Joint Sparsity Models for Distributed Compressed Sensing

Marco F. Duarte, Shriram Sarvotham, Michael B. Wakin, Dror Baron, and Richard G. Baraniuk

***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 new concept that we term the *joint sparsity* of a signal ensemble. We study in detail two simple models for jointly sparse signals, propose algorithms for joint recovery of multiple signals from incoherent projections, and characterize theoretically and empirically 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 one of our models, the results are asymptotically best-possible, meaning that both the upper and lower bounds match the performance of our practical algorithms. 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.**

I. 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* 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.

A. 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 [1]. 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 individual sensor's measurements. In such settings, *distributed source coding* that exploits both types of correlation might allow a substantial savings on communication costs [2–4].

A number of distributed coding algorithms have been developed that involve collaboration amongst the sensors [5, 6]. Any collaboration, however, involves some amount of inter-sensor communication overhead. The *Slepian-Wolf* framework for lossless distributed coding [2–4] 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 [3, 4] 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."

B. Compressed sensing (CS)

A new framework for single-signal sensing and compression has developed recently under the rubric of *Compressed Sensing* (CS) [7, 8]. 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* incoherent measurement basis. 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 [10]. However, while their scheme exploits inter-signal correlations, it ignores intra-signal correlations.

This work was supported by NSF, ONR, AFOSR, and the Texas Instruments Leadership University Program.

The authors are with the Department of Electrical and Computer Engineering at Rice University.

Email: {duarte, shri, wakin, drorb, richb}@rice.edu. Web: dsp.rice.edu/cs

[1] Roughly speaking, *incoherence* means that no element of one basis has a sparse representation in terms of the other basis [7–9].

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

C. 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 two models 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 see dramatic savings relative to the number measurements required for 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.

This paper is organized as follows. Section II provides background on CS. Section III outlines our two models for joint sparsity. Section IV overviews our results, which are highlighted in the talk, and Section V concludes.

II. COMPRESSED SENSING

Suppose that x is a signal and let $\Psi = \{\psi_1, \psi_2, \dots\}$ be a *dictionary* of vectors. When we say that x is sparse, we mean that x is well approximated by a linear combination of a small group of vectors from Ψ. That is, $x \approx \sum_{i=1}^{K} \theta_{n_i} \psi_{n_i}$ where K is small; we say that the signal x is K-sparse in Ψ. The CS theory states that it is possible to construct an $M \times N$ *measurement* matrix Φ, where $M \ll N$, yet the measurements $y = \Phi x$ preserve the essential information about x. For example, let Φ be a $cK \times N$ matrix with i.i.d. Gaussian entries, where $c = c(N,K) \approx \log_2(1 + N/K)$ is an *oversampling factor*. Using such a matrix it is possible, with high probability, to recover every signal that is K-sparse in the basis Ψ from its image under Φ. Moreover, for signals that are not K-sparse but *compressible*, meaning that their coefficient magnitudes decay exponentially, there are tractable algorithms that achieve not more than a multiple of the error of the best K-term approximation of the signal.

Several algorithms have been proposed for recovering x from the measurements y, each requiring a slightly different constant c. The canonical approach [7, 8] uses linear programming to solve the ℓ_1 minimization problem

$$\widehat{\theta} = \arg\min_{\theta} \|\theta\|_1 \quad \text{subject to} \quad \Phi\Psi\theta = y.$$

This problem can be solved in polynomial time but is somewhat slow. Additional methods have been proposed involving greedy pursuit methods. Examples include Matching Pursuit (MP) and Orthogonal Matching Pursuit (OMP), which tend to require fewer computations but at the expense of slightly more measurements [9].

III. JOINT SPARSITY 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*. We consider two different *joint sparsity models* (JSMs) that apply in different situations. In these models, each signal is itself sparse, and so we could use the CS framework from above to encode and decode each one separately. However, there also exists a framework wherein a *joint representation* for the ensemble uses fewer total vectors.

We use the following notation for our signal ensembles and measurement model. Denote the *signals* in the ensemble by x_j, $j \in \{1, 2, \dots, J\}$, and assume that each signal $x_j \in \mathbb{R}^N$. 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 *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 $M_j < N$ *incoherent measurements* of x_j.

A. JSM-1: Sparse common component + innovations

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

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

with

$$z = \Psi\theta_z, \ \ \|\theta_z\|_0 = K \qquad \text{and} \qquad z_j = \Psi\theta_j, \ \ \|\theta_j\|_0 = K_j.$$

Thus, the signal z is common to all of the x_j and has sparsity K in basis Ψ.[3] The signals z_j are the unique portions of the x_j and have sparsity K_j in the same basis.

A practical situation well-modeled by JSM-1 is a group of sensors measuring temperatures at a number of outdoor locations throughout the day. The temperature readings x_j

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

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. 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.

B. JSM-2: Common sparse supports

In this model, all signals are constructed from the same sparse set of basis vectors, but with different coefficients; that is,

$$x_j = \Psi\theta_j, \quad j \in \{1, 2, \ldots, J\},$$

where each θ_j is supported only on the same $\Omega \subset \{1, 2, \ldots, N\}$ with $|\Omega| = K$. Hence, all signals have ℓ_0 sparsity of K, and all are constructed from the same K basis elements, but with arbitrarily different coefficients.

A practical situation well-modeled by JSM-2 is where multiple sensors acquire the same signal but with phase shifts and attenuations caused by signal propagation. In many cases it is critical to recover each one of the sensed signals, such as in many acoustic localization and array processing algorithms. Another useful application for JSM-2 is MIMO communication [11].

IV. Overview of Results

For each of these models, we propose algorithms for joint signal recovery from incoherent projections and characterize theoretically and empirically the number of measurements per sensor required for accurate reconstruction. We now briefly overview these results, which will be highlighted in the talk. (See also [12–14] for more details on our recent work.)

A. JSM-1: Sparse common component + innovations

For this model, we propose an analytical framework inspired by principles of information theory. This allows us to characterize the measurement rates M_j required to *jointly* reconstruct the signals x_j. We see that the measurement rates relate directly to the signals' *conditional sparsities*, in parallel with the Slepian-Wolf theory. More specifically, we formalize the following intuition. Consider the simple case of $J = 2$ 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, yet only (*iii*) $(K + K_1 + K_2)c$ coefficients should 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. In addition, 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. Formalizing these arguments allows us to establish theoretical lower bounds on the required measurement rates at each sensor; Figure 1

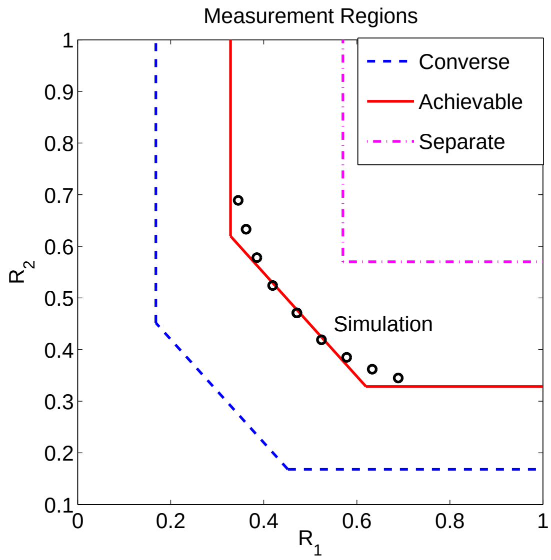

Fig. 1. *Converse bounds and achievable measurement rates for $J = 2$ signals with common sparse component and sparse innovations (JSM-1). The measurement rates $R_j := M_j/N$ reflect the number of measurements normalized by the signal length. The pink curve denotes the rates required for separable CS signal reconstruction.*

shows such a bound for the case of $J = 2$ signals, with signal lengths $N = 1000$ and sparsities $K = 200$, $K_1 = K_2 = 50$.

We also establish upper bounds on the required measurement rates M_j by proposing a specific algorithm for reconstruction [12]. The algorithm uses carefully designed measurement matrices Φ_j (in which some rows are identical and some differ) so that the resulting measurements can be combined to allow step-by-step recovery of the sparse components. We see that the theoretical rates M_j are below those required for separable CS recovery of each signal x_j (see Figure 1).

Finally, we propose a reconstruction technique based on a single execution of a linear program, which seeks the sparsest components $[z;\ z_1;\ \ldots\ z_J]$ that account for the observed measurements. Numerical experiments support such an approach (see Figure 1).

B. JSM-2: Common sparse supports

Under the JSM-2 signal ensemble model, independent recovery of each signal via ℓ_1 minimization would require cK measurements per signal. However, we propose algorithms inspired by conventional greedy pursuit algorithms (such as OMP [9]) that can substantially reduce this number. In the single-signal case, OMP iteratively constructs the sparse support set Ω; decisions are based on inner products between the columns of $\Phi\Psi$ and a residual. In the multi-signal case, there are more clues available for determining the elements of Ω.

To establish a theoretical justification for our approach, we first propose a simple One-Step Greedy Algorithm (OSGA) that combines all of the measurements and seeks the largest correlations with the columns of the $\Phi_j\Psi$. We have established that, assuming that Φ_j has i.i.d. Gaussian entries and that the nonzero coefficients in the θ_j are i.i.d. Gaussian, then

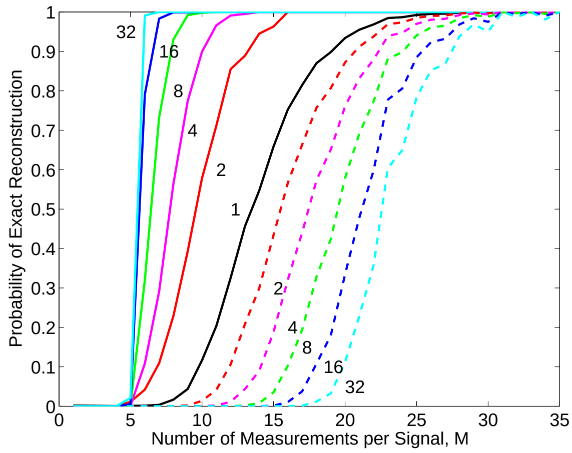

Fig. 2. *Reconstructing a signal ensemble with common sparse supports (JSM-2). We plot the probability of perfect reconstruction via DCS-SOMP (solid lines) and independent CS reconstruction (dashed lines) as a function of the number of measurements per signal M and the number of signals J. We fix the signal length to $N = 50$ and the sparsity to $K = 5$. An oracle encoder that knows the positions of the large signal expansion coefficients would use* 5 *measurements per signal.*

with $M \geq 1$ measurements per signal, OSGA recovers Ω with probability approaching 1 as $J \to \infty$. Moreover, with $M \geq K$ measurements per signal, OSGA recovers all x_j with probability approaching 1 as $J \to \infty$. This meets the theoretical lower bound for M_j.

In practice, OSGA can be improved upon by using an iterative greedy algorithm. We propose a simple variant of Simultaneous Orthogonal Matching Pursuit (SOMP) [11], which we term DCS-SOMP [14]. For this algorithm, Figure 2 plots the performance as the number of sensors varies from $J = 1$ to 32. We fix the signal lengths at $N = 50$ and the sparsity of each signal to $K = 5$. With DCS-SOMP, for perfect reconstruction of all signals the average number of measurements per signal decreases as a function of J. The trend suggests that, for very large J, close to K measurements per signal should suffice. On the contrary, with independent CS reconstruction, for perfect reconstruction of all signals the number of measurements per sensor *increases* as a function of J. This surprise is due to the fact that each signal will experience an independent probability $p \leq 1$ of successful reconstruction; therefore the overall probability of complete success is p^J. Consequently, each sensor must compensate by making additional measurements.

V. Extensions and Conclusions

We have taken the first steps towards extending the theory and practice of CS to multi-signal, distributed settings. Our simple joint sparsity models (JSMs) capture the essence of real physical scenarios, illustrate the basic analysis and algorithmic techniques, and indicate the gains to be realized from joint recovery. Additional investigations are ongoing, including additional models for joint sparsity, extensions to compressible signals, and examining the effect of noise and quantization in the measurements.

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

References

[1] 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.

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

[3] 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.

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

[5] 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.

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

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

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

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

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

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

[12] D. Baron, M. F. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, "An information-theoretic approach to distributed compressed sensing," in *Proc. 43rd Allerton Conf. Comm., Control, Comput.*, September 2005.

[13] M. B. Wakin, M. F. Duarte, S. Sarvotham, D. Baron, and R. G. Baraniuk, "Recovery of jointly sparse signals from few random projections," in *Proc. Neural Inform. Processing Systems – NIPS*, 2005.

[14] D. Baron, M. F. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Proc. 39th Asilomar Conf. Signals, Sys., Comput.*, November 2005.