on the scaling laws of Theorem 4.

Part 1) in Theorem 4 states that the group LASSO estimate $\hat{\beta}$ is nonuniqueness unless the subsystem realized by $(A, B_S, C)$ has zero delay. This is because, for $N > d > 0$, the submatrix $\Psi_{S,\{N:d\}}$ in (18) has full rank, but not $\Psi_S$. However, part 2) in Theorem 4 states that $\hat{S} \subseteq S$, for any optimal estimate $\hat{\beta}$ in (10). Thus, the estimated inputs restricted to the complement set are zero: $\hat{u}_{j,0} = 0$, for all $j \in S^c$. Thus, the nonuniqueness of the optimal solution does not affect the location consistency of the group LASSO estimator.

Part 4) in Theorem 4—a consequence of the $\ell_\infty$ norm bound in part 2)—states that for $\hat{S} = S$ to hold (i.e., to detect true inputs correctly), the true nonzero input signal strength should not be too small, precisely, smaller than $\beta_{\min}$ in (25). The probabilistic result in Theorem 4 also helps determine the number of measurements $(N)$ or sensors $(p)$ required to achieve a certain amount of performance. Let us simplify $\lambda_T$ in (24) to comment on its scaling. By substituting $T = p(N + 1)$ and assuming that $\log(m-m^*)/(N + 1) \gg c_1$, we have
\[
\lambda_T = O \left( \sqrt{\frac{\log(m-m^*)}{p(N + 1)}} + \frac{\delta}{2} \right).
\] (27)

For $N = 1$, $\lambda_T \approx O \left( \sqrt{\frac{\log(m-m^*)}{p}} \right)$, which is the optimal $\lambda_T$ for the standard LASSO problem [27]. Thus, $c_1(N + 1)$ in (24) quantifies the number of unknowns in $u_j^*$, and $N + 1$ in $T$ accounts for the number of measurements per sensor.

The choice of $\lambda_T$ plays an important role in determining if part 3) in Theorem 4 (that is, $\hat{S} = S$) holds. In fact, the smaller $\lambda_T$, the smaller the minimum threshold $g_{\min}(\lambda_T, \Psi)$. Interestingly, for $\lambda_T = 0$, which happens, say, when $\sigma = 0$, the optimization problem in (10) reduces to the standard ordinary least squares (OLS) problem. Thus, there is no shrinkage of input estimates toward zero. Furthermore, $\lambda_T$ does not depend on $c_{\min}$ in (22) but depends on the group normalization constant $C$ in (21) and the mutual incoherence parameter $\alpha$ in (23).

To study the role of the minimum singular value of $\Psi_S$, denoted by $\rho_{\min}(\Psi_S)$, on $g_{\min}(\lambda_T, \Psi)$, we assume that $\Psi_S$ has full rank. Then, from the standard norm bounds, we have
\[
k_1 + \frac{\lambda_T T \sqrt{\rho_2}}{\rho_{\min}(\Psi_S)} \geq g_{\min}(\lambda_T, \Psi) \geq k_1 + \frac{\lambda_T T \sqrt{\rho_2}}{\rho_{\min}(\Psi_S)}
\] where $k_1$ is the first term on the right side of the equality in (25) and $\rho_2 = (N + 1)m^*$ is the dimension of $u_S$. For fixed $C$ defined in (21) and $\rho_2$, from the preceding inequality, it is clear that larger $\sigma^2_{\min}(\Psi_S)$ requires smaller $g_{\min}(\lambda_T, \Psi)$ because the effective signal strength of $\Psi_S^2 u_S$ is large. Instead, smaller $\sigma^2_{\min}(\Psi_S)$ requires higher $g_{\min}(\lambda_T, \Psi)$, requiring $\rho_2$ to be large. If not, the strength of $\Psi_S u_S$ decreases. Finally, from (24), we observe that $\lambda_T$ is an increasing function of $\alpha \in [0, 1]$; thus, the higher $\alpha$, the larger is $g_{\min}(\lambda_T, \Psi)$. Recall that $\alpha$ is large if $J_j$, for $j \in S^c$, is highly correlated with $\Psi_S$.

We now comment on the $\ell_2$-error bound between $\beta^*_S,\{0:N-d\}$ and $\beta^*_S,\{0:N-d\}$ given in Corollary 5. First, the bound depends on the number of unknown parameters $n + t_S = n + (N - d + 1)m^*$, i.e., the dimension of the initial state and delayed input. Letting $T = p(N + 1) \gg n$, we observe that the first term of the bound in (26) scales as $O(2\sigma/\sqrt{\rho_{\min}} \sqrt{m^*/p + \delta})$. Thus, more sensors result in fewer errors. However, the bound is loose for large values of $\lambda_T$. To remedy this shortcoming, we consider the following OLS estimate:
\[
\beta^*_S,\{0:N-d\}^{(OLS)} = \Pi_{S,\{0:N-d\}} \left( \Psi_{\hat{S},\{0:N-d\}} y \right)
\] (28)
where $\Pi_{S,\{0:N-d\}}$ is defined similar to $\Pi_{S,\{0:N-d\}}$ in (19). We present the second main result of this section: an oracle bound on the error $\|\beta^*_S,\{0:N-d\} - \beta^*_S,\{0:N-d\}^{(OLS)}\|_2$.

Theorem 6 ($\ell_2$-consistency: oracle bounds): Suppose that the hypotheses in Theorem 4 hold. Then, for any $\delta, \delta_1 > 0$, with
the probability of at least \(1 - 4 \exp(-T \delta^2/2) - \delta_1\)

\[
\left\| \beta_{S,[0:N-d]}^S - \tilde{\beta}_{S,[0:N-d]}^{(OLS)} \right\|_2 \leq \frac{4\sigma}{\sqrt{c_{\min}}} \left\{ \sqrt{\frac{(n + 1)}{T}} \right\} + \frac{2\sigma}{\sqrt{c_{\min}}} \left\{ \sqrt{\frac{1}{T} \log \left( \frac{1}{\delta_1} \right)} \right\}.
\] (29)

The proof is in Appendix. Similar to the bound in Corollary 5, the first term in (29) is \(O(2\sigma/\sqrt{c_{\min}}\sqrt{m^*/p})\); however, the second term in (29) does not depend on \(\lambda_T\) and it approaches zero as \(T \to \infty\). Thus, the overall error is dictated by \(m^*/p\). We call the bound in (29) as the oracle because the bound holds for \(\tilde{\beta}_{S,[0:N-d]}^{(OLS)}\), albeit with probability \(1 - \delta_1\).

A. Mutual Incoherence: Frequency Domain

Thus far, we studied the location recovery- and estimation-consistency of the group LASSO estimator in (10) assuming that assumptions in (A1)–(A3) hold of which the first two are satisfied by stable dynamical systems with \((A, B_S, C)\) having no invariant zeros.\(^3\) However, (A3) might not hold for arbitrary systems, and moreover, verifying (23) can be computationally demanding when either \(N\) (the measurement horizon) or \(n\) (the dimension of system matrix \(A\)) is large. In what follows, we bound \(\max_{z \in S^c} \|T\Psi_S(z)\Psi_S(z)\|\) in (23) using a quantity that depends on the transfer function matrices associated with \((A, B_S, C)\) and \((A, b_j, C)\), for \(j \in S^c\). The advantage is that this upper bound can be computed efficiently, as it depends only on the lower-dimensional system matrices but not on \(N\).

To simplify the exposition, we let \(x_0 = 0\); thus, \(\Psi_S = J_S\). Let \(G_{S^c}[z] = C(zI_n - A)^{-1}B_S\); \(G_{S^c}[z] = C(zI_n - A)^{-1}B_{S^c}\); and \(G_j[z] = C(zI_n - A)^{-1}b_j\), where \(j \in S^c\) and \(B_{S^c}\) is the matrix composed of columns \(b_j\), with \(j \in S^c\).

**Theorem 7:** Assumption (A3) holds if nRank\(Z_S = n + m^*\) and

\[
\max_{j \in S^c} \max_{\{z \in C^* : \|z\| = 1\}} \|G_{S^c}^*[z]G_j[z]\|_2 \leq \frac{\alpha}{m^*} < 1.
\] (30)

**Proof:** See Appendix.

We refer to the expression in (30) as the frequency-domain MIC. To verify Assumption (A3), we need to check if the worst case gain of the matrix \(G_{S^c}^*[z]G_j[z]\) is bounded above by \(\alpha/m^*\); see Fig. 1. If computing (30) is prohibitive for each \(j \in S^c\), we can use the weaker condition: \(\max_{\{z \in C^* : \|z\| = 1\}} \|G_{S^c}^*[z]G_{S^c^c}[z]\| \leq \alpha/m^* < 1\). To appreciate the condition in (30), take the \(Z\)-transform of the system in (3)–(4)

\[
y[z] = G_S[z]u_S[z] + \sum_{j \in S^c} G_j[z]u_j[z] \quad \forall z \notin \text{spec}(A).
\]

By premultiplying the above identity with \(G_{S^c}^*[z]\), we have \(G_{S^c}^*[z]y[z] = z^{-d}u_S[z] + \sum_{j \in S^c} G_{S^c}^*[z]G_j[z]u_j[z]\), where we used the fact that \(G_S[z]\) is delay invertible, and hence, \(G_{S^c}^*[z]G_S[z] = z^{-d}I\). Thus, to recover \(u_S[z]\) accurately, the gain \(||G_{S^c}^*[z]G_j[z]\|_2\) or ||\(G_{S^c}^*[z]G_{S^c^c}[z]\||_2\) needs to be small.

\(^3\)Systems having invariant zeros lie in a zero-measure set [29].

![Fig. 1. Illustration of Theorem 7 for systems generated using MATLAB. The number of sources \(m = 10\). In both panels, the y-axis, \(FD_{MIC}\) - \(TD_{MIC}\), is the error between frequency- and time-domain MIC. (Left panel) We fix \(n = 20\) and plot \(FD_{MIC}\) - \(TD_{MIC}\) for several values of \(m^*\). (Right panel) For a large matrix \(A\), we fix \(m^*\) and \(p\), and plot \(FD_{MIC}\) - \(TD_{MIC}\) for several values of \(n\). In both panels, the error is positive and is monotone in \(N\) implying that \(FD_{MIC} \geq TD_{MIC}\), as predicted by Theorem 7.](image)

We highlight the following three cases where (30) holds.

i) \(R(G_{S^c}[z]) \subseteq R(G_{S^c}[z]) = R(G_{S^c}[z])\); that is, the columns of \(G_{S^c}[z]\) lie in the left nullspace of \(G_S[z]\).

ii) \(G[z] = G_S[z]G_{S^c}[z]\) is all-pass.\(^4\)

iii) Each column of \(G_{S^c^c}[z]\) is a scaled column of \(G_S[z]\) for some scaling factor \(\alpha \in [0, 1]\).

The first two cases are rather strong and do not allow columns of \(G_{S^c^c}[z]\) to be in the range space of \(G_S[z]\). Instead, case (iii) models another extreme where the range spaces of \(G_S[z]\) and \(G_{S^c^c}[z]\) are aligned with each. The latter case in the compressed sensing literature is referred to as overcomplete dictionaries [32].

Our results extend to the system in (3)–(4) driven by process noise (e.g., in the power system, the noise is load fluctuations); see our extended paper [25, Sec. IV.A] for details.

V. Simulations

We illustrate the performance of the group LASSO estimator on a large-scale power network and a random system. The following proposition states that the unknown input and initial state can be estimated in two stages. Consequently, we use off-the-shelf ADMM [10] to estimate the input first and then use this estimate to compute the initial state.

**Proposition 8:** Suppose that system in (3)–(4) is observable.

The optimization problem (10) is equivalent to

\[
\tilde{u} = \arg \min_{u \in \mathbb{R}^{m^*}} \frac{1}{2T} \|R(y - Ju)\|_2^2 + \lambda_T \sum_{j=1}^m \|u_j\|_2
\] (31)

\[
\tilde{x}_0 = O^\dagger (y - \tilde{J}u)
\] (32)

where \(O^\dagger = (O^T O)^{-1}O^T\) and \(R = I - O O^\dagger\).

The proof follows from the Karush–Kuhn–Tucker (KKT) conditions in (33)–(34). The inputs to the ADMM [10] are \((A, B, C)\), the measurement \(y\), and \(\lambda_T \geq 0\). The two-stage estimation method is one way to implement the group LASSO numerically. One may use other numerical algorithms to estimate \((\tilde{x}_0^*, u^*)\) in one shot.

\(^4\)A real rational transfer function matrix \(G[z]\) is all-pass if \(G[z]G[1/z] = I\).
We evaluate the group LASSO estimator’s localization performance using the false-positive rate (FPR): \( \frac{|S^c \cap \hat{S}|}{|S^c|} \), the false-negative rate (FNR): \( \frac{|S \cap S^c|}{|S^c|} \), and the exact recovery rate (ERR): \( \frac{|S \cap \hat{S}|}{|S^c \cap \hat{S}|} \). Thus, the FPR and FNR measure the proportion of falsely identified and left out inputs. Instead, we quantify the estimation performance using the metrics: \( \|x_0^* - \hat{x}_0\|_2/\|x_0\|_2 \) and \( \|u^* - \hat{u}\|_2/\|u\|_2 \). For the test cases ahead, the results are averaged over 50 runs.

(Power system): We apply our estimator in (31) to localize the sources of forced oscillatory inputs in the IEEE 68-bus, 16-machine system (see Fig. 2). Each machine (or generator) consists of 10 states, including rotor angle, speed, and the states of the automatic voltage regulator (AVR) and PSS. We model FOs as inputs injected by the AVRs and use bus voltage magnitudes as measurements. For the sampling time \( \delta t = 0.1 \), we obtained the system matrices \( A \in \mathbb{R}^{160 \times 160}, B \in \mathbb{R}^{160 \times 16}, \) and \( C \in \mathbb{R}^{p \times 160} \), where \( p \leq 68 \), using the Power System Toolbox [33].

Among \( m = 16 \) possible inputs, we assume \( m^* = 3 \) with the following inputs: \( u^*_1[k] = 0.5 \sin(2\pi f \delta t) + w[k], \) \( u^*_2[k] = 0.6 \sin(2\pi f \delta t) + w[k], \) and \( u^*_3[k] = 0.7 \sin(2\pi f \delta t) + w[k], \) where \( f = 1.5 U(0, 1) \) and \( w[k] \sim \mathcal{N}(0, 0.05^2) \). We set \( p = 4 \) and choose sensor locations arbitrarily with the only exception that these are noncollocated with inputs (shown in Fig. 2). Let \( x_0 = 0 \) (the nonzero case is considered in the subsequent case). Finally, we let \( N = 100 \) and the noise variance \( \sigma^2 = 0.01 \).

In Fig. 3, we plot the FPR, FNR, and ERR with respect to \( \lambda_T \). As expected, the FNR increases with \( \lambda_T \), whereas the FPR decreases with \( \lambda_T \), although not monotonically. From the bottom left panel, we can infer that values of \( \lambda_T \in (0.3, 0.4) \) yield maximum ERR. In the bottom right panel, note that for \( \lambda_T = 0.288 \), the group LASSO estimator accurately localized inputs among 40 out of 50 runs. In Fig. 4, for a measurement realization where the group LASSO estimator identified true locations, we plot the inputs estimated by the group LASSO and the reduced model-based OLS estimators.

(Large-scale random system): Following the work [22], we generate matrices as follows: \( A_{ij} \sim \mathcal{N}(0, 1/n) \); \( C_{ij} \sim \mathcal{N}(0, 1) \); and \( B^T = \begin{bmatrix} 1 & 0 \end{bmatrix} \). We let \( x_0 \sim \mathcal{N}(0, I_n) \) and the measurement noise variance parameter \( \sigma = 0.01 \). We set \( n = 50, m = 30, \) and \( m^* = 5 \). The active set \( S = \{1, 2, 3, 4, 5\} \) and \( u_j[k] \) is sampled uniformly on \([-2, 2]\) for all \( j \in S \) and \( k \in [N] \). The sensors measure the first \( p (\leq n) \) states. In Fig. 5, for \( p = 15 \), we plot the average estimation error metrics as a function of the measurement horizon \( N \). In both the panels, estimation errors remain uniform across \( N \) because the number of (to be estimated) inputs also increase with \( N \). Given the relation in (32), the estimation error of \( x_0^* \) is slightly higher than that of the unknown input. Finally, for greater estimation accuracy, one can always use the reduced model-based OLS estimator.

In Fig. 6, we show the average mutual incoherence in (23) as a function of \( p \) for \( x_0 = 0 \) and \( x_0 \neq 0 \). We computed both \( \ell_1 \) - and \( \ell_2 \)-norm-based MICs. As pointed out in Section IV and confirmed by our plots in the left panel of Fig. 6, \( \ell_2 \)-norm-based MIC assumption is stronger than the \( \ell_1 \)-norm. Furthermore,
when \( x_0 = 0 \), the MIC is satisfied (that is, less than one) for as few as \( p = 6 \) sensors. Instead, when \( x_0 \neq 0 \), we need at least \( p = 18 \) sensors to ensure that MIC is less than one.

### VI. Conclusion

We have studied a group LASSO estimator for localizing the sparse set of sources of forced inputs and estimating these inputs along with the initial state in \( d \)-delay left invertible linear dynamical systems. Under certain natural conditions, we showed that our estimator is well defined, and the underlying estimate is nonunique for \( d \geq 0 \). However, with high probability, we showed that the support of any optimal estimate recovers the true sparse set if 1) the subsystem associated with the sparse set of inputs has no invariant zeros, and 2) the observability and impulse response matrices of the overall system satisfy a mutual-incoherence type condition. In doing so, we have extended the existing theory of the group LASSO estimator for static regression models to the models generated by dynamical systems. Another key contribution of our work is that we derived a connection between the time- and frequency-domain MICs. The former imposes certain restrictions on the column space of the impulse response matrix; instead, the latter does so on the column space of the transfer function matrix. Furthermore, the frequency-domain condition is numerically easier to verify than its time-domain counterpart. Importantly, it provides insight into the structural aspects of transfer matrices associated with the zero and nonzero inputs. Finally, we have validated the performance of the group LASSO estimator on the IEEE 68-bus, 16-machine power system, and a large-scale synthetic model.

### APPENDIX A

#### A. KKT Conditions and PDW Construction

**Proposition 9 (KKT conditions):** A necessary and sufficient condition for \( (\hat{x}_0, \hat{u}) \), with \( \hat{u}^T = [\hat{u}^T_1, \ldots, \hat{u}^T_m] \), to be a solution of (10) is

\[
- \frac{1}{T} \hat{J}_1^T \begin{bmatrix} y - O\hat{x}_0 - \sum_{j=1}^m J_j \hat{u}_j \end{bmatrix} = 0
\]

(33)

\[
- \frac{1}{T} \hat{J}_1^T \begin{bmatrix} y - O\hat{x}_0 - \sum_{j=1}^m J_j \hat{u}_j \end{bmatrix} + \lambda_T \hat{z}_j = 0
\]

(34)

for all \( j \in \{1, \ldots, m\} \). Here, \( \hat{z}_j \) is the subgradient of \( ||\hat{u}_j||_2 \); that is, \( \hat{z}_j = \hat{u}_j/||\hat{u}_j||_2 \) if \( \hat{u}_j \neq 0 \), else \( \hat{z}_j \in \{q : ||q||_2 \leq 1\} \).

The proof follows by taking the derivative of the objective function in (10) with respect to \( (\hat{x}_0, \hat{u}) \) and using the subgradient characterization of the \( \| \cdot \|_2 \)-norm (see [35, Appendix B]). Without loss of generality let \( S = \{1, \ldots, m^*\} \) and \( S^c = \{m^* + 1, \ldots, m\} \). Let \( \hat{u}_j^T[k] = [\hat{u}_j[1], \ldots, \hat{u}_j[m]][k] \), for all \( k \in \{0, \ldots, N\} \), where \( \hat{u}_j[k] \) is the \( k \)-th entry of \( \hat{u}_j \). Define \( \hat{u}_S = [\hat{u}_S[0], \ldots, \hat{u}_S[N]] \). Thus

\[
[\hat{u}^T_S, \hat{u}^{m^*+1}_S]^T = P\hat{u}_S
\]

(35)

for some permutation matrix \( P \). Furthermore, we can verify that \( [J_1 \ldots J_m]P = J_S \) [as in (12)]. Let \( \hat{\beta}_S \triangleq [\hat{u}^T_0 \hat{u}^T_S]^T \). Then

\[
O\hat{x}_0 + \sum_{j \in S} J_j \hat{u}_j = \Psi_S \hat{\beta}_S.
\]

(36)

Let \( \tilde{J}_{S^c} = [J_{m^*+1} \ldots J_m] \). Then, (33)–(34) can be written as

\[
- \frac{1}{T} \hat{J}_{S^c}^T \begin{bmatrix} \Psi_{S^c}^T \hat{J}_{S^c} \end{bmatrix} \begin{bmatrix} y - O\hat{x}_0 - \sum_{j=1}^m J_j \hat{u}_j \end{bmatrix} + \lambda_T \begin{bmatrix} 0 \\ P^T \hat{z}_{S^c} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}
\]

(37)

where \( \hat{z}_{S^c}^T = [\hat{z}^T_{m^*+1} \ldots \hat{z}^T_m] \), and \( \hat{z}_L = [\hat{z}^T_{m^*+1} \ldots \hat{z}^T_m] \).

**PDW construction:** We prove Theorems 6 and 4 using the PDW method\(^5\) in [31]. Upon successful completion, this method returns a pair \( (\hat{\beta}, \hat{z}) \) that is primal–dual optimal, and it acts as a witness (or certificate) for the fact that the group LASSO estimate has the true support.

1) Set \( \hat{u}_i = 0 \), for all \( j \in S^c \).

2) Let \( (\hat{x}_0, \hat{u}_1, \ldots, \hat{u}_{m^*}) \) be the solution of the following subproblem:

\[
\min_{u_1, \ldots, u_{m^*}} \frac{1}{2T} \left\| y - O\hat{x}_0 - \sum_{j=1}^{m^*} J_j u_j \right\|_2^2 + \lambda_T \sum_{j=1}^{m^*} \| u_j \|_2.
\]

(38)

\(^5\)The PDW construction is not an algorithm for solving (10): This is because to solve the problem in step 2) of PDW, we need to know \( S \). However, PDW construction helps to prove theoretical results for the LASSO-type problems.
Choose the subgradient $\hat{z}_S = \left[ \hat{z}_1^T, \ldots, \hat{z}_{m'}^T \right]^T$ such that

$$-\frac{1}{T} \Psi_S^T \left[ y - O\hat{x}_0 - \sum_{j=1}^{m'} J_{j} u_j \right] + \lambda_T \left[ \begin{array}{c} 0 \\ P^T \hat{z}_S \end{array} \right] = 0.$$  

(39)

3) Solve $\hat{z}_{S'} = \left[ \hat{z}_{m'+1}^T, \ldots, \hat{z}_{m''}^T \right]^T$ using (37), and check if $\|\hat{z}_j\|_2 \leq 1$, for all $j \in S' = \{m'+1, \ldots, m''\}$. By construction, $(\hat{x}_0, \hat{u}_1, \ldots, \hat{u}_{m''}, \hat{z}_S, \hat{z}_{S'})$ that we determined in steps 1) and 2) satisfy conditions in (37). The PDW construction is said to be successful if $\hat{z}_{S'}$ satisfies the strict dual feasibility condition: $\|\hat{z}_j\|_2 \leq 1$, for all $j \in S'$. 

For the estimate in (38), define

$$\hat{\beta}_{PDW} = \left( \hat{x}_0, \hat{u}_1, \ldots, \hat{u}_{m''}, 0_{(N+1)}, \ldots, 0_{(N+1)} \right)_{m''-m'^{\prime}}.$$  

(40)

**Lemma 10:** Suppose that the PDW construction succeeds. Then, $\hat{\beta} = \hat{\beta}_{PDW}$ is an optimal solution of (10).

**Proof:** We adapt the proof technique provided in [31, Lemma 7.23]. Let $d \geq 0$. Because the PDW construction succeeds, $\hat{\beta}_{PDW}$ in (40) is an optimal solution of (10) satisfying the conditions in Proposition 9.

Let $u^T = [\hat{u}_1^T, \ldots, \hat{u}_m^T]$ and $F(u) \triangleq \phi \frac{1}{T} \|y - O\hat{x}_0 + \sum_{j=1}^{m} J_{j} u_j \|_2^2$. Let $\nabla F(u)$ be the gradient of $F(u)$ at $u$. Then, for any optimal solution $\hat{u}$ of (10), we have $F(\hat{u}) = \lambda_T \sum_{j=1}^{m} \|\hat{u}_j\|_2$, where $\hat{u} = [\hat{u}_1^T, \ldots, \hat{u}_m^T]$ is the subgradient and we used the fact that $\sum_{j=1}^{m} \|\hat{u}_j\|_2 = 0$ for $\hat{u}$ in $S^c$. Since $\|\hat{u}_j\|_2$ is the latter holds, which holds if $\hat{u}_j = 0$ or $\|\hat{u}_j\|_2 = \|\hat{u}_j\|_2$ for $\hat{u}$. Thus, $F(\hat{u}) = \lambda_T \sum_{j=1}^{m} \|\hat{u}_j\|_2$. Instead, from (33) and (34), we have $\lambda_T \hat{z} = -\nabla F(\hat{u})$. Thus

$$\nabla F(\hat{u}) = -\frac{1}{T} \Psi_S^T \left[ y - O\hat{x}_0 + \sum_{j=1}^{m} J_{j} u_j \right] + \lambda_T \left[ \begin{array}{c} 0 \\ P^T \hat{z}_S \end{array} \right].$$

By convexity of $F$, $F(\hat{u}) + \nabla F(\hat{u})^T (\hat{u} - \hat{u}) - F(\hat{u}) \leq 0$. Thus, $\sum_{j=1}^{m} \|\hat{u}_j\|_2 \leq \hat{z}_S^T \hat{u} = \sum_{j=1}^{m} \sum_{j=1}^{m} \|\hat{u}_j\|_2$, where $\hat{x}_0 = 0$. Since $\sum_{j=1}^{m} \|\hat{u}_j\|_2 = \sum_{j=1}^{m} \|\hat{u}_j\|_2$, we have $\|\hat{u}_j\|_2 < 1$, this equality holds only if $\hat{u}_j = 0$, for all $j \in S^c$. To see this note that $\sum_{j=1}^{m} \|\hat{u}_j\|_2 = \sum_{j=1}^{m} \sum_{j=1}^{m} \|\hat{u}_j\|_2$, $\sum_{j=1}^{m} \|\hat{u}_j\|_2 < 1$, $\hat{u}_j$ is the angle between $\hat{z}_j$ and $\hat{u}_j$, and $\|\hat{z}_j\|_2 \cos(\theta_j) \in (-1, 1)$. Thus, all optimal $\hat{\beta}$'s satisfy $\hat{\beta}_j = 0$ for $\hat{u} \in S^c$. 

**B. Proofs of Theorems and Corollaries in Section IV**

**Proof of Theorem 4:** Suppose the PDW construction succeeds. The proof of part 1) is given in Lemma 10. Furthermore, in view of Lemma 10, $\hat{\beta} = \hat{\beta}_{PDW}$ is an optimal solution of (10).

Thus, all the optimal input vectors are supported on the set $S$, i.e., $S \subset S$, where $S = \{ j : \hat{u}_j \neq 0 \}$; thus, part 2) holds.

We show that the PDW construction succeeds with the probability of at least $1 - 2 \exp(-\delta^2 T/2)$ by showing that $\|\hat{z}_j\|_2 \leq 1$, for all $j \in S'$. Here, $\hat{z}_j$ is determined in step 3) of PDW construction. Let $\hat{\beta}_S$ be as in (36). By substituting $y$ [given in (11)] and $\hat{u}_{S'} = 0$ in (37), we obtain

$$1 \left( \begin{array}{c} \Psi_S^T \Psi_S \Psi_S^T \hat{z}_S \\ J_S^T \Psi_S \hat{z}_S \\ J_S^T \hat{z}_S \end{array} \right) = \beta_S - \beta_{S'}$$

(41)

Using the second block equation of (41), solve for $\hat{z}_{S'}$ as

$$\hat{z}_{S'} = J_S^T \Psi_S \left[ \begin{array}{c} \beta_S - \beta_{S'} \\ \Psi_S^T \hat{z}_S \end{array} \right] + J_S^T \left( \frac{v}{\lambda_T T} \right),$$

(42)

where we used the fact $\Psi_S = \Psi_S^T \Psi_S$. On the other hand, from the top block equation in (41), we have

$$1 \left( \begin{array}{c} \Psi_S^T \Psi_S \Psi_S^T \hat{z}_S \\ J_S^T \Psi_S \hat{z}_S \\ J_S^T \hat{z}_S \end{array} \right) = \beta_S - \beta_{S'}$$

(43)

Prelimit both sides of the equality in (43) with $(\Psi_S^T \Psi_S)^\dagger$ and use the identity $\Psi_S = (\Psi_S^T \Psi_S)^\dagger \Psi_S$ (see [30]) to get

$$\Psi_S^T \Psi_S \left( \beta_S - \beta_{S'} \right) = -\Psi_S^T v + T \lambda_T \left( \Psi_S^T \Psi_S \right)^\dagger \left[ \begin{array}{c} 0 \\ P^T \hat{z}_S \end{array} \right].$$

(44)

Let $\Gamma_S = [I - (\Psi_S^T \Psi_S)]$. By substituting (44) in the first term of the second equality in (42), we can simplify $\hat{z}_{S'}$ as

$$\hat{z}_{S'} = J_S^T \left( \Psi_S^T \right)^\dagger \left[ \begin{array}{c} 0 \\ P^T \hat{z}_S \end{array} \right] + J_S^T \Gamma_S \left( \frac{v}{\lambda_T T} \right) \forall j \in S'.$$

(45)

By the submultiplicative property of norms, for any $j \in S'$

$$\left\| J_S^T \left( \Psi_S^T \right)^\dagger \left[ \begin{array}{c} 0 \\ P^T \hat{z}_S \end{array} \right] \right\|_2 \leq \max_{j \in S'} \left\| J_S \left( \Psi_S^T \right)^\dagger \right\|_2 \left\| \left[ \begin{array}{c} 0 \\ P^T \hat{z}_S \end{array} \right] \right\|_2 \leq \alpha \frac{1}{m^S} \| P^T \hat{z}_S \|_2 \leq \frac{1}{m^S} \sum_{j \in S'} \| \hat{z}_j \|_2 \leq \alpha \frac{m^S}{m^S} \| P^T \hat{z}_S \|_2 \leq \alpha \leq 1.$$  

(46)

(47)
Part 3): From Assumption (A2) and Proposition 1, we have
\[
\Psi^*_S \Psi_S = \text{Bildig} \left( I_n, I_{ts} \right) \Psi^*_S \Psi_S \left[ d-1:0 \right] \Psi_S \left[ d-1:0 \right]
\] (48)
where \( I_S = (N - d + 1) \). Thus, we have
\[
u_{S,[0:N-d]} - \tilde{\nu}_{S,[0:N-d]} = \Pi_S \tilde{\nu}_{S,[0:N-d]} \Psi_S \left( \beta^*_S - \hat{\beta}_S \right)
\] (49)
where \( \Pi_S \tilde{\nu}_{S,[0:N-d]} = [0_{t \times n} \ I_S \ 0_{t \times dm}] \) and
\[
\beta^*_S - \hat{\beta}_S = \left[ \begin{array}{c}
x_0^* - \tilde{x}_0 \\
u_{S,[0:N-d]} - \tilde{\nu}_{S,[0:N-d]} \\
u_{S,[N-d+1:0]} - \tilde{\nu}_{S,[N-d+1:0]}
\end{array} \right].
\] (50)
From (49) and (44), it now follows that
\[
\| \nu_{S,[0:N-d]} - \tilde{\nu}_{S,[0:N-d]} \| \leq \| \Pi_S \nu_{S,[0:N-d]} \| \leq \| \Pi_S \nu \| \leq \| \Pi_S \nu \| 
\] + \( \lambda_T \left\| \Pi_S \left[ d-1:0 \right] \left( \Psi^*_S \Psi_S / T \right) \right\| \| \Pi_S \nu \|
\] (51)
where we used the fact \( \| \tilde{S} \| \leq 1 \). The second term is deterministic. Instead, the first term is used, and, from Lemma 12, it is upper bounded by \( 2 \log (T) / \lambda_T \) with the probability of at least \( 1 - 2 \exp(-\delta^2 / 2) \). Finally, the left-hand side of (51) equals \( \max_j \| \nu_{S,j,[0:N-d]} - \tilde{\nu}_{S,j,[0:N-d]} \| \). Putting the pieces together, we have the inequality in (25).

Part 4): By the triangle inequality, for all \( j \), we have
\[
\| \nu_{S,j,[0:N-d]} \| \leq \| \nu_{S,j,[0:N-d]} - \tilde{\nu}_{S,j,[0:N-d]} + \tilde{\nu}_{S,j,[0:N-d]} \| \leq \| \nu_{S,j,[0:N-d]} - \tilde{\nu}_{S,j,[0:N-d]} \| + \| \tilde{\nu}_{S,j,[0:N-d]} \|
\] (52)
where (i) follows from part 3). Thus, \( \| \nu_{S,j,[0:N-d]} \| \leq \| \tilde{\nu}_{S,j,[0:N-d]} \| = g_{\text{min}}(\lambda_T, \Psi) \), this observation together with \( \tilde{S} \subseteq S \) implies that \( \tilde{S} = S \).

Finally, the probability stated in the theorem is obtained by taking the union bound over the event where the dual feasibility holds and the event where \( \epsilon \) bounds hold.

Proof of Corollary 5: First, let us recall that \( \Pi_S \tilde{\nu}_{S,[0:N-d]} = [0_{n+t} \ I_{n+t} \ 0_{n+t} \times dm^*] \). By proceeding similar to the steps outlined in the proof of Proposition 4, we get
\[
\left[ \begin{array}{c}
x_0^* - \tilde{x}_0 \\
u_{S,[0:N-d]} - \tilde{\nu}_{S,[0:N-d]}
\end{array} \right] = \Pi_S \tilde{\nu}_{S,[0:N-d]} \Psi_S \left( \beta^*_S - \hat{\beta}_S \right).
\]
Substituting this identity into (44) and followed by an application of triangle inequality yields us
\[
\left\| \beta^*_S - \hat{\beta}_S \right\| \leq \left\| \Pi_S \left[ d-1:0 \right] \Psi_S \nu \right\| \leq \left\| \Pi_S \nu \right\| \leq \left\| \Pi_S \nu \right\| 
\] + \( \lambda_T \left\| \Pi_S \left[ d-1:0 \right] \left( \Psi^*_S \Psi_S / T \right) \right\| \| \Pi_S \nu \|
\] (52)
Using the facts that \( \| \tilde{S} \| \leq \sqrt{\| \tilde{z}_1 \|^2 + \ldots + \| \tilde{z}_m \|^2} \leq m \), and \( \Pi \) and \( \Pi_S \tilde{\nu}_{S,[0:N-d]} \) are permutation and selection matrices, respectively, the second term in (52) can be bounded above as
\[
\lambda_T \left\| \left( \Psi^*_S \Psi_S / T \right) \right\| \| \Pi_S \nu \|
\] with the probability of at least \( 1 - \delta_1 \) for \( \delta_1 (0, 1) \). The statement of the theorem follows by taking a union bound over the events where (57) and (53) hold.

Proof of Theorem 7: Consider the auxiliary system \( x[k + 1] = Ax[k] + b_j u_j^*[k] \), where \( j \in S^c \) and \( u_j^*[k] = 0, k \geq N \). Let \( x[0] = 0 \). Thus, \( y[0] = J_j u_j^* \), where \( J_j \) is given by (7) and \( y \) and \( u_j^* \) as in (6). Let \( \Psi_S \) be as in (11), and consider
\[
\begin{bmatrix} y[0] \\ \vdots \\ y[N] \end{bmatrix} = \Psi_S \left[ J_j u_j^* \right] = \Psi_S \left[ J_j u_j^* \right].
\] (58)
By assumption we have \( \text{null}(\tilde{S}) = n + m^* \). Thus, for all \( z \in \text{null}(\tilde{S}) \) we have
where $c_N = 5^{N+1}$. The right-side term can be simplified as
\[
\exp \left( (N + 1) \log(5) + \log(m - m^*) - \frac{\bar{\alpha}^2 \lambda_T^2 T}{8\sigma^2 C^2} \right) .
\]

Substituting $\lambda_T$ \eqref{eq_lambda} and $\bar{\alpha} = 0.5(1 - \alpha)$ into \eqref{eq_bound}, and simplifying it gives us the required bound.

**Lemma 12:** With the notation and assumptions stated in Theorem 4, for $\delta \in [0, 1]$, we have $P \left[ \| \Pi S_{[0:N-d]} \psi_S \psi \| \geq \sigma / \sqrt{\min(\sqrt{2\log(t)/T} + \delta)} \right] \leq 2 \exp(-T\delta^2/2).

Proof: Let $e_t$ be the $l^\text{th}$ standard basis vector in $\mathbb{R}^{dz}$, and $tS = (N - d + 1)m^*$. Define $a_l = e_t^\text{T} \Pi S_{[0:N-d]} \psi_S \psi \in \psi_t$ to be the $l^\text{th}$ entry of $\Pi S_{[0:N-d]} \psi_S \psi \in \psi_t$. Then, because $\Pi S_{[0:N-d]} = [0_{tS \times N} I_{tS} 0_{tS \times m^*}]$, it follows that $\| \Pi S_{[0:N-d]} \psi_S \psi \| = \max_{l = 1 \ldots tS} |a_l|$. Thus, for any $\kappa \geq 0$, we have the bound
\[
P \left[ \max_{l \in 1 \ldots tS} |a_l| \geq \kappa \right] \leq \sum_{l=1}^{tS} P \left[ |a_l| \geq \kappa \right].
\]

We bound terms on the right-hand side by invoking standard concentration results. For compactness, let $\Pi S = \Pi S_{[0:N-d]}$. Because $\psi \sim \mathcal{N}(0, \sigma^2 I)$, we have $a_l \sim \mathcal{N}(0, \sigma^2)$. Then, $a_l \sim \mathcal{N}(0, \sigma^2)$. Therefore, $\sigma^2 = \sigma^2 \Pi S \psi_S \psi_T \psi_S^\text{T} \psi_T^\text{T} e_t$. Then, $\max_{l \in 1 \ldots tS} \sup_{|z| = 1} \| \psi_S \psi \| \| \psi_S \psi_T \psi_T^\text{T} \psi_T^\text{T} e_t \| = \sigma^2 \Pi S \psi_S \psi_T \psi_T^\text{T} \psi_T^\text{T} e_t$. Thus, $\| \psi_S \psi_T \psi_T^\text{T} \psi_T^\text{T} e_t \| = \sigma^2 / \sqrt{\min(\sqrt{2\log(t)/T} + \delta)}$. The second inequality follows from the interlacing property of singular values. The final inequality is shown in the proof of Theorem 6.

Since $a_l$ is Gaussian, from \cite{31}, page 22) and \eqref{eq_bound}, we have $P \left[ |z| \geq \kappa \right] \leq \exp(-\kappa^2 / (2\sigma^2)) \leq \exp(-\kappa^2 / (2\sigma^2))$. Substituting this inequality into \eqref{eq_bound}, we find that
\[
P \left[ \max_{l \in 1 \ldots tS} |z_l| \geq \kappa \right] \leq \exp \left( -\frac{\sigma^2 T \text{cmin}}{2\sigma^2} \right).
\]

The result follows by setting $\kappa = \sigma / \sqrt{\min(\sqrt{2\log(t)/T} + \delta}$ and simplifying terms in the exponential term.

**Lemma 13:** Let $p \sim \mathcal{N}(0, \Sigma)$, where $\Sigma \in \mathbb{R}^{l \times l}$ is a positive-definite matrix. Then, $P \left[ \| \psi \| \geq t \right] \leq 5^l \exp(-t^2 / (8\| \Sigma \|_2^2))$. Furthermore, $\| \psi \|_2 \leq 4 \sqrt{\| \Sigma \|_2^2 + 2\sqrt{\| \Sigma \|_2} \log(1/\delta)}$ with the probability of at least $1 - \delta$ for $\delta \in (0, 1)$.

Proof: See \cite{38}, \S 8.2, Th. 8.3.

**References**

[1] M. Ghorbaniparvar, “Survey on forced oscillations in power system,” J. Modern Power Syst. Clean Energy, vol. 5, no. 5, pp. 671–682, 2017.

[2] B. Wang and K. Sun, “Location methods of oscillation sources in power systems: A survey,” J. Modern Power Syst. Clean Energy, vol. 5, no. 2, pp. 151–159, 2017.

[3] T. Huang, N. M. Ferris, P. R. Kumar, and L. Xie, “A synchronphasor data-driven method for forced oscillation localization under resonance conditions,” IEEE Trans. Power Syst., vol. 35, no. 5, pp. 3927–3939, Sep. 2020.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
[4] S. C. Chevalier, V. Petr, and K. Turitsyn, “Using effective generator impedance for forced oscillation source location,” IEEE Trans. Power Syst., vol. 33, no. 6, pp. 6264–6277, Nov. 2018.

[5] S. Maslennikov et al., “A test cases library for methods locating the sources of sustained oscillations,” in Proc. IEEE Power Energy Soc. Gen. Meeting, 2016, pp. 1–5.

[6] K. Loumici, M. Pontil, S. V. D. Geer, and A. B. Tsybakov, “Oracle inequalities and optimal inference under group sparsity,” Ann. Statist., vol. 39, no. 4, pp. 2164–2204, 2011, doi: 10.1214/11-AOS986.

[7] N. Simon and R. Tibshirani, “Standardization and the group LASSO penalty,” Statistica Sinica, vol. 22, no. 3, pp. 983–1002, 2012.

[8] L. Meier, S. Van De Geer, and P. Bühlmann, “The group LASSO for logistic regression,” J. Roy. Statist. Soc., Ser. B (Stat. Methodol.), vol. 70, no. 1, pp. 53–71, 2008.

[9] S. Kirtikar, H. Palanathanlam-Madapusi, E. Zattoni, and D. S. Bernstein, “I-delay input and initial-state reconstruction for discrete-time linear systems,” Circuits Syst. Signal Process., vol. 30, pp. 233–262, 2011.

[10] S. Boyd et al., “Distributed optimization and statistical learning via the alternating direction method of multipliers,” Found. Trends Mach. Learn., vol. 3, no. 1, pp. 1–122, 2011.

[11] S. Chevalier, P. Vorobev, and K. Turitsyn, “A Bayesian approach to forced oscillation source location given uncertain generator parameters,” IEEE Trans. Power Syst., vol. 34, no. 2, pp. 1641–1649, Mar. 2019.

[12] U. Agrawal, J. W. Pierre, J. Follmann, D. Duan, D. Trudnowski, and M. Donnelly, “Locating the source of forced oscillations using PMU measurements and system model information,” in Proc. IEEE Power Energy Soc. Gen. Meeting, 2017, pp. 1–5.

[13] N. Zhou, M. Ghobaniparvaz, and S. Akhlaghi, “Locating sources of forced oscillations using transfer functions,” in Proc. IEEE Power Energy Conf. Illinois, 2017, pp. 1–8.

[14] Y. Meng, Z. Yu, N. Lu, and D. Shi, “Time series classification for locating forced oscillation sources,” IEEE Trans. Smart Grid, vol. 12, no. 2, pp. 1712–1721, Mar. 2021.

[15] F. Pasqualetti, F. Dörfler, and F. Bullo, “Attack detection and identification in cyber-physical systems,” IEEE Trans. Autom. Control, vol. 58, no. 11, pp. 2715–2729, Nov. 2013.

[16] F. Dörfler, F. Pasqualetti, and F. Bullo, “Continuous-time distributed observers with discrete communication,” IEEE J. Sel. Topics Signal Process., vol. 7, no. 2, pp. 296–304, Apr. 2013.

[17] M. Luan, D. Gan, Z. Wang, and H. Xin, “Application of unknown input observers to locate forced oscillation source,” Int. Trans. Elect. Energy Syst., vol. 29, no. 9, 2019, Art. no. e12600.

[18] H. Fawzi, P. Tabuada, and S. Diggavi, “Secure estimation and control for cyber-physical systems under adversarial attacks,” IEEE Trans. Autom. Control, vol. 59, no. 6, pp. 1454–1467, Jun. 2014.

[19] M. B. Wakin, B. M. Sanandaji, and T. L. Vincent, “On the observability of linear systems from random, compressive measurements,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 2010, pp. 4447–4450.

[20] G. Joseph and C. R. Murthy, “Measurement bounds for observability of linear dynamical systems under sparsity constraints,” IEEE Trans. Signal Process., vol. 67, no. 8, pp. 1992–2006, Apr. 2019.

[21] S. Fafat, N. J. Cowan, and R. Vidal, “Linear systems with sparse inputs: Observability and input recovery,” in Proc. Amer. Control Conf., 2015, pp. 525–529.

[22] S. M. Fosson, F. Garin, S. Gracy, A. Y. Kibangou, and D. Swart, “Input and state estimation exploiting input sparsity,” in Proc. 18th Eur. Control Conf., 2019, pp. 2344–2349.

[23] S. Z. Yong, M. Zhu, and E. Frazzoli, “A unified filter for simultaneous input and state estimation of linear discrete-time stochastic systems,” Automatica, vol. 63, pp. 321–329, 2016.

[24] S. Sundaram and C. Hadjicostis, “Distributed function calculation via linear iterative strategies in the presence of malicious agents,” IEEE Trans. Autom. Control, vol. 56, no. 7, pp. 1495–1508, Jul. 2011.

[25] R. Anguluri, L. Sankar, and O. Kosut, “Localization and estimation of unknown forced inputs: A group LASSO approach,” 2022, arXiv:2201.07907.

[26] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, “Block-sparse signals: Un- linear iterative strategies in the presence of malicious agents,” in Proc. IEEE Power Energy Conf., 2015, pp. 2715–2729.

[27] M. J. Wainwright, High-Dimensional Statistics: A Non-Asymptotic Viewpoint [See. Cambridge Series in Statistical and Probabilistic Mathematics], Cambridge, U.K.: Cambridge Univ. Press, 2019.

[28] J. J. Fuchs, “On sparse representations in arbitrary redundant bases,” IEEE Trans. Inf. Theory, vol. 50, no. 6, pp. 1341–1344, Jun. 2004.

[29] B. D. O. Anderson and M. Deistler, “Properties of zero-free spectral matrices,” IEEE Trans. Autom. Control, vol. 54, no. 10, pp. 2365–2375, Oct. 2009.

[30] H. Liu and J. Zhang, “On the $l_1-l_q$ regularized regression,” 2008.

[31] S. Boyd et al., “Distributed optimization and statistical learning via the alternating direction method of multipliers,” Found. Trends Mach. Learn., vol. 3, no. 1, pp. 1–122, 2011.

[32] A. Ansari, “Input and state estimation for discrete-time linear systems with application to target tracking and fault detection,” Ph.D. dissertation, Dept. Aerosp. Eng., Univ. Michigan Press, Ann Arbor, MI, USA, 2018.

[33] A. Rinaldo, “Advanced statistical theory, lecture 8,” 2019. [Online]. Available: http://www.stat.cmu.edu/~arinaldo/Teaching/36709/S19/Scribed_Lectures/Feb21_Shenghao.pdf

Rajasekhar Anguluri (Member, IEEE) received the B.Tech. degree in electrical engineering from the National Institute of Technology Warangal, Hanamkonda, India, in 2013, and the M.S. degree in statistics and the Ph.D. degree in mechanical engineering from the University of California at Riverside, Riverside, CA, USA, both in 2019.

He is currently a Postdoctoral Research Scholar with the School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ, USA. His research interests include high-dimensional statistics, statistical signal processing, estimation and control, and power systems.

Dr. Anguluri was the recipient of the Mistletoe Fellowship Award in 2022.

Oliver Kosut (Member, IEEE) received the B.S. degree in electrical engineering and mathematics from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2004, and the Ph.D. degree in electrical and computer engineering from Cornell University, Ithaca, NY, USA, in 2010.

Since 2012, he has been a Faculty Member with the School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ, USA, where he is currently an Associate Professor. From 2010 to 2012, he was a Postdoctoral Research Associate with the Laboratory for Information and Decision Systems, MIT. His research interests include information theory, cybersecurity, and power systems.

Dr. Kosut was the recipient of the NSF CAREER Award in 2015.

Laitha Sankar (Senior Member, IEEE) received the B.Eng. degree in physics IITB from the Indian Institute of Technology Bombay, Mumbai, India, in 1992, the master’s degree in electrical and computer engineering from the University of Maryland, Baltimore County, Baltimore, MD, USA, in 1994, and the Ph.D. degree in electrical and computer engineering from Rutgers University, New Brunswick, NJ, USA, in 2007.

She is currently an Associate Professor with the School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ, USA. She is currently with an NSF-Funded Institute focused on using data science to enable real-time integration of synchrophasor data into grid operations.

Dr. Sankar was the recipient of the NSF CAREER Award in 2014.