Distributed Compressed Sensing for Sensor Networks Using Thresholding

Mohammad Golbabaee and Pierre Vandergheynst

Signal Processing Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

ABSTRACT

Distributed compressed sensing is the extension of compressed sampling (CS) to sensor networks. The idea is to design a *CS joint decoding* scheme at a central decoder (base station) that exploits the inter-sensor correlations, in order to recover the whole observations from very few number of random measurements per node. In this paper, we focus on modeling the correlations and on the design and analysis of efficient joint recovery algorithms. We show, by extending earlier results of Baron et al.,[1] that a simple thresholding algorithm can exploit the full diversity offered by all channels to identify a common sparse support using a near optimal number of measurements.

Keywords: Distributed compressed sensing, Joint sparse signals, Thresholding, Rademacher sequence, Gaussian chaos sequence.

1. INTRODUCTION

Recently, Baron et al.[1] have studied the theoretical performances of Distributed Compressed Sensing by defining a joint sparsity model (i.e. a common sparsity support throughout the channels plus innovations) and proposing a simple joint recovery algorithm (OSGA or simply Thresholding) which shows a massive improvement by increasing the number of channels. However, since their analysis is limited to infinitely large networks and does not have a close form, it is not entirely clear how effectively their algorithm takes advantage of the available *diversity* from various channels. In this paper, for the same signal model, we propose a new analysis of joint recovery by thresholding. Using concentration of measure techniques, we bound its performance by a closed form expression that is valid in both asymptotical and finite size configurations.

In particular, we study two types of measurement matrices with i.i.d. Gaussian/Bernoulli entries. However, thanks to the generality of concentration arguments, our results are valid for a large class of random matrices including the subgaussian ones. We have proved that by increasing the number of channels, the recovery probability improves exponentially. As a consequence, thresholding exploits the *full diversity* from all channels, to reconstruct the common support from the minimum number of the measurements per node that one could ever hope to achieve.

2. SIGNAL MODEL

As mentioned above, we consider a sensor network having K nodes. Each node k observes a sparse signal $x(k) \in \mathbb{R}^N$. According to JSM2 in,[1] observations of all nodes have the same sparsity support Λ (set of nonzero entries of $x(k)$) of cardinality S and we also suppose that all signal components are positive, i.e $x(k) \geq 0$. Each node uses its own random sensing matrix $\Phi(k) \in \mathbb{R}^{M\times N}$ to collect M noisy measurements $y(k) \in \mathbb{R}^M$ as follows:

$$y(k) = \Phi(k)x(k) + z(k) \quad (1)$$

$$= \Phi_\Lambda(k)\theta(k) + z(k) \quad k = 1, ..., K \quad (2)$$

where $\Phi_\Lambda(.)$ and $\theta(.)$ (nonzero coefficients) are just restrictions of $\Phi(.)$ and $x(.)$ to the joint support and $z(k)$ models either innovations or non-sparse components that we will simply treat as perturbations here.

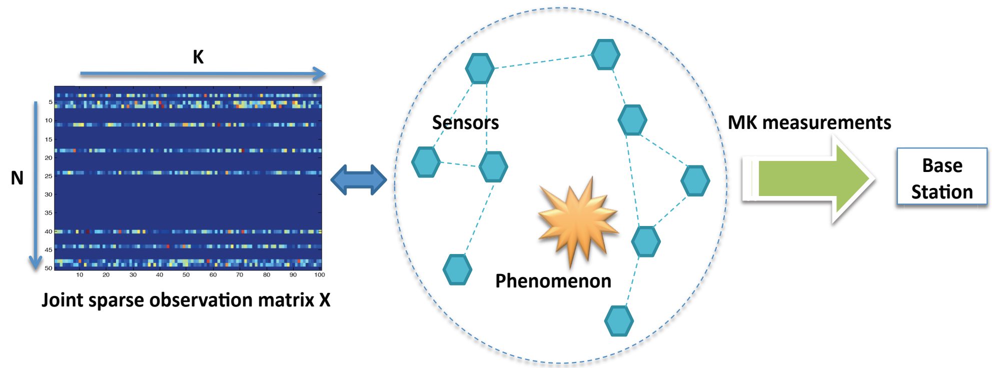


Figure 1: A sensor network with joint sparse correlated observations.

3. A CLASS OF JOINT DECODING ALGORITHMS

At the base station, the main idea is to design a CS joint decoding scheme which exploits the inter signal correlations (Joint sparsity assumption), in order to recover the whole joint support from very few number of measurements per node. Our algorithms are mainly inspired by *Thresholding*[2] (also known as Maximum correlation detection in[3]), which is the simplest greedy algorithm for sparse recovery and its extension to simultaneous sparse recovery, *p-thresholding*.[4] While these algorithms were initially proposed for joint recovery of sparse signals in a unique and deterministic dictionary, here we modify them for distributed CS applications, where dictionaries are random and also independent at each node. Both algorithms can be briefly summarized as follows.

Thresholding

1. For each node, compute $\rho \in \mathbb{R}^N$ which gives the correlations between its measurements and all columns of the corresponding sensing matrix,
$$\rho(k) = \Phi^*(k)y(k). \tag{3}$$
2. Build the vector of statistics ($f \in \mathbb{R}^N$) by summing up the rows of the correlation matrix $\Gamma \in \mathbb{R}^{\mathbb{N}\times\mathbb{K}}$ which contains all $\rho(.)$s as its columns,
$$f_i = \sum_{k=1}^{K} \rho_i(k) \quad i = 1, ..., N. \tag{4}$$
3. Sort out the S largest elements of the statistics vector. Their indices indicate the recovered support set $\hat{\Lambda}$.
4. For each channel, compute the nonzero coefficients by projecting its measurement vector to the corresponding recovered subspace, $\Phi_{\hat{\Lambda}}(.)$:
$$\hat{\theta}(k) = \Phi^{\dagger}_{\hat{\Lambda}}(k)y(k), \tag{5}$$
where by $A^{\dagger}$ we denote the pseudo inverse of the matrix A.

p-Thresholding

This algorithm is very similar to Thresholding, however to build the statistics in step 2, we use p-norming instead of summation:
$$f_i = \left(\sum_{k=1}^{K} |\rho_i(k)|^p\right)^{1/p} \qquad i = 1, ..., N. \tag{6}$$

Clearly this algorithm avoids the correlations with different signs to cancel out each other in the summation. This is a critical issue, specially, when the positivity assumption for nonzero coefficient does not hold.

4. MAIN RESULTS AND ANALYSIS

In this paper we focus on the analysis of the simple Thresholding algorithm, exploiting the positivity of the sparse signal components. Our main result is the following.

THEOREM 4.1. *Following the signal model described in Section 2 and using sensing matrices drawn from Bernoulli ensembles, the probability that Thresholding fails to recover perfectly the joint support is less than,*

$$P_f \leq N \exp\left(-\frac{R^2.SNR}{8(1+SNR)}\frac{MK}{S}\right), \tag{7}$$

where, SNR= $\frac{\sum_k \mathbb{E}\|\Phi_\Lambda(k)\theta(k)\|_2^2}{\sum_k \|z(k)\|_2^2} = \frac{\|\Theta\|_F^2}{\|Z\|_F^2}$ *and dynamic range* $R = \frac{\min_i \sum_j \theta_{ij}}{\sqrt{K/S}\|\Theta\|_F}$.

Corollary 1. To recover the joint support reliably ($P_f < \delta \ll 1$), it is sufficient that,

$$MK > \frac{8(1+SNR)}{R^2.SNR} S \log(N/\delta) \tag{8}$$

For the proof we use the following concentration inequality for sum of Rademacher sequences,[5]

THEOREM 4.2. *If* α *is an arbitrary real vector and* b *is a Rademacher sequence* ($b_i = \pm 1$ *with equal probability), then for any* $\tau > 0$

$$P\big(|\sum_i b_i \alpha_i| \geq \tau\big) \leq 2e^{-c\tau^2/\|\alpha\|_2^2} \tag{9}$$

where $c = 1/2$.

Proof of Theorem 4.1. We go sequentially through the following steps below:

1. All statistics concentrate around their averages. In order to show this, we rewrite statistic f_i as,

$$\begin{aligned} f_i &= \sum_k^K \langle \phi_i(k), y(k)\rangle && (10)\\ &= \sum_k^K \sum_n^N x_n(k)\langle \phi_i(k), \phi_n(k)\rangle + \sum_k^K \langle \phi_i(k), z(k)\rangle && (11)\\ &= \sum_k^K x_i(k)\|\phi_i(k)\|^2 + \sum_{k,n\neq i}^{K,N} x_n(k)\langle \phi_i(k), \phi_n(k)\rangle + \sum_k^K \langle \phi_i(k), z(k)\rangle && (12)\\ &= \sum_k^K x_i(k) + \frac{1}{M}\sum_{k,m,n\neq i}^{K,M,N} x_n(k)\underbrace{b_{mi}(k)b_{mn}(k)}_{b'^i_{mn}(k)} + \frac{1}{\sqrt{M}}\sum_{k,m}^{K,M} z_m(k)b_{mi}(k) && (13) \end{aligned}$$

where, b and b'^i are two independent Rademacher sequences. This comes from the fact that the product of two independent Bernoulli random variables is again a Bernoulli variable which is independent from both of them. As we can now see, each statistic can be written as a combination of its average, $\mathbb{E}f_i = \sum_k x_i(k)$, and a Rademacher series that comes from merging the rightmost terms of (13). Now, using Theorem 4.2, we can deduce their concentration around their averages,

$$\begin{aligned} P(|f_i - \mathbb{E}f_i| \geq \tau) &\leq 2\exp\left(-c\,\frac{M^2\tau^2}{\sum_{k,m,n\neq i} x_n^2(k) + M\sum_k z^2(k)}\right) && (14)\\ &\leq 2\exp\left(-cM\frac{\tau^2}{\|X\|_F^2 + \|Z\|_F^2}\right). && (15) \end{aligned}$$

Moreover, according to the signal model, since out-of-support coefficients are not active for all channels, $\|X\|_F = \|\Theta\|_F$ and the average term (which is the sum of the corresponding row of X) is zero for all out-of-support statistics, $\mathbb{E}f_{i\notin\Lambda} = 0$.

2. If we assume a minimum gap t between the average of in and out-of-support statistics, i.e.

$$t = \min_{i\in\Lambda} \mathbb{E}f_i - \max_{i\notin\Lambda} \mathbb{E}f_i \quad (16)$$

$$= \min_{i\in\Lambda} \sum_k x_i(k) > 0, \quad (17)$$

which is provided by the positivity of the components of Θ, then, for a successful recovery it is suffiecient that all statistics deviate from their averages by less than $t/2$:

$$P_s \geq P\big(\bigcap_i |f_i - \mathbb{E}f_i| < t/2\big). \quad (18)$$

Finally, together with the deviation bound (15), we can complete the proof as follows:

$$P_f \leq P\big(\bigcup_i |f_i - \mathbb{E}f_i| \geq t/2\big) \quad (19)$$

$$\leq \sum_i P(|f_i - \mathbb{E}f_i| \geq t/2) \quad (20)$$

$$\leq 2N \exp\left(-\frac{Mt^2}{8(\|\Theta\|_F^2 + \|Z\|_F^2)}\right). \quad (21)$$

Our next theorem extends this result to the recovery of joint sparse signals from noiseless measurements (i.e $z(k) = 0$) computed by Gaussian sensing matrices.

THEOREM 4.3. *Considering the signal model for the noiseless case together with using sensing matrices with Gaussian i.i.d entries, the probability that Thresholding fails to recover perfectly the joint support is less than,*

$$P_f \leq N \exp\left(-\frac{MK}{S}\frac{1}{c\ R^{-2} + c' R'^{-1}/\sqrt{S}}\right). \quad (22)$$

For some constants c, c' and dynamic ranges $R = \frac{\min_i \sum_j \theta_{ij}}{\sqrt{K/S}\|\Theta\|_F}$ and $R' = \frac{\min_i \sum_j \theta_{ij}}{K/\sqrt{S}\|\Theta\|_{2,\infty}}$. The matrix norm $\|\Theta\|_{2,\infty} = \max_k \|\Theta(k)\|$ denotes the maximum l_2 norm of the columns.

Corollary 2. In large problem setups, for a reliable recovery of the joint support ($P_f < \delta \ll 1$), it is sufficient that,

$$MK \gtrsim cR^{-2} S \log(N/\delta) \quad (23)$$

The key part of the proof uses the definition of Gaussian chaos random variables and large deviation inequalities for sum of their independent copies.

DEFINITION 4.4. *A Gaussian chaos random variable of order 2, h, is defined as,*

$$h = \sum_{i\neq j} \alpha_{ij} g_i g_j + \sum_i \alpha_{ii}(g_i^2 - 1) \quad (24)$$

where, α is a real/complex weight vector and g is an i.i.d. sequence of normal variables.

LEMMA 4.5. *Let h_k be a sequence of independent Gaussian chaos variables with corresponding weight vectors α_k, then for any $\tau > 0$,*

$$P\big(|\sum_k h_k| \geq \tau\big) \leq 2e^{-\frac{\tau^2}{V+U\tau}}. \tag{25}$$

Where, $V = 8e/\sqrt{6\pi}\sum_k \|\alpha_k\|^2$ and $U = 2\sqrt{2}e \max_k \|\alpha_k\|$.

The proof of this lemma is based on Bernstein's inequality for sum of independent zero mean random variables with a certain moments growth.[6] More details can be found in.[7]

Proof of Theorem 4.3. Like in Theorem 4.1 our proof has two steps, however, due to the similarities, we shall only discuss the first step and focus on showing how statistics are concentrating around their averages. Moreover, as before, averages are the corresponding row sum of the coefficient matrix X i.e. $\mathbb{E}f_i = \sum_k x_i(k)$. Regarding Definition 4.4, the subtraction of each statistic from its average can be expressed by the sum of independent Gaussian chaos variables (i.e. terms in the bracket below),

$$\begin{aligned} f_i - \mathbb{E}f_i &= \sum_k \sum_{n \neq i} x_n(k)\langle \phi_i(k), \phi_n(k)\rangle + \sum_k x_i(k)\left(\|\phi_i(k)\|^2 - 1\right) && (26)\\ &= \frac{1}{M}\sum_{k,m}\left\{\sum_{n\neq i} x_n(k) g_{mi}(k) g_{mn}(k) + x_i(k)(g_{mi}^2(k) - 1)\right\}. && (27)\end{aligned}$$

Hence, by Lemma 4.5, the deviations of statistics from their average can be bounded as follows,

$$\begin{aligned} P(|f_i - \mathbb{E}f_i| \geq \tau) &\leq 2\exp\left(-\frac{M^2\tau^2}{c\sum_{k,m}\|x(k)\|^2 + c' M\tau.\max_{k,m}\|x(k)\|}\right) && (28)\\ &\leq 2\exp\left(-\frac{M\tau^2}{c\|X\|_F^2 + c'\tau\|X\|_{2,\infty}}\right), && (29)\end{aligned}$$

where, $c = 8e/\sqrt{6\pi}$, $c' = 2\sqrt{2}e$. Now by carefully choosing τ as before, we bound the probability of failure by

$$P_f \leq 2N\exp\left(-\frac{Mt^2}{c\|\Theta\|_F^2 + c't\|\Theta\|_{2,\infty}}\right), \tag{30}$$

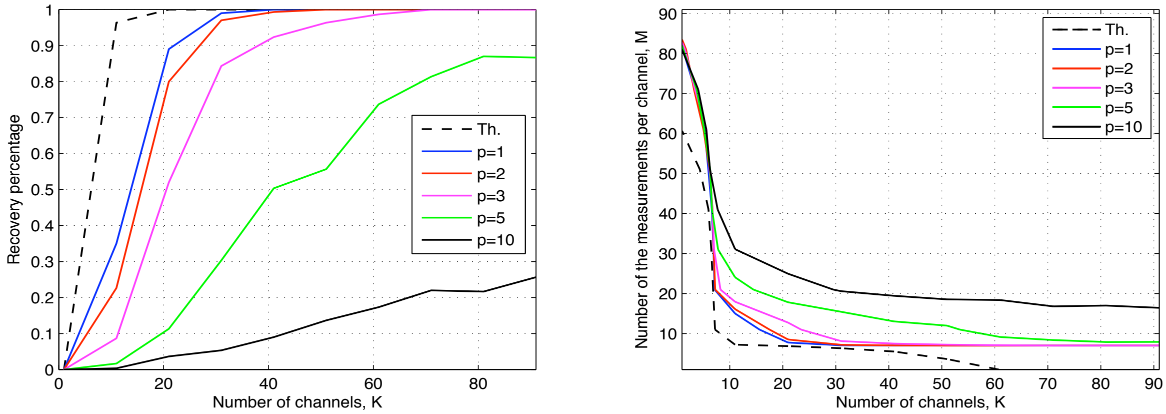
with $c = 32e/\sqrt{6\pi}$, $c' = 4\sqrt{2}e$. This, together with applying the definitions of R and R' completes the proof.

5. DISCUSSION

Theorem 4.1 shows that, by increasing the number of sensors, our algorithm becomes exponentially unlikely to fail. Moreover, from Corollary 1 we can deduce that a reliable recovery (i.e. $P_{fail} \leq \delta \ll 1$) requires the number of random measurements per node behaves as follows :

$$M \gtrsim O(S/K \log(N/\delta)). \tag{31}$$

This shows that, by growing the network size, each node needs to send much less measurements to our joint decoder, while maintaining correct support estimation. Unlike in,[1] this bound is not restricted only to the asymptotical cases ($K \to \infty$) : it points out clearly the tradeoff between K, M, S and N, for a reliable recovery, even in finite size sensor networks. It has to be noted that the positivity of the signal entries $x(k)$ was used technically in our main proof for the simple Thresholding algorithm. In case this assumption does not hold, one would have to resort to the p-thresholding algorithm. In this case, though we currently have no proof,



(a) (b)

Figure 2: Recovering the joint support, using Thresholding and p-thresholding. Simulation setup: Gaussian sensing matrices, signal length $N = 100$, sparsity $S = 5$, dynamic range $R = 1$ and 300 independent trials. (a) Recovery rate vs. Number of channels for $M = 10$ and several values p. (b) Recovery region for $P_s = 0.7$ and several values of p. The region above each line has recovery probability greater than P_s.

numerical results show that $p = 1$ seems to be an optimal choice and that results degrade for other choices (see Figure 2). However when positivity holds, Thresholding shows much faster improvement by adding more channels in comparison with p-thresholding. Further theoretical investigation of these observations is therefore necessary.

REFERENCES

1. D. Baron, M. B.Wakin, M. F.Duarte, S. Sarvotham, and R. G.Baraniuk, "Distributed compressed sensing for jointly sparse signals," in *Thirty-ninth Asilomar conference on signals, systems and computers*, **24**, pp. 1537–1541, 2005.
2. K. Schnass and P. Vandergheynst, "Average Performance Analysis for Thresholding," *IEEE Signal Processing Letters* **14**(11), pp. 828–831, 2007.
3. A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions on sparsity pattern recovery," *arXiv:0804.1839v1* , April 2008.
4. 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," *Journal of Fourier Analysis and Applications* **14**(5), pp. 655–687, 2008.
5. M. Talagrand and M. Ledoux, *Probability in Banach Spaces*, Springer-Verlag, Berlin, Heidelberg, NewYork, 1991.
6. G. Bennett, "Probability inequalities for sum of independent random variables," *Amer. Statist. Assoc.* **57**, pp. 33–45, 1962.
7. H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Transactions on Information Theory* **54**, pp. 2210–2219, May 2008.