Energy-Saving Gossip Algorithm for Compressed Sensing in Multi-Agent Systems

ENERGY-SAVING GOSSIP ALGORITHM FOR COMPRESSED SENSING IN MULTI-AGENT SYSTEMS

C. Ravazzi, S. M. Fosson, E. Magli

Department of Electronics and Telecommunications (DET), Politecnico di Torino, Italy

ABSTRACT

In this paper, we present a new recovery algorithm for in-network compressed sensing from measurements acquired in multi-agent systems. Each agent has to recover a common signal taking advantage of local communication and simple computations. Such distributed problem typically incurs a high energy cost due to inter-node communications. In this paper we propose an iterative distributed algorithm to address this problem, featuring pairwise gossip communications and updates. We propose some theoretical results on its dynamics and numerical comparisons with the most recent approaches proposed in literature. The performance turns out to be competitive in terms of reconstruction accuracy, complexity, and energy consumption required for convergence.

1. INTRODUCTION

Compressed sensing [1] is a universal technique used to represent sparse signals. It is possible to recover the original signal starting from few linear measurements, if they are acquired appropriately, using powerful optimization methods and reconstruction algorithms [2].

Compressed sensing is very appealing for applications in sensor networks and multi-agent systems, where distributed processing of the acquired data is needed due to large coverage [3, 4], and to little memory and available energy. In particular, although most distributed algorithms envisage a centralized reconstruction (*i.e.*, distributed measurements are collected at a central processing unit that performs the recovery), such solution is not always feasible, mainly for transmission costs and privacy reasons [5–7]. In this work, we consider in-network processing, *i.e.*, no central unit is available and the recovery task is performed by the agents, which can store a limited amount of information, perform a low number of operations, and communicate under some constraints.

The current literature offers various algorithms for distributed reconstruction. The most well known are the distributed subgradient algorithms (DSM [8–10]), the alternating direction method of multipliers (D-ADMM [4, 11, 12]), and the distributed iterative hard and soft thresholding (DIHT, CB-DIHT [13]; DISTA and DIHTA [6, 7]). In these algorithms, each agent keeps an approximation of the original signal and updates it iteratively according to both its own measurements and local information received from its neighbors. Although these methods are of significant interest, as they do away with a centralized processing unit, they still suffer from one or more of the following limitations [14]:

(a) *Synchronous updates*: the agents are assumed to send and process signals synchronously [6, 7, 9, 15].

(b) Selection of the *gradient stepsize*: we distinguish between constant and diminishing stepsizes. Constant stepsizes dot not guarantee convergence [15] or guarantee convergence only to neighborhoods of the optimal solution [7]. Larger stepsizes imply larger neighborhoods, while smaller ones produce slower convergence. On the other hand, diminishing stepsizes can ensure convergence, but a suitable design of them is very difficult and the convergence rate is always slow [16].

(c) *Spanning tree* and *multi-hop transmissions*: the agents first generate a spanning tree, which implies that they must be aware of the network's structure, in terms of position of the root and of their own roles (parents or children) with respect to their neighbors. Moreover, a routing protocol is necessary and multi-hop communication occurs at each iteration [13].

In order to overcome these limitations, in this paper we propose a gossip hard thresholding algorithm (GHT) which involves, at each time step, the activation of a randomly chosen pair of neighboring agents, requiring minimal coordination. In particular, GHT exhibits the following features: (a) is randomized and asynchronous; (b) despite being ruled by a constant stepsize, it converges deterministically to the original signal in the noise-free case, under some requirements on the number of measurements per node; (c) at each iteration step, it requires only local, single-hop communication.

In particular, in this paper we analyze the asymptotic behavior of GHT both in the noise-free and in the noisy settings. In the noise-free case, we propose a fixed-point analysis that provides insight into the algorithm's dynamics. In the noisy case, we show that the algorithm oscillates in time around a

This work is supported by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement n.279848.

certain limit value. Such oscillations can be smoothed by a suitable time-averaging process, which can be performed in an asynchronous way.

2. PROBLEM STATEMENT

We introduce some notations that we will use in this paper. We denote column vectors with small letters, and matrices with capital letters. If $x \in \mathbb{R}^n$ we denote its j-th element as x^j. Given a matrix X, X^T denotes its transpose and $X_{(v)}$ (or x_v) denotes the v-th column of X. For a square matrix $M \in \mathbb{R}^{n\times n}$, we consider the induced norm $\|M\|_2 = \sup_{z\neq 0} \|Mz\|_2/\|z\|_2$.

2.1. Sparse recovery problem in multi-agent systems

In our model, we consider a network of N agents, which may be sensors or nodes that collect measurements from different sensors. We assume that each agent $v \in \mathcal{V} = \{1, \dots, N\}$ in the network senses a common signal $x^\star \in \mathbb{R}^n$ and has available $m \leq n$ linear measurements of the form

$$y_v = A_v x^\star + \xi_v, \tag{1}$$

where $y_v \in \mathbb{R}^m$, $A_v \in \mathbb{R}^{m\times n}$, and $\xi_v \in \mathbb{R}^m$ is a bounded, unknown perturbation term. The agents seek to estimate $x^\star \in \mathbb{R}^n$ under the assumption that $x^\star$ is k-sparse (*i.e.*, it has at most k nonzero entries). It is thus natural to consider the following optimization problem in order to approximate $x^\star$:

$$\min_{x\in\mathbb{R}^n} \sum_{v\in\mathcal{V}} \|y_v - A_v x\|_2^2 \qquad \text{s.t. } \|x\|_0 \leq k. \tag{2}$$

It can be shown that in absence of noise (*i.e.* $\xi_v = 0, \forall v \in \mathcal{V}$), if for every index set $\Gamma \subseteq \{1, \dots, n\}$ with $|\Gamma| = 2k$ the columns of $A = (A_1^\mathsf{T}, \dots, A_N^\mathsf{T})^\mathsf{T}$ associated with Γ are linearly independent, then $x^\star$ is the unique solution to (2) [1].

The optimization problem in (2) is NP-hard [1], and the literature offers plenty of algorithms to approximate its solution (see [2] for a survey). Among them, we now review the iterative hard thresholding (IHT, [17, 18]), a centralized method upon which we will build our algorithm.

2.2. Iterative hard thresholding

Let $y = (y_1^\mathsf{T}, \dots, y_N^\mathsf{T})^\mathsf{T}$ and $A = (A_1^\mathsf{T}, \dots, A_N^\mathsf{T})^\mathsf{T}$. Given an initial estimate $x(0)$, the IHT iterate for $t = 0, 1, 2 \dots$ is

$$x(t+1) = \sigma_k[x(t) + \tau A^\mathsf{T}(y - Ax(t))] \tag{3}$$

where $\tau > 0$, and the operator $\sigma_k(x)$ is the best-k term approximation to vector x: $\sigma_k(x) = \text{argmin}\{\|x - z\|_2 : \|z\|_0 \leq k\}$. Convergence of this algorithm to a local minimum is proved in [17] under the assumption that $\|A\|^2 < 1/\tau$ and a stronger recovery result has recently been shown in [18]. In particular, if $x^\star$ is a k-sparse signal sensed according to (1) and A satisfies the restricted isometry property (RIP, [19]) with $\delta_{3k} < 1/\sqrt{32}$, then the sequence $\{x(t)\}_{t\in\mathbb{N}}$ generated by (3) with $\tau = 1$ is such that

$$\|x(t) - x^\star\|_2 \leq 2^{-t}\|x^\star\|_2 + 5\|\xi\|_2 \tag{4}$$

where $\xi = (\xi_1^\mathsf{T}, \dots, \xi_N^\mathsf{T})^\mathsf{T}$. In absence of noise, (4) guarantees the convergence to the original signal $x^\star$.

3. GOSSIP HARD THRESHOLDING (GHT)

3.1. Algorithm description

In this section, we introduce our proposed algorithm. From now on, we consider a connected network and we model it by an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{E} \in \mathcal{V} \times \mathcal{V}$ represents the set of the available communication links. The set of neighbors of $v \in \mathcal{V}$ is denoted as $\mathcal{N}_v = \{w \in \mathcal{V} : (v, w) \in \mathcal{E}\}$. We assume that $(v, v) \in \mathcal{E}$ for all $v \in \mathcal{V}$.

Algorithm 1 GHT

Initialization: $x_v(0) = 0 \in \mathbb{R}^n$ for any $v \in \mathcal{V}$, $\tau > 0$
for $t = 0, 1, \dots, \mathit{StopIter}$ **do**
  Select uniformly at random an edge $(v, w) \in \mathcal{E}$
  $x_v(t+1) = \sigma_k\left[\frac{x_v(t)+x_w(t)}{2} + \tau A_v^\mathsf{T}(y_v - A_v x_v(t))\right]$
  $x_w(t+1) = \sigma_k\left[\frac{x_v(t)+x_w(t)}{2} + \tau A_w^\mathsf{T}(y_w - A_w x_w(t))\right]$
  $x_h(t+1) = x_h(t)$ for any $h \neq v, w$
end for

The GHT algorithm, which is summarized in Algorithm 1, consists of the following iterative procedure. At each time step $t = 0, 1, \dots, \mathit{StopIter}$, each agent $v \in \mathcal{V}$ holds an estimate $x_v(t) \in \mathbb{R}^n$ of the original signal, starting from $x_v(0) = 0$. At time t, an edge $(v, w) \in \mathcal{E}$ is selected according to a discrete-time random process $\theta(t) \in \mathcal{E}$. More specifically, we assume that the sequence $\{\theta(t)\}_{t>0}$ is i.i.d., and that its probability distribution is uniform: $\mathbb{P}(\theta(t) = (v, w)) = 1/|\mathcal{E}|$ (see [20]). This choice does not entail any loss of generality; other distributions can be considered, provided that no disconnected communication clusters are generated. When edge (v, w) is selected, the agents v and w can communicate and share their own estimates. Once communication has been completed, both v and w compute the mean $\frac{x_v(t)+x_w(t)}{2}$, and they add to it their respective individual gradients $A_v^\mathsf{T}(y_v - A_v x_v(t))$ and $A_w^\mathsf{T}(y_w - A_w x_w(t))$ (multiplied by an opportune gradient parameter $\tau > 0$). The new signal's estimate for both v and w is obtained by taking the best k-term approximation of the result. At the same time t, all the other agents do not wake up and do not change their own estimates (see Figure 1 for an illustrative example). The procedure is repeated for a number of times that guarantees the complete spread of the information over the network.

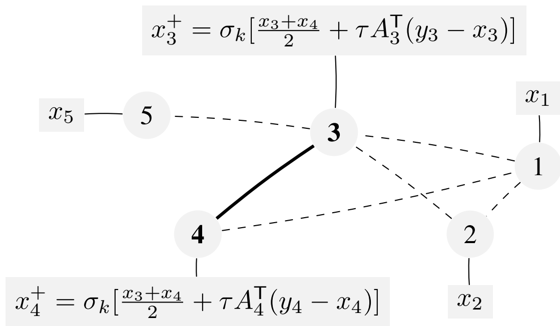

Fig. 1. Example of a network with 5 nodes, where $\theta = (3, 4)$, that is, the link $(3, 4)$ is activated, and only the states of nodes 3 and 4 evolve.

3.2. Limit behavior

3.2.1. Noise-free case

If $\xi_v = 0$ for any $v \in \mathcal{V}$, simulation results show that if the number of total measurements is large enough, then for any $v \in \mathcal{V}$, $x_v(t)$ asymptotically converges, in a deterministic sense, to the original signal $x^\star$. In Figure 2 we show such behavior in an example with $n = 150$, $k = 15$, $m = 10$, $N = 10$, over a complete graph.

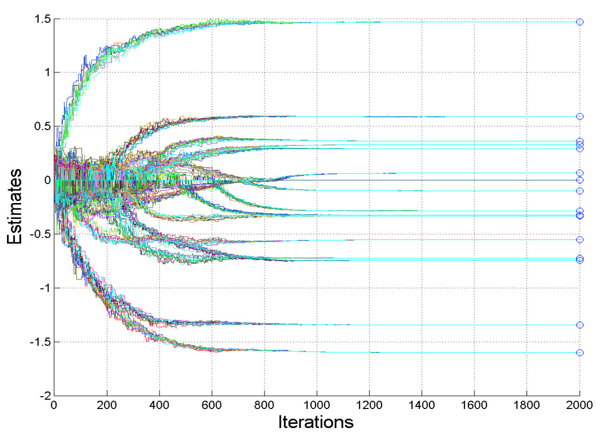

Fig. 2. Noise-free example ($n = 150$, $k = 15$, $m = 10$, $N = 10$, complete graph): for any $v \in \mathcal{V}$, the estimates $x_v(t)$ converge to $x^\star$ (whose components are marked by blue circles).

Let us consider for each $(v, w) \in \mathcal{E}$ the map $f_{(v,w)} : \mathbb{R}^{n \times N} \to \mathbb{R}^{n \times N}$ which acts on $X = (x_1, \dots, x_N)$ as in Algorithm 1, the set of maps $\mathfrak{F} = \{f_{(v,w)} : (v, w) \in \mathcal{E}\}$ and the set of fixed points of $\Gamma(\mathfrak{F}) = \{X \in \mathbb{R}^{n \times N} : f_{(v,w)}(X) = X, (v, w) \in \mathcal{E}\}$. It should be noticed that $x^\star \in \Gamma(\mathfrak{F})$. We now provide the conditions to guarantee that $x^\star$ is the unique fixed point of $\mathfrak{F}$. The proof is omitted for brevity.

Theorem 1 (Fixed points)**.** *If $\mathcal{G}$ is connected, then for any $X = \Gamma(\mathfrak{F})$, there exists $x \in \mathbb{R}^n$ such that $X = x\mathbb{1}^\intercal$, and*

$$x = \sigma_k \left(x + \frac{\tau}{N} A^\intercal (y - Ax) \right). \tag{5}$$

Theorem 2 (Uniqueness of fixed point)**.** *If $\mathcal{G}$ is connected and $A/\sqrt{N}$ satisfies the RIP [19] with restricted isometry constant $\delta_{2k} < 1/3$ and $2(1 + \delta_{2k}) < \tau^{-1} < \frac{8}{3}(1 - \delta_{2k})$ then $\Gamma(\mathfrak{F}) = \{x^\star \mathbb{1}^\intercal\}$.*

3.2.2. Noisy case

When noise occurs, *i.e.*, $\xi_v \neq 0$, the dynamics in Algorithm 1 oscillates and does not converge in a deterministic sense. This is not surprising, as $x^\star$ is no more a fixed point of $\mathfrak{F}$. However, the oscillations seem to asymptotically concentrate about a mean value, that approximates $x^\star$. These oscillations can be smoothed out by performing a time-averaging operation as in [21], which is reported in Algorithm 2 and must be applied after instruction 6 of Algorithm 1. This inner-loop just requires that each agent individually stores the number of times it has woken up in the variable $\kappa_v(t)$ and uses it to construct a time-averaged estimate $\widetilde{x}_v(t)$; no knowledge of global clocks or any other global variables is needed.

We show an example in Figure 3, where the smoothing effect can be appreciated: for any v, $\widetilde{x}_v(t)$ converges in a neighborhood of $x^\star$. The analysis of the ergodicity of the dynamics Algorithm 1 in case of noisy measurements and, consequently, the convergence of $\widetilde{x}_v(t)$ is left for future research. We refer to [22] for an overview of ergodic dynamics over networks.

Algorithm 2 Smoothing procedure

Require: $\theta(t) = (v, w)$, $x_v(t + 1)$, $x_w(t + 1)$
1: $\kappa_v(t + 1) = \kappa_v(t) + 1$
2: $\kappa_w(t + 1) = \kappa_w(t) + 1$
3: $\kappa_h(t + 1) = \kappa_h(t)$ for any $h \neq v, w$
4: $\widetilde{x}_v(t + 1) = \frac{1}{\kappa_v(t+1)}(\kappa_v(t)\widetilde{x}_v(t) + x_v(t + 1))$
5: $\widetilde{x}_w(t + 1) = \frac{1}{\kappa_w(t+1)}(\kappa_w(t)\widetilde{x}_w(t) + x_w(t + 1))$
6: $\widetilde{x}_h(t + 1) = \widetilde{x}_h(t)$ for any $h \neq v, w$

4. PERFORMANCE ANALYSIS

In this section, we present some experimental results that describe the performance of GHT in terms of reconstruction, convergence times and energy consumption. In the first case, the signal $x^\star$ to be recovered is generated choosing $k = 15$ nonzero components uniformly at random among the $n = 150$ elements and drawing the amplitude of each nonzero component from a standard Gaussian distribution. The sensing matrices $\{A_v\}_{v \in \mathcal{V}}$ are sampled from the Gaussian ensemble with m rows, n with zero mean

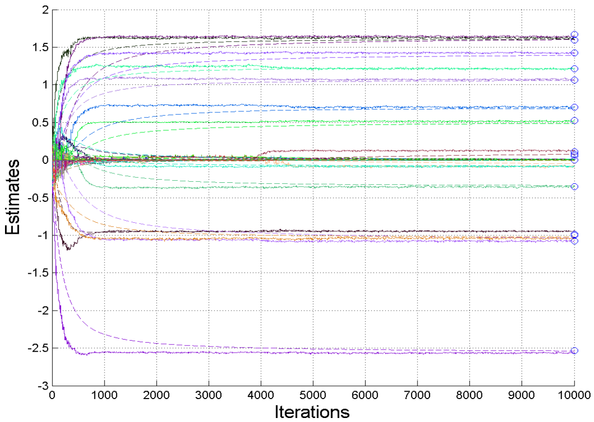

Fig. 3. Noisy case ($n = 150$, $k = 15$, $m = 10$, $N = 10$, ξ_v is a white Gaussian noise with variance 0.01, complete graph): the dynamics of one agent's estimate is depicted; the oscillations are ergodic and the estimates $\widetilde{x}_v(t)$ converge to a neighborhood of $x^\star$ (marked by blue circles).

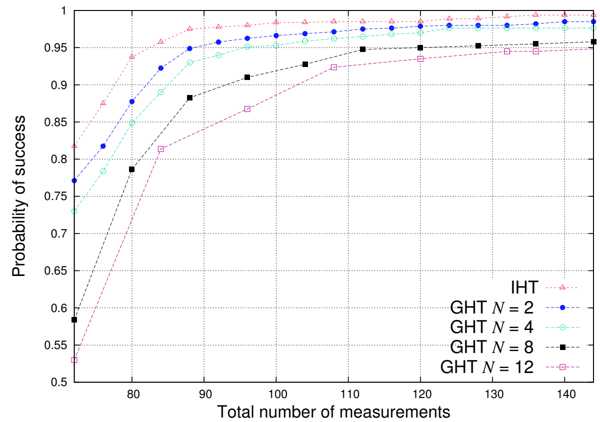

Fig. 4. Complete graph: empirical recovery success as a function of total number of measurements available in the network.

and variance $1/m$. We evaluate the performance of GHT in a noise-free setting, in terms of empirical recovery success rate, averaged over 800 experiments, as a function of the total number of measurements (see Figure 4). The recovery is considered successfully when the reconstruction error $\mathsf{E} := \sqrt{\sum_{v=1}^{N} \|x_v - x^\star\|_2^2} \Big/ \left(\sqrt{N}\|x^\star\|_2\right)$ (where x_v, $v = 1\ldots, N$, are the estimates given by GHT) is below the threshold 10^{-5}. We select a complete topology with network size $N \in \{1, 2, 4, 8, 12\}$ and set the stepsize τ approximately to $\|A_v\|_2^{-2}$. If $N = 1$, clearly GHT coincides with the IHT. Figure 4 shows a slight loss of performance as the network size increases. This suggests that, in addition to Theorem 2, some conditions on the number of measurements per sensor should hold to allow convergence to $x^\star$. We however observe that success rates are achieved by GHT even when the total number of measurements is very small; if $mN \approx 2n/3$, then successful recovery is achieved in 90% of cases.

As a second experiment, based on the recent study in [13], we consider the problem Sparco 7 [23], [13, Section V], where the signal to be estimated is a sign spike of length $n = 2560$, sparsity level $k = 20$, in a network of size $N = 40$ with $m = 15$ measurements per node (see [13, Table I]). GHT is implemented on different random graph topologies: Erdos-Rényi (ER, [24]) graphs with parameter $p = 0.25, 0.75$, and random geometric (RG, [24]) graphs with diameter 0.5 and 0.75. We fix $\tau = 0.05$ for ER, and $\tau = 0.07$ for RG. In Tables 1 and 2 we show the total number of transmissions required to achieve an accuracy level $\mathsf{E} < \epsilon$, where $\epsilon = 10^{-2}, 10^{-5}$, respectively. It is worth remarking that the number of transmission is directly related to energy consumption, since in sensor networks most of the energy is spent in over-the-air transmission. GHT is here compared with the state-of-the-art distributed methods: DIHT, CB-DIHT [13], D-ADMM [11, 12], and DSM [25] (the results of these methods are drawn from [13, Tables II and II]). We can notice that GHT significantly outperforms the other methods in terms of number of transmissions and hence energy consumption.

Graph	**GHT**	DIHT	D-ADMM	CB-DIHT	DSM
ER $p = 0.25$	$\mathbf{3.2 \cdot 10^4}$	$1.6 \cdot 10^7$	$3.1 \cdot 10^8$	$1.2 \cdot 10^{10}$	$> 3 \cdot 10^{11}$
ER $p = 0.75$	$\mathbf{3.9 \cdot 10^4}$	$1.6 \cdot 10^7$	$1.8 \cdot 10^9$	$3.8 \cdot 10^{10}$	$> 9 \cdot 10^{11}$
RG $d = 0.5$	$\mathbf{2.8 \cdot 10^4}$	$1.6 \cdot 10^7$	$1.2 \cdot 10^8$	$5.8 \cdot 10^{10}$	$> 1 \cdot 10^{11}$
RG $d = 0.75$	$\mathbf{3.1 \cdot 10^4}$	$1.6 \cdot 10^7$	$7.5 \cdot 10^8$	$2.6 \cdot 10^{10}$	$> 1 \cdot 10^{12}$

Table 1. Sparco 7: Total number of transmissions to get accuracy level 10^{-5}.

Graph	**GHT**	DIHT	D-ADMM	CB-DIHT	DSM
ER $p = 0.25$	$\mathbf{2.4 \cdot 10^4}$	$6.1 \cdot 10^6$	$2.5 \cdot 10^7$	$3.8 \cdot 10^9$	$> 3 \cdot 10^{11}$
ER $p = 0.75$	$\mathbf{3.0 \cdot 10^4}$	$6.1 \cdot 10^6$	$9.3 \cdot 10^8$	$1.2 \cdot 10^{10}$	$3.6 \cdot 10^{10}$
RG $d = 0.5$	$\mathbf{2.4 \cdot 10^4}$	$6.1 \cdot 10^6$	$1.2 \cdot 10^8$	$7.4 \cdot 10^9$	$> 1 \cdot 10^{11}$
RG $d = 0.75$	$\mathbf{2.8 \cdot 10^4}$	$6.1 \cdot 10^6$	$4.5 \cdot 10^8$	$8.0 \cdot 10^9$	$6.2 \cdot 10^{10}$

Table 2. Sparco 7: Total number of transmissions to get accuracy level 10^{-2}.

5. CONCLUDING REMARKS

In this paper, we have presented a distributed, gossip algorithm for in-network compressed sensing, which dramatically reduces the number of necessary transmissions and overcomes synchronization issues. The algorithm is shown to deterministically converge to the right solution under some conditions, which have been investigated in a number of simulations. Some analytical observations about the fixed points of the procedure have been provided, while a deeper insight on the convergence properties is the focus of our current research.

6. REFERENCES

[1] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207 – 1223, 2006.

[2] M. Fornasier, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, Radon Series on Computational and Applied Mathematics, 2010.

[3] J.-J. Xiao, A. Ribeiro, Z.-Q. Luo, and G.B. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Proc. Mag.*, vol. 23, no. 4, pp. 27 – 41, 2006.

[4] P. A. Forero, A. Cano, and G. B. Giannakis, "Consensus-based distributed support vector machines," *J. Mach. Learn. Res.*, vol. 99, pp. 1663 – 1707, 2010.

[5] G. Mateos, J. A. Bazerque, and G. B. Giannakis, "Distributed sparse linear regression," *IEEE Trans. Signal Proc.*, vol. 58, no. 10, pp. 5262 – 5276, 2010.

[6] C. Ravazzi, S. M. Fosson, and E. Magli, "Distributed soft thresholding for sparse signal recovery," in *Proc. of IEEE GLOBECOM*, 2013.

[7] C. Ravazzi, S. M. Fosson, and E. Magli, "Distributed iterative thresholding for l_0/l_1-regularized linear inverse problems," *submitted*, 2013.

[8] S. Sundhar Ram, A. Nedić, and V. V. Veeravalli, "Distributed subgradient projection algorithm for convex optimization," in *Proc. of IEEE ICASSP*, 2009, pp. 3653 – 3656.

[9] S. Sundhar Ram, A. Nedić, and V. V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *Journal of Optimization Theory and Applications*, pp. 516 – 545, 2010.

[10] A. Nedić, A. Ozdaglar, and P. A. Parrillo, "Constrained consensus and optimization in multi-agent networks," *IEEE Trans. Automat. Contr.*, vol. 55, pp. 922 – 938, 2010.

[11] J.F.C. Mota, J.M.F. Xavier, P.M.Q. Aguiar, and M. Puschel, "Distributed basis pursuit," *IEEE Trans. Signal Proc.*, vol. 60, no. 4, pp. 1942 – 1956, 2012.

[12] J.F.C. Mota, J.M.F. Xavier, P.M.Q. Aguiar, and M. Puschel, "D-ADMM: A communication-efficient distributed algorithm for separable optimization," *IEEE Trans. Signal Proc.*, vol. 61, no. 10, pp. 2718 – 2723, May 2013.

[13] S. Patterson, Y. C. Eldar, and I. Keidar, "Distributed compressed sensing for static and time-varying networks," *submitted*, 2013.

[14] J. Lu, C. Y. Tang, P. R. Regier, and T. D. Bow, "Gossip algorithms for convex consensus optimization over networks," *IEEE Trans. Automat. Contr.*, vol. 56, no. 12, pp. 2917 – 2923, 2011.

[15] S. Sundhar Ram, A. Nedić, and V. V. Veeravalli, "Asynchronous gossip algorithm for stochastic optimization: Constant stepsize analysis," in *Recent Advances in Optimization and its Applications in Engineering*, pp. 51 – 60. Springer Berlin Heidelberg, 2010.

[16] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847 – 1864, 2010.

[17] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 629 – 654, 2004.

[18] M. E. Blumensath, T.and Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265 – 274, 2009.

[19] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences. Paris, France, ser. I*, vol. 346, pp. 589 – 592, 2008.

[20] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2508 – 2530, 2006.

[21] B. Polyak and A. Juditsky, "Acceleration of stochastic approximation by averaging," *SIAM Journal on Control and Optimization*, vol. 30, no. 4, pp. 838–855, 1992.

[22] C. Ravazzi, P. Frasca, R. Tempo, and H. Ishii, "Ergodic randomized algorithms and dynamics over networks," *CoRR*, vol. abs/1309.1349, 2013.

[23] E. van den Berg, M. P. Friedlander, G. Hennenfent, F. Herrmann, R. Saab, and Ö. Yılmaz, "Sparco: A testing framework for sparse reconstruction," Tech. Rep. TR-2007-20, Dept. Computer Science, University of British Columbia, Vancouver, 2007.

[24] B. Bollobas, *Random Graphs*, London: Academic Press, 1985.

[25] I. Lobel and A. Ozdaglar, "Distributed subgradient methods for convex optimization over random networks," *IEEE Trans. Automat. Contr.*, vol. 56, no. 6, pp. 1291 – 1306, 2011.