Distributed Support Detection of Jointly Sparse Signals

DISTRIBUTED SUPPORT DETECTION OF JOINTLY SPARSE SIGNALS

Sophie M. Fosson[⋆] *Javier Matamoros*[†] *Carles Antón-Haro*[†] *Enrico Magli*[⋆]

[⋆] Department of Electronics and Telecommunications, Politecnico di Torino (Italy)
[†]Centre Tecnològic de Telecomunicacions de Catalunya, Barcelona (Spain)

ABSTRACT

In this paper, we address the problem of distributed support detection of multiple sparse signals with common support. Specifically, signals are acquired by the individual nodes of a network according to the so-called Joint Sparsity Model 2 (JSM-2). By leveraging on this model, we propose a distributed scheme for in-network signal recovery, i.e. not requiring data gathering and processing at a fusion center, based on distributed iterative thresholding and consensus strategies. For the proposed scheme, whose convergence properties we rigorously prove, no a priori knowledge on the non-zero number of entries in the signal vector is required.

Index Terms— Compressed sensing, joint sparsity, distributed algorithms, iterative thresholding, consensus.

1. INTRODUCTION

From Compressed Sensing (CS) theory [1], it is well known that an *individual* sensor in a network is able to recover the acquired signal from a reduced number of (its own) measurements vectors as long as its sensing matrix satisfy the Restricted Isometry Property [2]. This can be accomplished by resorting to (i) optimization-based approaches (see [3] for an overview); or (ii) greedy algorithms such as Orthogonal Matching Pursuit (OMP) [4], Subspace Pursuit (SP) [5] or Iterative Hard Thresholding (IHT) [6], to name a few. Greedy schemes are iterative in nature, exhibit lower complexity but, in general, are suboptimal.

However, we can do better by exploiting the inherent correlation of measurements over sensors. This results into the the so-called Multiple Measurement Vector (MMV) problem formulation, by virtue of which the measurements collected by all sensors are arranged as columns of a matrix. The Joint Sparsity Model 2 (JSM2) assumes that the acquired signals have a *common* support and, thus, only very few *rows* in that matrix have non-zero entries. For centralized settings, where all sensor measurements are collected by a Fusion Center (FC), the problem can be efficiently solved by resorting to e.g. row-based Lasso formulations [7] which introduce a penalty term on row energies in the minimization of the residual. Centralized settings, however, have a number of drawbacks: they imply the availability of such FC, sensors must convey their measurements and measurement matrices to it (which can be barely recommended for energy efficiency or privacy reasons) and, more importantly, a FC failure would prevents signal recovery. All this can be avoided by means of *in-network* approaches which allow for distributed support detection (and signal recovery) without the intervention of a FC. They leverage on *local* processing of sensor measurements and, possibly, (short-range) signaling with neighboring nodes. A number of distributed versions of the aforementioned optimization-based or greedy schemes have been proposed to date. For instance, [8] proposed a decentralized row-based Lasso algorithm. This iterative procedure, though, is suboptimal since it only approximates the exact minimum of the functional. In [9], instead, Sundman *et al* developed distributed versions of a number of greedy schemes such as OMP, SP or FROGS (Forward-Reverse Orthogonal Greedy Search). In these schemes (which in general require knowledge on the sparsity order), each node runs a greedy procedure, locally shares its support with its neighbors, and then a decision on which element to introduce in the support is made according to a voting process.

In this paper, we propose a decentralized scheme, referred to as Distributed iterative Thresholding (DiT), for in-network support detection. The approach is reminiscent of the iterative soft and hard thresholding methods of [10] in that it combines a gradient minimization procedure with an adaptive threshold update step. Unlike other approaches [8, 9], here we rigorously prove the convergence of the proposed scheme. Moreover and differently from [9], DiT does not require a priori knowledge on the sparsity order.

1.1. Notation

Before proceeding let us introduce some notation. Given $x \in \mathbb{R}^n$, we define the component-wise indicator function $\mathbb{1}(x)$ as $(\mathbb{1}(x))_i = 1$ if $x_i \neq 0$ and $(\mathbb{1}(x))_i = 0$ otherwise. The L^p-norm of x is denoted by $\|x\|_p$ for $p > 0$, whereas $\|x\|_0$ gives the number of non-zero elements of x. A graph $\mathcal{G}$ is defined as $\mathcal{G} := (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ and $\mathcal{E}$ stand for the set of

This work is supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications ♯ (Grant agreement n.318306) and by the European Research Council under FP7 / ERC Grant agreement n.279848.

vertices and edges respectively. Now assume that for each $v \in \mathcal{V}$ there is associated a value a_v, then, we will indicate by $\overline{a_v}$ the average over its neighborhood $\mathcal{N}_v$, that is $\overline{a_v} := \frac{1}{|\mathcal{N}_v|}\sum_{w\in\mathcal{N}_v} a_w$, with $\mathcal{N}_v := \{v' \in \mathcal{V} | (v,v') \in \mathcal{E}\}$ and $|\mathcal{N}_v|$ standing for the cardinality of $\mathcal{N}_v$. Likewise, we define the double average as $\overline{\overline{a_v}} := \frac{1}{|\mathcal{N}_v||\mathcal{N}_w|}\sum_{w\in\mathcal{N}_v}\sum_{u\in\mathcal{N}_w} a_u$.

2. SIGNAL MODEL

Consider a network composed of V nodes whose connectivity is described through the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $V = |\mathcal{V}|$. Accordingly, the sensor v can communicate with the sensor v' if and only if $\{v, v'\} \in \mathcal{E}$ or, in other words, v' belongs to its neighborhood set $\mathcal{N}_v$.

Each sensor observes a compressed version of a k-sparse signal $\{x_v\}_{v\in\mathcal{V}} \in \mathbb{R}^n$ through a set of linear and local measurements, namely

$$y_v = A_v x_v + \eta_v \quad ; \quad v \in \mathcal{V} \tag{1}$$

where $A_v \in \mathbb{R}^{m\times n}$ (with $m \ll n$) and $\eta_v \in \mathbb{R}^n$ stands for additive noise which may be due, for instance, to the acquisition process. We further assume that the observed signals $\{x_v\}_{v\in\mathcal{V}}$ share the same signal support Ω, that is $\Omega = \Omega_k := \{i | x_{v,i} \neq 0\}$. For convenience, we define the support vector $s \in \{0,1\}^n$ as $s_{v,i} = 1$ if $i \in \Omega$, and $s_{v,i} = 0$ otherwise.

The ultimate goal at each sensor is to locally reconstruct $\{x_v\}_{v\in\mathcal{V}}$. In this regard, it is well known that the challenge in CS problems is to identify the signal support, once this is accomplished, the estimate can be readily obtained by means of a LS operator. In view of this fact, this paper proposes a distributed iterative thresholding (DiT) algorithm that attempts to achieve consensus on the signal support to reconstruct each sensor observation while keeping privacy on the local measurements.

3. DISTRIBUTED ITERATIVE THRESHOLDING

For mathematical simplicity, we assume hereafter that the connectivity graph $\mathcal{G}$ is d-regular, that is, each node $v \in \mathcal{V} = \{1, 2, \ldots, |\mathcal{V}|\}$ has a neighborhood $\mathcal{N}_v$ of dimension $|\mathcal{N}_v| = d$ (included itself). Bearing this in mind, we attempt to minimize the following cost functional:

$$F(X) = \sum_{v\in\mathcal{V}} \Bigg\{ \tau_v \|y_v - A_v x_v\|_2^2 + \frac{\mu}{d}\sum_{w\in\mathcal{N}_v} \left\| \mathbb{1}(x_v) - \overline{\mathbb{1}(x_w)} \right\|_2^2 + \\ + \mu \|x_v\|_0 + 2\lambda \|x_v\|_1 \Bigg\}, \tag{2}$$

where $X = (x_1, \ldots, x_{|\mathcal{V}|})$, variables τ_v, μ and λ stand for positive weighting parameters, and $\overline{\mathbb{1}(x_v)} = \frac{1}{d}\sum_{w\in\mathcal{N}_v} \mathbb{1}(x_w) \in [0,1]$.

The rationale behind each term that composes the cost functional is the following:

1. The terms $\|y_v - A_v x_v\|_2^2$ account for the least square residuals. These terms need to be minimized in order to obtain consistent estimates with the local measurements.
2. The terms $\sum_{w\in\mathcal{N}_v} \left\| \mathbb{1}(x_v) - \overline{\mathbb{1}(x_w)} \right\|_2^2$ promote consensus in the signal support.
3. Finally, the terms $\|x_v\|_0$ and $\|x_v\|_1$ promote sparsity on the local estimates.

It is well known that classical iterative thresholding algorithms consider either the L^0-norm zero (in the "hard" version) or the L^1-norm (in the "soft" version) as penalization terms [11]. Here, instead, the functional in (2) includes both norms as penalization terms. This follows from the fact that soft thresholding is known to perform better than hard thresholding in terms of MSE, but it does not guarantee an exact recovery of the signal support. This makes the L^1-norm not suitable for consensus and motivates the use of the L^0-norm. Likewise, the choice of the weighting parameters, i.e. τ_v, μ and λ, follows analogous reasons.

3.1. Algorithm

In this section, we introduce the DiT algorithm which guarantees the minimization of the functional (2) at each iteration (see Section 3.2 for details).

The procedure, summarized in Algorithm 1, reads as follows. At time t, node $v \in \mathcal{V}$ stores two variables: the estimate of x_v, denoted by $\hat{x}_v(t)$, and the support estimate $\hat{s}_v \in [0,1]^n$. In particular, $\hat{s}_{v,i} = \alpha$ means that a fraction of α nodes in $\mathcal{N}_v$ agree that $i \in \Omega$. First, each node performs a gradient step with respect to its own residual (step 6) followed by a component-wise threshold operation (steps 7-10). It is worth noting that the local threshold, i.e. $h_{v,i}(t)$, changes from iteration to iteration and depends on the support variable $\hat{s}_v(t)$. Next, node v exchanges the support of the updated estimate (step 11), i.e., $\mathbb{1}(\hat{x}_v(t+1))$, with its neighbors. Then, the local average $\overline{\mathbb{1}(\hat{x}_v(t+1))}$ is computed (step 12) and sent over again to the neighbors. Finally, sensor v is ready to update the estimate of the support (step 14) as $\hat{s}_v(t+1) = \overline{\mathbb{1}(\hat{x}_v(t+1))}$. This procedure is repeated until a maximum number of iterations $T_{\max}$ is reached. The steps of the algorithm (and in particular the double exchange-average of steps 11-14, which is not intuitive) are theoretically motivated by Lemma 3 in next section.

It is worth noting that if $\hat{s}_{v,i}(t) = 1$, which means that all the neighbors of v agree that $i \in \Omega$, the algorithm turns out to be equivalent to classical soft-thresholding [11] with parameter λ. On the contrary, if $\hat{s}_{v,i}(t) < 1$, the threshold, as a consequence of the L^0-norm and the consensus terms in (2), becomes more selective, i.e., $h_{v,i} > \lambda$. Concluding, a component of the signal that has associated a large $\hat{s}_{v,i}(t)$ will have more chances to be selected than the rest.

Algorithm 1 Distributed iterative Thresholding (DiT)

for all $v \in \mathcal{V}$ **do**
  Initialize variables:
  $\hat{x}_v(0) = [0, 0, \ldots, 0]^T$; $\hat{s}_v(0) = [1, 1, \ldots, 1]^T$
end for
for $t = 0, 1, \ldots, T_{\mathsf{max}}$ **do**
  for all $v \in \mathcal{V}$ **do**
    $z_v(t) = \hat{x}_v(t) + \tau_v A^T \left(y_v - A\hat{x}_v(t)\right)$
    for all $i = 1, \ldots, n$ **do**
      Update the threshold:
      $h_{v,i}(t) = \lambda + \sqrt{2\mu\left(1 - \hat{s}_{v,i}(t)\right)}$
      Update the value estimate:
      $$\hat{x}_{v,i}(t+1) = \begin{cases} z_{v,i}(t) - \lambda & \text{if } z_{v,i}(t) > h_{v,i}(t) \\ z_{v,i}(t) + \lambda & \text{if } z_{v,i}(t) < -h_{v,i}(t) \\ 0 & \text{otherwise} \end{cases}$$
    end for
    Exchange $\mathbb{1}\left(\hat{x}_v(t+1)\right)$ with neighbors $\mathcal{N}_v$
    Compute $\overline{\mathbb{1}\left(\hat{x}_v(t+1)\right)}$
    Exchange $\mathbb{1}\left(\hat{x}_v(t+1)\right)$ with neighbors $\mathcal{N}_v$
    Update the support estimate:
    $\hat{s}_v(t+1) = \overline{\overline{\mathbb{1}(\hat{x}_v(t+1))}}$
  end for
end for

Regarding the communication aspects, the algorithm needs two communication rounds per iteration: firstly, to exchange the support of its current signal estimate and, secondly, to exchange the average support of the signal estimate of its neighborhood. Clearly, this helps to disseminate the information in the network while guaranteeing privacy on the local measurements.

3.2. Convergence

Let $\hat{X}(t) = (\hat{x}_1(t), \ldots, \hat{x}_{|\mathcal{V}|}(t)) \in \mathbb{R}^{n \times |\mathcal{V}|}$ and $\hat{S}(t) = (\hat{s}_1(t), \ldots, \hat{s}_{|\mathcal{V}|}(t)) \in [0,1]^{n \times |\mathcal{V}|}$ be the sequences generated by Algorithm 1.

Theorem 1. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a d-regular graph. If $\|A_v\|_2^2 < \frac{1}{\tau_v}$ for any $v \in \mathcal{V}$, then, for any $t \in \mathbb{N}$, $F(\hat{X}(t+1)) \leq F(\hat{X}(t))$. Moreover, $\hat{S}(t) = (\hat{s}_1(t), \ldots, \hat{s}_V(t))$ converges for $t \to \infty$.*

Proof. Due to space limitation, we only provide a sketch of the proof . We first prove that the functional F is decreasing. Let $X = (x_1, \ldots, x_{|\mathcal{V}|})$, $C = (c_1, \ldots, c_{|\mathcal{V}|})$, $B = (b_1, \ldots, b_{|\mathcal{V}|})$ be variables in $\mathbb{R}^{n \times |\mathcal{V}|}$ and define the surrogate functional

$$F^S(X, C, B) = \sum_{v \in \mathcal{V}} \Bigg\{ \tau_v \|y_v - A_v x_v\|_2^2 + \mu \sum_{w \in \mathcal{N}_v} \|\mathbb{1}(x_v) - c_w\|_2^2 + \mu \|x_v\|_0 + 2\lambda \|x_v\|_1 + \|x_v - b_v\|_2^2 - \tau_v \|A_v(x_v - b_v)\|_2^2 \Bigg\}.$$

In the following lemmas, we find the minimizers of F^S with respect to coordinates B, C and X.

First, for any $v \in \mathcal{V}$, if $\|A_v\|^2 < \frac{1}{\tau_v}$, then $\|x_v - b_v\|_2^2 - \tau_v \|A_v(x_v - b_v)\|_2^2 \geq 0$. Bearing this in mind, we obtain the following lemma.

Lemma 1. $X = \arg\min_{B \in \mathbb{R}^{n \times |\mathcal{V}|}} F^S(X, C, B)$.

We now remark that for any $\alpha_v, \beta_v \in \mathbb{R}$ related to $v \in \mathcal{V}$ of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, if $\mathcal{G}$ is d-regular, then $\sum_{v \in \mathcal{V}} \sum_{w \in \mathcal{N}_v} (\alpha_v - \beta_w)^2 = \sum_{v \in \mathcal{V}} \sum_{w \in \mathcal{N}_v} (\alpha_w - \beta_v)^2$, and its minimizer with respect to α_v (respectively β_v) is $\overline{\beta_v}$ (respectively $\overline{\alpha_v}$). This can be easily generalized to the multidimensional case, using $\|\cdot\|_2$. Leveraging such remark, Lemma 2 follows.

Lemma 2. $\overline{\mathbb{1}(X)} = \arg\min_{C \in \mathbb{R}^{n \times |\mathcal{V}|}} F^S(X, C, B)$, *where* $\overline{\mathbb{1}(X)} = (\overline{\mathbb{1}(x_1)}, \ldots, \overline{\mathbb{1}(x_{|\mathcal{V}|})})$.

Next, Lemma 3 obtains the minimum with respect to X.

Lemma 3. *Let $z_v = b_v + \tau_v A_v^T (y_v - A_v b_v)$. Then, $\widetilde{X} = \arg\min_{X \in \mathbb{R}^{n \times |\mathcal{V}|}} F^S(X, B, C)$ where the entries of $\widetilde{X}$ are $\widetilde{x}_{v,i} = z_{v,i} - \lambda$ if $z_{v,i} > \lambda + \sqrt{2\mu(1 - \overline{c_{v,i}})}$, $\widetilde{x}_{v,i} = z_{v,i} + \lambda$ if $z_{v,i} < -\lambda - \sqrt{2\mu(1 - \overline{c_{v,i}})}$, and null otherwise, $v \in \mathcal{V}$, $i \in \{1, \ldots, n\}$.*

This lemma can be proved noting that, for a suitable γ independent from x_v, $F^S(X, C, B) = \sum_{v \in \mathcal{V}} \|x_v - z_v\|_2^2 + \mu \sum_{w \in \mathcal{N}_v} \|\mathbb{1}(x_v) - c_w\|_2^2 + \mu \|x_v\|_0 + \lambda \|x_v\|_1 + \gamma$ and minimizing it with respect to the x_v's. Finally, from Lemmas 1, 2 and 3 we obtain

$$\begin{aligned} F(\hat{X}(t)) &= F^s(\hat{X}(t), \overline{\mathbb{1}(\hat{X}(t))}, \hat{X}(t)) \\ &\geq F^s(\hat{X}(t+1), \overline{\mathbb{1}(\hat{X}(t))}, \hat{X}(t)) \text{ (from Lemma 3)} \\ &\geq F^s(\hat{X}(t+1), \overline{\mathbb{1}(\hat{X}(t+1))}, \hat{X}(t)) \text{ (from Lemma 2)} \\ &\geq F^s(\hat{X}(t+1), \overline{\mathbb{1}(\hat{X}(t))}, \hat{X}(t+1)) \text{ (from Lemma 1)} \\ &= F(\hat{X}(t+1)). \end{aligned}$$

Moreover, algebraic steps lead to $F(\hat{X}(t)) - F(\hat{X}(t+1)) \geq \sum_{v \in \mathcal{V}} (1 - \tau_v \|A_v\|^2) \|x_v(t+1) - x_v(t)\|_2^2 \geq 0$. Since $F(\hat{X}(t))$ is decreasing and lower bounded, $F(\hat{X}(t)) - F(\hat{X}(t+1)) \to 0$, from the last inequality, $\|\hat{x}_v(t+1) - \hat{x}_v(t)\|_2^2 \to 0$ for any $v \in \mathcal{V}$. This fact prevents jumps after a certain time. Suppose that $\hat{x}_{v,i}(t) = 0$ for some $v \in \mathcal{V}, i \in \{1, \ldots, n\}$ (which implies $\hat{s}_{v,i}(t) < 1$, and $\sqrt{2\mu(1 - s_{v,i}(t))} \geq \sqrt{2\mu(1 - \frac{1}{d})} = \xi > 0$). Now, $\hat{x}_{v,i}(t+1) \neq 0$ would imply $|\hat{x}_{v,i}(t+1)| > \xi$, otherwise it would be cut to zero; but for t sufficiently large a jump of amplitude ξ is not possible. In conclusion, for a sufficiently large $t_0 \in \mathbb{N}$, $s_{v,i}(t_0) = 0$ implies $s_{v,i}(t) = 0$ for any $t > t_0$, which concludes the proof. □

4. NUMERICAL RESULTS

In the simulations, we consider signals $\{x_v\}_{v=1}^{|\mathcal{V}|}$ of length $n = 100$ with sparsity level $k = 10$. The support is generated uniformly at random, with non-zero elements drawn from a standard Gaussian distribution. As performance metric, we use the average support error $\varepsilon = \frac{1}{|\mathcal{V}|}\sum_{v\in\mathcal{V}} \frac{\|\hat{s}_v - s\|_0}{n}$. where $s = \mathbb{1}(x_v)$. Therefore, if k is known, the maximum error is $2k/n = 0.2$.

DiT is compared with the distributed greedy algorithms DiOMP and DiSP of [9] over a regular topology with degree $d = 5$, as well as with their corresponding complete-topology versions that we name DiT-c, DiOMP-c and DiSP-c respectively. We recall that the greedy algorithms benefit the knowledge of the sparsity level, which is not required by DiT. Results are obtained after averaging out 1000 different trials.

In Figure 1, we show the average support error as a function of the number of measurements per sensor, having fixed $|\mathcal{V}| = 10$. As it can be observed, for small number of measurements per sensor, DiT exhibits better performance than the greedy algorithms, whereas for large m the results are similar. The main difference between DiT and the greedy procedures resides in how the support is detected. While in the greedy algorithms, at each iteration a new element is added to the support, DiT starts from a situation of full support (note that typically $\lambda \ll 1$) and step by step eliminates elements from it. This is more cautious when a small number of measurements is taken and explains the superior performance of DiT in this regime. We finally remark that, in this experiment, both DiT and greedy methods always achieve support consensus for $m \geq 24$; no performance loss from the consensus viewpoint is then entailed by DiT.

Furthermore, Figure 2 shows the average support error as a function of the network size $|\mathcal{V}|$ having fixed $m = 20$. For small networks, that is when $\mathcal{V} < 7$, the greedy approaches outperform the DiT algorithm. In this regime, the knowledge of the sparsity level k pays off. On the contrary, for larger networks (i.e. $\mathcal{V} > 10$) DiT outperforms most of the greedy algorithms. This stems from the fact that in the greedy approaches the number of iterations is upper bounded by k, which limits the dissemination of information in the network. Although the price to be paid is the number of iterations, which in this case is 1×10^4, it is worth noting that the per-iteration complexity associated to DiT is low, this being in stark contrast with the greedy approaches that need to solve a L^2-norm minimization problem at each iteration.

We finally remark that for our simulations a good choice for τ_v is $\|A_v\|_2^{-2}$ (which is the upper bound fixed by Theorem 1), except when m is small: in such cases, $\|A_v\|_2^{-2}$ may result too large and it is better to fix $\tau_v = 8 \times 10^{-3}$. Also μ is adapted to m, while λ is always the same.

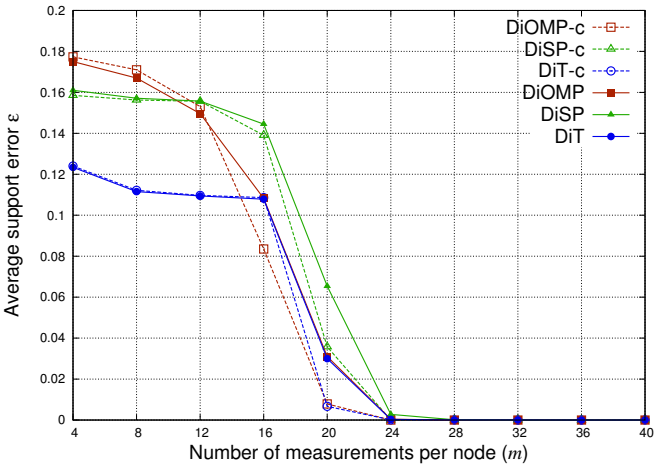

Fig. 1. Average support error (ε) vs number of measurements per sensor ($|\mathcal{V}| = 10$, $\lambda = 1\times10^{-4}$; if $m \leq 16$, $\mu = 5\times10^{-4}$, $\tau_v = 8 \times 10^{-3}$; if $m > 16$, $\mu = 1 \times 10^{-5}$, $\tau_v = \|A_v\|_2^{-2}$).

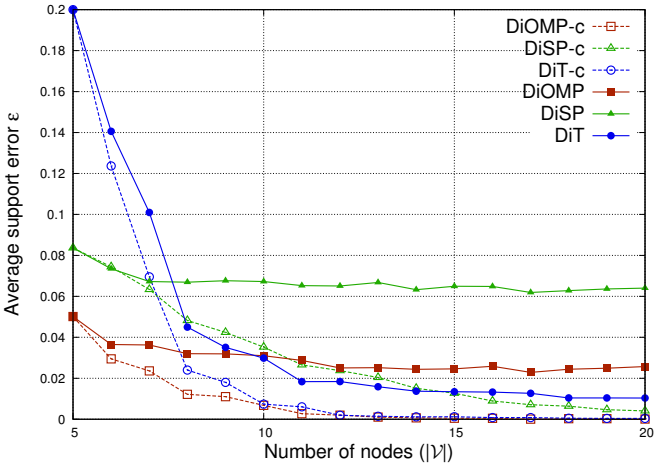

Fig. 2. Average support error (ε) vs number of sensors ($m = 20, \lambda = 1 \times 10^{-4}, \mu = 1 \times 10^{-5}, \tau_v = \|A_v\|_2^{-2}$).

5. CONCLUSIONS

In this paper, we have addressed the problem of distributed support detection of multiple sparse signals with common support. In particular, we have proposed and analyzed a distributed scheme, named DiT, for in-network signal recovery. DiT only requires message exchanges between adjacent nodes and, contrarily to the state-of-the-art methods, based on greedy procedures, it does not need a prior knowledge on the sparsity level. Besides, DiT has been proved to converge and numerical results have revealed that DiT achieves consensus in most cases. Finally, in those situations where the dissemination of the information matters, DiT outperforms the distributed greedy approaches.

6. REFERENCES

[1] D.L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] 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.

[3] J.A. Tropp and S.J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, 2010.

[4] M.F. Duarte and Y.C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Proc.*, vol. 59, no. 9, pp. 4053–4085, 2011.

[5] Wei Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *Information Theory, IEEE Transactions on*, vol. 55, no. 5, pp. 2230–2249, 2009.

[6] Thomas Blumensath and Mike E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–278, 2008.

[7] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.

[8] Q. Ling and Z. Tian, "Decentralized support detection of multiple measurement vectors with joint sparsity," in *IEEE ICASSP*, 2011, pp. 2996–2999.

[9] D. Sundman, S. Chatterjee, and M. Skoglund, "Distributed greedy pursuit algorithms," *submitted to EURASIP Signal Processing*, 2013.

[10] C. Ravazzi, S.M. Fosson, and E. Magli, "Distributed softh thresholding for sparse signal recovery," in *Globecom*, 2013.

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