An Information-Theoretic Approach to Distributed Compressed Sensing*

Dror Baron, Marco F. Duarte, Shriram Sarvotham,
Michael B. Wakin and Richard G. Baraniuk
Dept. of Electrical and Computer Engineering, Rice University, Houston, TX 77005

Abstract

Compressed sensing is an emerging field based on the revelation that a small group of linear projections of a sparse signal contains enough information for reconstruction. In this paper we introduce a new theory for *distributed compressed 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 concept that we term the *joint sparsity* of a signal ensemble. We study a model for jointly sparse signals, propose algorithms for joint recovery of multiple signals from incoherent projections, and characterize the number of measurements per sensor required for accurate reconstruction. We establish a parallel with the Slepian-Wolf theorem from information theory and establish upper and lower bounds on the measurement rates required for encoding jointly sparse signals. In some sense DCS is a framework for distributed compression of sources with memory, which has remained a challenging problem for some time. DCS is immediately applicable to a range of problems in sensor networks and arrays.

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. 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. Such *transform coders* [1] 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 [1]; the coding standards MP3, JPEG, and JPEG2000 directly exploit this sparsity.

1.1 Distributed source coding

While the theory and practice of compression have been well developed for individual signals, many applications involve multiple signals, for which there has been less progress. As a motivating example, consider a *sensor network*, in which a number of distributed nodes acquire data and report it to a central collection point [2]. In such networks, communication energy and bandwidth are often scarce resources, making 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* in each

*This work was supported by grants from NSF, NSF-NeTS, ONR, and AFOSR. Email: {drorb, duarte, shri, wakin, richb}@rice.edu; Web: dsp.rice.edu/cs.

individual sensor's measurements. In such settings, *distributed source coding* that exploits both types of correlation might allow a substantial savings on communication costs [3–6].

A number of distributed coding algorithms have been developed that involve collaboration amongst the sensors [7, 8]. Any collaboration, however, involves some amount of inter-sensor communication overhead. The *Slepian-Wolf* framework for lossless distributed coding [3–6] offers a collaboration-free approach in which each sensor node could communicate losslessly at its conditional entropy rate, rather than at its individual entropy rate. Unfortunately, however, most existing coding algorithms [5, 6] exploit only inter-signal correlations and not intra-signal correlations, and there has been only limited progress on distributed coding of so-called "sources with memory." The direct implementation for such sources would require huge lookup tables [3], and 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. Finally, a recent paper by Uyematsu [9] provides compression of spatially correlated sources with memory, but the solution is specific to lossless distributed compression and cannot be readily extended to lossy settings. We conclude that the design of distributed coding techniques for sources with both intra- and inter-signal correlation is a challenging problem with many potential applications.

1.2 Compressed sensing (CS)

A new framework for single-signal sensing and compression has developed recently under the rubric of *Compressed Sensing* (CS) [10, 11]. CS builds on the surprising revelation that a signal having a sparse representation in one basis can be recovered from a small number of projections onto a second basis that is *incoherent* with the first.[1] In fact, for an N-sample signal that is K-sparse,[2] roughly cK projections of the signal onto the incoherent basis are required to reconstruct the signal with high probability (typically $c \approx 3$ or 4). This has promising implications for applications involving sparse signal acquisition. Instead of sampling a K-sparse signal N times, only cK incoherent measurements suffice, where K can be orders of magnitude less than N. Moreover, the cK 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.[3] 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 the signal is sparse in (and thus the coding algorithm is independent of the sparsity-inducing basis) [11, 12]. A variety of algorithms have been proposed for signal recovery [10, 11, 14–16], each requiring a slightly different constant c (see Section 2.2).

While powerful, the CS theory at present is designed mainly to exploit intra-signal structures at a single sensor. To the best of our knowledge, the only work to date that applies CS in a multi-sensor setting is Haupt and Nowak [17]. However, while their scheme exploits inter-signal correlations, it ignores intra-signal correlations.

[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 [10–13].

[2] By K-sparse, we mean that the signal can be written as a sum of K basis functions.

[3] Since the "incoherent" measurement vectors must be known for signal recovery, in practice one may use a pseudorandom basis with a known random seed.

1.3 Distributed compressed sensing (DCS)

In this paper we introduce a new theory for *distributed compressed sensing* (DCS) that enables new distributed coding algorithms that exploit both intra- and inter-signal correlation structures. In a typical DCS scenario, a number of sensors measure signals (of any dimension) that are each individually sparse in some basis and also correlated from sensor to sensor. Each sensor *independently* encodes its signal by projecting it onto another, incoherent basis (such as a random one) and then transmits just a few of the resulting coefficients to a single collection point. Under the right conditions, a decoder at the collection point can *jointly* reconstruct all of the signals precisely.

The DCS theory rests on a concept that we term the *joint sparsity* of a signal ensemble. We study a model for jointly sparse signals, propose algorithms for joint recovery of multiple signals from incoherent projections, and characterize the number of measurements per sensor required for accurate reconstruction. While the sensors operate entirely without collaboration, we will see that the measurement rates relate directly to the signals' *conditional sparsities*, in parallel with the Slepian-Wolf theory. In certain scenarios, the savings in measurements can be substantial over separate CS decoding.

Our DCS coding schemes share many of the attractive and intriguing properties of CS, particularly when we employ random projections at the sensors. In addition to being universally incoherent, random measurements are also *future-proof*: if a better sparsity-inducing basis is found, then the same random measurements can be used to reconstruct an even more accurate view of the environment. Using a pseudorandom basis (with a random seed) effectively implements a weak form of *encryption*: the randomized measurements will themselves resemble noise and be meaningless to an observer who does not know the associated seed. Random coding is also *robust*: the randomized measurements coming from each sensor have equal priority, unlike transform coefficients in current coders. Thus they allow a *progressively better reconstruction* of the data as more measurements are obtained; one or more measurements can also be lost without corrupting the entire reconstruction. Finally, DCS distributes its computational complexity asymmetrically, placing most of it in the joint decoder, which will often have more substantial resources than any individual sensor node. The encoders are very simple; they merely compute incoherent projections with their signals and make no decisions.

We note that our aim in this paper is to minimize the overall sensor measurement rates in order to reduce communication costs. Characterizing quantization, noise, and rate-distortion aspects in the DCS setting are topics for future work (see Section 4).

This paper is organized as follows. Section 2 overviews the single-signal CS theory. Section 3 introduces our model for joint sparsity and presents our analysis and simulation results. We close with a discussion and conclusions in Section 4.

2 Compressed Sensing

Consider a length-N, real-valued signal x of any dimension (without loss of generality, we focus on one dimension) indexed as $x(n)$, $n \in \{1, 2, \ldots, N\}$. Suppose that the basis $\Psi = [\psi_1, \ldots, \psi_N]$ [1] provides a K-sparse representation of x; that is

$$x = \sum_{n=1}^{N} \theta(n)\, \psi_n = \sum_{\ell=1}^{K} \theta(n_\ell)\, \psi_{n_\ell},$$

where x is a linear combination of K vectors chosen from Ψ, $\{n_\ell\}$ are the indices of those vectors, and $\{\theta(n)\}$ are the coefficients; the concept is extendable to tight frames [1].

Alternatively, we can write $x = \Psi\theta$, 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,[4] we can write that $\|\theta\|_0 = K$. Various expansions, including wavelets, Gabor bases, curvelets, etc., are widely used for representation and compression of natural signals, images, and other data. In this paper, we will focus on exactly K-sparse signals and defer discussion of the more general situation where the coefficients decay rapidly but not to zero (see Section 4).

The standard procedure for compressing such signals, known as *transform coding*, is to (*i*) acquire the full N-point signal x; (*ii*) compute the complete set of transform coefficients $\{\theta(n)\}$; (*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 $\{\theta(n)\}$, 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 these locations will change with each signal.

2.1 Incoherent projections

These inefficiencies raise a simple question: For a given signal, is it possible to directly estimate the set of large $\theta(n)$'s that will not be discarded? While this seems improbable, the recent theory of *Compressed Sensing* (CS) [10–12] offers a solution. In CS, we do not measure or encode the K significant $\theta(n)$ directly. Rather, we measure and encode M projections $y(m) = \langle x, \phi_m^T \rangle$ of the signal onto a *second set* of basis 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 basis* matrix Φ is $M \times N$ with each row a basis vector ϕ_m.

The CS theory tells us that when certain conditions hold, namely that the basis $\{\phi_m\}$ cannot sparsely represent the elements of the basis $\{\psi_n\}$ (a condition known as *incoherence* of the two bases [10–13]) and the number of measurements M is large enough, then it is indeed possible to recover the set of large $\{\theta(n)\}$ (and thus the signal x) from a similarly sized set of measurements $\{y(m)\}$. This incoherence property holds for many pairs of bases, including for example, delta spikes and the sine waves of a Fourier basis, or significantly, between an arbitrary fixed basis/frame and a randomly generated one.

2.2 Signal recovery from incoherent projections

The recovery of the sparse set of significant coefficients $\{\theta(n)\}$ can be achieved using *optimization* by searching for the signal with ℓ_0-sparsest coefficients $\{\theta(n)\}$ that agrees with the M observed measurements in y (where presumably $M \ll N$):

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

(Assuming sufficient 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.) We will call the columns of $\Phi\Psi$ the *holographic basis*.

In principle, remarkably few incoherent measurements are required to ensure recovery a K-sparse signal via ℓ_0 minimization. Clearly, more than K measurements must be taken to avoid ambiguity. However, we have established that $K + 1$ random measurements

[4] The ℓ_0 norm $\|\theta\|_0$ merely counts the number of nonzero entries in the vector θ.

will suffice [18]. Unfortunately, solving this ℓ_0 optimization problem is prohibitively complex, requiring a combinatorial enumeration of the $\binom{N}{K}$ possible sparse subspaces; in fact it is NP-complete [14]. Yet another challenge is robustness; with little more than K measurements, 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.

The practical revelation that supports the new CS theory is that a much easier optimization problem yields an equivalent solution; we need only solve for the ℓ_1-sparsest coefficients θ that agree with the measurements y [10–12]

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

This optimization problem, also known as *Basis Pursuit* [19], 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; one typically requires $M \geq cK$ measurements to recover sparse signals via Basis Pursuit, where $c > 1$ is an *oversampling 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 1 [14–16] *Set $K = SN$ with $0 < S \ll 1$. Then there exists an oversampling factor $c(S) = O(\log(1/S))$, $c(S) > 1$, such that, for a K-sparse signal x in basis Ψ, the probability of recovering x via Basis Pursuit from $(c(S)+\epsilon)K$ random projections, $\epsilon > 0$, converges to 1 as $N \to \infty$. In contrast, the probability of recovering x via Basis Pursuit from $(c(S)-\epsilon)K$ random projections converges to 0 as $N \to \infty$.*

Donoho and Tanner [15, 16] have characterized this oversampling factor $c(S)$ precisely; we have discovered a useful *rule of thumb* that $c(S) \approx \log_2(1+S^{-1})$. In the remainder of the paper, we often use the abbreviated notation c to describe the oversampling factor in various settings even though $c(S)$ depends on the sparsity K and signal length N.

With appropriate oversampling, reconstruction via Basis Pursuit is robust to measurement noise and quantization error [10]. Iterative greedy algorithms have also been proposed [13], allowing even faster reconstruction at the expense of more measurements.

3 Joint Sparsity Model and Recovery Strategies

In the first part of 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*. In the second part, we investigate how *joint representations* can enable reconstruction of an ensemble of signals using fewer measurements per (separate) encoder. We characterize which *measurement rates* are feasible and describe effective reconstruction algorithms.

3.1 Additive common component + innovations model

Notation: We use $x_j(n)$ to denote sample n in signal j where $j \in \{1, 2, \ldots, J\}$, $x_j \in \mathbb{R}^N$, and we assume that there exists a known *sparse basis* Ψ for $\mathbb{R}^N$ in which the x_j can be sparsely represented. Denote by Φ_j the $M_j \times N$ *measurement matrix* for signal j. Thus, $y_j = \Phi_j x_j$ consists of $M_j < N$ *incoherent measurements* of x_j. We will emphasize random i.i.d. Gaussian matrices Φ_j, but other schemes are possible.

Additive model: In our model, all signals share a *common* sparse component while each individual signal contains a sparse *innovation* component; that is,

$$x_j = z + z_j, \quad j \in \{1, 2, \ldots, J\}$$

with $z = \Psi\theta_z$, $\|\theta_z\|_0 = K$, $z_j = \Psi\theta_j$, and $\|\theta_j\|_0 = K_j$. Thus, the signal z is common to all of the x_j and has sparsity K in basis Ψ. The signals z_j are the unique portions of the x_j and have sparsity K_j in the same basis.

To give ourselves a firm footing for analysis, we consider a specific *stochastic* generative model for the jointly sparse signals. We randomly pick K indices from $\{1, 2, ..., N\}$ for which the corresponding coefficients in θ_z are nonzero. We pick the indices such that each configuration has an equal likelihood of being selected. In a similar fashion we pick the K_j indices that correspond to the nonzero indices of θ_j, independently across all $J + 1$ components (including θ_z). The values of the nonzero coefficients are then generated from an i.i.d. Gaussian distribution. Though our results are specific to this context, they can be expected to generalize to other similar scenarios.

Applications: A practical situation well-suited to this model 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 that is both common to all sensors and structured enough to permit sparse representation. 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 sensor networks recording other phenomena. In such scenarios, we measure physical processes that change smoothly in time and space and thus are highly correlated.

3.2 Information theory framework and notions of sparsity rate

Consider first the simplest case where a *single joint encoder* processes $J = 2$ signals. By employing the CS machinery, we might expect that (i) $(K + K_1)c$ coefficients suffice to reconstruct x_1, (ii) $(K+K_2)c$ coefficients suffice to reconstruct x_2, and (iii) $(K+K_1+K_2)c$ coefficients suffice to reconstruct both x_1 and x_2, because we have $K + K_1 + K_2$ nonzero elements in x_1 and x_2.[5] Next, consider the case where $J = 2$ signals are processed by *separate encoders*. Given the $(K + K_1)c$ measurements for x_1 as side information, and assuming that the partitioning of x_1 into z and z_1 is known, cK_2 measurements that describe z_2 should allow reconstruction of x_2 . Similarly, conditioned on x_2 we should need only cK_1 measurements to reconstruct x_1.

These observations seem related to various types of entropy from information theory. We thus define notions for sparsity that are similar to existing notions of entropy. As a motivating example, suppose that the signals x_j, $j \in \{1, 2, \ldots, J\}$ are generated by sources X_j, $j \in \{1, 2, \ldots, J\}$. Assuming that in some sense the sparsity is linear in the signal length, as the signal-length N is incremented one by one, the sources provide new values for $z(N)$ and $z_j(N)$ and the sparsity levels gradually increase. If the sources are ergodic, then we can define the *sparsity rate* of X_j as the limit of the proportion of coefficients that need to be specified to reconstruct it given the vector indices, that is,

$$S(X_j) \triangleq \lim_{N\to\infty} \frac{K + K_j}{N}, \quad j \in \{1, 2, \ldots, J\}.$$

We also define the *joint sparsity* $S(X_{j_1}, X_{j_2})$ of x_{j_1} and x_{j_2} as the proportion of coefficients that need to be specified to reconstruct both signals given the vector indices:

$$S(X_{j_1}, X_{j_2}) \triangleq \lim_{N\to\infty} \frac{K + K_{j_1} + K_{j_2}}{N}, \quad j_1, j_2 \in \{1, 2, \ldots, J\}.$$

[5] We use the same notation c for the oversampling factors for coding x_1, x_2, or both sequences.

Finally, the *conditional sparsity* of x_{j_1} given x_{j_2} is the proportion of coefficients that need to be specified to reconstruct x_{j_1}, where x_{j_2} and the vector indices for x_{j_1} are available:

$$S(X_{j_1}|X_{j_2}) \triangleq \lim_{N\to\infty} \frac{K_{j_1}}{N}, \quad j_1, j_2 \in \{1, 2, \ldots, J\}.$$

The joint and conditional sparsities extend naturally to problems with $J > 2$ signals. Note also that the ergodicity of the source implies that these various limits exist, yet we can still consider signals in a deterministic framework where K and K_j are fixed.

The common component z can also be considered to be generated by a common source Z, in which case its sparsity rate is given by $S(Z) \triangleq \lim_{N\to\infty} \frac{K}{N}$. Alternatively, the sparsity rate of z can be analyzed in a manner analogous to the *mutual information* [3] of traditional information theory, for example, $S(Z) = I(Z_1; Z_2) = S(X_1) + S(X_2) - S(X_1, X_2)$, which we denote by S. These definitions offer a framework for joint sparsity with notions similar to the entropy, conditional entropy, and joint entropy of information theory.

3.3 Sparsity reduction

We now confine our attention to $J = 2$ signals where the innovation sparsity rates are equal, and we denote them by $S_I \triangleq S(Z_1) = S(Z_2)$. We consider an interesting outcome of our stochastic model when the supports of the common and innovation parts overlap. Consider the n'th coefficient $\theta_z(n)$ of the common component z and the corresponding innovation coefficients $\theta_1(n)$ and $\theta_2(n)$. Suppose that all three coefficients are nonzero. Clearly, the same signals x_1 and x_2 could have been generated using at most two nonzero values among the three, for example by adding the current value of $\theta_z(n)$ to current values of $\theta_1(n)$ and $\theta_2(n)$. Indeed, when all three original coefficients are nonzero, we can represent them equivalently by any subset of two coefficients. In this case, there exists a sparser representation than we might expect given K, K_1, and K_2.

Because the coefficient amplitudes are Gaussian, *sparsity reduction* is possible with positive probability only in the case where three corresponding nonzero coefficients are changed to two nonzero coefficients. Because the locations of the nonzero coefficients are uniform, the probability that all three are nonzero is $S^* \triangleq S(S_I)^2$.

3.4 Measurement rate region

To characterize the performance in our setup, we introduce a *measurement rate region*. Let M_1 and M_2 be the number of measurements taken for x_1 and x_2, respectively. We define the measurement rates R_1 and R_2 in an asymptotic manner as

$$R_1 \triangleq \lim_{N\to\infty} \frac{M_1}{N} \quad \text{and} \quad R_2 \triangleq \lim_{N\to\infty} \frac{M_2}{N}.$$

For a measurement rate pair (R_1, R_2), we wish to see whether we can reconstruct the signals with vanishing probability as N increases. In this case, we say then that the measurement rate pair is *achievable*.

For signals that are jointly sparse under our model, the individual sparsity rate of signal x_j is $S(X_j) = S + S_I - SS_I$. Separate recovery via ℓ_0 minimization would require a measurement rate $R_j = S(X_j)$. Separate recovery via ℓ_1 minimization would require an oversampling factor $c(S(X_j))$. To improve upon these figures, we adapt the standard machinery of CS to the joint recovery problem. Using this machinery and ℓ_0 reconstruction, we have provided a precise characterization of the measurement rate region [18]. We omit the detailed results and focus instead on more tractable algorithms.

3.5 Joint recovery via ℓ_1 minimization

In non-distributed compressed sensing (Section 2.2), ℓ_1 minimization can be implemented via linear programming but requires an oversampling factor of $c(S)$ (Theorem 1). We now study what penalty must be paid for ℓ_1 reconstruction of jointly sparse signals.

3.5.1 Bounds on performance of ℓ_1 signal recovery

We begin with a converse theorem, which describes what measurement rate pairs *cannot* be achieved via ℓ_1 recovery. Before proceeding, we shed some light on the notion of a converse region in this computational scenario. We focus on the setup where random measurements for signal x_j are performed via multiplication by an M_j by N matrix and reconstruction of the $J = 2$ signals is performed via application of ℓ_1 techniques on subsets of the 2 signals. Within this setup, a converse region is a set of measurement rates for which any such reconstruction techniques fail with probability 1. We now present our converse theorem. For compactness, we define the measurement function $c'(S) \triangleq S \cdot c(S)$ based on Donoho and Tanner's oversampling factor c [15, 16].

Theorem 2 *Let $J = 2$ and fix the sparsity rate of the common part $S(Z) = S$ and the innovation sparsity rates $S(Z_1) = S(Z_2) = S_I$. The following conditions are necessary to enable reconstruction with vanishing probability of error:*

$$\begin{aligned} R_1 &\geq c'\left(S_I - SS_I - (S_I)^2 + S^*\right), \\ R_2 &\geq c'\left(S_I - SS_I - (S_I)^2 + S^*\right). \end{aligned}$$

Proof sketch: If x_2 is completely available, then we must still measure and reconstruct z_1. Even under the most stringent sparsity reductions via overlap with z and z_2, the sparsity rate of z_1 is lower bounded by $S_I - SS_I - (S_I)^2 + S^*$, which leads to the requisite bound on the measurement rate for R_1. The argument for R_2 is analogous. □

The theorem provides a converse region such that, if (R_1, R_2) violate these conditions and we perform $M_1 = \lceil (R_1 - \epsilon)N \rceil$ measurements for x_1 and $M_2 = \lceil (R_2 - \epsilon)N \rceil$ measurements for x_2, then the probability of incorrect reconstruction will converge to 1 as N increases. We also anticipate the following bound on the sum measurement rate

$$R_1 + R_2 \geq c'(S + 2S_I - 2SS_I - (S_I)^2 + S^*),$$

which appears in Figure 1.

Having ruled out part of the measurement region, we wish to specify regions where joint reconstruction can be performed. The following theorem, proved in the full paper [18], provides a first step in this direction.

Theorem 3 *Let $J = 2$ and fix the sparsity rate of the common part $S(Z) = S$ and the innovation sparsity rates $S(Z_1) = S(Z_2) = S_I$. Then there exists an ℓ_1 reconstruction technique (along with a measurement strategy) if the measurement rates satisfy:*

$$\begin{aligned} R_1, R_2 &\geq c'(2S_I - S^*), \\ R_1 + R_2 &\geq c'(2S_I - S^*) + c'(S + 2S_I - 2SS_I - (S_I)^2 + S^*). \end{aligned}$$

Furthermore, as $S_I \to 0$ the sum measurement rate approaches $c'(S)$.

The proof of Theorem 3 [18] describes a constructive reconstruction algorithm, which is very insightful. We construct measurement matrices Φ_1 and Φ_2, which each consist of two parts. The first part in each measurement matrix is common to both, and is used

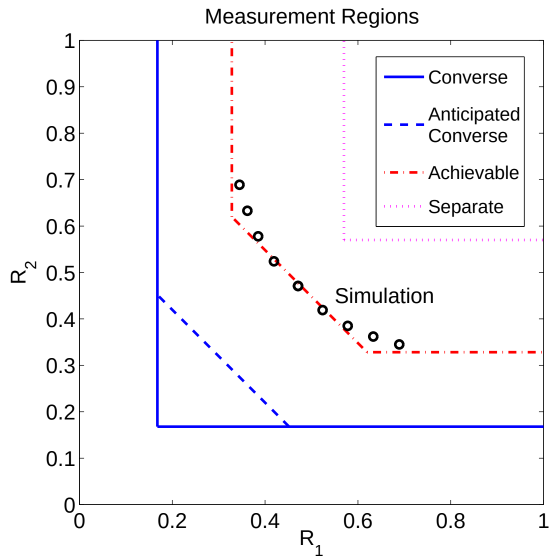

Figure 1: *Rate region for distributed compressed sensing. We chose a common sparsity rate $S = 0.2$ and innovation sparsity rates $S_I = S_1 = S_2 = 0.05$. Our simulation results use eq. (1) and signals of length $N = 1000$.*

to reconstruct $x_1 - x_2 = z_1 - z_2$. The second parts of the matrices are different and enable the reconstruction of $z + 0.5(z_1 + z_2)$. Once these two components have been reconstructed, the computation of x_1 and x_2 is straightforward. The measurement rate can be computed by considering both common and different parts of Φ_1 and Φ_2.

Our measurement rate bounds are strikingly similar to those in the Slepian-Wolf theorem [4], where each signal must be encoded above its conditional entropy rate, and the ensemble must be coded above the joint entropy rate. Yet despite these advantages, the achievable measurement rate region of Theorem 3 is loose with respect to the converse region of Theorem 2, as shown in Figure 1. We are attempting to tighten these bounds in our ongoing work and have promising preliminary results.

3.5.2 Joint reconstruction with a single linear program

We now present a reconstruction approach based on a single execution of a linear program. Although we have yet to characterize the performance of this approach theoretically, our simulation tests indicate that it slightly outperforms the approach of Theorem 3 (Figure 1). In our approach, we wish to *jointly* recover the sparse signals using a single linear program, and so we define the following joint matrices and vectors:

$$\theta = \begin{bmatrix} \theta_z \\ \theta_1 \\ \theta_2 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad \Phi = \begin{bmatrix} \Phi_1 & 0 \\ 0 & \Phi_2 \end{bmatrix}, \quad \widetilde{\Psi} = \begin{bmatrix} \Psi & \Psi & 0 \\ \Psi & 0 & \Psi \end{bmatrix}.$$

Using the frame $\widetilde{\Psi}$, we can represent x sparsely using the coefficient vector θ, which contains $K + \sum_j K_j$ nonzero coefficients, to obtain $x = \widetilde{\Psi}\theta$. The concatenated measurement vector y is computed from individual measurements, where the joint measurement basis is Φ and the joint holographic basis is then $A = \Phi\widetilde{\Psi}$. With sufficient oversampling, we can recover the vector θ, and thus x_1 and x_2, by solving the linear program

$$\widehat{\theta} = \arg\min \|\theta\|_1 \text{ s.t. } y = \Phi\widetilde{\Psi}\theta.$$

In practice, we find it helpful to modify the Basis Pursuit algorithm to account for the special structure of DCS recovery. In the linear program, we use a modified ℓ_1 penalty

$$\gamma_z||\theta_z||_1 + \gamma_1||\theta_1||_1 + \gamma_2||\theta_2||_1, \tag{1}$$

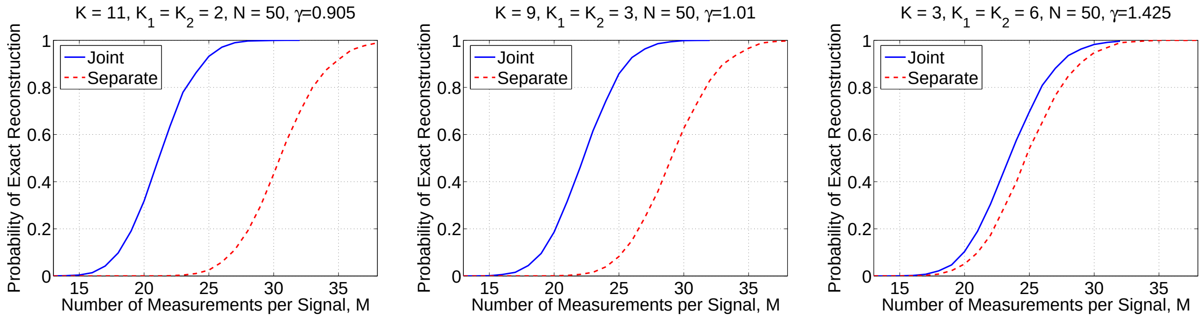

Figure 2: *Comparison of joint reconstruction using eq. (1) and separate reconstruction . The advantage of using joint instead of separate reconstruction depends on the common sparsity.*

where $\gamma_z, \gamma_1, \gamma_2 \geq 0$. We note in passing that if $K_1 = K_2$, then we set $\gamma_1 = \gamma_2$. In this scenario, without loss of generality, we define $\gamma_1 = \gamma_2 = 1$ and set $\gamma_z = \gamma$.

Practical considerations: Joint reconstruction with a single linear program has several disadvantages relative to the approach of Theorem 3. In terms of computation, the linear program must reconstruct the $J+1$ vectors $z, z_1, \ldots, z_J$. Because the complexity of linear programming is roughly cubic, the computational burden scales with J^3. In contrast, Theorem 3 first reconstructs $J(J-1)/2$ signal differences of the form $x_{j_1} - x_{j_2}$ and then reconstructs the common part $z + \frac{1}{J}(z_1 + \ldots + z_J)$. Each such reconstruction is only for a length-N signal, making the computational load lighter by an $O(J)$ factor.

Another disadvantage of the modified ℓ_1 reconstruction is that the optimal choice of γ_z, γ_1, and γ_2 depends on the relative sparsities K, K_1, and K_2. At this stage we have not been able to determine these optimal values analytically. Instead, we rely on a numerical optimization, which is computationally intense. For a discussion of the tradeoffs that affect these values, see the full paper [18].

3.6 Numerical examples

Reconstructing two signals with symmetric measurement rates: We consider $J = 2$ signals generated from components z, z_1, and z_2 having sparsities K, K_1, and K_2 in the basis $\Psi = I$.[6] We consider signals of length $N = 50$ with sparsities chosen such that $K_1 = K_2$ and $K + K_1 + K_2 = 15$; we assign Gaussian values to the nonzero coefficients. Finally, we focus on symmetric measurement rates $M = M_1 = M_2$ and use the joint ℓ_1 decoding method, as described in Section 3.5.2.

For each value of M, we optimize the choice of γ numerically and run several thousand trials to determine the probability of correctly recovering x_1 and x_2. The simulation results are summarized in Figure 2. The degree to which joint decoding outperforms separate decoding is directly related to the amount of shared information K. For $K = 11$, $K_1 = K_2 = 2$, M is reduced by approximately 30%. For smaller K, joint decoding barely outperforms separate decoding.

Reconstructing two signals with asymmetric measurement rates: In Figure 1, we compare separate CS reconstruction with the converse bound of Theorem 2, the achievable bound of Theorem 3, and numerical results. We use $J = 2$ signals and choose a common sparsity rate $S = 0.2$ and innovation sparsity rates $S_I = S_1 = S_2 = 0.05$. Several different asymmetric measurement rates are considered. In each such setup, we constrain M_2 to have the form $M_2 = \alpha M_1$, where $\alpha \in \{1, 1.25, 1.5, 1.75, 2\}$. By swapping M_1 and M_2, we obtain additional results for $\alpha \in \{1/2, 1/1.75, 1/1.5, 1/1.25\}$. In the

[6]We expect the results to hold for an arbitrary basis Ψ.

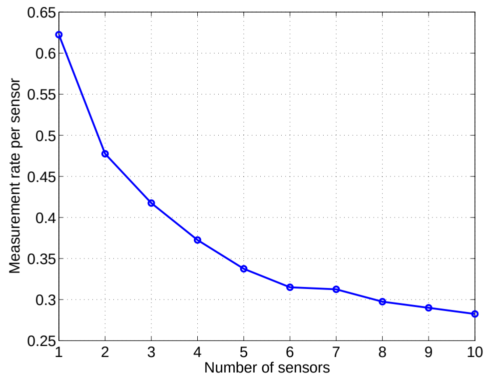

Figure 3: *Multi-sensor measurement results for our Joint Sparsity Model using eq. (1). We choose a common sparsity rate* $S = 0.2$ *and innovation sparsity rates* $S_I = 0.05$*. Our simulation results use signals of length* $N = 400$.

simulation itself, we first find the optimal γ numerically using $N = 40$ to accelerate the computation, and then simulate larger problems of size $N = 1000$. The results plotted indicate the smallest pairs (M_1, M_2) for which we always succeeded reconstructing the signal over 100 simulation runs.

Reconstructing multiple signals with symmetric measurement rates: The reconstruction techniques of this section are especially promising when $J > 2$ sensors are used, because the measurements for the common part are split among more sensors. These savings may be especially 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 $S = 0.2$ and $S_I = 0.05$ as in the asymmetric rate simulations; here we use symmetric measurement rates. We first find the optimal γ numerically using $N = 40$ to accelerate the computation, and then simulate larger problems of size $N = 400$. The results of Figure 3 describe the smallest symmetric measurement rates for which we always succeeded reconstructing the signal over 100 simulation runs. As J is increased, lower rates can be used.

4 Discussion and Conclusions

In this paper we have taken the first steps towards extending the theory and practice of CS to multi-signal, distributed settings. Our joint sparsity model captures the essence of real physical scenarios, illustrates the basic analysis and algorithmic techniques, and indicates the gains to be realized from joint recovery. We have provided a measurement rate region analogous to the Slepian-Wolf theorem [4], as well as appealing numerical results.

There are many opportunities for extensions of our ideas. *Compressible signals*: Natural signals are not exactly ℓ_0 sparse but rather can be better modeled as ℓ_p sparse with $0 < p \leq 1$. *Quantized and noisy measurements*: Our (random) measurements will be real numbers; quantization will gradually degrade the reconstruction quality as it becomes coarser [20]. Moreover, noise will often corrupt the measurements, making them not strictly sparse in any basis. *Fast algorithms*: In some applications, linear programming could prove too computationally intense. We leave these extensions for future work.

Finally, our model for sparse common and innovation components is useful, but we have also studied additional ways in which joint sparsity may occur [18]. *Common sparse supports:* In this model, all signals are constructed from the same sparse set of basis vectors, but with different coefficients. Examples of such scenarios include MIMO communication and audio signal arrays; the signals may be sparse in the Fourier domain, for example, yet multipath effects cause different attenuations among the frequency components. *Nonsparse common component + sparse innovations*: We extend our current model so that the common component need no longer be sparse in any basis. Since the common component is not sparse, no individual signal contains enough structure to permit efficient compression or CS; in general N measurements would be required for each individual N-sample signal. We demonstrate, however, that the common structure shared by the signals permits a dramatic reduction in the required measurement rates.

Acknowledgments. We thank Emmanuel Candès, Hyeokho Choi, Joel Tropp, Robert Nowak, Jared Tanner, and Anna Gilbert for informative and inspiring conversations. And thanks to Ryan King for invaluable help enhancing our computational capabilities.

References

[1] S. Mallat, *A wavelet tour of signal processing*, Academic Press, San Diego, CA, USA, 1999.

[2] 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.

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

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

[5] 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.

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

[7] H. Luo and G. Pottie, "Routing explicit side information for data compression in wireless sensor networks," in *Int. Conf. on Distirbuted Computing in Sensor Systems (DCOSS)*, Marina Del Rey, CA, June 2005.

[8] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loeve transform," *IEEE Trans. Inform. Theory*, Nov. 2004, Submitted.

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

[10] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, 2004, Submitted.

[11] D. Donoho, "Compressed sensing," 2004, Preprint.

[12] E. Candès and T. Tao, "Near optimal signal recovery from random projections and universal encoding strategies," *IEEE Trans. Inform. Theory*, 2004, Submitted.

[13] J. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," Apr. 2005, Preprint.

[14] E. Candès and T. Tao, "Error correction via linear programming," *Found. of Comp. Math.*, 2005, Submitted.

[15] D. Donoho and J. Tanner, "Neighborliness of randomly-projected simplices in high dimensions," Mar. 2005, Preprint.

[16] D. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," Jan. 2005, Preprint.

[17] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inform. Theory*, 2005, Submitted.

[18] D. Baron, M. B. Wakin, S. Sarvotham, M. F. Duarte, and R. G. Baraniuk, "Distributed compressed sensing," Tech. Rep., Available at http://www.dsp.rice.edu.

[19] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. on Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.

[20] E. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Annals of Statistics*, 2005, Submitted.