PERFORMANCE LIMITS FOR JOINTLY SPARSE SIGNALS VIA GRAPHICAL MODELS

Marco F. Duarte[r], Shriram Sarvotham[r], Dror Baron[m], Michael B. Wakin[u], and Richard G. Baraniuk[r]

[r] Rice University, Electrical and Computer Engineering. Houston, TX
[m] Menta Capital LLC. San Francisco, CA
[u] University of Michigan, Electrical Engineering and Computer Science. Ann Arbor, MI

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.

Index Terms— Distributed compressed sensing, jointly sparse signals, graphical models.

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 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.

This work was supported by NSF-CCF, NSF-NeTS, ONR, and AFOSR. E-mails: {duarte, shri, richb}@rice.edu, barondror@gmail.com, wakin@umich.edu.

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

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

3.1. Algebraic framework

Our framework enables analysis of a given ensemble $x_1, x_2, \ldots, x_J$ 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 $1 \le K' \le N$. 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 $x_j = z_C + z_j, \ j \in \Lambda$. 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

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

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 $\Theta = [\theta_C^T \ \theta_1^T \ \theta_2^T \ \ldots \ \theta_J^T]^T$, 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 $d \in \{1,\dots,D\}$ 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 $n \in \{1,\dots,N\}$; 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 $m \in \{1,\dots,M_j\}$. 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.

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 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,*

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

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

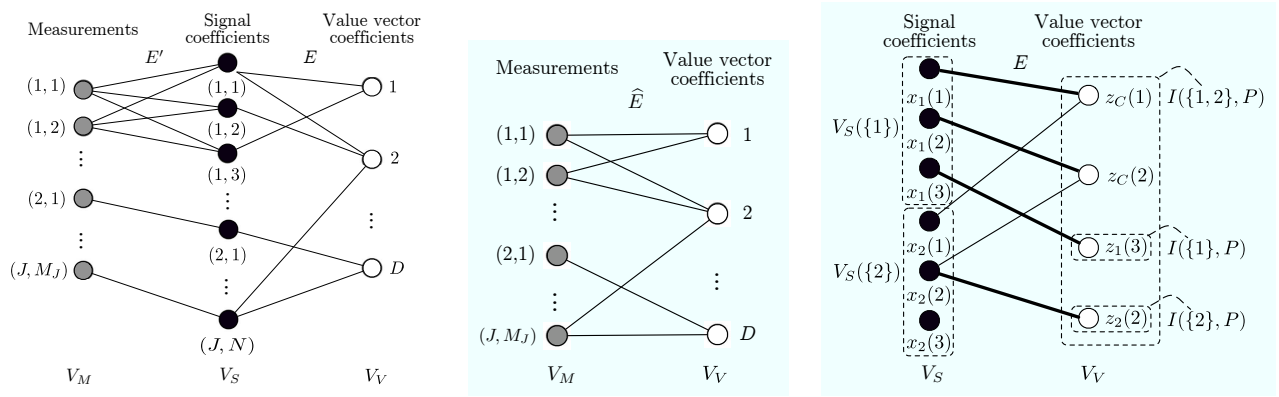

Fig. 1. *Bipartite graphs for distributed compressed sensing. (a) $G = (V_S, V_V, E)$ connects the entries of each signal with the value vector coefficients they depend on; $G' = (V_M, V_S, E')$ connects the measurements at each sensor with observed signal entries. The matrix Φ is a dense Gaussian random 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.*

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

A converse and achievable bound on the number of measurements necessary for recovery is given below.

Theorem 1 *(Achievable) Assume a signal ensemble X is obtained from a common/innovation component JSM where $\mathcal{P}$ and Φ_j are random Gaussian for all $j \in \Lambda$. If there exists a location matrix $P \in \mathcal{P}_M(X)$ such that*

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

for all $\Gamma \subseteq \Lambda$, then X can be uniquely recovered with probability one from Y using ℓ_0 minimization.
(Converse) If for each $P \in \mathcal{P}_M(X)$, for some $\Gamma \subseteq \Lambda$

$$\sum_{j \in \Gamma} M_j < |I(\Gamma, P)| + K_{C,\Gamma}(P), \qquad (3)$$

then the signal ensemble X cannot be uniquely recovered from Y, regardless of the measurement matrices Φ_j.

Theorem 1 is proved in [6]. 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 (2-3).

Discussion: The theorem 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 $\sum_{j \in \{1,\ldots,N\}} M_j \geq D + 1$.

5. 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] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Theoretic performance limits for jointly sparse signals via graphical models," Tech. Rep. TR0802, Rice University, Houston, TX, April 2008.