Theoretical performance limits for jointly sparse signals via graphical models

Marco F. Duarte[r], Shriram Sarvotham[h], Dror Baron[m],
Michael B. Wakin[c], and Richard G. Baraniuk[r] *
[r] Rice University, Electrical and Computer Engineering. Houston, TX
[h] Halliburton. Houston, TX
[m] Menta Capital LLC. San Francisco, CA
[c] University of Michigan, Division of Engineering. Ann Arbor, MI
Technical Report TREE0802

July 29, 2008

Abstract

The compressed sensing (CS) framework has been proposed for efficient acquisition of sparse and compressible signals through incoherent measurements. In our recent work, we introduced a new concept of joint sparsity of a signal ensemble. For several specific joint sparsity models, we demonstrated distributed CS schemes. This paper considers joint sparsity via graphical models that link the sparse underlying coefficient vector, signal entries, and measurements. Our converse and achievable bounds establish that the number of measurements required in the noiseless measurement setting is closely related to the dimensionality of the sparse coefficient vector. Single signal and joint (single-encoder) CS are special cases of joint sparsity, and their performance limits fit into our graphical model framework for distributed (multi-encoder) CS.

1 Introduction

A framework for single-signal sensing and compression has recently emerged under the rubric of *Compressed Sensing* (CS). CS builds on the work of Candès, Romberg, and Tao [1] and Donoho [2], and relies on tractable signal recovery procedures that provide exact recovery of a signal of length N and sparsity K as long as cK projections are used to recover the signal (typically $c \approx 3$ or 4). While powerful, the CS theory is mainly designed to exploit intra-signal structures at a single sensor. Certain schemes have been proposed to apply CS in a multi-sensor setting [3, 4], but they ignore intra-signal correlations.

In our recent work [5], we introduced a theory for *distributed compressed sensing* (DCS) that enables new distributed coding algorithms to exploit both intra- and inter-signal correlation structures. In a typical DCS scenario, multiple sensors measure signals that are each individually sparse in some basis and also correlated among sensors. Each sensor *independently* encodes its signal by

*This work was supported by NSF-CCF, NSF-NeTS, ONR, and AFOSR.
Email: {duarte, shri, richb}@rice.edu, barondror@gmail.com, wakin@umich.edu; Web: dsp.rice.edu/cs

projecting it onto another, incoherent basis (such as a random one) and then transmits just a few of the resulting coefficients to a single collection point. Under the right conditions, a decoder at the collection point can recover each of the signals precisely.

The DCS theory relies on the *joint sparsity* of a signal ensemble. Unlike the single-signal definition of sparsity, however, there are numerous plausible ways in which joint sparsity could be defined. In this paper, we provide a general framework for joint sparsity using graphical models. Using this framework, we derive upper and lower bounds for the number of noiseless measurements required for recovery. Our results are also applicable to cases where the signal ensembles are measured jointly, as well as to the single signal case.

2 Compressed Sensing Background

Consider a length-N, real-valued signal $x \in \mathbb{R}^N$ and a sparsifying basis Ψ, which provides a K-sparse representation $\theta = \Psi^T x$ of x. Using $\|\cdot\|_p$ to denote the ℓ_p norm,[1] we have $\|\theta\|_0 = K$. Various expansions, including Fourier and wavelets, are widely used for representation and compression of natural signals, and other data.

In CS we do not measure or encode the sparse vector θ directly. Rather, we take $M < N$ projections of the signal onto a *second set* of random functions. Using matrix notation, we measure $y = \Phi x$, where $y \in \mathbb{R}^M$ column vector and the *measurement matrix* $\Phi \in \mathbb{R}^{M \times N}$ with i.i.d. Gaussian entries. Since $M < N$, recovery of the signal x from the measurements y is ill-posed in general. However, the assumption of signal sparsity makes recovery possible and computationally tractable.

The sparse set of significant coefficients θ can be recovered by searching for the signal with ℓ_0-sparsest coefficients $\widehat{\theta}$ that agrees with y:

$$\widehat{\theta} = \arg\min \|\theta\|_0 \quad \text{s.t. } y = \Phi\Psi\theta. \tag{1}$$

In principle, remarkably few incoherent measurements are required to perfectly recover a K-sparse signal using (1). Although it is necessary to take more than K measurements to avoid ambiguity, $K + 1$ measurements will suffice [5]. Thus, one measurement separates the *achievable region*, where perfect recovery is possible with probability one, from the *converse region*, where recovery is impossible. Unfortunately, (1) is prohibitively complex. In fact, it is NP-complete. Recovery methods such as ℓ_1 minimization provide computationally tractable signal recovery at the expense of a moderate increase in the number of measurements M [1, 2].

3 Joint Sparsity Models

In this section, we generalize the notion of a signal being sparse in some basis to *joint sparsity* within a signal ensemble. We begin with basic notation. Let $\Lambda := \{1, 2, \ldots, J\}$ be the set of signal indices. Denote the signals in the ensemble by $x_j \in \mathbb{R}^N$, where $j \in \Lambda$. We use $x_j(n)$ to denote sample n in signal j, and assume for the sake of illustration that these signals are sparse in the canonical basis, i.e., $\Psi = \mathbf{I}$. The entries of the signal can take arbitrary real values, and the framework is extendable to arbitrary Ψ.

We denote by Φ_j the measurement matrix for signal j; Φ_j is $M_j \times N$ and, in general, entries of Φ_j are different for each j. Thus, $y_j = \Phi_j x_j$ consists of $M_j < N$ *random measurements* of x_j. We emphasize random Gaussian matrices Φ_j in the following, but other measurement matrices are possible.

[1] The ℓ_0 "norm" $\|\theta\|_0$ counts the number of nonzero entries in θ.

To compactly represent the signal and measurement ensembles, we define $X = [x_1^T \ \dots \ x_J^T]^T \in \mathbb{R}^{JN}$ and $Y = [y_1^T \ \dots \ y_J^T]^T \in \mathbb{R}^{\sum M_j}$. Finally, we also define $\Phi = \text{diag}(\Phi_1, \dots, \Phi_J)$, where diag denotes a matrix diagonal concatenation, to get $Y = \Phi X$.

3.1 Algebraic framework

Our framework enables analysis of a given ensemble $x_1, x_2, \dots, x_J$ in a "jointly sparse" sense, as well as a metric for the complexities of different signal ensembles. It is based on a factored representation of the signal ensemble, and decouples location and value information. We begin by illustrating the single signal case.

Single signal case: Consider a sparse $x \in \mathbb{R}^N$ with $K < N$ nonzero entries. Alternatively, we can write $x = P\theta$, where $\theta \in \mathbb{R}^K$ contains the nonzero values of x, and P is an *identity submatrix*, i.e., P contains K columns of the $N \times N$ identity matrix $\mathbf{I}$. To model the set of all possible sparse signals, let $\mathcal{P}$ be the set of all identity submatrices of all possible sizes $N \times K'$, with $1 \leq K' \leq N$. We refer to $\mathcal{P}$ as a *sparsity model*. Given a signal x, one may consider all possible factorizations $x = P\theta$, with $P \in \mathcal{P}$. Among them, the smallest dimensionality for θ indicates the *sparsity* of x under the model $\mathcal{P}$.

Multiple signal case: For multiple signals, consider factorizations of the form $X = P\Theta$ where $X \in \mathbb{R}^{JN}$ as above, $P \in \mathbb{R}^{JN \times D}$, and $\Theta \in \mathbb{R}^D$. We refer to P and Θ as the *location matrix* and *value vector*, respectively. A *joint sparsity model* (JSM) is defined in terms of a set $\mathcal{P}$ of admissible location matrices P with varying numbers of columns. Unlike the single signal case, there are multiple choices for what matrices P belong to a joint sparsity model $\mathcal{P}$.

Minimal sparsity: For a given ensemble X, let $\mathcal{P}_F(X)$ denote the set of feasible location matrices $P \in \mathcal{P}$ for which a factorization $X = P\Theta$ exists. Among the feasible location matrices, we let $\mathcal{P}_M(X) \subseteq \mathcal{P}_F(X)$ denote the matrices P having the minimal number of columns. The number of columns D for each $P \in \mathcal{P}_M(X)$ is called the *joint sparsity level* of X under the model $\mathcal{P}$. Generally speaking, the minimal location matrices $\mathcal{P}_M(X)$ permit the most efficient factorizations of the signal ensemble; we show in Section 4 that these matrices dictate the number of measurements.

We restrict our attention in this paper to scenarios where each signal x_j is generated as a combination of two components: (i) a common component z_C, which is present in all signals, and (ii) an innovation component z_j, which is unique to each signal. These combine additively, giving $x_j = z_C + z_j, \quad j \in \Lambda$. However, individual components might be zero-valued in specific scenarios.

3.2 Example Joint Sparsity Model: JSM-1

In the sparse common and innovations (JSM-1) model [5], the common component z_C and each innovation component z_j are *sparse* with respective sparsities K_C and K_j. Within our algebraic framework, the class of JSM-1 signals correspond to the set of all matrices

$$P = \begin{bmatrix} P_C & P_1 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ P_C & \mathbf{0} & \dots & P_J \end{bmatrix},$$

where P_C and $\{P_j\}_{j \in \Lambda}$ are arbitrary identity submatrices of sizes $N \times K_C$ and $N \times K_j$, respectively, and $\mathbf{0}$ denotes a zero matrix of appropriate size. Given $X = P\Theta$, we can partition the value vector $\Theta = [\theta_C^T \ \theta_1^T \ \theta_2^T \ \dots \ \theta_J^T]^T$, where $\theta_C \in \mathbb{R}^{K_C}$ and each $\theta_j \in \mathbb{R}^{K_j}$. When generating a signal according

to this model, we have $z_C = P_C\theta_C$, $z_j = P_j\theta_j$, $j \in \Lambda$. If $P \in \mathcal{P}_M(X)$, then the joint sparsity is $D = K_C + \sum_{j\in\Lambda} K_j$.

Sparsity reduction: If a signal ensemble $X = P\Theta$, $\Theta \in \mathbb{R}^D$, were to be generated by a selection of P_C and $\{P_j\}_{j\in\Lambda}$, where all $J+1$ identity submatrices share a common column vector, then $P \notin \mathcal{P}_M(X)$. By removing the instance of this column in P_C, one obtains $Q \in \mathcal{P}$ such that there exists $\Theta' \in \mathbb{R}^{D-1}$ with $X = Q\Theta'$. We term this phenomenon *sparsity reduction*, since it reduces the effective joint sparsity of a signal ensemble.

4 Bound on Measurement Rates

We seek conditions on the number of measurements from each sensor that guarantee perfect recovery of X given Y. Within our algebraic framework, recovering X involves determining a value vector Θ and location matrix P such that $X = P\Theta$. Two challenges are present. First, a given measurement depends only on some of the components of Θ, and the measurement budget should be adjusted between the sensors in order to gather sufficient information on all components of Θ. Second, the decoder must identify a feasible location matrix $P \in \mathcal{P}_F(X)$ from the set $\mathcal{P}$ and the measurements Y. In this section, we develop tools to address these challenges and characterize the number of measurements needed by them.

4.1 Graphical model framework

We introduce a graphical representation that captures the dependencies between the measurements in Y and the value vector Θ, represented by Φ and P. Consider a feasible decomposition of X into $P \in \mathcal{P}_F(X)$ and the corresponding Θ. We define the following sets of vertices, illustrated in Figure 1(a): (*i*) the set of *value vertices* V_V has elements with indices $d \in \{1, \ldots, D\}$ representing entries of the value vector $\theta(d)$; (*ii*) the set of *signal vertices* V_S has elements with indices (j, n) representing the signal entries $x_j(n)$, with $j \in \Lambda$ and $n \in \{1, \ldots, N\}$; and (*iii*) the set of *measurement vertices* V_M has elements with indices (j, m) representing the measurements $y_j(m)$, with $j \in \Lambda$ and $m \in \{1, \ldots, M_j\}$. The cardinalities of these sets are $|V_V| = D$, $|V_S| = JN$ and $|V_M| = \sum_{j\in\Lambda} M_j$.

Let P be partitioned into *location submatrices* P^j, $j \in \Lambda$, so that $x_j = P^j\Theta$; here P^j is the restriction of P to the rows that generate the signal x_j. We then define the bipartite graph $G = (V_S, V_V, E)$, determined by P, where there exists an edge connecting (j, n) and d if and only if $P^j(n, d) \neq 0$.

A similar bipartite graph $G' = (V_M, V_S, E')$, illustrated in Figure 1(a), connects between the measurement vertices $\{(j, m)\}$ and the signal vertices $\{(j, n)\}$; there exists an edge in G' connecting $(j, n) \in V_S$ and $(j, m) \in V_M$ if $\Phi_j(m, n) \neq 0$. When the measurements matrices Φ_j are dense, which occurs with probability one for i.i.d. Gaussian random matrices, the vertices corresponding to entries of a given signal x_j in V_S are all connected to all vertices corresponding to the measurements y_j in V_V. Figure 1 shows an example for dense measurement matrices: each measurement vertex $(j, \cdot)$ is connected to each signal vertex $(j, \cdot)$.

The graphs G and G' can be merged into $\widehat{G} = (V_M, V_V, \widehat{E})$ that relates entries of the value vector to measurements. Figure 1(b) shows the example composition of the previous two bipartite graphs. $\widehat{G}$ is used to recover Θ from the measurement ensemble Y when P is known.

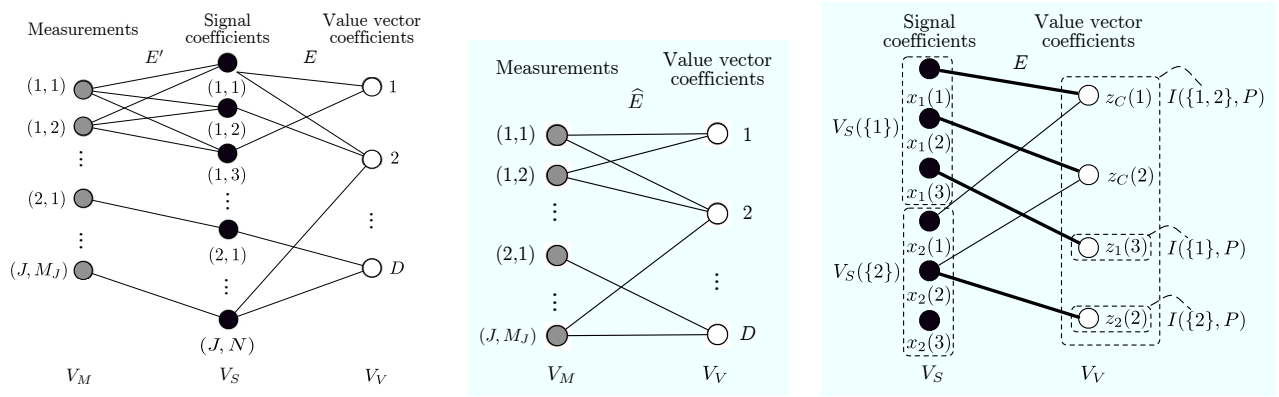

Figure 1: *Bipartite graphs for distributed compressed sensing. (a) $G = (V_S, V_V, E)$ connects the entries of each signal with the value vector coefficients they depend on; $G' = (V_M, V_S, E')$ connects the measurements at each sensor with observed signal entries. The matrix Φ is a dense Gaussian random matrix, as shown in the graph. (b) $\widehat{G} = (V_M, V_V, \widehat{E})$ is the composition of G and G', and relates between value vector coefficients and measurements. (c) Sets of exclusive indices for our example.*

4.2 Quantifying dependencies and redundancies

We now define the subset of the value vector entries that is measured exclusively by a subset Γ of the sensors in the ensemble; the cardinality of this set will help determine the number of measurements the sensors in Γ should perform. We denote by $E(V)$ the neighbors of a set of vertices V through the edge set E.

Definition 1 *Let $G = (V_S, V_V, E)$ be the bipartite graph determined by P, let $\Gamma \subseteq \Lambda$, and let $V_S(\Gamma)$ be the set of vertices $V_S(\Gamma) = \{(j,n) \in V_S : j \in \Gamma, n \in \{1,\dots,N\}\}$. We define the* set of exclusive indices *for Γ given P, denoted $I(\Gamma, P)$, as the largest subset of $\{1,\dots,D\}$ such that $E(I(\Gamma,P)) \subseteq V_S(\Gamma)$.*

$I(\Gamma, P)$ is significant in our distributed measurement setting, because it contains the coefficients of θ that only affect the signals in the set Γ and, therefore, can only be measured by those sensors. Figure 1(c) shows an example setting of two signals of length $N = 3$ generated by a matrix P from the JSM-1 model, with the sets $I(\{1\}, P)$ and $I(\{2\}, P)$ defined as the vertices in V_V that connect exclusively with $V_S(\{1\})$ and $V_S(\{2\})$, respectively.

Overlaps: When overlaps between common and innovation components are present in a signal, we cannot recover the overlapped portions of both components from the measurements of this signal alone; we need to recover the common component's coefficients using measurements of other signals that do not feature the same overlap. Furthermore, these coefficients of the value vector are not included in $I(\Gamma, P)$. We thus quantify the size of the overlap for all subsets of signals $\Gamma \subset \Lambda$ under a feasible representation given by P and Θ.

Definition 2 *The* overlap size *for the set of signals $\Gamma \subset \Lambda$, denoted $K_{C,\Gamma}$, is the number of indices in which there is overlap between the common and the innovation component supports at the signals $j \notin \Gamma$; more formally,*

$$K_{C,\Gamma}(P) = |\{n \in \{1,\dots,N\} : z_C(n) \neq 0,\ \forall j \notin \Gamma, z_j(n) \neq 0\}|.$$

For the entire set of signals, the overlap size $K_{C,\Lambda} = 0$.

For $\Gamma \neq \Lambda$, $K_{C,\Gamma}(P)$ provides a penalty term due to the need for recovery of common component coefficients that are overlapped by innovations in all other signals $j \notin \Gamma$. The definition of $K_{C,\Lambda}$ accounts for the fact that all the coefficients of Θ are included in $I(\Lambda, P)$.

4.3 Main Result

Converse and achievable bounds on the number of measurements necessary for recovery are given below.

Theorem 1 (Achievable, known P) *Assume that a signal ensemble X is obtained from a common/innovation component JSM $\mathcal{P}$. Let $\{M_j\}_{j\in\Lambda}$ be a measurement tuple. Suppose there exists a full rank location matrix $P \in \mathcal{P}_F(X)$ such that*

$$\sum_{j\in\Gamma} M_j \geq |I(\Gamma, P)| + K_{C,\Gamma}(P) \tag{2}$$

for all $\Gamma \subseteq \Lambda$. If the Φ_j are random matrices having M_j rows of i.i.d. Gaussian entries for each $j \in \Lambda$, and if $Y = \Phi X$, then with probability one over Φ, there is a unique solution $\widehat{\Theta}$ to the system of equations $Y = \Phi P\widehat{\Theta}$, and hence the signal ensemble X can be uniquely reconstructed as $X = P\widehat{\Theta}$.

Theorem 2 (Achievable, unknown P) *Assume that a signal ensemble X is obtained from a common/innovation component JSM $\mathcal{P}$, and let Φ_j be random matrices having M_j rows of i.i.d. Gaussian entries for each $j \in \Lambda$. If there exists a location matrix $P^* \in \mathcal{P}_F(X)$ such that*

$$\sum_{j\in\Gamma} M_j \geq |I(\Gamma, P^*)| + K_{C,\Gamma}(P^*) + |\Gamma| \tag{3}$$

for all $\Gamma \subseteq \Lambda$, then X can be uniquely recovered from Y with probability one over Φ.

Theorem 3 (Converse) *Assume that a signal ensemble X is obtained from a common/innovation component JSM $\mathcal{P}$. Let $\{M_j\}_{j\in\Lambda}$ be a measurement tuple. Suppose there exists a full rank location matrix $P \in \mathcal{P}_F(X)$ such that*

$$\sum_{j\in\Gamma} M_j < |I(\Gamma, P)| + K_{C,\Gamma}(P) \tag{4}$$

for some $\Gamma \subseteq \Lambda$. Let Φ_j be any set of measurement matrices having M_j rows for each $j \in \Lambda$, and let $Y = \Phi X$. Then there exists a solution $\widehat{\Theta}$ such that $Y = \Phi P\widehat{\Theta}$ but $\widehat{X} := P\widehat{\Theta} \neq X$.

These theorems are proved in Appendices A, C, and D, respectively. The number of measurements needed for recovery depends on the number of value vector coefficients that are observed only by the sensors in Γ. The identication of a feasible location matrix P causes the 2 measurement-per-sensor gap between the converse and achievable bounds (3-4). The algorithm used in Theorem 2 essentially performs an ℓ_0 minimization to acquire Θ, where the correct P is identified using an additional cross-validation step.

Discussion: The theorems can also be applied to the single sensor and joint measurement settings. In the single signal setting, we will have $x = P\theta$ with $\theta \in \mathbb{R}^K$, and $\Lambda = \{1\}$; the theorem provides the requirement $M \geq K + 1$, which matches the existing requirements for reconstruction.

The joint measurement setting is equivalent to the single signal setting with a dense measurement matrix, as all measurements are dependent on all signal entries. In this case, however, the distribution of the measurements among the available sensors is irrelevant. Therefore, we only obtain a condition on the total number of measurements obtained by the group of sensors as $\sum_{j\in\{1,\ldots,N\}} M_j \geq D + 1$.

A Proof of Theorem 1

Because $P \in \mathcal{P}_F(X)$, there exists $\Theta \in \mathbb{R}^\Xi$ such that $X = P\Theta$. Because $Y = \Phi X$, then Θ is a solution to $Y = \Phi P\Theta$. We will argue that, with probability one over Φ,

$$\Upsilon := \Phi P$$

has rank Ξ, and thus Θ is the unique solution to the equation $Y = \Phi P\Theta = \Upsilon\Theta$.

We recall that, under our common/innovation model, P has the form

$$P = \begin{bmatrix} P_C & P_1 & \mathbf{0} & \ldots & \mathbf{0} \\ P_C & \mathbf{0} & P_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_C & \mathbf{0} & \mathbf{0} & \ldots & P_J \end{bmatrix},$$

where P_C is an $N \times K_C$ submatrix of the $N \times N$ identity, and each P_j, $j \in \Lambda$, is an $N \times K_j$ submatrix of the $N \times N$ identity. We let

$$\Xi := K_C + \sum_{j\in\Lambda} K_j = K_C + \sum_{j\in\Lambda} |I(\{j\}, P)| \tag{5}$$

denote the number of columns in P.

To prove that Υ has rank Ξ, we will require the following lemma, which we prove in Appendix B.

Lemma 1 *If (2) holds, then there exists a mapping $\mathcal{M} : \{1, 2, \ldots, K_C\} \to \Lambda$, assigning each element of the common component to one of the sensors, such that for each $\Gamma \subseteq \Lambda$,*

$$\sum_{j\in\Gamma} M_j \geq |I(\Gamma, P)| + \sum_{k=1}^{K_C} 1_{\mathcal{M}(k)\in\Gamma} \tag{6}$$

and such that for each $k \in \{1, 2, \ldots, K_C\}$, the k^{th} column of P_C does not also appear as a column of $P_{\mathcal{M}(k)}$.

Intuitively, the existence of such a mapping suggests that *(i)* each sensor has taken enough measurements to cover its own innovation (requiring $|I(\{j\}, P)|$ measurements) and perhaps some of the common component, *(ii)* for any $\Gamma \subseteq \Lambda$, the sensors in Γ have collectively taken enough extra measurements to cover the requisite $K_{C,\Gamma}(P)$ elements of the common component, and *(iii)*

the extra measurements are taken at sensors where the common and innovation components do not overlap. Formally, we will use the existence of such a mapping to prove that Υ has rank Ξ.

We proceed by noting that Υ has the form

$$\Upsilon = \begin{bmatrix} \Phi_1 P_C & \Phi_1 P_1 & \mathbf{0} & \dots & \mathbf{0} \\ \Phi_2 P_C & \mathbf{0} & \Phi_2 P_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Phi_J P_C & \mathbf{0} & \mathbf{0} & \dots & \Phi_J P_J \end{bmatrix},$$

where each $\Phi_j P_C$ (respectively, $\Phi_j P_j$) is an $M_j \times K_C$ (respectively, $M_j \times K_j$) submatrix of Φ_j obtained by selecting columns from Φ_j according to the nonzero entries of P_C (respectively, P_j). In total, Υ has Ξ columns (5). To argue that Υ has rank Ξ, we will consider a sequence of three matrices Υ_0, Υ_1, and Υ_2 constructed from small modifications to Υ.

We begin by letting Υ_0 denote the "partially zeroed" matrix obtained from Υ using the following construction. We first let $\Upsilon_0 = \Upsilon$ and then make the following adjustments:

1. Let $k = 1$.
2. For each j such that P_j has a column that matches column k of P_C (note that by Lemma 1 this cannot happen if $\mathcal{M}(k) = j$), let k' represent the column index of the full matrix P where this column of P_j occurs. Subtract column k' of Υ_0 from column k of Υ_0. This forces to zero all entries of Υ_0 formerly corresponding to column k of the block $\Phi_j P_C$.
3. If $k < K_C$, add one to k and go to step 2.

The matrix Υ_0 is identical to Υ everywhere except on the first K_C columns, where any portion of a column overlapping with a column of $\Phi_j P_j$ to its right has been set to zero. Thus, Υ_0 satisfies the following two properties, which will be inherited by matrices Υ_1 and Υ_2 that we subsequently define:

P1. Each entry of Υ_0 is either zero or a Gaussian random variable.

P2. All Gaussian random variables in Υ_0 are i.i.d.

Finally, because Υ_0 was constructed only by subtracting columns of Υ from one another,

$$\operatorname{rank}(\Upsilon_0) = \operatorname{rank}(\Upsilon). \tag{7}$$

We now let Υ_1 be the matrix obtained from Υ_0 using the following construction. For each $j \in \Lambda$, we select $|I(\{j\}, P)| + \sum_{k=1}^{K_C} 1_{\mathcal{M}(k)=j}$ arbitrary rows from the portion of Υ_0 corresponding to sensor j. Using (5), the resulting matrix Υ_1 has

$$\sum_{j \in \Lambda} \left(|I(\{j\}, P)| + \sum_{k=1}^{K_C} 1_{\mathcal{M}(k)=j} \right) = \left(\sum_{j \in \Lambda} |I(\{j\}, P)| \right) + K_C = \Xi$$

rows. Also, because Υ_1 was obtained by selecting a subset of rows from Υ_0, it has Ξ columns (5) and satisfies

$$\operatorname{rank}(\Upsilon_1) \leq \operatorname{rank}(\Upsilon_0). \tag{8}$$

We now let Υ_2 be the $\Xi \times \Xi$ matrix obtained by permuting columns of Υ_1 using the following construction:

1. Let $\Upsilon_2 = [\,]$, and let $j = 1$.

2. For each k such that $\mathcal{M}(k) = j$, let $\Upsilon_1(k)$ denote the k^{th} column of Υ_1, and concatenate $\Upsilon_1(k)$ to Υ_2, i.e., let $\Upsilon_2 \leftarrow [\Upsilon_2\ \Upsilon_1(k)]$. There are $\sum_{k=1}^{K_C} 1_{\mathcal{M}(k)=j}$ such columns.

3. Let Υ_1' denote the columns of Υ_1 corresponding to the entries of $\Phi_j P_j$ (the innovation components of sensor j), and concatenate Υ_1' to Υ_2, i.e., let $\Upsilon_2 \leftarrow [\Upsilon_2\ \Upsilon_1']$. There are $|I(\{j\}, P)|$ such columns.

4. If $j < J$, let $j \leftarrow j + 1$ and go to Step 2.

Because Υ_1 and Υ_2 share the same columns up to reordering, it follows that

$$\text{rank}(\Upsilon_2) = \text{rank}(\Upsilon_1). \tag{9}$$

Based on its dependence on Υ_0, and following from Lemma 1, the square matrix Υ_2 meets properties P1 and P2 defined above in addition to a third property:

P3. All diagonal entries of Υ_2 are Gaussian random variables.

This follows because for each j, $|I(\{j\}, P)| + \sum_{k=1}^{K_C} 1_{\mathcal{M}(k)=j}$ rows of Υ_1 are assigned in its construction, while $|I(\{j\}, P)| + \sum_{k=1}^{K_C} 1_{\mathcal{M}(k)=j}$ columns of Υ_2 are assigned in its construction. Thus, each diagonal element of Υ_2 will either be an entry of some $\Phi_j P_j$, which remains Gaussian throughout our constructions, or it will be an entry of some k^{th} column of some $\Phi_j P_C$ for which $\mathcal{M}(k) = j$. In the latter case, we know by Lemma 1 and the construction of Υ_0 that this entry remains Gaussian throughout our constructions.

Having identified these three properties satisfied by Υ_2, we will prove by induction that, with probability one over Φ, such a matrix has full rank.

Lemma 2 *Let $\Upsilon^{(d-1)}$ be a $(d-1) \times (d-1)$ matrix having full rank. Construct a $d \times d$ matrix $\Upsilon^{(d)}$ as follows:*

$$\Upsilon^{(d)} := \begin{bmatrix} \Upsilon^{(d-1)} & v_1 \\ v_2^t & \omega \end{bmatrix}$$

where $v_1, v_2 \in \mathbb{R}^{d-1}$ are vectors with each entry being either zero or a Gaussian random variable, ω is a Gaussian random variable, and all random variables are i.i.d. and independent of $\Upsilon^{(d-1)}$. Then with probability one, $\Upsilon^{(d)}$ has full rank.

Applying Lemma 2 inductively Ξ times, the success probability remains one. It follows that with probability one over Φ, $\text{rank}(\Upsilon_2) = \Xi$. Combining this last result with (7-9), we obtain $\text{rank}(\Upsilon) = \Xi$ with probability one over Φ. It remains to prove Lemma 2.

Proof of Lemma 2: When $d = 1$, $\Upsilon^{(d)} = [\omega]$, which has full rank if and only if $\omega \neq 0$, which occurs with probability one.

When $d > 1$, using expansion by minors, the determinant of $\Upsilon^{(d)}$ satisfies

$$\det(\Upsilon^{(d)}) = \omega \cdot \det(\Upsilon^{(d-1)}) + C,$$

where $C = C(\Upsilon^{(d-1)}, v_1, v_2)$ is independent of ω. The matrix $\Upsilon^{(d)}$ has full rank if and only if $\det(\Upsilon^{(d)}) \neq 0$, which is satisfied if and only if

$$\omega \neq \frac{-C}{\det(\Upsilon^{(d-1)})}.$$

By assumption, $\det(\Upsilon^{(d-1)}) \neq 0$ and ω is a Gaussian random variable that is independent of C and $\det(\Upsilon^{(d-1)})$. Thus, $\omega \neq \frac{-C}{\det(\Upsilon^{(d-1)})}$ with probability one. $\square$

B Proof of Lemma 1

To prove this lemma, we apply tools from graph theory. We begin by specifying a bipartite graph $\widetilde{G} = (V_V, V_M, \widetilde{E})$ that depends on the structure of the location matrix $P \in \mathcal{P}_F(X)$. The graph $\widetilde{G}$ has two sets of vertices V_V and V_M and a collection of edges $\widetilde{E}$ joining elements of V_V to V_M. The set V_V has vertices with indices $k \in \{1, 2, \ldots, \Xi\}$, which are known as *value vertices* and represent entries of the value vector Θ (equivalently, columns of the matrix P). The set V_M has vertices with indices (j, m), with $j \in \Lambda$ and $m \in \{1, 2, \ldots, M_j\}$, which are known as *measurement vertices* and represent entries $y_j(m)$ of the measurement vectors (equivalently, rows of the matrix Φ). The edges $\widetilde{E}$ are specified as follows:

- For every $k \in \{1, 2, \ldots, K_C\} \subseteq V_V$ and $j \in \Lambda$ such that column k of P_C does not also appear as a column of P_j, we have an edge connecting k to each vertex $(j, m) \in V_M$ for $1 \leq m \leq M_j$.
- For every $k \in \{K_C+1, K_C+2, \ldots, \Xi\} \subseteq V_V$, we consider the sensor j associated with column k of P, and we have an edge connecting k to each vertex $(j, m) \in V_M$ for $1 \leq m \leq M_j$.

This graph $\widetilde{G}$ is a subgraph of the graph $\widehat{G}$ shown in Figure 1(c), from which we remove the edges going from common component vertices in V_V to measurement vertices in V_M that have incoming edges from innovation component vertices in V_V.

We seek a matching within this graph, i.e., a subgraph $(V_V, V_M, \bar{E})$ with $\bar{E} \subseteq \widetilde{E}$ that pairs each element of V_V with a unique element of V_M. Such a matching will immediately give us the desired mapping $\mathcal{M}$ as follows: for each $k \in \{1, 2, \ldots, K_C\} \subseteq V_V$, we let $(j, m) \in V_M$ denote the single node matched to k by an edge in $\widetilde{E}$, and we set $\mathcal{M}(k) = j$.

To prove the existence of such a matching within the graph, we invoke a version of Hall's marriage theorem for bipartite graphs [6]. Hall's theorem states that within a bipartite graph (V_1, V_2, E), there exists a matching that assigns each element of V_1 to a unique element of V_2 if for any collection of elements $\Pi \subseteq V_1$, the set $E(\Pi)$ of neighbors of Π in V_2 has cardinality $|E(\Pi)| \geq |\Pi|$.

In the context of our lemma, Hall's condition requires that for any set of entries in the value vector, $\Pi \subseteq V_V$, the set $\widetilde{E}(\Pi)$ of neighbors of Π in V_M has size $|\widetilde{E}(\Pi)| \geq |\Pi|$. We will prove that if (2) is satisfied, then Hall's condition is satisfied, and thus a matching must exist.

Let us consider an arbitrary set $\Pi \subseteq V_V$. We let $\widetilde{E}(\Pi)$ denote the set of neighbors of Π in V_M joined by edges in $\widetilde{E}$, and we let $S_\Pi = \{j \in \Lambda : (j, m) \in \widetilde{E}(\Pi) \text{ for some } m\}$. Thus, $S_\Pi \subseteq \Lambda$ denotes the set of signal indices whose measurement nodes have edges that connect to Π. It follows that $|\widetilde{E}(\Pi)| = \sum_{j \in S_\Pi} M_j$. Thus, in order to satisfy Hall's condition for Π, we require

$$\sum_{j \in S_\Pi} M_j \geq |\Pi|. \tag{10}$$

We would now like to show that $|I(S_\Pi, P)| + K_{C,S_\Pi}(P) \geq |\Pi|$, and thus if (2) is satisfied for all $\Gamma \subseteq \Lambda$, then (10) is satisfied in particular for $S_\Pi \subseteq \Lambda$.

In general, the set Π may contain vertices for both common components and innovation components. We write $\Pi = \Pi_I \cup \Pi_C$ to denote the disjoint union of these two sets.

By construction, $|I(S_\Pi, P)| = |\Pi_I|$ because $I(S_\Pi, P)$ counts all innovations with neighbors in S_Π, and because S_Π contains all neighbors for nodes in Π_I. We will also argue that $K_{C,S_\Pi}(P) \geq |\Pi_C|$ as follows. By definition, for a set $\Gamma \subseteq \Lambda$, $K_{C,\Gamma}(P)$ counts the number of columns in P_C that also appear in P_j for all $j \notin \Gamma$. By construction, for each $k \in \Pi_C$, node k has no connection to nodes (j, m) for $j \notin S_\Pi$; thus it must follow that the k^{th} column of P_C is present in P_j for all $j \notin S_\Pi$, due to the construction of the graph G. Consequently, $K_{C,S_\Pi}(P) \geq |\Pi_C|$.

Thus, $|I(S_\Pi, P)| + K_{C,S_\Pi}(P) \geq |\Pi_I| + |\Pi_C| = |\Pi|$, and so (2) implies (10) for any Π, and so Hall's condition is satisfied, and a matching exists. Because in such matching a set of vertices in V_M matches to a set in V_V of lower or equal cardinality, we have in particular that (6) holds for each $\Gamma \subseteq \Lambda$. □

C Proof of Theorem 2

Given the measurements Y and measurement matrix Φ, we show that it is possible to recover some $P \in \mathcal{P}_F(X)$ and a corresponding vector Θ such that $X = P\Theta$ using the following algorithm:

- Take the last measurement of each sensor for verification, and sum these J measurements to obtain a single *global* test measurement $\bar{y}$. Similarly, add the corresponding rows of Φ into a single row $\bar{\phi}$.
- Group all the remaining $\sum_{j\in\Lambda} M_j - J$ measurements into a vector $\bar{Y}$ and a matrix $\bar{\Phi}$.
- For each matrix $P \in \mathcal{P}$
  - choose a single solution Θ_P to $\bar{Y} = \bar{\Phi}P\Theta_P$ independently of $\bar{\phi}$ – if no solution exists, skip the next two steps;
  - define $X_P = P\Theta_P$;
  - cross-validate: check if $\bar{y} = \bar{\phi}X_P$; if so, return the estimate (P, Θ_P); if not, continue with the next matrix.

We begin by showing that, with probability one over Φ, the algorithm only terminates when it gets a correct solution – in other words, that for each $P \in \mathcal{P}$ the cross-validation measurement $\bar{y}$ can determine whether $X_P = X$. We note that all entries of the vector $\bar{\phi}$ are i.i.d. Gaussian, and independent from $\bar{\Phi}$. Assume for the sake of contradiction that there exists a matrix $P \in \mathcal{P}$ such that $\bar{y} = \bar{\phi}X_P$, but $X_P = P\Theta_P \neq X$; this implies $\bar{\phi}(X - X_P) = 0$, which occurs with probability zero over Φ. Thus, if $X_P \neq X$, then $\bar{\phi}X_P \neq \bar{y}$ with probability one over Φ. Since we only need to search over a finite number of matrices $P \in \mathcal{P}$, cross validation will determine whether each matrix $P \in \mathcal{P}$ gives the correct solution with probability one.

We now show that there is a matrix in $\mathcal{P}$ for which the algorithm will terminate with the correct solution. We know that the matrix $P^* \in \mathcal{P}_F(X) \subseteq \mathcal{P}$ will be part of our search, and that the unique solution Θ_{P^*} to $\bar{Y} = \bar{\Phi}P^*\Theta_{P^*}$ yields $X = P^*\Theta_{P^*}$ when (3) holds for P^*, as shown in Theorem 1. Thus, the algorithm will find at least one matrix P and vector Θ_P such that $X = P\Theta_P$; when such matrix is found the cross-validation step will return this solution and end the algorithm. □

Remark. Consider the algorithm used in the proof: if the matrices in $\mathcal{P}$ are sorted by number of columns, the algorithm is akin to ℓ_0 minimization on Θ with an additional cross-validation step.

D Proof of Theorem 3

We let Ξ denote the number of columns in P. Because $P \in \mathcal{P}_F(X)$, there exists $\Theta \in \mathbb{R}^{\Xi}$ such that $X = P\Theta$. Because $Y = \Phi X$, then Θ is a solution to $Y = \Phi P \Theta$. We will argue for $\Upsilon := \Phi P$ that $\text{rank}(\Upsilon) < \Xi$, and thus there exists $\widehat{\Theta} \neq \Theta$ such that $Y = \Upsilon\Theta = \Upsilon\widehat{\Theta}$. Moreover, since P has full rank, it follows that $\widehat{X} := P\widehat{\Theta} \neq P\Theta = X$.

We let Υ_0 be the "partially zeroed" matrix obtained from Υ using the identical procedure detailed in Appendix A. Again, because Υ_0 was constructed only by subtracting columns of Υ from one another, it follows that $\text{rank}(\Upsilon_0) = \text{rank}(\Upsilon)$.

Suppose $\Gamma \subseteq \Lambda$ is a set for which (4) holds. We let Υ_1 be the submatrix of Υ_0 obtained by selecting the following columns:

- For any $k \in \{1, 2, \ldots, K_C\}$ such that column k of P_C also appears as a column in all P_j for $j \notin \Gamma$, we include column k of Υ_0 as a column in Υ_1. There are $K_{C,\Gamma}(P)$ such columns k.
- For any $k \in \{K_C + 1, K_C + 2, \ldots, \Xi\}$ such that column k of P corresponds to an innovation for some sensor $j \in \Gamma$, we include column k of Υ_0 as a column in Υ_1. There are $|I(\Gamma, P)|$ such columns k.

This submatrix has $|I(\Gamma, P)| + K_{C,\Gamma}(P)$ columns. Because Υ_0 has the same size as Υ, and in particular has only Ξ columns, then in order to have that $\text{rank}(\Upsilon_0) = \Xi$, it is necessary that all $|I(\Gamma, P)| + K_{C,\Gamma}(P)$ columns of Υ_1 be linearly independent.

Based on the method described for constructing Υ_0, it follows that Υ_1 is zero for all measurement rows not corresponding to the set Γ. Therefore, let us consider the submatrix Υ_2 of Υ_1 obtained by selecting only the measurement rows corresponding to the set Γ. Because of the zeros in Υ_1, it follows that $\text{rank}(\Upsilon_1) = \text{rank}(\Upsilon_2)$. However, since Υ_2 has only $\sum_{j\in\Gamma} M_j$ rows, we invoke (4) and have that $\text{rank}(\Upsilon_1) = \text{rank}(\Upsilon_2) \leq \sum_{j\in\Gamma} M_j < |I(\Gamma, P)| + K_{C,\Gamma}(P)$. Thus, all $|I(\Gamma, P)| + K_{C,\Gamma}(P)$ columns of Υ_1 cannot be linearly independent, and so Υ does not have full rank. □

References

[1] E. J. Candès, "Compressive sampling," in *Proc. International Congress of Mathematicians*, vol. 3, (Madrid, Spain), pp. 1433–1452, 2006.

[2] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1289–1306, Apr. 2006.

[3] W. U. Bajwa, J. D. Haupt, A. M. Sayeed, and R. D. Nowak, "Joint source-channel communication for distributed estimation in sensor networks," *IEEE Trans. Inform. Theory*, vol. 53, pp. 3629–3653, Oct. 2007.

[4] W. Wang, M. Garofalakis, and K. Ramchandran, "Distributed sparse random projections for refinable approximation," in *Proc. 6th International Workshop on Inf. Processing in Sensor Networks*, (Cambridge, MA), 2007.

[5] D. Baron, M. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing," Tech. Rep. TREE0612, Rice University, Houston, TX, Nov. 2006. Available at `http://dsp.rice.edu/cs/`.

[6] D. B. West, *Introduction to Graph Theory*. Prentice Hall, 1996.