Distributed Compressive Sensing

Dror Baron,[1] *Marco F. Duarte,*[2] *Michael B. Wakin,*[3]
Shriram Sarvotham,[4] *and Richard G. Baraniuk*[2] *

[1]Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa, Israel
[2]Department of Electrical and Computer Engineering, Rice University, Houston, TX
[3]Division of Engineering, Colorado School of Mines, Golden, CO
[4]Halliburton, Houston, TX

This paper is dedicated to the memory of Hyeokho Choi, our colleague, mentor, and friend.

Abstract

Compressive sensing is a signal acquisition framework based on the revelation that a small collection of linear projections of a sparse signal contains enough information for stable recovery. In this paper we introduce a new theory for *distributed compressive sensing* (DCS) that enables new distributed coding algorithms for multi-signal ensembles that exploit both intra- and inter-signal correlation structures. The DCS theory rests on a new concept that we term the *joint sparsity* of a signal ensemble. Our theoretical contribution is to characterize the fundamental performance limits of DCS recovery for jointly sparse signal ensembles in the noiseless measurement setting; our result connects single-signal, joint, and distributed (multi-encoder) compressive sensing. To demonstrate the efficacy of our framework and to show that additional challenges such as computational tractability can be addressed, we study in detail three example models for jointly sparse signals. For these models, we develop practical algorithms for joint recovery of multiple signals from incoherent projections. In two of our three models, the results are asymptotically best-possible, meaning that both the upper and lower bounds match the performance of our practical algorithms. Moreover, simulations indicate that the asymptotics take effect with just a moderate number of signals. DCS is immediately applicable to a range of problems in sensor arrays and networks.

Keywords: Compressive sensing, distributed source coding, sparsity, random projection, random matrix, linear programming, array processing, sensor networks.

1 Introduction

A core tenet of signal processing and information theory is that signals, images, and other data often contain some type of *structure* that enables intelligent representation and processing. The notion of structure has been characterized and exploited in a variety of ways for a variety of purposes. In this paper, we focus on exploiting signal *correlations* for the purpose of *compression*.

*This work was supported by the grants NSF CCF-0431150 and CCF-0728867, DARPA HR0011-08-1-0078, DARPA/ONR N66001-08-1-2065, ONR N00014-07-1-0936 and N00014-08-1-1112, AFOSR FA9550-07-1-0301, ARO MURI W311NF-07-1-0185, and the Texas Instruments Leadership University Program. Preliminary versions of this work have appeared at the Allerton Conference on Communication, Control, and Computing [1], the Asilomar Conference on Signals, Systems and Computers [2], the Conference on Neural Information Processing Systems [3], and the Workshop on Sensor, Signal and Information Processing [4].
E-mail: drorb@ee.technion.ac.il, {duarte, shri, richb}@rice.edu, mwakin@mines.edu; Web: dsp.rice.edu/cs

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JAN 2009		**00-00-2009 to 00-00-2009**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Distributed Compressive Sensing	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Rice University ,Department of Electrical and Computer Engineering,Houston,TX,77005**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Preprint, Jan. 2009

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**42**	

Current state-of-the-art compression algorithms employ a decorrelating transform such as an exact or approximate Karhunen-Loève transform (KLT) to compact a correlated signal's energy into just a few essential coefficients [5–7]. Such *transform coders* exploit the fact that many signals have a *sparse* representation in terms of some basis, meaning that a small number K of adaptively chosen transform coefficients can be transmitted or stored rather than $N \gg K$ signal samples. For example, smooth signals are sparse in the Fourier basis, and piecewise smooth signals are sparse in a wavelet basis [8]; the commercial coding standards MP3 [9], JPEG [10], and JPEG2000 [11] directly exploit this sparsity.

1.1 Distributed source coding

While the theory and practice of compression have been well developed for individual signals, distributed sensing applications involve multiple signals, for which there has been less progress. Such settings are motivated by the proliferation of complex, multi-signal acquisition architectures, such as acoustic and RF sensor arrays, as well as sensor networks. These architectures sometimes involve battery-powered devices, which restrict the communication energy, and high aggregate data rates, limiting bandwidth availability; both factors make the reduction of communication critical.

Fortunately, since the sensors presumably observe related phenomena, the ensemble of signals they acquire can be expected to possess some joint structure, or *inter-signal correlation*, in addition to the *intra-signal correlation* within each individual sensor's measurements. In such settings, *distributed source coding* that exploits both intra- and inter-signal correlations might allow the network to save on the communication costs involved in exporting the ensemble of signals to the collection point [12–15]. A number of distributed coding algorithms have been developed that involve collaboration amongst the sensors [16–19]. Note, however, that any collaboration involves some amount of inter-sensor communication overhead.

In the *Slepian-Wolf* framework for lossless distributed coding [12–15], the availability of correlated side information at the decoder (collection point) enables each sensor node to communicate losslessly at its conditional entropy rate rather than at its individual entropy rate, as long as the sum rate exceeds the joint entropy rate. Slepian-Wolf coding has the distinct advantage that the sensors need not collaborate while encoding their measurements, which saves valuable communication overhead. Unfortunately, however, most existing coding algorithms [14, 15] exploit only inter-signal correlations and not intra-signal correlations. To date there has been only limited progress on distributed coding of so-called "sources with memory." The direct implementation for sources with memory would require huge lookup tables [12]. Furthermore, approaches combining pre- or post-processing of the data to remove intra-signal correlations combined with Slepian-Wolf coding for the inter-signal correlations appear to have limited applicability, because such processing would alter the data in a way that is unknown to other nodes. Finally, although recent papers [20–22] provide compression of spatially correlated sources with memory, the solution is specific to lossless distributed compression and cannot be readily extended to lossy compression settings. We conclude that the design of constructive techniques for distributed coding of sources with both intra- and inter-signal correlation is a challenging problem with many potential applications.

1.2 Compressive sensing (CS)

A new framework for single-signal sensing and compression has developed under the rubric of *compressive sensing* (CS). CS builds on the work of Candès, Romberg, and Tao [23] and Donoho [24], who showed that if a signal has a sparse representation in one basis then it can be recovered from

a small number of projections onto a second basis that is *incoherent* with the first.[1] CS relies on tractable recovery procedures that can provide exact recovery of a signal of length N and sparsity K, i.e., a signal that can be written as a sum of K basis functions from some known basis, where K can be orders of magnitude less than N.

The implications of CS are promising for many applications, especially for sensing signals that have a sparse representation in some basis. Instead of sampling a K-sparse signal N times, only $M = O(K \log N)$ incoherent measurements suffice, where K can be orders of magnitude less than N. Moreover, the M measurements need not be manipulated in any way before being transmitted, except possibly for some quantization. Finally, independent and identically distributed (i.i.d.) Gaussian or Bernoulli/Rademacher (random ± 1) vectors provide a useful *universal* basis that is incoherent with all others. Hence, when using a random basis, CS is universal in the sense that the sensor can apply the same measurement mechanism no matter what basis sparsifies the signal [27].

While powerful, the CS theory at present is designed mainly to exploit intra-signal structures at a *single* sensor. In a multi-sensor setting, one can naively obtain *separate measurements* from each signal and recover them separately. However, it is possible to obtain measurements that each depend on all signals in the ensemble by having sensors collaborate with each other in order to combine all of their measurements; we term this process a *joint measurement setting*. In fact, initial work in CS for multi-sensor settings used standard CS with joint measurement and recovery schemes that exploit inter-signal correlations [28–32]. However, by recovering sequential time instances of the sensed data individually, these schemes ignore intra-signal correlations.

1.3 Distributed compressive sensing (DCS)

In this paper we introduce a new theory for *distributed compressive sensing* (DCS) to enable new distributed coding algorithms that exploit *both* intra- and inter-signal correlation structures. In a typical DCS scenario, a number of sensors measure signals that are each individually sparse in some basis and also correlated from sensor to sensor. Each sensor *separately* 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. Unlike the joint measurement setting described in Section 1.2, DCS requires no collaboration between the sensors during signal acquisition. Nevertheless, we are able to exploit the inter-signal correlation by using all of the obtained measurements to recover all the signals *simultaneously*. Under the right conditions, a decoder at the collection point can recover each of the signals precisely.

The DCS theory rests on a concept that we term the *joint sparsity* — the sparsity of the entire signal ensemble. The joint sparsity is often smaller than the aggregate over individual signal sparsities. Therefore, DCS offers a reduction in the number of measurements, in a manner analogous to the rate reduction offered by the Slepian-Wolf framework [13]. Unlike the single-signal definition of sparsity, however, there are numerous plausible ways in which joint sparsity could be defined. In this paper, we first provide a general framework for joint sparsity using algebraic formulations based on a graphical model. Using this framework, we derive bounds for the number of measurements necessary for recovery under a given signal ensemble model. Similar to Slepian-Wolf coding [13], the number of measurements required for each sensor must account for the minimal features unique to that sensor, while at the same time features that appear among multiple sensors must be amortized over the group. Our bounds are dependent on the dimensionality of the subspaces in which each group of signals reside; they afford a reduction in the number of measurements that we quantify through the notions of *joint* and *conditional* sparsity, which are conceptually related to joint and

[1] Roughly speaking, *incoherence* means that no element of one basis has a sparse representation in terms of the other basis. This notion has a variety of formalizations in the CS literature [24–26].

conditional entropies. The common thread is that dimensionality and entropy both quantify the volume that the measurement and coding rates must cover. Our results are also applicable to cases where the signal ensembles are measured jointly, as well as to the single-signal case.

While our general framework does not by design provide insights for computationally efficient recovery, we also provide interesting models for joint sparsity where our results carry through from the general framework to realistic settings with low-complexity algorithms. In the first model, each signal is itself sparse, and so we could use CS to separately encode and decode each signal. However, there also exists a framework wherein a joint sparsity model for the ensemble uses fewer total coefficients. In the second model, all signals share the locations of the nonzero coefficients. In the third model, no signal is itself sparse, yet there still exists a joint sparsity among the signals that allows recovery from significantly fewer than N measurements per sensor. For each model we propose tractable algorithms for joint signal recovery, followed by theoretical and empirical characterizations of the number of measurements per sensor required for accurate recovery. We show that, under these models, joint signal recovery can recover signal ensembles from significantly fewer measurements than would be required to recover each signal individually. In fact, for two of our three models we obtain best-possible performance that could not be bettered by an oracle that knew the the indices of the nonzero entries of the signals.

This paper focuses primarily on the basic task of reducing the number of measurements for recovery of a signal ensemble in order to reduce the communication cost of source coding that ensemble. Our emphasis is on noiseless measurements of strictly sparse signals, where the optimal recovery relies on ℓ_0-norm optimization,[2] which is computationally intractable. In practical settings, additional criteria may be relevant for measuring performance. For example, the measurements will typically be real numbers that must be quantized, which gradually degrades the recovery quality as the quantization becomes coarser [33, 34]. Characterizing DCS in light of practical considerations such as rate-distortion tradeoffs, power consumption in sensor networks, etc., are topics of future research [31, 32].

1.4 Paper organization

Section 2 overviews the single-signal CS theories and provides a new result on CS recovery. While some readers may be familiar with this material, we include it to make the paper self-contained. Section 3 introduces our general framework for joint sparsity models and proposes three example models for joint sparsity. We provide our detailed analysis for the general framework in Section 4; we then address the three models in Section 5. We close the paper with a discussion and conclusions in Section 6. Several appendices contain the proofs.

2 Compressive Sensing Background

2.1 Transform coding

Consider a real-valued signal[3] $x \in \mathbb{R}^N$ indexed as $x(n)$, $n \in \{1, 2, \ldots, N\}$. Suppose that the basis $\Psi = [\psi_1, \ldots, \psi_N]$ provides a K-sparse representation of x; that is,

$$x = \sum_{n=1}^{N} \vartheta(n)\, \psi_n = \sum_{k=1}^{K} \vartheta(n_k)\, \psi_{n_k}, \tag{1}$$

[2]The ℓ_0 "norm" $\|x\|_0$ merely counts the number of nonzero entries in the vector x.

[3]Without loss of generality, we will focus on one-dimensional signals (vectors) for notational simplicity; the extension to multi-dimensional signal, e.g., images, is straightforward.

where x is a linear combination of K vectors chosen from Ψ, $\{n_k\}$ are the indices of those vectors, and $\{\vartheta(n)\}$ are the coefficients; the concept is extendable to tight frames [8]. Alternatively, we can write in matrix notation $x = \Psi\vartheta$, where x is an $N \times 1$ column vector, the *sparse basis* matrix Ψ is $N \times N$ with the basis vectors ψ_n as columns, and ϑ is an $N \times 1$ column vector with K nonzero elements. Using $\|\cdot\|_p$ to denote the ℓ_p norm, we can write that $\|\vartheta\|_0 = K$; we can also write the set of nonzero indices $\Omega \subseteq \{1, \ldots, N\}$, with $|\Omega| = K$. Various expansions, including wavelets [8], Gabor bases [8], curvelets [35], etc., are widely used for representation and compression of natural signals, images, and other data.

The standard procedure for compressing sparse and nearly-sparse signals, known as *transform coding*, is to (*i*) acquire the full N-sample signal x; (*ii*) compute the complete set of transform coefficients $\{\vartheta(n)\}$; (*iii*) locate the K largest, significant coefficients and discard the (many) small coefficients; (*iv*) encode the *values and locations* of the largest coefficients. This procedure has three inherent inefficiencies: First, for a high-dimensional signal, we must start with a large number of samples N. Second, the encoder must compute *all* of the N transform coefficients $\{\vartheta(n)\}$, even though it will discard all but K of them. Third, the encoder must encode the locations of the large coefficients, which requires increasing the coding rate since the locations change with each signal.

We will focus our theoretical development on exactly K-sparse signals and defer discussion of the more general situation of *compressible* signals where the coefficients decay rapidly with a power law but not to zero. Section 6 contains additional discussion on real-world compressible signals, and [36] presents simulation results.

2.2 Incoherent projections

These inefficiencies raise a simple question: For a given signal, is it possible to directly estimate the set of large $\vartheta(n)$'s that will not be discarded? While this seems improbable, Candès, Romberg, and Tao [23, 25] and Donoho [24] have shown that a reduced set of projections can contain enough information to recover sparse signals. A framework to acquire sparse signals, often referred to as *compressive sensing* (CS) [37], has emerged that builds on this principle.

In CS, we do not measure or encode the K significant $\vartheta(n)$ directly. Rather, we measure and encode $M < N$ projections $y(m) = \langle x, \phi_m^T \rangle$ of the signal onto a *second set* of functions $\{\phi_m\}, m = 1, 2, \ldots, M$, where ϕ_m^T denotes the transpose of ϕ_m and $\langle \cdot, \cdot \rangle$ denotes the inner product. In matrix notation, we measure $y = \Phi x$, where y is an $M \times 1$ column vector and the *measurement matrix* Φ is $M \times N$ with each row a *measurement vector* ϕ_m. Since $M < N$, recovery of the signal x from the measurements y is ill-posed in general; however the additional assumption of signal *sparsity* makes recovery possible and practical.

The CS theory tells us that when certain conditions hold, namely that the basis $\{\psi_n\}$ cannot sparsely represent the vectors $\{\phi_m\}$ (a condition known as *incoherence* [24–26]) and the number of measurements M is large enough (proportional to K), then it is indeed possible to recover the set of large $\{\vartheta(n)\}$ (and thus the signal x) from the set of measurements $\{y(m)\}$ [24, 25]. This incoherence property holds for many pairs of bases, including for example, delta spikes and the sine waves of a Fourier basis, or the Fourier basis and wavelets. Signals that are sparsely represented in frames or unions of bases can be recovered from incoherent measurements in the same fashion. Significantly, this incoherence also holds with high probability between *any* arbitrary fixed basis or frame and a randomly generated one. In the sequel, we will focus our analysis to such *random measurement* procedures.

2.3 Signal recovery via ℓ_0-norm minimization

The recovery of the sparse set of significant coefficients $\{\vartheta(n)\}$ can be achieved using *optimization* by searching for the signal with the sparsest coefficient vector $\{\widehat{\vartheta}(n)\}$ that agrees with the M observed measurements in y (recall that $M < N$). Recovery relies on the key observation that, under mild conditions on Φ and Ψ, the coefficient vector ϑ is the unique solution to the ℓ_0-norm minimization

$$\widehat{\vartheta} = \arg\min \|\vartheta\|_0 \quad \text{s.t. } y = \Phi\Psi\vartheta \tag{2}$$

with overwhelming probability. (Thanks to the incoherence between the two bases, if the original signal is sparse in the ϑ coefficients, then no other set of sparse signal coefficients ϑ' can yield the same projections y.)

In principle, remarkably few incoherent measurements are required to recover a K-sparse signal via ℓ_0-norm minimization. More than K measurements must be taken to avoid ambiguity; the following theorem, proven in Appendix A, establishes that $K+1$ random measurements will suffice. Similar results were established by Venkataramani and Bresler [38].

Theorem 1 *Let Ψ be an orthonormal basis for $\mathbb{R}^N$, and let $1 \leq K < N$. Then:*

1. *Let Φ be an $M \times N$ measurement matrix with i.i.d. Gaussian entries with $M \geq 2K$. Then all signals $x = \Psi\vartheta$ having expansion coefficients $\vartheta \in \mathbb{R}^N$ that satisfy $\|\vartheta\|_0 = K$ can be recovered uniquely from the M-dimensional measurement vector $y = \Phi x$ via the ℓ_0-norm minimization (2) with probability one over Φ.*

2. *Let $x = \Psi\vartheta$ such that $\|\vartheta\|_0 = K$. Let Φ be an $M \times N$ measurement matrix with i.i.d. Gaussian entries (notably, independent of x) with $M \geq K + 1$. Then x can be recovered uniquely from the M-dimensional measurement vector $y = \Phi x$ via the ℓ_0-norm minimization (2) with probability one over Φ.*

3. *Let Φ be an $M \times N$ measurement matrix, where $M \leq K$. Then, aside from pathological cases (specified in the proof), no signal $x = \Psi\vartheta$ with $\|\vartheta\|_0 = K$ can be uniquely recovered from the M-dimensional measurement vector $y = \Phi x$.*

Remark 1 *The second statement of the theorem differs from the first in the following respect: when $K < M < 2K$, there will necessarily exist K-sparse signals x that cannot be uniquely recovered from the M-dimensional measurement vector $y = \Phi x$. However, these signals form a set of measure zero within the set of* all *K-sparse signals and can safely be avoided with high probability if Φ is randomly generated independently of x.*

Comparing the second and third statements of Theorem 1, we see that one measurement separates the *achievable region*, where perfect recovery is possible with probability one, from the *converse region*, where with overwhelming probability recovery is impossible. Moreover, Theorem 1 provides a *strong converse measurement region* in a manner analogous to the strong channel coding converse theorems of information theory [12].

Unfortunately, solving the ℓ_0-norm minimization problem is prohibitively complex, requiring a combinatorial enumeration of the $\binom{N}{K}$ possible sparse subspaces. In fact, the ℓ_0-norm minimization problem in general is known to be NP-hard [39]. Yet another challenge is robustness; in the setting of Theorem 1, the recovery may be very poorly conditioned. In fact, *both* of these considerations (computational complexity and robustness) can be addressed, but at the expense of slightly more measurements.

2.4 Signal recovery via ℓ_1-norm minimization

The practical revelation that supports the new CS theory is that it is not necessary to solve the ℓ_0-norm minimization to recover the set of significant $\{\vartheta(n)\}$. In fact, a much easier problem yields an equivalent solution (thanks again to the incoherence of the bases); we need only solve for the smallest ℓ_1-norm coefficient vector ϑ that agrees with the measurements y [24, 25]:

$$\widehat{\vartheta} = \arg\min \|\vartheta\|_1 \quad \text{s.t. } y = \Phi\Psi\vartheta. \tag{3}$$

This optimization problem, also known as *Basis Pursuit*, is significantly more approachable and can be solved with traditional linear programming techniques whose computational complexities are polynomial in N.

There is no free lunch, however; according to the theory, more than $K + 1$ measurements are required in order to recover sparse signals via Basis Pursuit. Instead, one typically requires $M \geq cK$ measurements, where $c > 1$ is an *overmeasuring factor*. As an example, we quote a result asymptotic in N. For simplicity, we assume that the sparsity scales linearly with N; that is, $K = SN$, where we call S the *sparsity rate*.

Theorem 2 [39–41] *Set $K = SN$ with $0 < S \ll 1$. Then there exists an overmeasuring factor $c(S) = O(\log(1/S))$, $c(S) > 1$, such that, for a K-sparse signal x in basis Ψ, the following statements hold:*

1. *The probability of recovering x via ℓ_1-norm minimization from $(c(S)+\epsilon)K$ random projections, $\epsilon > 0$, converges to one as $N \to \infty$.*
2. *The probability of recovering x via ℓ_1-norm minimization from $(c(S)-\epsilon)K$ random projections, $\epsilon > 0$, converges to zero as $N \to \infty$.*

In an illuminating series of papers, Donoho and Tanner [40–42] have characterized the overmeasuring factor $c(S)$ precisely. In our work, we have noticed that the overmeasuring factor is quite similar to $\log_2(1 + S^{-1})$. We find this expression a useful rule of thumb to approximate the precise overmeasuring ratio. Additional overmeasuring is proven to provide robustness to measurement noise and quantization error [25].

Throughout this paper we use the abbreviated notation c to describe the overmeasuring factor required in various settings even though $c(S)$ depends on the sparsity K and signal length N.

2.5 Signal recovery via greedy pursuit

Iterative greedy algorithms have also been developed to recover the signal x from the measurements y. The Orthogonal Matching Pursuit (OMP) algorithm, for example, iteratively selects the vectors from the matrix $\Phi\Psi$ that contain most of the energy of the measurement vector y. The selection at each iteration is made based on inner products between the columns of $\Phi\Psi$ and a residual; the residual reflects the component of y that is orthogonal to the previously selected columns. The algorithm has been proven to successfully recover the acquired signal from incoherent measurements with high probability, at the expense of slightly more measurements, [26, 43]. Algorithms inspired by OMP, such as regularized orthogonal matching pursuit [44], CoSaMP [45], and Subspace Pursuit [46] have been shown to attain similar guarantees to those of their optimization-based counterparts. In the following, we will exploit both Basis Pursuit and greedy algorithms for recovering jointly sparse signals from incoherent measurements.

2.6 Properties of random measurements

In addition to offering substantially reduced measurement rates, CS has many attractive and intriguing properties, particularly when we employ random projections at the sensors. Random measurements are *universal* in the sense that any sparse basis can be used, allowing the same encoding strategy to be applied in different sensing environments. Random measurements are also *future-proof*: if a better sparsity-inducing basis is found for the signals, then the same measurements can be used to recover a more accurate view of the environment. Random coding is also *robust*: the measurements coming from each sensor have equal priority, unlike Fourier or wavelet coefficients in current coders. Finally, random measurements allow a *progressively better recovery* of the data as more measurements are obtained; one or more measurements can also be lost without corrupting the entire recovery.

2.7 Related work

Several researchers have formulated *joint measurement settings* for CS in sensor networks that exploit inter-signal correlations [28–32]. In their approaches, each sensor $n \in \{1, 2, \ldots, N\}$ simultaneously records a single reading $x(n)$ of some spatial field (temperature at a certain time, for example).[4] Each of the sensors generates a pseudorandom sequence $r_n(m), m = 1, 2, \ldots, M$, and modulates the reading as $x(n)r_n(m)$. Each sensor n then transmits its M numbers in sequence to the collection point where the measurements are aggregated, obtaining M measurements $y(m) = \sum_{n=1}^{N} x(n)r_n(m)$. Thus, defining $x = [x(1), x(2), \ldots, x(N)]^T$ and $\phi_m = [r_1(m), r_2(m), \ldots, r_N(m)]$, the collection point automatically receives the measurement vector $y = [y(1), y(2), \ldots, y(M)]^T = \Phi x$ after M transmission steps. The samples $x(n)$ of the spatial field can then be recovered using CS provided that x has a sparse representation in a known basis. These methods have a major limitation: since they operate at a single time instant, they exploit only inter-signal and not intra-signal correlations; that is, they essentially assume that the sensor field is i.i.d. from time instant to time instant. In contrast, we will develop signal models and algorithms that are agnostic to the spatial sampling structure and that exploit both inter- *and* intra-signal correlations.

Recent work has adapted DCS to the finite rate of innovation signal acquisition framework [47] and to the continuous-time setting [48]. Since the original submission of this paper, additional work has focused on the analysis and proposal of recovery algorithms for jointly sparse signals [49, 50].

3 Joint Sparsity Signal Models

In this section, we generalize the notion of a signal being sparse in some basis to the notion of an ensemble of signals being *jointly sparse*.

3.1 Notation

We will use the following notation for signal ensembles and our measurement model. Let $\Lambda := \{1, 2, \ldots, J\}$ denote the set of indices for the J signals in the ensemble. Denote the *signals* in the ensemble by x_j, with $j \in \Lambda$ and assume that each signal $x_j \in \mathbb{R}^N$. We use $x_j(n)$ to denote sample n in signal j, and assume for the sake of illustration — but without loss of generality — that these signals are sparse in the canonical basis, i.e., $\Psi = \mathbf{I}$. The entries of the signal can take arbitrary real values.

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

[4] Note that in Section 2.7 only, N refers to the number of sensors, since each sensor acquires a signal sample.

of x_j. We will emphasize random i.i.d. Gaussian matrices Φ_j in the following, but other schemes are possible, including random ± 1 Bernoulli/Rademacher matrices, and so on.

To compactly represent the signal and measurement ensembles, we denote $\overline{M} = \sum_{j \in \Lambda} M_j$ and define $X \in \mathbb{R}^{JN}$, $Y \in \mathbb{R}^{\overline{M}}$, and $\Phi \in \mathbb{R}^{\overline{M} \times JN}$ as

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_J \end{bmatrix}, \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_J \end{bmatrix}, \qquad \text{and} \qquad \Phi = \begin{bmatrix} \Phi_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \Phi_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \Phi_J \end{bmatrix}, \tag{4}$$

with $\mathbf{0}$ denoting a matrix of appropriate size with all entries equal to 0. We then have $Y = \Phi X$. Equation (4) shows that separate measurement matrices have a characteristic block-diagonal structure when the entries of the sparse vector are grouped by signal.

Below we propose a general framework for *joint sparsity models* (JSMs) and three example JSMs that apply in different situations.

3.2 General framework for joint sparsity

We now propose a general framework to quantify the sparsity of an ensemble of correlated signals $x_1, x_2, \dots, x_J$, which allows us to compare the complexities of different signal ensembles and to quantify their measurement requirements. The framework is based on a factored representation of the signal ensemble that decouples its location and value information.

To motivate this factored representation, we begin by examining the structure of a single sparse signal, where $x \in \mathbb{R}^N$ with $K \ll N$ nonzero entries. As an alternative to the notation used in (1), we can decouple the location and value information in x by writing $x = P\theta$, where $\theta \in \mathbb{R}^K$ contains only the nonzero entries of x, and P is an *identity submatrix*, i.e., P contains K columns of the $N \times N$ identity matrix $\mathbf{I}$. Any K-sparse signal can be written in similar fashion. To model the set of all possible sparse signals, we can then 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*. Whether a signal is sufficiently sparse is defined *in the context of this model*: given a signal x, one can consider all possible factorizations $x = P\theta$ with $P \in \mathcal{P}$. Among these factorizations, the unique representation with smallest dimensionality for θ equals the *sparsity level* of the signal x under the model $\mathcal{P}$.

In the signal ensemble case, we consider factorizations of the form $X = P\Theta$ where $X \in \mathbb{R}^{JN}$ as above, $P \in \mathbb{R}^{JN \times \delta}$, and $\Theta \in \mathbb{R}^{\delta}$ for various integers δ. 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; we specify below additional conditions that the matrices P must satisfy for each model. For a given ensemble X, we let $\mathcal{P}_F(X) \subseteq \mathcal{P}$ denote the set of feasible location matrices $P \in \mathcal{P}$ for which a factorization $X = P\Theta$ exists. We define the *joint sparsity level* of the signal ensemble as follows.

Definition 1 *The* joint sparsity level *D of the signal ensemble X is the number of columns of the smallest matrix $P \in \mathcal{P}_F(X)$.*

In contrast to the single-signal case, there are several natural choices for what matrices P should be members of a joint sparsity model $\mathcal{P}$. We restrict our attention in the sequel to what we call *common/innovation component JSMs*. In these models 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

$$x_j = z_C + z_j, \quad j \in \Lambda.$$

Note, however, that the individual components might be zero-valued in specific scenarios. We can express the component signals as

$$z_C = P_C \theta_C, \quad z_j = P_j \theta_j, \quad j \in \Lambda,$$

where $\theta_C \in \mathbb{R}^{K_C}$ and each $\theta_j \in \mathbb{R}^{K_j}$ have nonzero entries. Each matrix $P \in \mathcal{P}$ that can express such signals $\{x_j\}$ has the form

$$P = \begin{bmatrix} P_C & P_1 & \mathbf{0} & \dots & \mathbf{0} \\ P_C & \mathbf{0} & P_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_C & \mathbf{0} & \mathbf{0} & \dots & P_J \end{bmatrix}, \tag{5}$$

where $P_C, \{P_j\}_{j \in \Lambda}$ are identity submatrices. We define the value vector as $\Theta = [\theta_C^T \; \theta_1^T \; \theta_2^T \; \dots \; \theta_J^T]^T$, where $\theta_C \in \mathbb{R}^{K_C}$ and each $\theta_j \in \mathbb{R}^{K_j}$, to obtain $X = P\Theta$. Although the values of K_C and K_j are dependent on the matrix P, we omit this dependency in the sequel for brevity, except when necessary for clarity.

If a signal ensemble $X = P\Theta$, $\Theta \in \mathbb{R}^{\delta}$ 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 would not be full rank. In other cases, we may observe a vector Θ that has zero-valued entries; i.e., we may have $\theta_j(k) = 0$ for some $1 \le k \le K_j$ and some $j \in \Lambda$, or $\theta_C(k) = 0$ for some $1 \le k \le K_C$. In both of these cases, by removing one instance of this column from any of the identity submatrices, one can obtain a matrix Q with fewer columns for which there exists $\Theta' \in \mathbb{R}^{\delta - 1}$ that gives $X = Q\Theta'$. If $Q \in \mathcal{P}$, then we term this phenomenon *sparsity reduction*. Sparsity reduction, when present, reduces the effective joint sparsity of a signal ensemble. As an example of sparsity reduction, consider $J = 2$ signals of length $N = 2$. Consider the coefficient $z_C(1) \neq 0$ of the common component z_C and the corresponding innovation coefficients $z_1(1), z_2(1) \neq 0$. Suppose that all other coefficients are zero. The location matrix P that arises is

$$P = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

The span of this location matrix (i.e., the set of signal ensembles X that it can generate) remains unchanged if we remove any one of the columns, i.e., if we drop any entry of the value vector Θ. This provides us with a lower-dimensional representation Θ' of the same signal ensemble X under the JSM $\mathcal{P}$; the joint sparsity of X is $D = 2$.

3.3 Example joint sparsity models

Since different real-world scenarios lead to different forms of correlation within an ensemble of sparse signals, we consider several possible designs for a JSM $\mathcal{P}$. The distinctions among our three JSMs concern the differing sparsity assumptions regarding the common and innovation components.

3.3.1 JSM-1: Sparse common component + innovations

In this model, we suppose that each signal contains a common component z_C that is *sparse* plus an innovation component z_j that is also *sparse*. Thus, this joint sparsity model (JSM-1) $\mathcal{P}$ is represented by the set of all matrices of the form (5) with K_C and all K_j smaller than N. Assuming that sparsity reduction is not possible, the joint sparsity $D = K_C + \sum_{j\in\Lambda} K_j$.

A practical situation well-modeled by this framework is a group of sensors measuring temperatures at a number of outdoor locations throughout the day. The temperature readings x_j have both temporal (intra-signal) and spatial (inter-signal) correlations. Global factors, such as the sun and prevailing winds, could have an effect z_C that is both common to all sensors and structured enough to permit sparse representation. More local factors, such as shade, water, or animals, could contribute localized innovations z_j that are also structured (and hence sparse). A similar scenario could be imagined for a network of sensors recording light intensities, air pressure, or other phenomena. All of these scenarios correspond to measuring properties of physical processes that change smoothly in time and in space and thus are highly correlated [51, 52].

3.3.2 JSM-2: Common sparse supports

In this model, the common component z_C is equal to zero, each innovation component z_j is *sparse*, and the innovations $\{z_j\}$ share the *same sparse support* but have different nonzero coefficients. To formalize this setting in a joint sparsity model (JSM-2) we let $\mathcal{P}$ represent the set of all matrices of the form (5), where $P_C = \emptyset$ and $P_j = \overline{P}$ for all $j \in \Lambda$. Here $\overline{P}$ denotes an arbitrary identity submatrix of size $N \times K$, with $K \ll N$. For a given $X = P\Theta$, we may again partition the value vector $\Theta = [\theta_1^T\ \theta_2^T\ \ldots\ \theta_J^T]^T$, where each $\theta_j \in \mathbb{R}^K$. It is easy to see that the matrices P from JSM-2 are full rank. Therefore, when sparsity reduction is not possible, the joint sparsity $D = JK$.

The JSM-2 model is immediately applicable to acoustic and RF sensor arrays, where each sensor acquires a replica of the same Fourier-sparse signal but with phase shifts and attenuations caused by signal propagation. In this case, it is critical to recover each one of the sensed signals. Another useful application for this framework is MIMO communication [53].

Similar signal models have been considered in the area of *simultaneous sparse approximation* [53–55]. In this setting, a collection of sparse signals share the same expansion vectors from a redundant dictionary. The sparse approximation can be recovered via greedy algorithms such as *Simultaneous Orthogonal Matching Pursuit* (SOMP) [53, 54] or *MMV Order Recursive Matching Pursuit* (M-ORMP) [55]. We use the SOMP algorithm in our setting (Section 5.2) to recover from incoherent measurements an ensemble of signals sharing a common sparse structure.

3.3.3 JSM-3: Nonsparse common component + sparse innovations

In this model, we suppose that each signal contains an arbitrary common component z_C and a sparse innovation component z_j; this model extends JSM-1 by relaxing the assumption that the common component z_C has a sparse representation. To formalize this setting in the JSM-3 model, we let $\mathcal{P}$ represent the set of all matrices (5) in which $P_C = I$, the $N \times N$ identity matrix. This implies each K_j is smaller than N while $K_C = N$; thus, we obtain $\theta_C \in \mathbb{R}^N$ and $\theta_j \in \mathbb{R}^{K_j}$. Assuming that sparsity reduction is not possible, the joint sparsity $D = N + \sum_{j\in\Lambda} K_j$. We also consider the specific case where the supports of the innovations are shared by all signals, which extends JSM-2; in this case we will have $P_j = \overline{P}$ for all $j \in \Lambda$, with $\overline{P}$ an identity submatrix of size $N \times K$. It is easy to see that in this case sparsity reduction is possible, and so the the joint sparsity can drop to $D = N + (J-1)K$. Note that *separate* CS recovery is impossible in JSM-3 with any fewer than N measurements per sensor, since the common component is not sparse. However, we will demonstrate that *joint* CS recovery can indeed exploit the common structure.

A practical situation well-modeled by this framework is where several sources are recorded by different sensors together with a background signal that is not sparse in any basis. Consider, for example, a verification system in a component production plant, where cameras acquire snapshots of each component to check for manufacturing defects. While each image could be extremely complicated, and hence nonsparse, the ensemble of images will be highly correlated, since each camera is observing the same device with minor (sparse) variations.

JSM-3 can also be applied in non-distributed scenarios. For example, it motivates the compression of data such as video, where the innovations or differences between video frames may be sparse, even though a single frame may not be very sparse. In this case, JSM-3 suggests that we encode each video frame separately using CS and then decode all frames of the video sequence jointly. This has the advantage of moving the bulk of the computational complexity to the video decoder. The PRISM system proposes a similar scheme based on Wyner-Ziv distributed encoding [56].

There are many possible joint sparsity models beyond those introduced above, as well as beyond the common and innovation component signal model. Further work will yield new JSMs suitable for other application scenarios; an example application consists of multiple cameras taking digital photos of a common scene from various angles [57]. Extensions are discussed in Section 6.

4 Theoretical Bounds on Measurement Rates

In this section, we seek conditions on $\mathcal{M} = (M_1, M_2, \ldots, M_J)$, the tuple of number of measurements from each sensor, such that we can guarantee perfect recovery of X given Y. To this end, we provide a graphical model for the general framework provided in Section 3.2. This graphical model is fundamental in the derivation of the number of measurements needed for each sensor, as well as in the formulation of a combinatorial recovery procedure. Thus, we generalize Theorem 1 to the distributed setting to obtain fundamental limits on the number of measurements that enable recovery of sparse signal ensembles.

Based on the models presented in Section 3, recovering X requires determining a value vector Θ and location matrix P such that $X = P\Theta$. Two challenges immediately present themselves. First, a given measurement depends only on some of the components of Θ, and the measurement budget should be adjusted between the sensors according to the information that can be gathered on the components of Θ. For example, if a component $\Theta(d)$ does not affect any signal coefficient $x_j(\cdot)$ in sensor j, then the corresponding measurements y_j provide no information about $\Theta(d)$. Second, the decoder must identify a location matrix $P \in \mathcal{P}_F(X)$ from the set $\mathcal{P}$ and the measurements Y.

4.1 Modeling dependencies using bipartite graphs

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 a full-rank matrix $P \in \mathcal{P}_F(X)$ and the corresponding Θ; the matrix P defines the sparsities of the common and innovation components K_C and K_j, $1 \leq j \leq J$, as well as the joint sparsity $D = K_C + \sum_{j=1}^{J} K_j$. Define the following sets of vertices: (i) the set of *value vertices* V_V has elements with indices $d \in \{1, \ldots, D\}$ representing the entries of the value vector $\Theta(d)$, and (ii) 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 for these sets are $|V_V| = D$ and $|V_M| = \overline{M}$, respectively.

We now introduce a bipartite graph $G = (V_V, V_M, E)$, that represents the relationships between the entries of the value vector and the measurements (see [4] for details). The set of edges E is defined as follows:

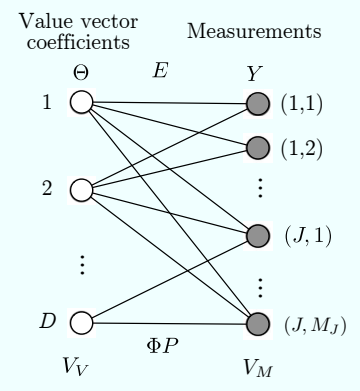

Figure 1: *Bipartite graph for distributed compressive sensing (DCS). The bipartite graph* $G = (V_V, V_M, E)$ *indicates the relationship between the value vector coefficients and the measurements.*

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

In words, we say that $y_j(m)$, the m^{th} measurement of sensor j, *measures* $\Theta(d)$ if the vertex $d \in V_V$ is linked to the vertex $(j, m) \in V_M$ in the graph G. An example graph for a distributed sensing setting is shown in Figure 1.

4.2 Quantifying redundancies

In order to obtain sharp bounds on the number of measurements needed, our analysis of the measurement process must account for redundancies between the locations of the nonzero coefficients in the common and innovation components. To that end, we consider the overlaps between common and innovation components in each signal. When we have $z_c(n) \neq 0$ and $z_j(n) \neq 0$ for a certain signal j and some index $1 \leq n \leq N$, we cannot recover the values of both coefficients from the measurements of this signal alone; therefore, we will need to recover $z_c(n)$ using measurements of other signals that do not feature the same overlap. 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, as described in Section 3.2.

Definition 2 *The* overlap size *for the set of signals* $\Gamma \subset \Lambda$*, denoted* $K_C(\Gamma, P)$*, is the number of indices in which there is overlap between the common and the innovation component supports at all signals* $j \notin \Gamma$*:*

$$K_C(\Gamma, P) = \left|\{n \in \{1, \ldots, N\} : z_C(n) \neq 0 \text{ and } \forall\ j \notin \Gamma, z_j(n) \neq 0\}\right|. \tag{6}$$

We also define $K_C(\Lambda, P) = K_C(P)$ *and* $K_C(\emptyset, P) = 0$.

For $\Gamma \subset \Lambda$, $K_C(\Gamma, P)$ provides a penalty term due to the need for recovery of common component coefficients that are overlapped by innovations in all other signals $j \notin \Gamma$. Intuitively, for each entry counted in $K_C(\Gamma, P)$, some sensor in Γ must take one measurement to account for that entry of the common component — it is impossible to recover such entries from measurements made by sensors outside of Γ. When all signals $j \in \Lambda$ are considered, it is clear that all of the common component coefficients must be recovered from the obtained measurements.

4.3 Measurement bounds

Converse and achievable bounds for the number of measurements necessary for DCS recovery are given below. Our bounds consider each subset of sensors $\Gamma \subseteq \Lambda$, since the cost of sensing the common component can be amortized across sensors: it may be possible to reduce the rate at one sensor $j_1 \in \Gamma$ (up to a point), as long as other sensors in Γ offset the rate reduction. We quantify the reduction possible through the following definition.

Definition 3 *The* conditional sparsity *of the set of signals Γ is the number of entries of the vector Θ that must be recovered by measurements y_j, $j \in \Gamma$:*

$$K_{\text{cond}}(\Gamma, P) = \left(\sum_{j \in \Gamma} K_j(P) \right) + K_C(\Gamma, P).$$

The joint sparsity gives the number of degrees of freedom for the signals in Λ, while the conditional sparsity gives the number of degrees of freedom for signals in Γ when the signals in $\Lambda \setminus \Gamma$ are available as side information. Note also that Definition 1 for joint sparsity can be extended to a subset of signals Γ by considering the number of entries of Θ that affect these signals:

$$K_{\text{joint}}(\Gamma, P) = D - K_{\text{cond}}(\Lambda - \Gamma, P) = \left(\sum_{j \in \Gamma} K_j(P) \right) + K_C(P) - K_C(\Lambda \setminus \Gamma, P).$$

Note that $K_{\text{cond}}(\Lambda, P) = K_{\text{joint}}(\Lambda, P) = D$.

The bipartite graph introduced in Section 4.1 is the cornerstone of Theorems 3, 4, and 5, which consider whether a perfect matching can be found in the graph; see the proofs in Appendices B, D, and E, respectively, for detail.

Theorem 3 (Achievable, known P) *Assume that a signal ensemble X is obtained from a common/innovation component JSM $\mathcal{P}$. Let $\mathcal{M} = (M_1, M_2, \ldots, M_J)$ be a measurement tuple, let $\{\Phi_j\}_{j \in \Lambda}$ be random matrices having M_j rows of i.i.d. Gaussian entries for each $j \in \Lambda$, and write $Y = \Phi X$. Suppose there exists a full rank location matrix $P \in \mathcal{P}_F(X)$ such that*

$$\sum_{j \in \Gamma} M_j \geq K_{\text{cond}}(\Gamma, P) \tag{7}$$

for all $\Gamma \subseteq \Lambda$. Then with probability one over $\{\Phi_j\}_{j \in \Gamma}$, there exists a unique solution $\widehat{\Theta}$ to the system of equations $Y = \Phi P \widehat{\Theta}$; hence, the signal ensemble X can be uniquely recovered as $X = P\widehat{\Theta}$.

Theorem 4 (Achievable, unknown P) *Assume that a signal ensemble X and measurement matrices $\{\Phi_j\}_{j \in \Lambda}$ follow the assumptions of Theorem 3. Suppose there exists a full rank location matrix $P^* \in \mathcal{P}_F(X)$ such that*

$$\sum_{j \in \Gamma} M_j \geq K_{\text{cond}}(\Gamma, P^*) + |\Gamma| \tag{8}$$

for all $\Gamma \subseteq \Lambda$. Then X can be uniquely recovered from Y with probability one over $\{\Phi_j\}_{j \in \Gamma}$.

Theorem 5 (Converse) *Assume that a signal ensemble X and measurement matrices $\{\Phi_j\}_{j \in \Lambda}$ follow the assumptions of Theorem 3. Suppose there exists a full rank location matrix $P \in \mathcal{P}_F(X)$ such that*

$$\sum_{j \in \Gamma} M_j < K_{\text{cond}}(\Gamma, P) \tag{9}$$

for some $\Gamma \subseteq \Lambda$*. Then there exists a solution* $\widehat{\Theta}$ *such that* $Y = \Phi P \widehat{\Theta}$ *but* $\widehat{X} := P\widehat{\Theta} \neq X$*.*

The identification of a feasible location matrix P causes the one measurement per sensor gap that prevents (8)–(9) from being a tight converse and achievable bound pair. We note in passing that the signal recovery procedure used in Theorem 4 is akin to ℓ_0-norm minimization on X; see Appendix D for details.

4.4 Discussion

The bounds in Theorems 3–5 are dependent on the dimensionality of the subspaces in which the signals reside. The number of noiseless measurements required for ensemble recovery is determined by the dimensionality $\dim(\mathcal{S})$ of the subspace $\mathcal{S}$ in the relevant signal model, because dimensionality and sparsity play a volumetric role akin to the entropy H used to characterize rates in source coding. Whereas in source coding each bit resolves between two options, and 2^{NH} typical inputs are described using NH bits [12], in CS we have $M = \dim(\mathcal{S}) + O(1)$. Similar to Slepian-Wolf coding [13], the number of measurements required for each sensor must account for the minimal features unique to that sensor, while at the same time features that appear among multiple sensors must be amortized over the group.

Theorems 3–5 can also be applied to the single sensor and joint measurement settings. In the single-signal setting (Theorem 1), we will have $x = P\theta$ with $\theta \in \mathbb{R}^K$, and $\Lambda = \{1\}$; Theorem 4 provides the requirement $M \geq K + 1$. It is easy to show that the joint measurement is equivalent to the single-signal setting: we stack all the individual signals into a single signal vector, and in both cases all measurements are dependent on all the entries of the signal vector. However, the distribution of the measurements among the available sensors is irrelevant in a joint measurement setting. Therefore, we only obtain a necessary condition $\sum_j M_j \geq D + 1$ on the total number of measurements required.

5 Practical Recovery Algorithms and Experiments

Although we have provided a unifying theoretical treatment for the three JSM models, the nuances warrant further study. In particular, while Theorem 4 highlights the basic tradeoffs that must be made in partitioning the measurement budget among sensors, the result does not by design provide insight into tractable algorithms for signal recovery. We believe there is additional insight to be gained by considering each model in turn, and while the presentation may be less unified, we attribute this to the fundamental diversity of problems that can arise under the umbrella of jointly sparse signal representations. In this section, we focus on tractable recovery algorithms for each model and, when possible, analyze the corresponding measurement requirements.

5.1 Recovery strategies for sparse common + innovations (JSM-1)

We first characterize the sparse common signal and innovations model JSM-1 from Section 3.3.1. For simplicity, we limit our description to $J = 2$ signals, but describe extensions to multiple signals as needed.

5.1.1 Measurement bounds for joint recovery

Under the JSM-1 model, separate recovery of the signal x_j via ℓ_0-norm minimization would require $K_{\text{joint}}(\{j\}) + 1 = K_{\text{cond}}(\{j\}) + 1 = K_C + K_j - K_C(\Lambda \setminus \{j\}) + 1$ measurements, where $K_C(\Lambda \setminus \{j\})$ accounts for sparsity reduction due to overlap between z_C and z_j. We apply Theorem 4 to the JSM-1 model to obtain the corollary below. To address the possibility of sparsity reduction, we denote

by K_R the number of indices in which the common component z_C and all innovation components z_j, $j \in \Lambda$ overlap; this results in sparsity reduction for the common component.

Corollary 1 *Assume the measurement matrices $\{\Phi_j\}_{j\in\Lambda}$ contain i.i.d. Gaussian entries. Then the signal ensemble X can be recovered with probability one if the following conditions hold:*

$$\begin{aligned}\sum_{j\in\Gamma} M_j &\geq \left(\sum_{j\in\Gamma} K_j\right) + K_C(\Gamma) + |\Gamma|, \quad \Gamma \neq \Lambda,\\ \sum_{j\in\Lambda} M_j &\geq K_C + \left(\sum_{j\in\Lambda} K_j\right) + J - K_R.\end{aligned}$$

Our joint recovery scheme provides a significant savings in measurements, because the common component can be measured as part of any of the J signals.

5.1.2 Stochastic signal model for JSM-1

To give ourselves a firm footing for analysis, in the remainder of Section 5.1 we use a stochastic process for JSM-1 signal generation. This framework provides an information theoretic setting where we can scale the size of the problem and investigate which measurement rates enable recovery. We generate the common and innovation components as follows. For $n \in \{1, \ldots, N\}$ the decision whether $z_C(n)$ and $z_j(n)$ is zero or not is an i.i.d. Bernoulli process, where the probability of a nonzero value is given by parameters denoted S_C and S_j, respectively. The values of the nonzero coefficients are then generated from an i.i.d. Gaussian distribution. The outcome of this process is that z_C and z_j have sparsities $K_C \sim \text{Binomial}(N, S_C)$ and $K_j \sim \text{Binomial}(N, S_j)$. The parameters S_j and S_C are *sparsity rates* controlling the random generation of each signal. Our model resembles the Gaussian spike process [58], which is a limiting case of a Gaussian mixture model.

Likelihood of sparsity reduction and overlap: This stochastic model can yield signal ensembles for which the corresponding generating matrices P allow for sparsity reduction; specifically, there might be overlap between the supports of the common component z_C and all the innovation components z_j, $j \in \Lambda$. For $J = 2$, the probability that a given index is present in all supports is $S_R := S_C S_1 S_2$. Therefore, the distribution of the cardinality of this overlap is $K_R \sim \text{Binomial}(N, S_R)$. We must account for the reduction obtained from the removal of the corresponding number of columns from the location matrix P when the total number of measurements $M_1 + M_2$ is considered. In the same way we can show that the distributions for the number of indices in the overlaps required by Corollary 1 are $K_C(\{1\}) \sim \text{Binomial}(N, S_{C,\{1\}})$ and $K_C(\{2\}) \sim \text{Binomial}(N, S_{C,\{2\}})$, where $S_{C,\{1\}} := S_C(1 - S_1)S_2$ and $S_{C,\{2\}} := S_C S_1(1 - S_2)$.

Measurement rate region: To characterize DCS recovery performance, we introduce a *measurement rate region*. We define the measurement rate R_j in an asymptotic manner as

$$R_j := \lim_{N\to\infty} \frac{M_j}{N}, \ j \in \Lambda.$$

Additionally, we note that

$$\lim_{N\to\infty} \frac{K_C}{N} = S_C \text{ and } \lim_{N\to\infty} \frac{K_j}{N} = S_j, \ j \in \Lambda$$

Thus, we also set $S_{X_j} = S_C + S_j - S_C S_j$, $j \in \{1, 2\}$. For a measurement rate pair (R_1, R_2) and sources X_1 and X_2, we evaluate whether we can recover the signals with vanishing probability of error as N increases. In this case, we say that the measurement rate pair is *achievable*.

For jointly sparse signals under JSM-1, separate recovery via ℓ_0-norm minimization would require a measurement rate $R_j = S_{X_j}$. Separate recovery via ℓ_1-norm minimization would require an overmeasuring factor $c(S_{X_j})$, and thus the measurement rate would become $R_j = S_{X_j} \cdot c(S_{X_j})$. To improve upon these figures, we adapt the standard machinery of CS to the joint recovery problem.

5.1.3 Joint recovery via ℓ_1-norm minimization

As discussed in Section 2.3, solving an ℓ_0-norm minimization is NP-hard, and so in practice we must relax our ℓ_0 criterion in order to make the solution tractable. We now study what penalty must be paid for ℓ_1-norm recovery of jointly sparse signals. Using the vector and frame

$$Z := \begin{bmatrix} z_C \\ z_1 \\ z_2 \end{bmatrix} \quad \text{and} \quad \widetilde{\Phi} := \begin{bmatrix} \Phi_1 & \Phi_1 & 0 \\ \Phi_2 & 0 & \Phi_2 \end{bmatrix}, \tag{11}$$

we can represent the concatenated measurement vector Y sparsely using the concatenated coefficient vector Z, which contains $K_C + K_1 + K_2 - K_R$ nonzero coefficients, to obtain $Y = \widetilde{\Phi} Z$. With sufficient overmeasuring, we have seen experimentally that it is possible to recover a vector $\widehat{Z}$, which yields $x_j = \widehat{z}_C + \widehat{z}_j$, $j = 1, 2$, by solving the weighted ℓ_1-norm minimization

$$\widehat{Z} = \arg\min \gamma_C ||z_C||_1 + \gamma_1 ||z_1||_1 + \gamma_2 ||z_2||_1 \quad \text{s.t. } y = \widetilde{\Phi} Z, \tag{12}$$

where $\gamma_C, \gamma_1, \gamma_2 \geq 0$. We call this the *$\gamma$-weighted ℓ_1-norm formulation*; our numerical results (Section 5.1.6 and our technical report [59]) indicate a reduction in the requisite number of measurements via this enhancement. If $K_1 = K_2$ and $M_1 = M_2$, then without loss of generality we set $\gamma_1 = \gamma_2 = 1$ and numerically search for the best parameter γ_C. We discuss the asymmetric case with $K_1 = K_2$ and $M_1 \neq M_2$ in the technical report [59].

5.1.4 Converse bound on performance of γ-weighted ℓ_1-norm minimization

We now provide a converse bound that describes what measurement rate pairs *cannot* be achieved via the γ-weighted ℓ_1-norm minimization. Our notion of a converse focuses on the setting where each signal x_j is measured via multiplication by the M_j by N matrix Φ_j and joint recovery is performed via our γ-weighted ℓ_1-norm formulation (12). Within this setting, a converse region is a set of measurement rates for which the recovery fails with overwhelming probability as N increases.

We assume that $J = 2$ sources have innovation sparsity rates that satisfy $S_1 = S_2 = S_I$. Our first result, proved in Appendix F, provides deterministic necessary conditions to recover the components z_C, z_1, and z_2, using the γ-weighted ℓ_1-norm formulation (12). We note that the lemma holds for all such combinations of components that generate the same signals $x_1 = z_C + z_1$ and $x_2 = z_C + z_2$.

Lemma 1 *Consider any γ_C, γ_1, and γ_2 in the γ-weighted ℓ_1-norm formulation (12). The components z_C, z_1, and z_2 can be recovered using measurement matrices Φ_1 and Φ_2 only if (i) z_1 can be recovered via ℓ_1-norm minimization (3) using Φ_1 and measurements $\Phi_1 z_1$; (ii) z_2 can be recovered via ℓ_1-norm minimization using Φ_2 and measurements $\Phi_2 z_2$; and (iii) z_C can be recovered via ℓ_1-norm minimization using the joint matrix $[\Phi_1^T \;\; \Phi_2^T]^T$ and measurements $[\Phi_1^T \;\; \Phi_2^T]^T z_C$.*

Lemma 1 can be interpreted as follows. If M_1 and M_2 are not large enough individually, then the innovation components z_1 and z_2 cannot be recovered. This implies a converse bound on the individual measurement rates R_1 and R_2. Similarly, combining Lemma 1 with the converse bound

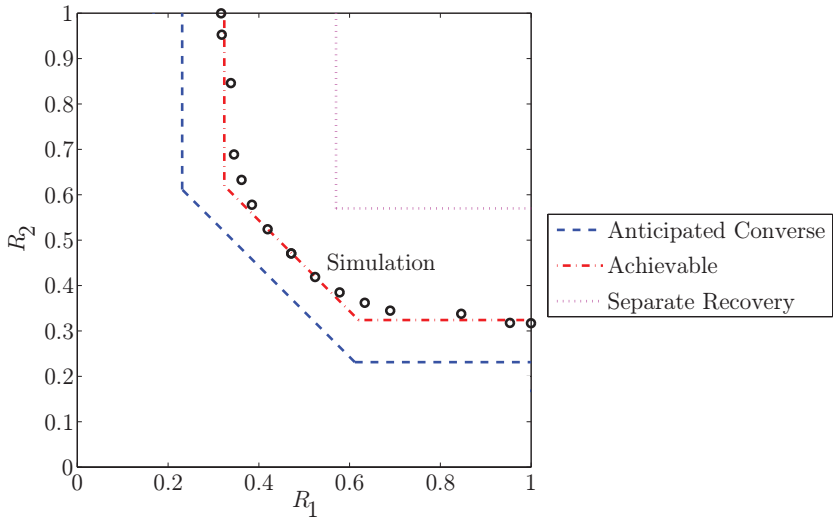

Figure 2: *Recovering a signal ensemble with sparse common + innovations (JSM-1). We chose a common component sparsity rate $S_C = 0.2$ and innovation sparsity rates $S_I = S_1 = S_2 = 0.05$. Our simulation results use the γ-weighted ℓ_1-norm formulation (12) on signals of length $N = 1000$; the measurement rate pairs that achieved perfect recovery over 100 simulations are denoted by circles.*

of Theorem 2 for single-source ℓ_1-norm minimization of the common component z_C implies a lower bound on the sum measurement rate $R_1 + R_2$.

Anticipated converse: As shown in Corollary 1, for indices n such that $x_1(n)$ and $x_2(n)$ differ and are nonzero, each sensor must take measurements to account for one of the two coefficients. In the case where $S_1 = S_2 = S_I$, the joint sparsity rate is $S_C + 2S_I - S_C S_I^2$. We define the measurement function $c'(S) := S \cdot c(S)$ based on Donoho and Tanner's oversampling factor $c(S)$ (Theorem 2). It can be shown that the function $c'(\cdot)$ is concave; in order to minimize the sum rate bound, we "explain" as many of the sparse coefficients in one of the signals and as few as possible in the other. From Corollary 1, we have $R_1, R_2 \geq S_I + S_C S_I - S_C S_I^2$. Consequently, one of the signals must "explain" this sparsity rate, whereas the other signal must explain the rest:

$$[S_C + 2S_I - S_C S_I^2] - [S_I + S_C S_I - S_C S_I^2] = S_C + S_I - S_C S_I.$$

Unfortunately, the derivation of $c'(S)$ relies on Gaussianity of the measurement matrix, whereas in our case Φ has a block matrix form. Therefore, the following conjecture remains to be proved rigorously.

Conjecture 1 *Let $J = 2$ and fix the sparsity rate of the common component S_C and the innovation sparsity rates $S_1 = S_2 = S_I$. Then the following conditions on the measurement rates are necessary to enable recovery with probability one:*

$$\begin{aligned} R_j &\geq c'\left(S_I + S_C S_I - S_C S_I^2\right), \ j = 1, 2, \\ R_1 + R_2 &\geq c'\left(S_I + S_C S_I - S_C S_I^2\right) + c'\left(S_C + S_I - S_C S_I\right). \end{aligned}$$

5.1.5 Achievable bound on performance of ℓ_1-norm minimization

We have not yet characterized the performance of γ-weighted ℓ_1-norm formulation (12) analytically. Instead, Theorem 6 below uses an alternative ℓ_1-norm based recovery technique. The proof describes a constructive recovery algorithm. We construct measurement matrices Φ_1 and Φ_2, each consisting

of two parts. The first parts of the matrices are identical and recover $x_1 - x_2$. The second parts of the matrices are different and enable the recovery of $\frac{1}{2}x_1 + \frac{1}{2}x_2$. Once these two components have been recovered, the computation of x_1 and x_2 is straightforward. The measurement rate can be computed by considering both identical and different parts of the measurement matrices.

Theorem 6 *Let $J = 2$, $N \to \infty$ and fix the sparsity rate of the common component S_C and the innovation sparsity rates $S_1 = S_2 = S_I$. If the measurement rates satisfy the following conditions:*

$$R_j > c'(2S_I - S_I^2),\ j = 1, 2, \tag{13a}$$

$$R_1 + R_2 > c'(2S_I - S_I^2) + c'(S_C + 2S_I - 2S_C S_I - S_I^2 + S_C S_I^2), \tag{13b}$$

then we can design measurement matrices Φ_1 and Φ_2 with random Gaussian entries and an ℓ_1-norm minimization recovery algorithm that succeeds with probability approaching one as N increases. Furthermore, as $S_I \to 0$ the sum measurement rate approaches $c'(S_C)$.

The theorem is proved in Appendix G. The recovery algorithm of Theorem 6 is based on linear programming. It can be extended from $J = 2$ to an arbitrary number of signals by recovering all signal differences of the form $x_{j_1} - x_{j_2}$ in the first stage of the algorithm and then recovering $\frac{1}{J}\sum_j x_j$ in the second stage. In contrast, our γ-weighted ℓ_1-norm formulation (12) recovers a length-JN signal. Our simulation experiments (Section 5.1.6) indicate that the γ-weighted formulation can recover using fewer measurements than the approach of Theorem 6.

The achievable measurement rate region of Theorem 6 is loose with respect to the region of the anticipated converse Conjecture 1 (see Figure 2). We leave for future work the characterization of a tight measurement rate region for computationally tractable (polynomial time) recovery techniques.

5.1.6 Simulations for JSM-1

We now present simulation results for several different JSM-1 settings. The γ-weighted ℓ_1-norm formulation (12) was used throughout, where the optimal choice of γ_C, γ_1, and γ_2 depends on the relative sparsities K_C, K_1, and K_2. The optimal values have not been determined analytically. Instead, we rely on a numerical optimization, which is computationally intense. A detailed discussion of our intuition behind the choice of γ appears in the technical report [59].

Recovering two signals with symmetric measurement rates: Our simulation setting is as follows. The signal components z_C, z_1, and z_2 are assumed (without loss of generality) to be sparse in $\Psi = I_N$ with sparsities K_C, K_1, and K_2, respectively. We assign random Gaussian values to the nonzero coefficients. We restrict our attention to the symmetric setting in which $K_1 = K_2$ and $M_1 = M_2$, and consider signals of length $N = 50$ where $K_C + K_1 + K_2 = 15$.

In our joint decoding simulations, we consider values of M_1 and M_2 in the range between 10 and 40. We find the optimal γ_C in the γ-weighted ℓ_1-norm formulation (12) using a line search optimization, where simulation indicates the "goodness" of specific γ_C values in terms of the likelihood of recovery. With the optimal γ_C, for each set of values we run several thousand trials to determine the empirical probability of success in decoding z_1 and z_2. The results of the simulation are summarized in Figure 3. The savings in the number of measurements M can be substantial, especially when the common component K_C is large (Figure 3). For $K_C = 11$, $K_1 = K_2 = 2$, M is reduced by approximately 30%. For smaller K_C, joint decoding barely outperforms separate decoding, since most of the measurements are expended on innovation components. Additional results appear in [59].

Recovering two signals with asymmetric measurement rates: In Figure 2, we compare separate CS recovery with the anticipated converse bound of Conjecture 1, the achievable bound of Theorem 6, and numerical results.

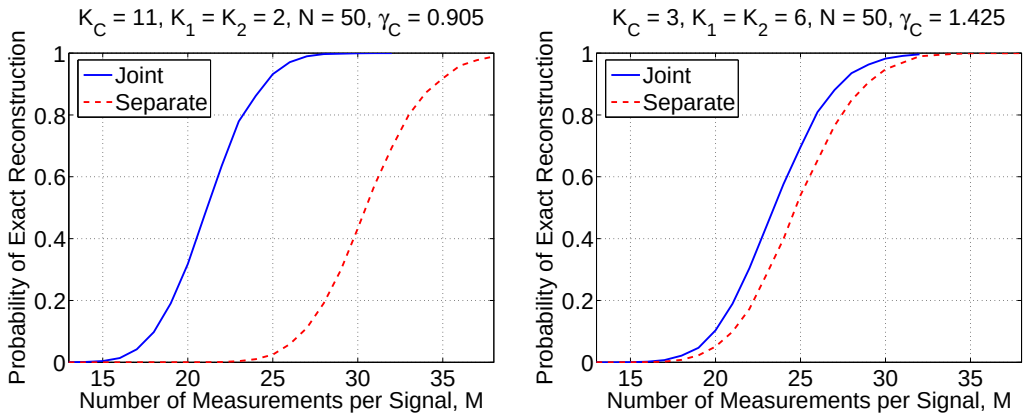

Figure 3: *Comparison of joint decoding and separate decoding for JSM-1. The advantage of joint over separate decoding depends on the common component sparsity.*

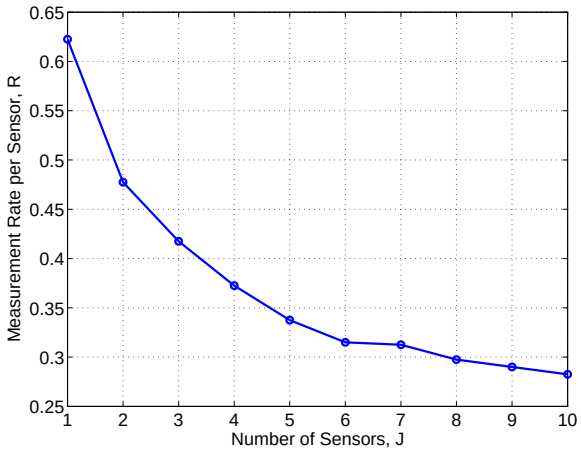

Figure 4: *Multi-sensor measurement results for JSM-1. We choose a common component sparsity rate $S_C = 0.2$, innovation sparsity rates $S_I = 0.05$, and signals of length $N = 500$; our results demonstrate a reduction in the measurement rate per sensor as the number of sensors J increases.*

We use $J = 2$ signals and choose a common component sparsity rate $S_C = 0.2$ and innovation sparsity rates $S_I = S_1 = S_2 = 0.05$. We consider several different asymmetric measurement rates. In each such setting, we constrain M_2 to have the form $M_2 = \alpha M_1$ for some α, with $N = 1000$. The results plotted indicate the smallest pairs (M_1, M_2) for which we always succeeded recovering the signal over 100 simulation runs. In some areas of the measurement rate region our γ-weighted ℓ_1-norm formulation (12) requires fewer measurements than the achievable approach of Theorem 6.

Recovering multiple signals with symmetric measurement rates: The γ-weighted ℓ_1-norm recovery technique of this section is especially promising when $J > 2$ sensors are used. These savings may be valuable in applications such as sensor networks, where data may contain strong spatial (inter-source) correlations.

We use $J \in \{1, 2, \ldots, 10\}$ signals and choose the same sparsity rates $S_C = 0.2$ and $S_I = 0.05$ as the asymmetric rate simulations; here we use symmetric measurement rates and let $N = 500$. The results of Figure 4 describe the smallest symmetric measurement rates for which we always succeeded recovering the signal over 100 simulation runs. As J increases, lower measurement rates can be used; the results compare favorably with the lower bound from Conjecture 1, which gives $R_j \approx 0.232$ as $J \to \infty$.

5.2 Recovery strategies for common sparse supports (JSM-2)

Under the JSM-2 signal ensemble model from Section 3.3.2, separate recovery of each signal via ℓ_0-norm minimization would require $K+1$ measurements per signal, while separate recovery via ℓ_1-norm minimization would require cK measurements per signal. When Theorems 4 and 5 are applied in the context of JSM-2, the bounds for joint recovery match those of individual recovery using ℓ_0-norm minimization. Within this context, it is also possible to recover one of the signals using $K+1$ measurements from the corresponding sensor, and then with the prior knowledge of the support set Ω, recover all other signals from K measurements per sensor; thus providing an additional savings of $J-1$ measurements [60]. Surprisingly, we will demonstrate below that for large J, the common support set can actually be recovered using only one measurement per sensor and algorithms that are computationally tractable.

The algorithms we propose are inspired by conventional greedy pursuit algorithms for CS (such as OMP [26]). In the single-signal case, OMP iteratively constructs the sparse support set Ω; decisions are based on inner products between the columns of Φ and a residual. In the multi-signal case, there are more clues available for determining the elements of Ω.

5.2.1 Recovery via Trivial Pursuit (TP)

When there are many correlated signals in the ensemble, a simple non-iterative greedy algorithm based on inner products will suffice to recover the signals jointly. For simplicity but without loss of generality, we assume that an equal number of measurements $M_j = M$ are taken of each signal. We write Φ_j in terms of its columns, with $\Phi_j = [\phi_{j,1}, \phi_{j,2}, \ldots, \phi_{j,N}]$.

Trivial Pursuit (TP) Algorithm for JSM-2

1. **Get greedy:** Given all of the measurements, compute the test statistics

$$\xi_n = \frac{1}{J}\sum_{j=1}^{J} \langle y_j, \phi_{j,n}\rangle^2, \qquad n \in \{1, 2, \ldots, N\}, \tag{14}$$

and estimate the elements of the common coefficient support set by

$$\widehat{\Omega} = \{n \text{ having one of the } K \text{ largest } \xi_n\}.$$

When the sparse, nonzero coefficients are sufficiently generic (as defined below), we have the following surprising result, which is proved in Appendix H.

Theorem 7 *Let Ψ be an orthonormal basis for $\mathbb{R}^N$, let the measurement matrices Φ_j contain i.i.d. Gaussian entries, and assume that the nonzero coefficients in the θ_j are i.i.d. Gaussian random variables. Then with $M \geq 1$ measurements per signal, TP recovers Ω with probability approaching one as $J \to \infty$.*

In words, with *fewer* than K measurements per sensor, it is actually possible to recover the sparse support set Ω under the JSM-2 model.[5] Of course, this approach does not recover the K coefficient values for each signal; at least K measurements per sensor are required for this.

Corollary 2 *Assume that the nonzero coefficients in the θ_j are i.i.d. Gaussian random variables. Then the following statements hold:*

[5] One can also show the somewhat stronger result that, as long as $\sum_j M_j \gg N$, TP recovers Ω with probability approaching one. We have omitted this additional result for brevity.

1. *Let the measurement matrices Φ_j contain i.i.d. Gaussian entries, with each matrix having an overmeasuring factor of $c = 1$ (that is, $M_j = K$ for each measurement matrix Φ_j). Then TP recovers all signals from the ensemble $\{x_j\}$ with probability approaching one as $J \to \infty$.*

2. *Let Φ_j be a measurement matrix with overmeasuring factor $c < 1$ (that is, $M_j < K$), for some $j \in \Lambda$. Then with probability one, the signal x_j cannot be uniquely recovered by any algorithm for any value of J.*

The first statement is an immediate corollary of Theorem 7; the second statement follows because each equation $y_j = \Phi_j x_j$ would be underdetermined even if the nonzero indices were known. Thus, under the JSM-2 model, the TP algorithm asymptotically performs as well as an oracle decoder that has prior knowledge of the locations of the sparse coefficients. From an information theoretic perspective, Corollary 2 provides tight achievable and converse bounds for JSM-2 signals. We should note that the theorems in this section have a slightly different flavor than Theorem 4 and 5, which ensure recovery of *any* sparse signal ensemble, given a suitable set of measurement matrices. Theorem 7 and Corollary 2 above, in contrast, rely on a random signal model and do not guarantee simultaneous performance for all sparse signals under any particular measurement ensemble. Nonetheless, we feel this result is worth presenting to highlight the strong subspace concentration behavior that enables the correct identification of the common support.

In the technical reports [59, 61], we derive an approximate formula for the probability of error in recovering the common support set Ω given J, K, M, and N. While theoretically interesting and potentially practically useful, these results require J to be large. Our numerical experiments show that the number of measurements required for recovery using TP decreases quickly as J increases. However, in the case of small J, TP performs poorly. Hence, we propose next an alternative recovery technique based on simultaneous greedy pursuit that performs well for small J.

5.2.2 Recovery via iterative greedy pursuit

In practice, the common sparse support among the J signals enables a fast iterative algorithm to recover all of the signals jointly. Tropp and Gilbert have proposed one such algorithm, called *Simultaneous Orthogonal Matching Pursuit* (SOMP) [53], which can be readily applied in our DCS framework. SOMP is a variant of OMP that seeks to identify Ω one element at a time. A similar simultaneous sparse approximation algorithm has been proposed using convex optimization [62]. We dub the DCS-tailored SOMP algorithm DCS-SOMP.

To adapt the original SOMP algorithm to our setting, we first extend it to cover a different measurement matrix Φ_j for each signal x_j. Then, in each DCS-SOMP iteration, we select the column index $n \in \{1, 2, \ldots, N\}$ that accounts for the greatest amount of residual energy across *all* signals. As in SOMP, we orthogonalize the remaining columns (in each measurement matrix) after each step; after convergence we obtain an expansion of the measurement vector y_j on an orthogonalized subset of the columns of basis vectors. To obtain the expansion coefficients in the sparse basis, we then reverse the orthogonalization process using the QR matrix factorization. Finally, we again assume that $M_j = M$ measurements per signal are taken.

DCS-SOMP Algorithm for JSM-2

1. **Initialize:** Set the iteration counter $\ell = 1$. For each signal index $j \in \Lambda$, initialize the orthogonalized coefficient vectors $\widehat{\beta}_j = 0$, $\widehat{\beta}_j \in \mathbb{R}^M$; also initialize the set of selected indices $\widehat{\Omega} = \emptyset$. Let $r_{j,\ell}$ denote the residual of the measurement y_j remaining after the first ℓ iterations, and initialize $r_{j,0} = y_j$.

2. **Select** the dictionary vector that maximizes the value of the sum of the magnitudes of the projections of the residual, and add its index to the set of selected indices

$$\begin{aligned} n_\ell &= \arg\max_{n\in\{1,\ldots,N\}} \sum_{j=1}^{J} \frac{|\langle r_{j,\ell-1}, \phi_{j,n}\rangle|}{\|\phi_{j,n}\|_2}, \\ \widehat{\Omega} &= [\widehat{\Omega} \;\; n_\ell]. \end{aligned}$$

3. **Orthogonalize** the selected basis vector against the orthogonalized set of previously selected dictionary vectors

$$\gamma_{j,\ell} = \phi_{j,n_\ell} - \sum_{t=0}^{\ell-1} \frac{\langle \phi_{j,n_\ell}, \gamma_{j,t}\rangle}{\|\gamma_{j,t}\|_2^2}\gamma_{j,t}.$$

4. **Iterate:** Update the estimate of the coefficients for the selected vector and residuals

$$\begin{aligned} \widehat{\beta}_j(\ell) &= \frac{\langle r_{j,\ell-1}, \gamma_{j,\ell}\rangle}{\|\gamma_{j,\ell}\|_2^2}, \\ r_{j,\ell} &= r_{j,\ell-1} - \frac{\langle r_{j,\ell-1}, \gamma_{j,\ell}\rangle}{\|\gamma_{j,\ell}\|_2^2}\gamma_{j,\ell}. \end{aligned}$$

5. **Check for convergence:** If $\|r_{j,\ell}\|_2 > \epsilon\|y_j\|_2$ for all j, then increment ℓ and go to Step 2; otherwise, continue to Step 6. The parameter ϵ determines the target error power level allowed for algorithm convergence.

6. **De-orthogonalize:** Consider the relationship between $\Gamma_j = [\gamma_{j,1}, \gamma_{j,2}, \ldots, \gamma_{j,M}]$ and the Φ_j given by the QR factorization $\Phi_{j,\widehat{\Omega}} = \Gamma_j R_j$, where $\Phi_{j,\widehat{\Omega}} = [\phi_{j,n_1}, \phi_{j,n_2}, \ldots, \phi_{j,n_M}]$ is the so-called *mutilated basis*.[6] Since $y_j = \Gamma_j\beta_j = \Phi_{j,\widehat{\Omega}}x_{j,\widehat{\Omega}} = \Gamma_j R_j x_{j,\widehat{\Omega}}$, where $x_{j,\widehat{\Omega}}$ is the mutilated coefficient vector, we can compute the signal estimates $\{\widehat{x}_j\}$ as

$$\widehat{x}_{j,\widehat{\Omega}} = R_j^{-1}\widehat{\beta}_j,$$

where $\widehat{x}_{j,\widehat{\Omega}}$ is the mutilated version of the sparse coefficient vector $\widehat{x}_j$.

In practice, we obtain $\widehat{c}K$ measurements from each signal x_j for some value of $\widehat{c}$. We then use DCS-SOMP to recover the J signals jointly. We orthogonalize because as the number of iterations approaches M the norms of the residues of an orthogonal pursuit decrease faster than for a non-orthogonal pursuit; indeed, due to Step 3 the algorithm can only run for up to M iterations. The computational complexity of this algorithm is $O(JNM^2)$, which matches that of separate recovery for each signal while reducing the required number of measurements.

Thanks to the common sparsity structure among the signals, we believe (but have not proved) that DCS-SOMP will succeed with $\widehat{c} < c(S)$. Empirically, we have observed that a small number of measurements proportional to K suffices for a moderate number of sensors J. Based on our observations, described in Section 5.2.3, we conjecture that $K+1$ measurements per sensor suffice as $J \to \infty$. Thus, this efficient greedy algorithm enables an overmeasuring factor $\widehat{c} = (K+1)/K$ that approaches 1 as J, K, and N increase.

[6] We define a *mutilated basis* Φ_Ω as a subset of the basis vectors from $\Phi = [\phi_1, \phi_2, \ldots, \phi_N]$ corresponding to the indices given by the set $\Omega = \{n_1, n_2, \ldots, n_M\}$, that is, $\Phi_\Omega = [\phi_{n_1}, \phi_{n_2}, \ldots, \phi_{n_M}]$. This concept can be extended to vectors in the same manner.

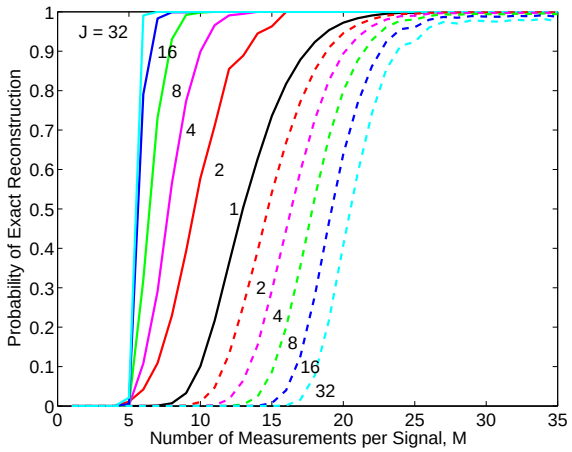

Figure 5: *Recovering a signal ensemble with common sparse supports (JSM-2). We plot the probability of perfect recovery via DCS-SOMP (solid lines) and separate CS recovery (dashed lines) as a function of the number of measurements per signal M and the number of signals J. We fix the signal length to $N = 50$, the sparsity to $K = 5$, and average over 1000 simulation runs. An oracle encoder that knows the positions of the large signal expansion coefficients would use 5 measurements per signal.*

5.2.3 Simulations for JSM-2

We now present simulations comparing separate CS recovery versus joint DCS-SOMP recovery for a JSM-2 signal ensemble. Figure 5 plots the probability of perfect recovery corresponding to various numbers of measurements M as the number of sensors varies from $J = 1$ to 32, over 1000 trials in each case. We fix the signal lengths at $N = 50$ and the sparsity of each signal to $K = 5$.

With DCS-SOMP, for perfect recovery of all signals the average number of measurements per signal decreases as a function of J. The trend suggests that for large J close to K measurements per signal should suffice. On the contrary, with separate CS recovery, for perfect recovery of all signals the number of measurements per sensor *increases* as a function of J. This occurs because each signal experiences an independent probability $p \leq 1$ of successful recovery; therefore the overall probability of complete success is p^J. Consequently, each sensor must compensate by making additional measurements. This phenomenon further motivates joint recovery under JSM-2.

Finally, we note that we can use algorithms other than DCS-SOMP to recover the signals under the JSM-2 model. Cotter et al. [55] have proposed additional algorithms (such as M-FOCUSS) that iteratively eliminate basis vectors from the dictionary and converge to the set of sparse basis vectors over which the signals are supported. We hope to extend such algorithms to JSM-2 in future work.

5.3 Recovery strategies for nonsparse common component + sparse innovations (JSM-3)

The JSM-3 signal ensemble model from Section 3.3.3 provides a particularly compelling motivation for joint recovery. Under this model, no individual signal x_j is sparse, and so recovery of each signal separately would require fully N measurements per signal. As in the other JSMs, however, the commonality among the signals makes it possible to substantially reduce this number. Again, the potential for this savings is evidenced by specializing Theorem 4 to the context of JSM-3.

Corollary 3 *If Φ_j is a random Gaussian matrix for all $j \in \Lambda$, $\mathcal{P}$ is defined by JSM-3, and*

$$\sum_{j\in\Gamma} M_j \geq \sum_{j\in\Gamma} K_j + K_C(\Gamma, P) + |\Gamma|, \quad \Gamma \in \Lambda, \tag{15}$$

$$\sum_{j\in\Lambda} M_j \geq \sum_{j\in\Lambda} K_j + N + J - K_R, \tag{16}$$

then the signal ensemble X can be uniquely recovered from Y with probability one.

This suggests that the number of measurements of an individual signal can be substantially decreased, as long as the total number of measurements is sufficiently large to capture enough information about the nonsparse common component z_C. The term K_R denotes the number of indices where the common and all innovation components overlap, and appears due to the sparsity reduction that can be performed at the common component before recovery. We also note that when the supports of the innovations are independent, as $J \to \infty$, it becomes increasingly unlikely that a given index will be included in all innovations, and thus the terms $K_C(\Gamma, P)$ and K_R will go to zero. On the other hand, when the supports are completely matched (implying $K_j = K$, $j \in \Lambda$), we will have $K_R = K$, and after sparsity reduction has been addressed, $K_C(\Gamma, P) = 0$ for all $\Gamma \subseteq \Lambda$.

5.3.1 Recovery via Transpose Estimation of Common Component (TECC)

Successful recovery of the signal ensemble $\{x_j\}$ requires recovery of both the nonsparse common component z_C and the sparse innovations $\{z_j\}$. To help build intuition about how we might accomplish signal recovery using far fewer than N measurements per sensor, consider the following thought experiment.

If z_C were known, then each innovation z_j could be estimated using the standard single-signal CS machinery on the adjusted measurements $y_j - \Phi_j z_C = \Phi_j z_j$. While z_C is not known in advance, it can be *estimated* from the measurements. In fact, across all J sensors, a total of $\sum_{j\in\Lambda} M_j$ random projections of z_C are observed (each corrupted by a contribution from one of the z_j). Since z_C is not sparse, it cannot be recovered via CS techniques, but when the number of measurements is sufficiently large ($\sum_{j\in\Lambda} M_j \gg N$), z_C can be estimated using standard tools from linear algebra. A key requirement for such a method to succeed in recovering z_C is that each Φ_j be different, so that their rows combine to span all of $\mathbb{R}^N$. In the limit (again, assuming the sparse innovation coefficients are well-behaved), the common component z_C can be recovered while still allowing each sensor to operate at the minimum measurement rate dictated by the $\{z_j\}$. A prototype algorithm is listed below, where we assume that each measurement matrix Φ_j has i.i.d. $\mathcal{N}(0, \sigma_j^2)$ entries.

TECC Algorithm for JSM-3

1. **Estimate common component:** Define the matrix $\widehat{\Phi}$ as the vertical concatenation of the regularized individual measurement matrices $\widehat{\Phi}_j = \frac{1}{M_j \sigma_j^2}\Phi_j$, that is, $\widehat{\Phi} = [\widehat{\Phi}_1^T, \widehat{\Phi}_2^T, \ldots, \widehat{\Phi}_J^T]^T$. Calculate the estimate of the common component as $\widehat{z}_C = \frac{1}{J}\widehat{\Phi}^T Y$.

2. **Estimate measurements generated by innovations:** Using the previous estimate, subtract the contribution of the common part from the measurements and generate estimates for the measurements caused by the innovations for each signal: $\widehat{y}_j = y_j - \Phi_j \widehat{z}_C$.

3. **Recover innovations:** Using a standard single-signal CS recovery algorithm,[7] obtain estimates of the innovations $\widehat{z}_j$ from the estimated innovation measurements $\widehat{y}_j$.

[7]For tractable analysis of the TECC algorithm, the proof of Theorem 8 employs a least-squares variant of ℓ_0-norm minimization.

4. **Obtain signal estimates:** Estimate each signal as the sum of the common and innovations estimates; that is, $\widehat{x}_j = \widehat{z}_C + \widehat{z}_j$.

The following theorem, proved in Appendix I, shows that asymptotically, by using the TECC algorithm, each sensor needs to only measure at the rate dictated by the sparsity K_j.

Theorem 8 *Assume that the nonzero expansion coefficients of the sparse innovations z_j are i.i.d. Gaussian random variables and that their locations are uniformly distributed on $\{1, 2, \ldots, N\}$. Let the measurement matrices Φ_j contain i.i.d. $\mathcal{N}(0, \sigma_j^2)$ entries with $M_j \geq K_j + 1$. Then each signal x_j can be recovered using the TECC algorithm with probability approaching one as $J \to \infty$.*

For large J, the measurement rates permitted by Theorem 8 are the best possible for *any* recovery strategy for JSM-3 signals, even neglecting the presence of the nonsparse component. These rates meet the minimum bounds suggested by Corollary 3, although again Theorem 8 is of a slightly different flavor, as it does not provide a uniform guarantee for all sparse signal ensembles under any particular measurement matrix collection. The CS technique employed in Theorem 8 involves combinatorial searches that estimate the innovation components; we have provided the theorem simply as support for our intuitive development of the TECC algorithm. More efficient techniques could also be employed (including several proposed for CS in the presence of noise [25, 63, 64]). It is reasonable to expect similar behavior; as the error in estimating the common component diminishes, these techniques should perform similarly to their noiseless analogues.

5.3.2 Recovery via Alternating Common and Innovation Estimation (ACIE)

The preceding analysis demonstrates that the number of required measurements in JSM-3 can be substantially reduced through joint recovery. While Theorem 8 shows theoretical gains as $J \to \infty$, practical gains can also be realized with a moderate number of sensors. In particular, suppose in the TECC algorithm that the initial estimate $\widehat{z}_C$ is not accurate enough to enable correct identification of the sparse innovation supports $\{\Omega_j\}$. In such a case, it may still be possible for a rough approximation of the innovations $\{z_j\}$ to help refine the estimate $\widehat{z}_C$. This in turn could help to refine the estimates of the innovations.

The Alternating Common and Innovation Estimation (ACIE) algorithm exploits the observation that once the basis vectors comprising the innovation z_j have been identified in the index set Ω_j, their effect on the measurements y_j can be removed to aid in estimating z_C. Suppose that we have an estimate for these innovation basis vectors in $\widehat{\Omega}_j$. We can then partition the measurements into two parts: the projection into $\text{span}(\{\phi_{j,n}\}_{n \in \widehat{\Omega}_j})$ and the component orthogonal to that span. We build a basis for the $\mathbb{R}^{M_j}$ where y_j lives:

$$B_j = [\Phi_{j,\widehat{\Omega}_j} \; Q_j],$$

where $\Phi_{j,\widehat{\Omega}_j}$ is the mutilated matrix Φ_j corresponding to the indices in $\widehat{\Omega}_j$, and the $M_j \times (M_j - |\widehat{\Omega}_j|)$ matrix Q_j has orthonormal columns that span the orthogonal complement of $\Phi_{j,\widehat{\Omega}_j}$.

This construction allows us to remove the projection of the measurements into the aforementioned span to obtain measurements caused exclusively by vectors not in $\widehat{\Omega}_j$:

$$\widetilde{y}_j = Q_j^T y_j \text{ and } \widetilde{\Phi}_j = Q_j^T \Phi_j. \tag{17}$$

These modifications enable the sparse decomposition of the measurement, which now lives in $\mathbb{R}^{M_j - |\widehat{\Omega}_j|}$, to remain unchanged:

$$\widetilde{y_j} = \sum_{n=1}^{N} \alpha_j \widetilde{\phi}_{j,n}.$$

Thus, the modified measurements $\widetilde{Y} = \left[\widetilde{y}_1^T \ \widetilde{y}_2^T \ \ldots \ \widetilde{y}_J^T\right]^T$ and modified measurement matrix $\widetilde{\Phi} = \left[\widetilde{\Phi}_1^T \ \widetilde{\Phi}_2^T \ \ldots \ \widetilde{\Phi}_J^T\right]^T$ can be used to refine the estimate of the common component of the signal,

$$\widetilde{z_C} = \widetilde{\Phi}^\dagger \widetilde{Y}, \tag{18}$$

where $A^\dagger = (A^T A)^{-1} A^T$ denotes the pseudoinverse of matrix A.

When the innovation support estimate is correct ($\widehat{\Omega}_j = \Omega_j$), the measurements $\widetilde{y_j}$ will describe only the common component z_C. If this is true for every signal j and the number of remaining measurements $\sum_{j=1}^{J}(M_j - K_j) \geq N$, then z_C can be perfectly recovered via (18). However, it may be difficult to obtain correct estimates for all signal supports in the first iteration of the algorithm, and so we find it preferable to refine the estimate of the support by executing several iterations.

ACIE Algorithm for JSM-3

1. **Initialize:** Set $\widehat{\Omega}_j = \emptyset$ for each j. Set the iteration counter $\ell = 1$.
2. **Estimate common component:** Update estimate $\widetilde{z_C}$ according to (17)–(18).
3. **Estimate innovation supports:** For each sensor j, after subtracting the contribution $\widetilde{z_C}$ from the measurements, $\widehat{y}_j = y_j - \Phi_j \widetilde{z_C}$, estimate the support of each signal innovation $\widehat{\Omega}_j$.
4. **Iterate:** If $\ell < L$, a preset number of iterations, then increment ℓ and return to Step 2. Otherwise proceed to Step 5.
5. **Estimate innovation coefficients:** For each j, estimate the coefficients for the indices in $\widehat{\Omega}_j$,
$$\widehat{z}_{j,\widehat{\Omega}_j} = \Phi^\dagger_{j,\widehat{\Omega}_j}(y_j - \Phi_j \widetilde{z_C}),$$
where $\widehat{z}_{j,\widehat{\Omega}_j}$ is a mutilated version of the innovation's sparse coefficient vector estimate $\widehat{z}_j$.
6. **Recover signals:** Compute the estimate of each signal as $\widehat{x}_j = \widetilde{z_C} + \widehat{z}_j$.

Estimation of the supports in Step 3 can be accomplished using a variety of techniques. We propose to run a fixed number of iterations of OMP; if the supports of the innovations are known to match across signals — as in JSM-2 — then more powerful algorithms like SOMP can be used. The ACIE algorithm is similar in spirit to other iterative estimation algorithms, such as turbo decoding [65].

5.3.3 Simulations for JSM-3

We now present simulations of JSM-3 recovery for the following scenario. Consider J signals of length $N = 50$ containing a common white noise component $z_C(n) \sim \mathcal{N}(0, 1)$ for $n \in \{1, \ldots, N\}$. Each innovations component z_j has sparsity $K = 5$ (once again in the time domain), resulting in $x_j = z_C + z_j$. The signals are generated according to the model used in Section 5.1.6.

We study two different cases. The first is an extension of JSM-1: we select the supports for the various innovations separately and then apply OMP to each signal in Step 3 of the ACIE algorithm in order to estimate its innovations component. The second case is an extension of JSM-2: we select one common support for all of the innovations across the signals and then apply the DCS-SOMP algorithm (Section 5.2.2) to estimate the innovations in Step 3. In both cases we use $L = 10$

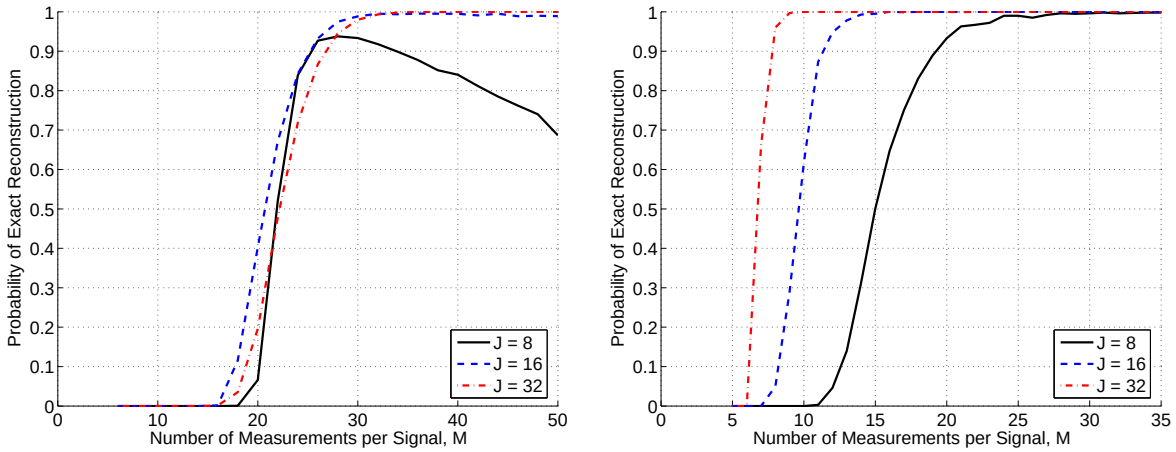

Figure 6: *Recovering a signal ensemble with nonsparse common component and sparse innovations (JSM-3) using ACIE. (a) recovery using OMP separately on each signal in Step 3 of the ACIE algorithm (innovations have arbitrary supports). (b) recovery using DCS-SOMP jointly on all signals in Step 3 of the ACIE algorithm (innovations have identical supports). Signal length* $N = 50$*, sparsity* $K = 5$*. The common structure exploited by DCS-SOMP enables dramatic savings in the number of measurements. We average over 1000 simulation runs.*

iterations of ACIE. We test the algorithms for different numbers of signals J and calculate the probability of correct recovery as a function of the (same) number of measurements per signal M.

Figure 6(a) shows that, for sufficiently large J, we can recover all of the signals with significantly fewer than N measurements per signal. As J grows, it becomes more difficult to perfectly recover all J signals. We believe this is inevitable, because even if z_C were known without error, then perfect ensemble recovery would require the successful execution of J *independent* runs of OMP. Second, for small J, the probability of success can decrease at high values of M. We believe this behavior is due to the fact that initial errors in estimating z_C may tend to be somewhat sparse (since $\widehat{z}_C$ roughly becomes an average of the signals $\{x_j\}$), and these sparse errors can mislead the subsequent OMP processes. For more moderate M, it seems that the errors in estimating z_C (though greater) tend to be less sparse. We expect that a more sophisticated algorithm could alleviate such a problem; the problem is also mitigated at higher J.

Figure 6(b) shows that when the sparse innovations share common supports we see an even greater savings. As a point of reference, a traditional approach to signal acquisition would require 1600 total measurements to recover these $J = 32$ nonsparse signals of length $N = 50$. Our approach requires only approximately 10 random measurements per sensor — a total of 320 measurements — for high probability of recovery.

6 Discussion and Conclusions

In this paper we have extended the theory and practice of compressive sensing to multi-signal, distributed settings. The number of noiseless measurements required for ensemble recovery is determined by the dimensionality of the subspace in the relevant signal model, because dimensionality and sparsity play a volumetric role akin to the entropy used to characterize rates in source coding. Our three example joint sparsity models (JSMs) for signal ensembles with both intra- and inter-signal correlations capture the essence of real physical scenarios, illustrate the basic analysis and algorithmic techniques, and indicate the significant gains to be realized from joint recovery. In some sense, distributed compressive sensing (DCS) is a framework for distributed compression of

sources with memory, which has remained a challenging problem for some time.

In addition to offering substantially reduced measurement rates, the DCS-based distributed source coding schemes we develop here share the properties of CS mentioned in Section 2. Two additional properties of DCS make it well-matched to distributed applications such as sensor networks and arrays [51, 52]. First, each sensor encodes its measurements separately, which eliminates the need for inter-sensor communication. Second, DCS distributes its computational complexity asymmetrically, placing most of it in the joint decoder, which will often have more computational resources than any individual sensor node. The encoders are very simple; they merely compute incoherent projections with their signals and make no decisions.

There are many opportunities for applications and extensions of these ideas. First, natural signals are not exactly sparse but rather can be better modeled as ℓ_p-compressible with $0 < p \leq 1$. Roughly speaking, a signal in a *weak*-ℓ_p ball has coefficients that decay as $n^{-1/p}$ once sorted according to magnitude [24]. The key concept is that the ordering of these coefficients is important. For JSM-2, we can extend the notion of simultaneous sparsity for ℓ_p-sparse signals whose sorted coefficients obey roughly the same ordering [66]. This condition could perhaps be enforced as an ℓ_p constraint on the composite signal

$$\left\{ \sum_{j=1}^{J} |x_j(1)|, \quad \sum_{j=1}^{J} |x_j(2)|, \quad \ldots, \quad \sum_{j=1}^{J} |x_j(N)| \right\}.$$

Second, (random) measurements are real numbers; quantization gradually degrades the recovery quality as the quantization becomes coarser [34, 67, 68]. Moreover, in many practical situations some amount of measurement noise will corrupt the $\{x_j\}$, making them not exactly sparse in any basis. While characterizing these effects and the resulting rate-distortion consequences in the DCS setting are topics for future work, there has been work in the single-signal CS literature that we should be able to leverage, including variants of Basis Pursuit with Denoising [63, 69], robust iterative recovery algorithms [64], CS noise sensitivity analysis [25, 34], the Dantzig Selector [33], and one-bit CS [70].

Third, in some applications, the linear program associated with some DCS decoders (in JSM-1 and JSM-3) could prove too computationally intense. As we saw in JSM-2, efficient iterative and greedy algorithms could come to the rescue, but these need to be extended to the multi-signal case. Recent results on recovery from a union of subspaces give promise for efficient, model-based algorithms [66].

Finally, we focused our theory on models that assign common and innovation components to the signals in the ensemble. Other models tailored to specific applications can be posed; for example, in hyperspectral imaging applications, it is common to obtain strong correlations only across spectral slices within a certain neighborhood. It would be then appropriate to pose a common/innovation model with separate common components that are localized to a subset of the spectral slices obtained. Results similar to those obtained in Section 4 are simple to derive for models with full-rank location matrices.

A Proof of Theorem 1

Statement 2 is an application of the achievable bound of Theorem 4 to the case of $J = 1$ signal. It remains then to prove Statements 1 and 3.

Statement 1 (Achievable, $M \geq 2K$): We first note that, if $K \geq N/2$, then with probability one, the matrix Φ has rank N, and there is a unique (correct) recovery. Thus we assume that $K < N/2$. With probability one, all subsets of up to $2K$ columns drawn from Φ are linearly independent. Assuming this holds, then for two index sets $\Omega \neq \widehat{\Omega}$ such that $|\Omega| = |\widehat{\Omega}| = K$, $\text{colspan}(\Phi_\Omega) \cap \text{colspan}(\Phi_{\widehat{\Omega}})$ has dimension equal to the number of indices common to both Ω and $\widehat{\Omega}$. A signal projects to this common space only if its coefficients are nonzero on exactly these (fewer than K) common indices; since $\|\theta\|_0 = K$, this does not occur. Thus every K-sparse signal projects to a unique point in $\mathbb{R}^M$.

Statement 3 (Converse, $M \leq K$): If $M < K$, there is insufficient information in the vector y to recover the K nonzero coefficients of θ; thus we assume $M = K$. In this case, there is a single explanation for the measurements only if there is a single set Ω of K linearly independent columns *and* the nonzero indices of θ are the elements of Ω. Aside from this pathological case, the rank of subsets $\Phi_{\widehat{\Omega}}$ will generally be less than K — which would prevent robust recovery of signals supported on $\widehat{\Omega}$, or will be equal to K — which would give ambiguous solutions among all such sets $\widehat{\Omega}$. □

B Proof of Theorem 3

We let

$$D := K_C + \sum_{j \in \Lambda} K_j \tag{19}$$

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

$$\Upsilon := \Phi P$$

has rank D, 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 (5), 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. To prove that Υ has rank D, we will require the following lemma, which we prove in Appendix C.

Lemma 2 *If (7) holds, then there exists a mapping $\mathcal{C} : \{1, 2, \ldots, K_C\} \to \Lambda$, assigning each element of the common component to one of the sensors, such that for each $\Gamma \subseteq \Lambda$,*

$$\sum_{j \in \Gamma} M_j \geq \sum_{j \in \Gamma} K_j + \sum_{k=1}^{K_C} 1_{\mathcal{C}(k) \in \Gamma} \tag{20}$$

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

Intuitively, the existence of such a mapping suggests that (i) each sensor has taken enough measurements to cover its own innovation (requiring K_j 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 D.

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 D columns (19). To argue that Υ has rank D, 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 2 this cannot happen if $\mathcal{C}(k) = j$), 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 next 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,

$$\operatorname{rank}(\Upsilon_0) = \operatorname{rank}(\Upsilon). \tag{21}$$

We now let Υ_1 be the matrix obtained from Υ_0 using the following construction. For each $j \in \Lambda$, we select $K_j + \sum_{k=1}^{K_C} 1_{\mathcal{C}(k)=j}$ arbitrary rows from the portion of Υ_0 corresponding to sensor j. Using (19), the resulting matrix Υ_1 has

$$\sum_{j \in \Lambda} \left(K_j + \sum_{k=1}^{K_C} 1_{\mathcal{C}(k)=j} \right) = \sum_{j \in \Lambda} K_j + K_C = D$$

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

$$\operatorname{rank}(\Upsilon_1) \leq \operatorname{rank}(\Upsilon_0). \tag{22}$$

We now let Υ_2 be the $D \times D$ 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 $\mathcal{C}(k) = j$, 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 $\sum_{k=1}^{K_C} 1_{\mathcal{C}(k)=j}$ 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 K_j 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

$$\text{rank}(\Upsilon_2) = \text{rank}(\Upsilon_1). \tag{23}$$

Based on its dependency on Υ_0, and following from Lemma 2, 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, $K_j + \sum_{k=1}^{K_C} 1_{\mathcal{C}(k)=j}$ rows of Υ_1 are assigned in its construction, while $K_j + \sum_{k=1}^{K_C} 1_{\mathcal{C}(k)=j}$ 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 $\mathcal{C}(k) = j$. In the latter case, we know by Lemma 2 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 3 *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 3 inductively D times, the success probability remains one. It follows that with probability one over Φ, $\text{rank}(\Upsilon_2) = D$. Combining this last result with (21-23), we obtain $\text{rank}(\Upsilon) = D$ with probability one over Φ. It remains to prove Lemma 3.

Proof of Lemma 3: 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. □

C Proof of Lemma 2

To prove this lemma, we apply tools from graph theory.

We seek a matching within the graph $G = (V_V, V_M, E)$ from Figure 1, i.e., a subgraph $(V_V, V_M, \overline{E})$ with $\overline{E} \subseteq E$ that pairs each element of V_V with a unique element of V_M. Such a matching will immediately give us the desired mapping $\mathcal{C}$ 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 E, and we set $\mathcal{C}(k) = j$.

To prove the existence of such a matching within the graph, we invoke a version of Hall's marriage theorem for bipartite graphs [71]. 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 $E(\Pi)$ of neighbors of Π in V_M has size $|E(\Pi)| \geq |\Pi|$. We will prove that if (7) 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 $E(\Pi)$ denote the set of neighbors of Π in V_M joined by edges in E, and we let $S_\Pi = \{j \in \Lambda : (j, m) \in E(\Pi) \text{ for some } m\}$. Thus, $S_\Pi \subseteq \Lambda$ denotes the set of signal indices whose measurement nodes have edges that connect to Π. It follows that $|E(\Pi)| = \sum_{j \in S_\Pi} M_j$. Thus, in order to satisfy Hall's condition for Π, we require

$$\sum_{j \in S_\Pi} M_j \geq |\Pi|. \tag{24}$$

We would now like to show that $\sum_{j \in S_\Pi} K_j + K_C(S_\Pi, P) \geq |\Pi|$, and thus if (7) is satisfied for all $\Gamma \subseteq \Lambda$, then (24) 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, $|\Pi_I| = \sum_{j \in S_\Pi} K_j$ because we count 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, $\sum_{j \in S_\Pi} K_j + K_C(S_\Pi, P) \geq |\Pi_I| + |\Pi_C| = |\Pi|$, and so (7) implies (24) for any Π, and so Hall's condition is satisfied, and a matching exists. Because in such a 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 (20) holds for each $\Gamma \subseteq \Lambda$. □

D Proof of Theorem 4

Given the measurements Y and measurement matrix Φ, we will 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 $\overline{Y}$ and a matrix $\overline{\Phi}$.

- For each matrix $P \in \mathcal{P}$:
  - choose a single solution Θ_P to $\overline{Y} = \overline{\Phi} P \Theta_P$ 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 $\overline{\Phi}$. 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 $\overline{Y} = \overline{\Phi} P^* \Theta_{P^*}$ yields $X = P^* \Theta_{P^*}$ when (8) holds for P^*, as shown in Theorem 3. 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 2 *Consider the algorithm used in the proof: if the matrices in $\mathcal{P}$ are sorted by number of columns, then the algorithm is akin to ℓ_0-norm minimization on Θ with an additional cross-validation step. The ℓ_0-norm minimization algorithm is known to be optimal for recovery of strictly sparse signals from noiseless measurements.*

E Proof of Theorem 5

We let D denote the number of columns in P. Because $P \in \mathcal{P}_F(X)$, there exists $\Theta \in \mathbb{R}^D$ 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 $\operatorname{rank}(\Upsilon) < D$, 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 B. Again, because Υ_0 was constructed only by subtracting columns of Υ from one another, it follows that $\operatorname{rank}(\Upsilon_0) = \operatorname{rank}(\Upsilon)$.

Suppose $\Gamma \subseteq \Lambda$ is a set for which (9) 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 $k \in \{K_C + 1, K_C + 2, \ldots, D\}$ 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 $\sum_{j \in \Gamma} K_j$ such columns k.

This submatrix has $\sum_{j \in \Gamma} K_j + K_C(\Gamma, P)$ columns. Because Υ_0 has the same size as Υ, and in particular has only D columns, then in order to have that $\operatorname{rank}(\Upsilon_0) = D$, it is necessary that all $\sum_{j \in \Gamma} K_j + K_C(\Gamma, P)$ 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, 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 (9) and have that $\text{rank}(\Upsilon_1) = \text{rank}(\Upsilon_2) \leq \sum_{j\in\Gamma} M_j < \sum_{j\in\Gamma} K_j + K_C(\Gamma, P)$. Thus, all $\sum_{j\in\Gamma} K_j + K_C(\Gamma, P)$ columns of Υ_1 cannot be linearly independent, and so Υ does not have full rank. □

F Proof of Lemma 1

Necessary conditions on innovation components: We begin by proving that in order to recover z_C, z_1, and z_2 via the γ-weighted ℓ_1-norm formulation it is necessary that z_1 can be recovered via single-signal ℓ_1-norm minimization using Φ_1 and measurements $\bar{y_1} = \Phi_1 z_1$.

Consider the single-signal ℓ_1-norm minimization problem

$$\bar{z_1} = \arg\min \|z_1\|_1 \quad \text{s.t. } \bar{y_1} = \Phi_1 z_1.$$

Suppose that this ℓ_1-norm minimization for z_1 fails; that is, there exists $\bar{z_1} \neq z_1$ such that $\bar{y_1} = \Phi_1 \bar{z_1}$ and $\|\bar{z_1}\|_1 \leq \|z_1\|_1$. Therefore, substituting $\bar{z_1}$ instead of z_1 in the γ-weighted ℓ_1-norm formulation (12) provides an alternate explanation for the measurements with a smaller or equal modified ℓ_1-norm penalty. Consequently, recovery of z_1 using (12) will fail and we will recover x_1 incorrectly. We conclude that the single-signal ℓ_1-norm minimization of z_1 using Φ_1 is necessary for successful recovery using the γ-weighted ℓ_1-norm formulation. A similar condition for ℓ_1-norm minimization of z_2 using Φ_2 and measurements $\Phi_2 z_2$ can be proved in an analogous manner.

Necessary condition on common component: We now prove that in order to recover z_C, z_1, and z_2 via the γ-weighted ℓ_1-norm formulation it is necessary that z_C can be recovered via single-signal ℓ_1-norm minimization using the joint matrix $[\Phi_1^T \;\; \Phi_2^T]^T$ and measurements $[\Phi_1^T \;\; \Phi_2^T]^T z_C$.

The proof is very similar to the previous proof for the innovation component z_1. Consider the single-signal ℓ_1-norm minimization

$$\bar{z}_C = \arg\min \|z_C\|_1 \quad \text{s.t. } \bar{y}_C = [\Phi_1^T \;\; \Phi_2^T]^T z_C.$$

Suppose that this ℓ_1-norm minimization for z_C fails; that is, there exists $\bar{z}_C \neq z_C$ such that $\bar{y}_C = [\Phi_1^T \;\; \Phi_2^T]^T \bar{z}_C$ and $\|\bar{z}_C\|_1 \leq \|z_C\|_1$. Therefore, substituting $\bar{z}_C$ instead of z_C in the γ-weighted ℓ_1-norm formulation (12) provides an alternate explanation for the measurements with a smaller modified ℓ_1-norm penalty. Consequently, the recovery of z_C using the γ-weighted ℓ_1-norm formulation (12) will fail, and thus we will recover x_1 and x_2 incorrectly. We conclude that the single-signal ℓ_1-norm minimization of z_C using $[\Phi_1^T \;\; \Phi_2^T]^T$ is necessary for successful recovery using the γ-weighted ℓ_1-norm formulation. □

G Proof of Theorem 6

We construct measurement matrices Φ_1 and Φ_2 that consist of two sets of rows. The first set of rows is identical in both and recovers the signal *difference* $x_1 - x_2$. The second set is different and recovers the signal *average* $\frac{1}{2}x_1 + \frac{1}{2}x_2$. Let the submatrix formed by the identical rows for the signal difference be Φ_D, and let the submatrices formed by unique rows for the signal average be $\Phi_{A,1}$ and $\Phi_{A,2}$. Thus the measurement matrices Φ_1 and Φ_2 are of the following form:

$$\Phi_1 = \begin{bmatrix} \Phi_D \\ \Phi_{A,1} \end{bmatrix} \text{ and } \Phi_2 = \begin{bmatrix} \Phi_D \\ \Phi_{A,2} \end{bmatrix}.$$

The submatrices Φ_D, $\Phi_{A,1}$, and $\Phi_{A,2}$ contain i.i.d. Gaussian entries. Once the difference $x_1 - x_2$ and average $\frac{1}{2}x_1 + \frac{1}{2}x_2$ have been recovered using the above technique, the computation of x_1 and x_2 is straightforward. The measurement rate can be computed by considering both parts of the measurement matrices.

Recovery of signal difference: The submatrix Φ_D is used to recover the signal difference. By subtracting the product of Φ_D with the signals x_1 and x_2, we have

$$\Phi_D x_1 - \Phi_D x_2 = \Phi_D(x_1 - x_2).$$

In the original representation we have $x_1 - x_2 = z_1 - z_2$ with sparsity rate $2S_I$. But $z_1(n) - z_2(n)$ is nonzero only if $z_1(n)$ is nonzero or $z_2(n)$ is nonzero. Therefore, the sparsity rate of $x_1 - x_2$ is equal to the sum of the individual sparsities reduced by the sparsity rate of the overlap, and so we have $S(X_1 - X_2) = 2S_I - (S_I)^2$. Therefore, any measurement rate greater than $c'(2S_I - (S_I)^2)$ for each Φ_D permits recovery of the length N signal $x_1 - x_2$. (As always, the probability of correct recovery approaches one as N increases.)

Recovery of average: Once $x_1 - x_2$ has been recovered, we have

$$x_1 - \frac{1}{2}(x_1 - x_2) = \frac{1}{2}x_1 + \frac{1}{2}x_2 = x_2 + \frac{1}{2}(x_1 - x_2).$$

At this stage, we know $x_1 - x_2$, $\Phi_D x_1$, $\Phi_D x_2$, $\Phi_{A,1}x_1$, and $\Phi_{A,2}x_2$. We have

$$\begin{aligned}
\Phi_D x_1 - \frac{1}{2}\Phi_D(x_1 - x_2) &= \Phi_D\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right), \\
\Phi_{A,1}x_1 - \frac{1}{2}\Phi_{A,1}(x_1 - x_2) &= \Phi_{A,1}\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right), \\
\Phi_{A,2}x_2 + \frac{1}{2}\Phi_{A,2}(x_1 - x_2) &= \Phi_{A,2}\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right),
\end{aligned}$$

where $\Phi_D(x_1 - x_2)$, $\Phi_{A,1}(x_1 - x_2)$, and $\Phi_{A,2}(x_1 - x_2)$ are easily computable because $(x_1 - x_2)$ has been recovered. The signal $\frac{1}{2}x_1 + \frac{1}{2}x_2$ is of length N; its sparsity rate is equal to the sum of the individual sparsities $S_C + 2S_I$ reduced by the sparsity rate of the overlaps, and so we have $S(\frac{1}{2}X_1 + \frac{1}{2}X_2) = S_C + 2S_I - 2S_C S_I - (S_I)^2 + S_C(S_I)^2$. Therefore, any measurement rate greater than $c'(S_C + 2S_I - 2S_C S_I - (S_I)^2 + S_C(S_I)^2)$ aggregated over the matrices Φ_D, $\Phi_{A,1}$, and $\Phi_{A,2}$ enables recovery of $\frac{1}{2}x_1 + \frac{1}{2}x_2$.

Computation of measurement rate: By considering the requirements on Φ_D, the individual measurement rates R_1 and R_2 must satisfy (13a). Combining the measurement rates required for $\Phi_{A,1}$ and $\Phi_{A,2}$, the sum measurement rate satisfies (13b). We complete the proof by noting that $c'(\cdot)$ is continuous and that $\lim_{S\to 0} c'(S) = 0$. Thus, as S_I goes to zero, the limit of the sum measurement rate is $c'(S)$. □

H Proof of Theorem 7

We assume that Ψ is an orthonormal matrix. Like Φ_j itself, the matrix $\Phi_j\Psi$ also has i.i.d. $\mathcal{N}(0,1)$ entries, since Ψ is orthonormal. For convenience, we assume $\Psi = I_N$. The results presented can be easily extended to a more general orthonormal matrix Ψ by replacing Φ_j with $\Phi_j\Psi$.

Assume without loss of generality that $\Omega = \{1, 2, \ldots, K\}$ for convenience of notation. Thus, the correct estimates are $n \leq K$, and the incorrect estimates are $n \geq K + 1$. Now consider the statistic ξ_n in (14). This is the sample mean of J i.i.d. variables. The variables $\langle y_j, \phi_{j,n}\rangle^2$ are i.i.d.

since each $y_j = \Phi_j x_j$, and Φ_j and x_j are i.i.d. Furthermore, these variables have a finite variance.[8] Therefore, we invoke the Law of Large Numbers (LLN) to argue that ξ_n, which is a sample mean of $\langle y_j, \phi_{j,n}\rangle^2$, converges to $E[\langle y_j, \phi_{j,n}\rangle^2]$ as J grows large. We now compute $E[\langle y_j, \phi_{j,n}\rangle^2]$ under two cases. In the first case, we consider $n \geq K+1$ (we call this the "bad statistics case"), and in the second case, we consider $n \leq K$ ("good statistics case").

Bad statistics: Consider one of the bad statistics by choosing $n = K+1$ without loss of generality. We have

$$\begin{aligned}
E[\langle y_j, \phi_{j,K+1}\rangle^2] &= E\left[\sum_{n=1}^{K} x_j(n)\langle \phi_{j,n}, \phi_{j,K+1}\rangle\right]^2 \\
&= E\left[\sum_{n=1}^{K} x_j(n)^2 \langle \phi_{j,n}, \phi_{j,K+1}\rangle^2\right] \\
&\quad + E\left[\sum_{n=1}^{K}\sum_{\ell=1,\ell\neq n}^{K} x_j(\ell)x_j(n)\langle \phi_{j,\ell}, \phi_{j,K+1}\rangle\langle \phi_{j,n}, \phi_{j,K+1}\rangle\right] \\
&= \sum_{n=1}^{K} E\left[x_j(n)^2\right] E\left[\langle \phi_{j,n}, \phi_{j,K+1}\rangle^2\right] \\
&\quad + \sum_{n=1}^{K}\sum_{\ell=1,\ell\neq n}^{K} E[x_j(\ell)]E[x_j(n)]E\left[\langle \phi_{j,\ell}, \phi_{j,K+1}\rangle\langle \phi_{j,n}, \phi_{j,K+1}\rangle\right],
\end{aligned}$$

since the terms are independent. We also have $E[x_j(n)] = E[x_j(\ell)] = 0$, and so

$$\begin{aligned}
E[\langle y_j, \phi_{j,K+1}\rangle^2] &= \sum_{n=1}^{K} E\left[x_j(n)^2\right] E\left[\langle \phi_{j,n}, \phi_{j,K+1}\rangle^2\right] \\
&= \sum_{n=1}^{K} \sigma^2 E\left[\langle \phi_{j,n}, \phi_{j,K+1}\rangle^2\right]. \qquad (25)
\end{aligned}$$

To compute $E\left[\langle \phi_{j,n}, \phi_{j,K+1}\rangle^2\right]$, let $\phi_{j,n}$ be the column vector $[a_1, a_2, \ldots, a_M]^T$, where each element in the vector is i.i.d. $\mathcal{N}(0,1)$. Likewise, let $\phi_{j,K+1}$ be the column vector $[b_1, b_2, \ldots, b_M]^T$ where the elements are i.i.d. $\mathcal{N}(0,1)$. We have

$$\begin{aligned}
\langle \phi_{j,n}, \phi_{j,K+1}\rangle^2 &= (a_1b_1 + a_2b_2 + \ldots + a_Mb_M)^2 \\
&= \sum_{m=1}^{M} a_m^2 b_m^2 + 2\sum_{m=1}^{M-1}\sum_{r=m+1}^{M} a_m a_r b_m b_r.
\end{aligned}$$

[8] In [61], we evaluate the variance of $\langle y_j, \phi_{j,n}\rangle^2$ as

$$\mathrm{Var}[\langle y_j, \phi_{j,n}\rangle^2] = \begin{cases} M\sigma^4(34MK + 6K^2 + 28M^2 + 92M + 48K + 90 + 2M^3 + 2MK^2 + 4M^2K), & n \in \Omega \\ 2MK\sigma^4(MK + 3K + 3M + 6), & n \notin \Omega. \end{cases}$$

For finite M, K and σ, the above variance is finite.

Taking the expected value, we have

$$\begin{aligned}
E\left[\langle\phi_{j,n},\phi_{j,K+1}\rangle^2\right] &= E\left[\sum_{m=1}^{M} a_m^2 b_m^2\right] + 2E\left[\sum_{m=1}^{M-1}\sum_{r=m+1}^{M} a_m a_r b_m b_r\right] \\
&= \sum_{m=1}^{M} E\left[a_m^2 b_m^2\right] + 2\sum_{m=1}^{M-1}\sum_{r=q+1}^{M} E\left[a_m a_r b_m b_r\right] \\
&= \sum_{m=1}^{M} E\left[a_m^2\right] E\left[b_m^2\right] + 2\sum_{m=1}^{M-1}\sum_{r=m+1}^{M} E\left[a_m\right] E\left[a_r\right] E\left[b_m\right] E\left[b_r\right] \\
&\qquad\qquad \text{(since the random variables are independent)} \\
&= \sum_{m=1}^{M}(1) + 0 \qquad \text{(since } E\left[a_m^2\right] = E\left[b_m^2\right] = 1 \text{ and } E\left[a_m\right] = E\left[b_m\right] = 0\text{)} \\
&= M,
\end{aligned}$$

and thus

$$E\left[\langle\phi_{j,n},\phi_{j,K+1}\rangle^2\right] = M. \tag{26}$$

Combining this result with (25), we find that

$$E[\langle y_j,\phi_{j,K+1}\rangle^2] = \sum_{n=1}^{K}\sigma^2 M = MK\sigma^2.$$

Thus we have computed $E[\langle y_j,\phi_{j,K+1}\rangle^2]$ and can conclude that as J grows large, the statistic ξ_{K+1} converges to

$$E[\langle y_j,\phi_{j,K+1}\rangle^2] = MK\sigma^2. \tag{27}$$

Good statistics: Consider one of the good statistics, and without loss of generality choose $n = 1$. Then, we have

$$\begin{aligned}
E[\langle y_j,\phi_{j,1}\rangle^2] &= E\left[\left(x_j(1)\|\phi_{j,1}\|^2 + \sum_{n=2}^{K} x_j(n)\langle\phi_{j,n},\phi_{j,1}\rangle\right)^2\right] \\
&= E\left[\left(x_j(1)\right)^2\|\phi_{j,1}\|^4\right] + E\left[\sum_{n=2}^{K} x_j(n)^2\langle\phi_{j,n},\phi_{j,1}\rangle^2\right] \\
&\qquad\qquad \text{(all other cross terms have zero expectation)} \\
&= E\left[x_j(1)^2\right] E\left[\|\phi_{j,1}\|^4\right] + \sum_{n=2}^{K} E\left[x_j(n)^2\right] E\left[\langle\phi_{j,n},\phi_{j,1}\rangle^2\right] \qquad \text{(by independence)} \\
&= \sigma^2 E\left[\|\phi_{j,1}\|^4\right] + \sum_{n=2}^{K}\sigma^2 E\left[\langle\phi_{j,n},\phi_{j,1}\rangle^2\right]. \qquad (28)
\end{aligned}$$

Extending the result from (26), we can show that $E\langle\phi_{j,n},\phi_{j,1}\rangle^2 = M$. Using this result in (28),

$$E[\langle y_j,\phi_{j,1}\rangle^2] = \sigma^2 E\|\phi_{j,1}\|^4 + \sum_{n=2}^{K}\sigma^2 M. \tag{29}$$

To evaluate $E\left[\|\phi_{j,1}\|^4\right]$, let $\phi_{j,1}$ be the column vector $[c_1, c_2, \ldots, c_M]^T$, where the elements of the vector are random $\mathcal{N}(0,1)$. Define the random variable $Z = \|\phi_{j,1}\|^2 = \sum_{m=1}^{M} c_m^2$. Note that (i) $E\left[\|\phi_{j,1}\|^4\right] = E\left[Z^2\right]$ and (ii) Z is chi-squared distributed with M degrees of freedom. Thus, $E\left[\|\phi_{j,1}\|^4\right] = E\left[Z^2\right] = M(M+2)$. Using this result in (29), we have

$$\begin{aligned} E[\langle y_j, \phi_{j,1}\rangle^2] &= \sigma^2 M(M+2) + (K-1)\sigma^2 M \\ &= M(M+K+1)\sigma^2. \end{aligned}$$

We have computed the variance of $\langle y_j, \phi_{j,1}\rangle$ and can conclude that as J grows large, the statistic ξ_1 converges to

$$E[\langle y_j, \phi_{j,1}\rangle^2] = (M+K+1)M\sigma^2. \tag{30}$$

Conclusion: From (27) and (30) we conclude that

$$\lim_{J\to\infty} \xi_n = E[\langle y_j, \phi_{j,n}\rangle^2] = \begin{cases} (M+K+1)M\sigma^2, & n \in \Omega \\ KM\sigma^2, & n \notin \Omega. \end{cases}$$

For any $M \geq 1$, these values are distinct — their ratio is $\frac{M+K+1}{K}$. Therefore, as J increases we can distinguish between the two expected values of ξ_n with overwhelming probability. □

I Proof of Theorem 8

Our proof has two parts. First we argue that $\lim_{J\to\infty} \widehat{z}_C = z_C$. Then we show that this implies vanishing probability of error in recovering each innovation z_j.

Part 1: We can write our estimate as

$$\widehat{z}_C = \frac{1}{J}\widehat{\Phi}^T y = \frac{1}{J}\sum_{j=1}^{J} \widehat{\Phi}_j^T y_j = \frac{1}{J}\sum_{j=1}^{J} \frac{1}{M_j\sigma_j^2}\Phi_j^T\Phi_j x_j = \frac{1}{J}\sum_{j=1}^{J} \frac{1}{M_j\sigma_j^2}\sum_{m=1}^{M_j} (\phi_{j,m}^R)^T \phi_{j,m}^R x_j,$$

where $\phi_{j,m}^R$ denotes the m-th row of Φ_j, that is, the m-th measurement vector for node j. Since the elements of each Φ_j are Gaussians with variance σ_j^2, the product $(\phi_{j,m}^R)^T\phi_{j,m}^R$ has the property

$$E[(\phi_{j,m}^R)^T\phi_{j,m}^R] = \sigma_j^2 I_N.$$

It follows that

$$E[(\phi_{j,m}^R)^T\phi_{j,m}^R x_j] = \sigma_j^2 E[x_j] = \sigma_j^2 E[z_C + z_j] = \sigma_j^2 z_C$$

and, similarly, that

$$E\left[\frac{1}{M_j\sigma_j^2}\sum_{m=1}^{M_j} (\phi_{j,m}^R)^T\phi_{j,m}^R x_j\right] = z_C.$$

Thus, $\widehat{z}_C$ is a sample mean of J independent random variables with mean z_C. From the law of large numbers, we conclude that $\lim_{J\to\infty}\widehat{z}_C = z_C$.

Part 2: Consider recovery of the innovation z_j from the adjusted measurement vector $\widehat{y_j} = y_j - \Phi_j\widehat{z}_C$. As a recovery scheme, we consider a combinatorial search over all K-sparse index sets drawn from $\{1, 2, \ldots, N\}$. For each such index set Ω', we compute the distance from $\widehat{y}$ to the column span of $\Phi_{j,\Omega'}$, denoted by $d(\widehat{y}, \text{colspan}(\Phi_{j,\Omega'}))$, where $\Phi_{j,\Omega'}$ is the matrix obtained by sampling the columns Ω' from Φ_j. (This distance can be measured using the pseudoinverse of $\Phi_{j,\Omega'}$.)

For the correct index set Ω, we know that $d(\widehat{y_j}, \text{colspan}(\Phi_{j,\Omega})) \to 0$ as $J \to \infty$. For any other index set Ω', we know from the proof of Theorem 1 that $d(\widehat{y_j}, \text{colspan}(\Phi_{j,\Omega'})) > 0$. Let

$$\zeta := \min_{\Omega' \neq \Omega} d(\widehat{y}_j, \mathrm{colspan}(\Phi_{i,\Omega'})).$$

With probability one, $\zeta > 0$. Thus for sufficiently large J, we will have $d(\widehat{y}_j, \mathrm{colspan}(\Phi_{j,\Omega})) < \zeta/2$, and so the correct index set Ω can be correctly identified. Since $\lim_{J\to\infty} \widehat{z}_C = z_C$, the innovation estimates $\widehat{z}_j = z_j$ for each j and for J large enough. □

Acknowledgments

Thanks to Emmanuel Candès, Albert Cohen, Ron DeVore, Anna Gilbert, Illya Hicks, Robert Nowak, Jared Tanner, and Joel Tropp for informative and inspiring conversations. Special thanks to to Mark Davenport for a thorough critique of the manuscript. MBW thanks his former affiliated institutions Caltech and the University of Michigan, where portions of this work were performed. Final thanks to Ryan King for supercharging our computational capabilities.

References

[1] D. Baron, M. F. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, "An information-theoretic approach to distributed compressed sensing," in *Allerton Conf. Communication, Control, and Computing*, (Monticello, IL), Sept. 2005.

[2] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Asilomar Conf. Signals, Systems and Computers*, (Pacific Grove, CA), Nov. 2005.

[3] M. B. Wakin, S. Sarvotham, M. F. Duarte, D. Baron, and R. G. Baraniuk, "Recovery of jointly sparse signals from few random projections," in *Workshop on Neural Info. Proc. Sys. (NIPS)*, (Vancouver, Canada), Nov. 2005.

[4] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Performance limits for jointly sparse signals via graphical models," in *Workshop on Sensor, Signal and Information Proc. (SENSIP)*, (Sedona, AZ), May 2008.

[5] R. A. DeVore, B. Jawerth, and B. J. Lucier, "Image compression through wavelet transform coding," *IEEE Trans. Inform. Theory*, vol. 38, pp. 719–746, Mar. 1992.

[6] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Proc.*, vol. 41, pp. 3445–3462, Dec. 1993.

[7] Z. Xiong, K. Ramchandran, and M. T. Orchard, "Space-frequency quantization for wavelet image coding," *IEEE Trans. Image Proc.*, vol. 6, no. 5, pp. 677–693, 1997.

[8] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego: Academic Press, 1999.

[9] K. Brandenburg, "MP3 and AAC explained," in *AES 17th International Conference on High-Quality Audio Coding*, Sept. 1999.

[10] W. Pennebaker and J. Mitchell, "JPEG: Still image data compression standard," *Van Nostrand Reinhold*, 1993.

[11] D. S. Taubman and M. W. Marcellin, *JPEG 2000: Image Compression Fundamentals, Standards and Practice*. Kluwer, 2001.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[13] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, pp. 471–480, July 1973.

[14] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inform. Theory*, vol. 49, pp. 626–643, Mar. 2003.

[15] Z. Xiong, A. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Proc. Mag.*, vol. 21, pp. 80–94, Sept. 2004.

[16] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "On network correlated data gathering," in *IEEE INFOCOM*, (Hong Kong), Mar. 2004.

[17] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loeve transform," *IEEE Trans. Inform. Theory*, vol. 52, pp. 5177–5196, Dec. 2006.

[18] R. Wagner, V. Delouille, H. Choi, and R. G. Baraniuk, "Distributed wavelet transform for irregular sensor network grids," in *IEEE Statistical Signal Proc. (SSP) Workshop*, (Bordeaux, France), July 2005.

[19] A. Ciancio and A. Ortega, "A distributed wavelet compression algorithm for wireless multihop sensor networks using lifting," in *IEEE Int. Conf. Acoustics, Speech, Signal Proc. (ICASSP)*, (Philadelphia), Mar. 2005.

[20] T. Uyematsu, "Universal coding for correlated sources with memory," in *Canadian Workshop Inform. Theory*, (Vancouver, Canada), June 2001.

[21] J. Garcia-Frias and W. Zhong, "LDPC codes for compression of multi-terminal sources with hidden Markov correlation," *IEEE Communications Letters*, vol. 7, pp. 115–117, Mar. 2003.

[22] C. Chen, X. Ji, Q. Dai, and X. Liu, "Slepian-Wolf coding of binary finite memory sources using Burrows Wheeler transform," in *Data Compression Conference*, (Snowbird, UT), Mar. 2009.

[23] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, pp. 489–509, Feb. 2006.

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

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

[26] J. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," *IEEE Trans. Info. Theory*, vol. 53, pp. 4655–4666, Dec. 2007.

[27] R. G. Baraniuk, M. Davenport, R. A. DeVore, and M. B. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, pp. 253–263, Dec. 2008.

[28] 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.

[29] M. Rabbat, J. D. Haupt, A. Singh, and R. D. Nowak, "Decentralized compression and predistribution via randomized gossiping," in *Int. Workshop on Info. Proc. in Sensor Networks (IPSN)*, (Nashville, TN), Apr. 2006.

[30] W. Wang, M. Garofalakis, and K. Ramchandran, "Distributed sparse random projections for refinable approximation," in *Int. Workshop on Info. Proc. in Sensor Networks (IPSN)*, (Cambridge, MA), 2007.

[31] S. Aeron, M. Zhao, and V. Saligrama, "On sensing capacity of sensor networks for the class of linear observation, fixed SNR models," 2007. Preprint.

[32] S. Aeron, M. Zhao, and V. Saligrama, "Fundamental limits on sensing capacity for sensor networks and compressed sensing," 2008. Preprint.

[33] E. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Annals of Statistics*, vol. 35, pp. 2313–2351, Dec. 2007.

[34] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Measurements vs. bits: Compressed sensing meets information theory," in *Allerton Conf. Communication, Control, and Computing*, (Monticello, IL), Sept. 2006.

[35] E. Candès and D. Donoho, "Curvelets — A surprisingly effective nonadaptive representation for objects with edges," *Curves and Surfaces*, 1999.

[36] M. F. Duarte, M. B. Wakin, D. Baron, and R. G. Baraniuk, "Universal distributed sensing via random projections," in *Int. Workshop on Info. Proc. in Sensor Networks (IPSN)*, (Nashville, TN), Apr. 2006.

[37] "Compressed sensing website." `http://dsp.rice.edu/cs/`.

[38] R. Venkataramani and Y. Bresler, "Further results on spectrum blind sampling of 2D signals," in *IEEE Int. Conf. Image Proc. (ICIP)*, vol. 2, (Chicago), Oct. 1998.

[39] E. Candès, M. Rudelson, T. Tao, and R. Vershynin, "Error correction via linear programming," in *Annual IEEE Symposium on Foundations of Computer Science*, (Pittsburgh, PA), pp. 295–308, Oct. 2005.

[40] D. Donoho and J. Tanner, "Neighborliness of randomly projected simplices in high dimensions," *Proc. National Academy of Sciences*, vol. 102, no. 27, pp. 9452–457, 2005.

[41] D. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," *Discrete and Computational Geometry*, vol. 35, pp. 617–652, Mar. 2006.

[42] D. L. Donoho and J. Tanner, "Counting faces of randomly projected polytopes when the projection radically lowers dimension," *Journal of the American Mathematical Society*, vol. 22, pp. 1–53, Jan. 2009.

[43] A. Cohen, W. Dahmen, and R. A. DeVore, "Near optimal approximation of arbitrary vectors from highly incomplete measurements," 2007. Preprint.

[44] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," Dec. 2007. Preprint.

[45] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, June 2008. To appear.

[46] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing: Closing the gap between performance and complexity," Mar. 2008. Preprint.

[47] A. Hormati and M. Vetterli, "Distributed compressed sensing: Sparsity models and reconstruction algorithms using anihilating filter," in *IEEE Int. Conf. Acoustics, Speech, Signal Proc. (ICASSP)*, (Las Vegas, NV), pp. 5141–5144, Apr. 2008.

[48] M. Mishali and Y. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," Feb. 2008. Preprint.

[49] M. Fornassier and H. Rauhut, "Recovery algorithms for vector valued data with joint sparsity constrains," *SIAM Journal on Numerical Analysis*, vol. 46, no. 2, pp. 577–613, 2008.

[50] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, "Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms," 2007. Preprint.

[51] D. Estrin, D. Culler, K. Pister, and G. Sukhatme, "Connecting the physical world with pervasive networks," *IEEE Pervasive Computing*, vol. 1, no. 1, pp. 59–69, 2002.

[52] G. J. Pottie and W. J. Kaiser, "Wireless integrated network sensors," *Comm. ACM*, vol. 43, no. 5, pp. 51–58, 2000.

[53] J. Tropp, A. C. Gilbert, and M. J. Strauss, "Simulataneous sparse approximation via greedy pursuit," in *IEEE Int. Conf. Acoustics, Speech, Signal Proc. (ICASSP)*, (Philadelphia), Mar. 2005.

[54] V. N. Temlyakov, "A remark on simultaneous sparse approximation," *East J. Approx.*, vol. 100, pp. 17–25, 2004.

[55] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Proc.*, vol. 51, pp. 2477–2488, July 2005.

[56] R. Puri and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," in *Allerton Conf. Communication, Control, and Computing*, (Monticello, IL), Oct. 2002.

[57] R. Wagner, R. G. Baraniuk, and R. D. Nowak, "Distributed image compression for sensor networks using correspondence analysis and super-resolution," in *Data Comp. Conf. (DCC)*, Mar. 2000.

[58] C. Weidmann and M. Vetterli, "Rate-distortion analysis of spike processes," in *Data Compression Conference (DCC)*, (Snowbird, UT), pp. 82–91, Mar. 1999.

[59] 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/`.

[60] R. D. Nowak. Personal Communication, 2006.

[61] S. Sarvotham, M. B. Wakin, D. Baron, M. F. Duarte, and R. G. Baraniuk, "Analysis of the DCS one-stage greedy algoritm for common sparse supports," Tech. Rep. TREE0503, Rice University, Houston, TX, Oct. 2005. Available at `http://www.ece.rice.edu/~shri/docs/TR0503.pdf`.

[62] J. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Proc.*, vol. 86, pp. 589–602, Mar. 2006.

[63] D. Donoho and Y. Tsaig, "Extensions of compressed sensing," *Signal Proc.*, vol. 86, pp. 533–548, Mar. 2006.

[64] J. D. Haupt and R. D. Nowak, "Signal reconstruction from noisy random projections," *IEEE Transactions on Information Theory*, vol. 52, pp. 4036–4048, Sept. 2006.

[65] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error correcting coding and decoding: Turbo codes," *IEEE Int. Conf. Communications*, pp. 1064–1070, May 1993.

[66] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," 2008. Preprint.

[67] P. T. Boufounos and R. G. Baraniuk, "Quantization of sparse representations," Technical Report TREE0701, Rice University, Houston, TX, Mar. 2007.

[68] A. K. Fletcher, S. Rangan, and V. K. Goyal, "On the rate-distortion performance of compressed sensing," in *IEEE Int. Conf. Acoustics, Speech, Signal Proc. (ICASSP)*, (Honolulu, HI), pp. III–885–888, Apr. 2007.

[69] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. on Pure and Applied Math.*, vol. 59, pp. 1207–1223, Aug. 2006.

[70] P. T. Boufounos and R. G. Baraniuk, "1-bit compressive sensing," in *Conf. Information Sciences and Systems (CISS)*, (Princeton, NJ), pp. 16–21, Mar. 2008.

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