Politecnico di Torino

Porto Institutional Repository

[Proceeding] Joint recovery algorithms using difference of innovations for distributed compressed sensing

(Article begins on next page)

Joint recovery algorithms using difference of innovations for distributed compressed sensing

Diego Valsesia
Politecnico di Torino
Email: diego.valsesia@polito.it

Giulio Coluccia
Politecnico di Torino
Email: giulio.coluccia@polito.it

Enrico Magli
Politecnico di Torino
Email: enrico.magli@polito.it

Abstract—Distributed compressed sensing is concerned with representing an ensemble of jointly sparse signals using as few linear measurements as possible. Two novel joint reconstruction algorithms for distributed compressed sensing are presented in this paper. These algorithms are based on the idea of using one of the signals as side information; this allows to exploit joint sparsity in a more effective way with respect to existing schemes. They provide gains in reconstruction quality, especially when the nodes acquire few measurements, so that the system is able to operate with fewer measurements than is required by other existing schemes. We show that the algorithms achieve better performance with respect to the state-of-the-art.

I. INTRODUCTION

Distributed compressed sensing (DCS) has recently attracted great interest as an efficient technique for acquiring data in a distributed fashion [1]–[3]. DCS relies on the theory of compressed sensing (CS) to reduce the dimensionality of the signal acquired by each node of the distributed network, supposed to be sparse under some basis, by means of random projections. On the other hand, it also exploits the inter-correlation among the different signals in the ensemble to lower the number of measurements that each node needs to acquire, without requiring cooperation among nodes. Sensor networks represent the foremost application that can benefit from this technique because of their need of simple signal representations while meeting strict low complexity constraints (*e.g.*, see [4]). A joint reconstruction algorithm can outperform separate recovery of the sensors' signals exploiting the correlation among them. Joint sparsity models have been proposed to account for sparsity patterns appearing in an ensemble of signals, e.g. signals acquired by nodes of a sensor network. According to the correlation model different algorithms have been proposed, such as the *Texas Hold 'Em* [5], and the *Sort* and *Intersect* algorithms [6] for the JSM-1 model, in which signals are composed of a sparse common component plus a sparse innovation component. Concerning other models, we can find the DCS-SOMP [2] method for JSM-2 (common sparse supports) and the TECC and ACIE algorithms for JSM-3 [2] (nonsparse common component, sparse innovations).

This paper proposes two novel joint reconstruction algorithms. The algorithms are based on the idea of using one signal in the ensemble as side information. This allows to devise recovery schemes that attempt to reconstruct the difference between a signal and the side information, rather than an individual signal. In particular, we show that the proposed algorithms obtain lower reconstruction error with respect to other existing algorithms [5] [6], and minimize the number of measurements that have to be collected by each node.

II. BACKGROUND

A. Compressed sensing

Compressed sensing is a novel theory for signal sampling and acquisition [7], [8]. It is able to acquire signals in an already compressed fashion, i.e., using fewer coefficients than is dictated by the classical Nyquist-Shannon theory. Let us consider a signal $\mathbf{x} \in \mathbb{R}^n$, having a sparse representation under basis $\Psi \in \mathbb{R}^{n \times n}$: $\mathbf{x} = \Psi\boldsymbol{\theta}$ with $\|\boldsymbol{\theta}\|_0 = k \ll n$, being $\|\boldsymbol{\theta}\|_0$ the l_0 norm of $\boldsymbol{\theta}$, *i.e.*, the number of its nonzero entries. We acquire measurements as a vector of random projections $\mathbf{y} = \Phi\mathbf{x} = \Phi\Psi\boldsymbol{\theta}$, $\mathbf{y} \in \mathbb{R}^m$, using a sensing matrix $\Phi \in \mathbb{R}^{m \times n}$. The best way to recover the original signal from its measurements is by solving an optimization problem trying to minimize the l_0 norm of the signal in the sparsity domain. However, this problem is computationally intractable due to its NP-hard complexity, so it is common to consider a relaxed form using the l_1 norm, which can be solved by means of linear programming techniques:

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

This method is equivalent to l_0 norm minimization provided that the sensing matrix satisfies the Restricted Isometry Property (RIP) with constant $\delta_{2k} < \sqrt{2} - 1$ [9]. The number of measurements to be acquired is typically $m = O\left(k \log \frac{n}{k}\right)$. A quadratically-constrained variant of the previous optimization problem is often used when dealing with noise with norm bounded by ε affecting the measurements:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_1 \quad \text{subject to} \quad \|\Phi\Psi\boldsymbol{\theta} - \mathbf{y}\|_2 \leq \varepsilon. \qquad (1)$$

B. Distributed compressed sensing

In a distributed scenario, an ensemble of signals with both intra- and inter-sensor correlations is considered. The notion of joint sparsity has been introduced in [2] for the framework of DCS. Among the joint sparsity models discussed in [2], we focus on the JSM-1 and JSM-3 models, according to which the J signals in the ensemble have sparse innovation components and sparse or nonsparse common component, respectively.

$$\boldsymbol{\theta}_j = \boldsymbol{\theta}_C + \boldsymbol{\theta}_{I,j} \quad \forall j \in [1, J]$$

$$\|\boldsymbol{\theta}_C\|_0 = k_C \quad \text{and} \quad \|\boldsymbol{\theta}_{I,j}\|_0 = k_{I,j} \quad \forall j \in [1, J]$$

A joint reconstruction algorithm can leverage the structure of the joint sparsity model to improve performance, namely to achieve higher quality for the same number of measurements or decrease the number of measurements needed to achieve the same quality.

Some of the existing joint reconstruction algorithms for the JSM-1 model include the *weighted l_1 minimization* proposed in [2],

which requires numerical optimization of the weights, the *Texas Hold 'Em* [5] and the *Sort* and *Intersect* algorithms [6]. *Texas Hold 'Em* averages a subset of the measurements of all the sensors to estimate the common component, which is then subtracted from the measurements to recover the innovations. *Sort* and *Intersect* also estimate the common component, in a nonlinear way. In particular, in *Sort* the coefficients of a first estimate of the signal are sorted by decreasing magnitude and compared to ones of the side information to decide which positions belong to the common support, while in *Intersect* the supports of a first estimate of the signal and of the side information are intersected to find the support of the common component.

As far as the JSM-3 model is concerned, the common component is not sparse, hence a joint recovery algorithm must be used in order to acquire fewer than n measurements per signal. The Transpose Estimation of Common Component (TECC) algorithm described in [2] estimates the common component by stacking all the measurements from all nodes in a single problem and then recover the innovations alone by cancelling the measurements of the estimated common component. A key requirement for TECC to work is having different sensing matrices Φ_j so that, when stacked, they span the whole $\mathbb{R}^n$. The Alternating Common and Innovation Estimation (ACIE) algorithm [2] is an improvement over TECC, based on an iterative scheme that alternates improvements on the estimate of the common component with improvements in the estimate of the innovations, at the expense of a great computational complexity. As in TECC, each node must have a different sensing matrix in order to work properly. This is different from the scenario we are considering, in which all nodes use the same sensing matrix.

III. PROPOSED ALGORITHMS

The proposed algorithms focus on the JSM-1 and JSM-3 models discussed in section II-B. We also suppose that side information is available at the decoder in the form of perfect knowledge of one of the signals. From now on we suppose, without loss of generality, that the known signal is $\mathbf{x}_1$. From a practical perspective, the requirement of side information is not a limitation; under the JSM-1 model the SI signal is sparse, so we can acquire $m_1 < n$ measurements, with m_1 large enough to get the desired accuracy. Under the JSM-3 model the SI signal is not sparse and has to be acquired uncompressed or compressed using a standard technique. However, as the number of nodes increases, the overhead due to side information becomes negligible. The savings in the number of measurements to be acquired by the other nodes outweigh the small overhead due to the acquisition of side information, thus making the framework interesting even for the JSM-3 model, where the SI signal is not compressed.

A. Difference-Of-Innovations (DOI) algorithm

Algorithm 1 *DOI* algorithm

Require: $A = \Phi\Psi$
  for j in $2 : J$ **do**
    Compute $\mathbf{y}_{\text{diff},j} = \mathbf{y}_j - \mathbf{y}_1$
    Recover $\boldsymbol{\theta}_{\text{diff},j}$ from $\mathbf{y}_{\text{diff},j}$
    $\hat{\boldsymbol{\theta}}_j = \boldsymbol{\theta}_1 + \boldsymbol{\theta}_{\text{diff},j}$
  end for

The underlying idea of this first algorithm is to exploit side information to eliminate the need to recover the common component.

Figure 1(a) presents a schematic representation of the algorithm, also outlined in Alg.1. Proceeding pairwise by using the side information and each of the J signals in the ensemble, it is possible to compute the difference between the measurements of the side information and those of signal j. This removes exactly any component that is common to the two signals, so we are left with measurements of the difference of the innovation components. It is then possible to recover the difference signal from these measurements using any recovery procedure such as l_1 minimization. Once the difference signal is recovered, it is then sufficient to add the side information to fully recover signal j. It is interesting to notice that, unlike [5], this algorithm does not introduce any error due to an inexact estimation of the common component. However, the difference signal is, in general, less sparse than the individual innovation component, so we can expect performance gains only when innovations are significantly sparser than the common component. In particular, as a rule of thumb, we expect gains for $k_C \geq 2k_{I,j}$. This condition is easily satisfied for highly correlated ensembles where there is a dominant common component and much sparser innovations. Also note that the *DOI* algorithm can be readily implemented in a parallel manner since any recovery only involves the side information and the measurements of the signal to be recovered.

B. Performance bound

In this section we evaluate a bound to the reconstruction error of the *DOI* algorithm. In particular, it can be shown that:

$$\left\| \hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j \right\|_2 = \left\| \boldsymbol{\theta}_{I,1} - \boldsymbol{\theta}_{I,j} + \boldsymbol{\theta}_{\text{diff},j} \right\|_2$$
$$= \left\| \boldsymbol{\theta}_{\text{diff},j} - (\boldsymbol{\theta}_{I,j} - \boldsymbol{\theta}_{I,1}) \right\|_2 \leq C\varepsilon \qquad (2)$$

where $\varepsilon = \left\| \mathbf{y}_{\text{diff},j} - \Phi (\mathbf{x}_{I,1} - \mathbf{x}_{I,j}) \right\|_2$ is the norm of the noise affecting the measurements of the difference signal and C is a constant that depends on the method used for reconstruction and on the RIP constant of the sensing matrix. When there is no quantization, or other sources of noise, we have $\varepsilon = 0$ and, provided that there are enough measurements available, reconstruction is perfect. This means that the *DOI* algorithm can achieve perfect reconstruction, unlike [5] which is limited by the residual noise in the averaging procedure.

C. Texas DOI *algorithm*

Algorithm 2 *Texas DOI* algorithm

Require: $J, A = \Phi\Psi$, k_I
  $\hat{\mathbf{y}}_C = \frac{1}{J} \sum\limits_{j=1}^{J} \mathbf{y}_j$
  $\hat{\mathbf{y}}_{I,1} = \mathbf{y}_1 - \hat{\mathbf{y}}_C$
  Recover $\hat{\boldsymbol{\theta}}_{I,1}$ from $\hat{\mathbf{y}}_{I,1}$
  for j in $2 : J$ **do**
    $\mathbf{y}_{\text{diff},j} = \mathbf{y}_j - \mathbf{y}_1$
    $\hat{\mathbf{y}}_{I,j} = \mathbf{y}_{\text{diff},j} + \hat{\mathbf{y}}_{I,1}$
    Recover $\hat{\boldsymbol{\theta}}_{I,j}$ from $\hat{\mathbf{y}}_{I,j}$
    $\hat{\boldsymbol{\theta}}_j = \boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_{I,1} + \hat{\boldsymbol{\theta}}_{I,j}$
  end for

The second proposed algorithm, called *Texas DOI*, attempts at overcoming the drawbacks of *DOI* and [5]. Figure 1(b) presents a schematic representation of the algorithm, also outlined in Alg.2. Albeit it maintains the idea of [5] of averaging a fraction of the collected measurements, it avoids any explicit reconstruction of the common component, but rather employs the measurements of

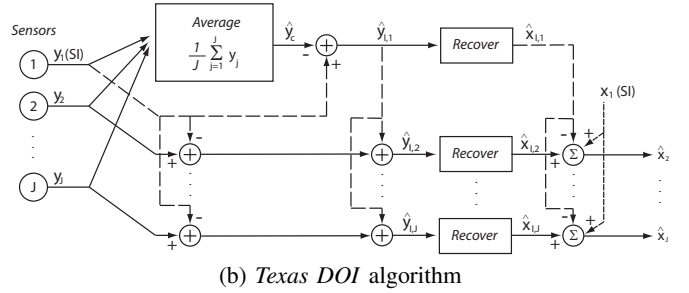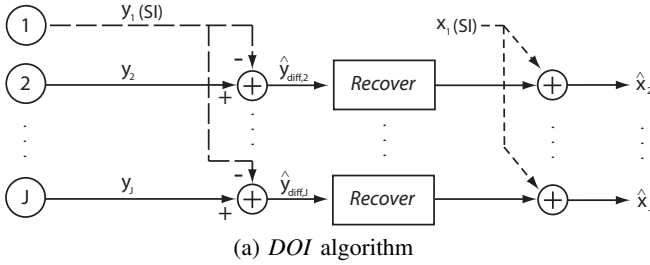(a) *DOI* algorithm   (b) *Texas DOI* algorithm

Fig. 1. Joint reconstruction algorithms

the common component, and combines this with the use of side information in a fashion similar to the *DOI* algorithm. In particular, side information is used to get differences of innovation components, but the measurements of the innovation component of the side information ($y_{I,1}$) can be obtained subtracting the output of the averaging procedure from the original SI measurements. This allows to perform recovery on the innovation component of a single signal by removing $y_{I,1}$. The algorithm can be implemented in a parallel manner if the averaging procedure and the recovery of $x_{I,1}$ are computed first. In fact all the remaining operations for the recovery of the signals only involve the available quantities related to the side information and the measurements of the signal to be recovered.

D. Performance bound

In this section we show a bound to the reconstruction error of the *Texas DOI* algorithm, whose proof is reported in the appendix. In particular, it can be shown that:

$$\left\| \hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j \right\|_2 \leq 2C \cdot \frac{\sqrt{1 + \delta_k}}{\sqrt{J}} \eta, \tag{3}$$

being δ_k the RIP constant of matrix $A = \Phi\Psi$, and $\|\boldsymbol{\theta}_{I,j}\|_2 = \eta$ for all $j \in [1, J]$. This analysis points out some interesting properties of the algorithm. Even if quantization or other sources of noise are not considered by the analysis, the algorithm is still affected by a certain reconstruction error, in the same way as [5]. This does not happen with *DOI*, which we showed in section III-B to be exact. Here, and in [5], the limiting factor lies in the averaging procedure that imposes a floor on the reconstruction error, which cannot be overcome by adjustments on the rate. However, this error floor decreases as $\frac{1}{\sqrt{J}}$, so the performance of an ensemble with a large number of signals may indeed be limited by the quantization rate or other sources of noise rather than the averaging procedure.

IV. EXPERIMENTAL RESULTS

We compared the two proposed algorithms with some of the existing joint reconstruction algorithms in the literature for the JSM-1 model such as *Texas Hold 'Em* [5], *Intersection* and *Sort* [6]. The simulations have been performed using a sensing matrix with i.i.d. zero-mean Gaussian entries with unit-norm columns and a JSM-1 ensemble of signals sparse in the identity basis. The amplitude of the nonzero entries of the signal is randomly generated according to a standard Gaussian distribution. Each measurement is quantized using R bits. For the *Texas Hold 'Em* algorithm all the measurements of each node are considered community measurements, thus contributing to the estimation of the common component. Figure IV shows that the proposed algorithms achieve lower MSE when few measurements

are available. *Texas DOI* can achieve the best performance but is not able to improve when m increases due to the error floor in (3). The *DOI* algorithm leverages the side information to remove the common component, but the recovery step is performed on the difference of measurements of the innovation components, hence it is outperformed by *Texas DOI* for low m. In fact, *Texas DOI* is able to run the compressed sensing recovery procedures only for the individual innovation components, whereas *DOI* has to recover the difference between two innovation signals, which is typically less sparse. When few measurements are available the algorithms in [6] may have trouble in recovering the common component because they rely on an initial estimate of the unknown signals. Hence the greatest gains are achieved when a limited number of measurements is available. The *Texas DOI* algorithm borrows ideas from both the *Texas Hold 'Em* strategy and the *DOI* procedure. *Texas DOI* inherits the averaging procedure from *Texas Hold 'Em*, which is very efficient when the number of nodes is large. However exploiting side information allows to improve over *Texas Hold 'Em* when the signals are highly correlated and few measurements are available because *Texas Hold 'Em* may have difficulties in recovering the common component. This allows each node to work closer to the minimum number of measurements needed by CS reconstruction for successful recovery.

It should be noted that *Texas Hold 'Em* does not use side information. However, the overhead due to acquiring more measurements to recover the side information with the desired accuracy is negligible in our simulations. As an example, in the case of $J = 100$, the side information can be recovered from $m_1 \cong 5k = 125$ measurements quantized with rate $R_1 = 8$. This means that the total number of bits is $(J - 1)mR + m_1 R_1$ for *DOI* and *Texas DOI* and $Jm'R$ for *Texas Hold 'Em*, with $m' > m$ in order to achieve $Jm'R = (J - 1)mR + m_1 R_1$. However, typically $m' - m < 1$, so no extra measurement has to be assigned to *Texas Hold 'Em* to compensate for the overhead of SI, which is completely negligible.

The algorithms have also been tested on the JSM-3 model of distributed compressed sensing and compared against the TECC algorithm presented in [2]. As explained in section III the proposed algorithms rely on the usage of the same sensing matrix for all nodes, while TECC requires different matrices. Figure VI shows the MSE as a function of the number of measurements acquired by each node. The proposed algorithms are able to outperform TECC and confirm the behaviour presented for the JSM-1 model.

V. CONCLUSIONS

We proposed two novel joint reconstruction algorithms for the JSM-1 and JSM-3 models in distributed compressed sensing. Thanks to the use of side information, it is possible to devise methods that
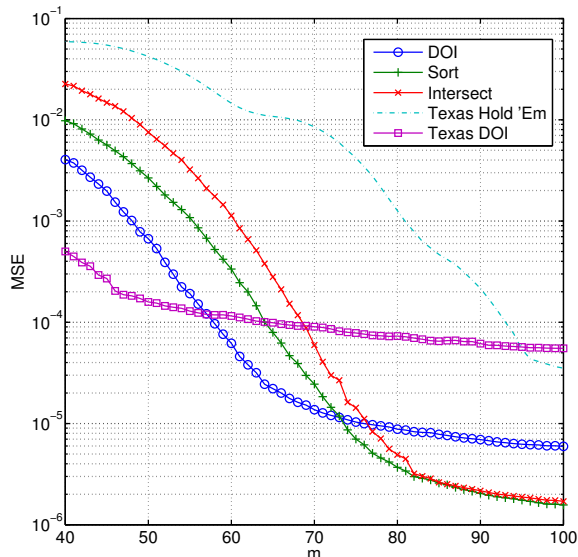
Fig. 2. JSM-1. Mean square error vs. number of measurements. $J = 100$, $k_C = 20$, $k_I = 5$, $n = 256$, $R = 8bps$

Fig. 3. JSM-3. Mean square error vs. number of measurements. $J = 100$, $k_I = 20$, $n = 256$, $R = 8bps$

avoid the need to reconstruct the common component, thus allowing to deal with the case of a non-sparse common component in a straightforward manner. The algorithms provide performance gains over other existing techniques, especially when few measurements are available, thus allowing to decrease the number of measurements needed to achieve a target quality in the reconstruction or to improve quality for the same number of measurements.

VI. APPENDIX

PROOF OF THE PERFORMANCE BOUND OF *Texas DOI*

$$\left\| \hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j \right\|_2 = \left\| \left(\boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_{I,1} + \hat{\boldsymbol{\theta}}_{I,j} \right) - \boldsymbol{\theta}_j \right\|_2$$
$$= \left\| \boldsymbol{\theta}_{I,1} - \hat{\boldsymbol{\theta}}_{I,1} - \boldsymbol{\theta}_{I,j} + \hat{\boldsymbol{\theta}}_{I,j} \right\|_2$$
$$= \left\| \left(\boldsymbol{\theta}_{I,1} - \hat{\boldsymbol{\theta}}_{I,1} \right) + \left(\hat{\boldsymbol{\theta}}_{I,j} - \boldsymbol{\theta}_{I,j} \right) \right\|_2$$
$$\leq \left\| \boldsymbol{\theta}_{I,1} - \hat{\boldsymbol{\theta}}_{I,1} \right\|_2 + \left\| \hat{\boldsymbol{\theta}}_{I,j} - \boldsymbol{\theta}_{I,j} \right\|_2 \quad (4)$$

Let us analyse how the innovation components are recovered.

$$\hat{\mathbf{y}}_{I,1} = \mathbf{y}_1 - \hat{\mathbf{y}}_C = \mathbf{y}_{I,1} - \frac{1}{J}\sum_{l=1}^{J}\mathbf{y}_{I,l}$$

$$\hat{\mathbf{y}}_{I,j} = \mathbf{y}_j - \mathbf{y}_1 + \hat{\mathbf{y}}_{I,1} = \mathbf{y}_{I,j} - \frac{1}{J}\sum_{l=1}^{J}\mathbf{y}_{I,l}$$

Let $\mathbf{n} = \frac{1}{J}\sum_{l=1}^{J}\mathbf{y}_{I,l}$ denote the error in the estimation of the common component, due to imperfect cancellation of the innovations. Hence, $\hat{\mathbf{y}}_{I,j} = A\boldsymbol{\theta}_{I,j} - \mathbf{n}$ for $j \in [1, J]$. Suppose that we use a reconstruction procedure (e.g. BPDN (1)) from noisy measurements that has a performance guarantee ensuring that the reconstruction error is proportional to the noise norm with a constant C depending on the reconstruction method and on the RIP constant of A. Assuming that enough measurements have been acquired so that A satisfies the RIP of order k with constant δ_k with high probability, that $\|\boldsymbol{\theta}_{I,j}\|_2 = \eta$ for all $j \in [1, J]$ and that the $\boldsymbol{\theta}_{I,j}$'s are pairwise orthogonal (i.e.

$\boldsymbol{\theta}_{I,j}^T \boldsymbol{\theta}_{I,k} = 0$ for $j \neq k$), we can write:

$$\left\| \hat{\boldsymbol{\theta}}_{I,j} - \boldsymbol{\theta}_{I,j} \right\|_2 \leq C \left\| \mathbf{n} \right\|_2 = C \left\| \frac{1}{J}\sum_{j=1}^{J} A\boldsymbol{\theta}_{I,j} \right\|_2$$
$$\leq C \frac{\sqrt{1+\delta_k}}{\sqrt{J}} \|\boldsymbol{\theta}_{I,j}\|_2 = C \frac{\sqrt{1+\delta_k}}{\sqrt{J}}\eta \quad (5)$$

where δ_k is the RIP constant of matrix $A = \Phi\Psi$. Finally, we can plug this result in (4) and we obtain:

$$\left\| \hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j \right\|_2 \leq \left\| \boldsymbol{\theta}_{I,1} - \hat{\boldsymbol{\theta}}_{I,1} \right\|_2 + \left\| \hat{\boldsymbol{\theta}}_{I,j} - \boldsymbol{\theta}_{I,j} \right\|_2$$
$$\leq 2C \cdot \frac{\sqrt{1+\delta_k}}{\sqrt{J}}\eta \quad (6)$$

REFERENCES

[1] M.R. Duarte, M.B. Wakin, D. Baron, and R.G. Baraniuk, "Universal distributed sensing via random projections," in *Information Processing in Sensor Networks, 2006. IPSN 2006. The Fifth International Conference on*, 0-0 2006, pp. 177 –185.

[2] D. Baron, M.F. Duarte, M.B. Wakin, S. Sarvotham, and R.G. Baraniuk, "Distributed compressive sensing," *Preprint*, Jan. 2009.

[3] M.F. Duarte, M.B. Wakin, D. Baron, S. Sarvotham, and R.G. Baraniuk, "Measurement bounds for sparse signal ensembles via graphical models," *Information Theory, IEEE Transactions on*, vol. 59, no. 7, pp. 4280–4289, 2013.

[4] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Proceedings of the 5th international conference on Information processing in sensor networks*, New York, NY, USA, 2006, IPSN '06, pp. 134–142, ACM.

[5] S.R. Schnelle, J.N. Laska, C. Hegde, M.F. Duarte, M.A. Davenport, and R.G. Baraniuk, "Texas hold 'em algorithms for distributed compressive sensing," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, march 2010, pp. 2886 –2889.

[6] G. Coluccia, E. Magli, A. Roumy, and V. Toto-Zarasoa, "Lossy compression of distributed sparse sources: a practical scheme," in *2011 European Signal Processing Conference (EUSIPCO-2011)*, Barcelona, Spain, Sept. 2011.

[7] E.J. Candes and M.B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[8] D.L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289 –1306, april 2006.

[9] E.J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 910, pp. 589 – 592, 2008.