DISTRIBUTED COMPRESSED SENSING OF NON-NEGATIVE SIGNALS USING SYMMETRIC ALPHA-STABLE DISTRIBUTIONS

George Tzagkarakis and Panagiotis Tsakalides

Department of Computer Science, University of Crete & Institute of Computer Science - FORTH
FORTH-ICS, P.O. Box 1385, 711 10 Heraklion, Crete, Greece
e-mail: gtzag@csd.uoc.gr, tsakalid@ics.forth.gr

ABSTRACT

Sensor networks gather an enormous amount of data over space and time to derive an estimate of a parameter or function. Several constraints, such as limited power, bandwidth, and storage capacity, motivate the need for a new paradigm for sensor data processing in order to extend the network's lifetime, while also obtaining accurate estimates. In a companion paper [1], we proposed a novel iterative algorithm for reconstructing non-negative sparse signals in highly impulsive background by modeling their prior distribution using symmetric alpha-stable distributions. In the present work, we extend this algorithm in the framework of distributed compressed sensing using duality theory and the method of subgradients for the optimization of the associated cost function. The experimental results show that our proposed distributed method maintains the reconstruction performance of its centralized counterpart, while also achieving a highly sparse basis configuration, thus reducing the total amount of data handled by each sensor.

1. INTRODUCTION

A major challenge in designing wireless sensor network (WSN) systems and algorithms is that transmitting data from a sensor to a central processing node may set a significant exhaustion of communication and energy resources. Such concerns may place undesirable limits on the amount of data collected and processed by sensor networks. Thus, it is natural to seek distributed algorithms for processing the data gathered by the nodes of a sensor network.

Distributed compressed sensing (DCS) [2] enables a potentially significant reduction in sampling and computation costs at a sensing system with limited capabilities. In particular, an ensemble of signals having a jointly sparse representation in a transform domain (e.g., wavelets, sinusoids) can be reconstructed from a small set of projections onto a second, measurement basis that is incoherent with the first one. In a WSN scenario, compressive wireless sensing (CWS) [3] appears to be able to reduce the latency of data gathering in a single-hop network by delivering linear projections of sensor readings through synchronized amplitude-modulated analog transmissions or in a distributed fashion.

On the other hand, the majority of the previous CS algorithms are based on a Gaussian assumption for the signal and/or noise statistics, which is violated in several distinct environments, such as in underwater acoustics [4] and in sonar/radar [5], where the associated signals and/or noise take large-amplitude values much more frequently than what a Gaussian model implies. In addition, these studies, as well as several other recent works, show that the family of alpha-stable distributions, and particularly the class of *symmetric alpha-stable* ($S\alpha S$) distributions, is a powerful statistical tool for modelling highly impulsive, and thus highly sparse, source signals. For this purpose, in a companion study [1] we developed a new iterative greedy algorithm for CS reconstruction of sparse signals corrupted by additive heavy-tailed noise, by modelling the impulsive behavior using members from the family of $S\alpha S$ distributions.

Moreover, there are cases where the information may not be available in a single node of a WSN. In a WSN, power and storage resources are limited enough such that the communication of an increased amount of data to a central node (fusion center) would affect significantly the network's lifetime. Working in a distributed framework, it is well known that a famous basis selection method, namely, the basis pursuit (BP) can be reformulated as a distributed linear program [6]. However, the resulting approach requires a fully connected network in the sense that, at each iteration, every sensor must be able to communicate with all the remaining ones. Motivated by this, we develop a DCS algorithm for reconstructing a non-negative sparse signal corrupted by additive heavy-tailed noise, while requiring a less demanding network topology. The prior belief for a highly sparse signal and heavy-tailed noise is modelled by employing members from the family of $S\alpha S$ distributions.

The paper is organized as follows: in Section 2, we briefly review for completeness the main properties of the family of $S\alpha S$ distributions exploited in the development of the proposed method. In Section 3, the distributed $S\alpha S$-based CS algorithm is described by employing a dual non-linear method based on subgradients. In Section 4, we compare the performance of the proposed approach with its centralized counterpart, while we conclude in Section 5.

2. STATISTICAL SIGNAL MODEL

According to the CS theory, if a given signal $\vec{f} \in \mathbb{R}^N$ is L-sparse in a suitable transform domain, then it is possible to be reconstructed directly using a compressed set of (noisy) measurements $\vec{g}$, obtained through incoherent random projections: $\vec{g} = \mathbf{\Phi}\mathbf{\Psi}^T \vec{f} + \vec{\eta} = \mathbf{\Phi}\vec{w} + \vec{\eta}$, where $\mathbf{\Phi} = [\vec{\phi}_1, \ldots, \vec{\phi}_M]^T$ is a $M \times N$ ($M < N$) random measurement matrix, $\mathbf{\Psi}$ is a $N \times N$ transform matrix, whose columns are the transform basis functions and must be incoherent with the rows of $\mathbf{\Phi}$, and $\vec{w} \in \mathbb{R}^N$ is the sparse weight vector with L non-zero components (or equivalently, the transform-domain representation of $\vec{f}$). We also note that in practice $\vec{f}$ is not strictly L-sparse but *compressible*, that is, the re-ordered components of $\vec{w}$ decay at a power-law.

In the present study, the prior belief that the unknown signal $\vec{w}$ is highly sparse and the noise $\vec{\eta}$ (with unknown variance σ_η^2) is heavy-tailed is exploited by using a $S\alpha S$ distribution as their prior. We also note that at the present study we consider only a measurement noise without assuming any kind of quantization. In the following, we consider that the signal and noise components are also jointly $S\alpha S$. The use of this family is motivated by the fact that the tails of a $S\alpha S$ distribution decay at an algebraic rate, which is in agreement with the rate of decay of the re-ordered components of a compressible vector $\vec{w}$.

2.1 The family of $S\alpha S$ distributions

For convenience, we introduce briefly the family of univariate $S\alpha S$ distributions, as well as some of their fundamental statistical properties exploited in the proposed distributed CS method. A $S\alpha S$ distribution is best defined by its characteristic function [7]:

This work was funded by the Greek General Secretariat for Research and Technology under Program ΠΕΝΕΔ-Code 03ΕΔ69 and by the Marie Curie TOK-DEV "ASPIRE" grant (MTKD-CT-2005-029791) within the 6th European Community Framework Program.

 ISSN 2076-1465

$$\phi(t) = \exp(i\delta t - \gamma^{\alpha}|t|^{\alpha}), \tag{1}$$

where α ($0 < \alpha \leq 2$) is the *characteristic exponent*, which is a shape parameter controlling the "thickness" of the tails of the density function, $\delta \in \mathbb{R}$ is the *location parameter* and $\gamma > 0$ is the *dispersion*, which determines the spread of the distribution around its location parameter, similar to the variance of the Gaussian. The smaller the α, the heavier the tails of a $S\alpha S$ density function. A $S\alpha S$ distribution is called *standard* if $\delta = 0$ and $\gamma = 1$. With $X \sim f_{\alpha}(\gamma, \delta)$ we denote a $S\alpha S$ random variable X with parameters α, γ, δ.

In general, no closed-form expressions exist for most $S\alpha S$ density functions except for the Gaussian ($\alpha = 2$) and the Cauchy ($\alpha = 1$). Unlike the Gaussian density, which has exponential tails, stable densities have algebraic tails: $\Pr\{X > x\} \sim Cx^{-\alpha}$, as $x \to \infty$, where C is a constant depending on the model parameters. Hence, $S\alpha S$ random variables with small α values are highly impulsive.

An important characteristic of $S\alpha S$ distributions is the lack of second-order moments. Instead, all moments of order $p < \alpha$ do exist and are called the *Fractional Lower-Order Moments* (FLOMs). In particular, the FLOMs of $X \sim f_{\alpha}(\gamma, \delta = 0)$ are given by [7]:

$$\mathbb{E}\{|X|^p\} = \big(C(p,\alpha)\cdot\gamma\big)^p \ , \quad 0 < p < \alpha, \tag{2}$$

where $\big(C(p,\alpha)\big)^p = \frac{\Gamma\left(1-\frac{p}{\alpha}\right)}{\cos\left(\frac{\pi}{2}p\right)\Gamma(1-p)}$. The $S\alpha S$ model parameters (α, γ) can be estimated using the consistent Maximum Likelihood (ML) method described by Nolan [8], which gives reliable estimates and provides the tightest possible confidence intervals. By restricting ourselves to the case $1 \leq \alpha \leq 2$, the *covariation norm* of $X \sim f_{\alpha}(\gamma, 0)$ is defined by

$$\|X\|_{\alpha} = \gamma_X \quad , 0 < p < \alpha \ , \tag{3}$$

where γ_X is given by solving (2) with respect to γ_X.

The concept of covariance, which is fundamental in the second-order moment theory, is not valid in the $S\alpha S$ case. Instead, a quantity called *covariation*, which plays an analogous role for $S\alpha S$ random variables to the one played by the covariance in the Gaussian case, has been proposed. Let X, Y be jointly $S\alpha S$ random variables with $1 < \alpha \leq 2$, zero location parameters and dispersions γ_X and γ_Y, respectively. Then, the covariation of X with Y is defined by [7]:

$$[X,Y]_{\alpha} = \frac{\mathbb{E}\{XY^{<p-1>}\}}{\mathbb{E}\{|Y|^p\}}\|Y\|_{\alpha}^{\alpha} \ , \tag{4}$$

where for any $z \in \mathbb{R}$ and $a \geq 0$ we use the notation $z^{<a>} = |z|^a \mathrm{sign}(z)$, while for a real vector $\vec{z} \in \mathbb{R}^N$ and $a \geq 0$ we write $\vec{z}^{<a>} = [|z_1|^a \mathrm{sign}(z_1), \ldots, |z_N|^a \mathrm{sign}(z_N)]$. The covariation satisfies the following (pseudo-)linearity properties in the first and second argument, respectively: If X_1, X_2, Y are jointly $S\alpha S$, then for any constants $a, b \in \mathbb{R}$ we have:

$$[aX_1 + bX_2, Y]_{\alpha} = a[X_1,Y]_{\alpha} + b[X_2,Y]_{\alpha} \tag{5}$$

$$[Y, aX_1 + bX_2]_{\alpha} = a^{<\alpha-1>}[Y,X_1]_{\alpha} + b^{<\alpha-1>}[Y,X_2]_{\alpha} \ . \tag{6}$$

Let $X \sim f_{\alpha}(\gamma_X, 0)$ and $Y \sim f_{\alpha}(\gamma_Y, 0)$ be independent $S\alpha S$ random variables. Then, $cX \sim f_{\alpha}(|c|\gamma_X, 0)$ ($c \neq 0$) and $X + Y \sim f_{\alpha}\big((\gamma_X^{\alpha} + \gamma_Y^{\alpha})^{1/\alpha}, 0\big)$. Thus, for the noisy CS measurements $\vec{g} = \mathbf{\Phi}\vec{w} + \vec{\eta}$, if $\{w_i \sim f_{\alpha}(\gamma_i, 0)\}_{i=1}^{N}$ and $\{\eta_j \sim f_{\alpha}(\gamma_{\eta}, 0)\}_{j=1}^{M}$, then

$$g_j \sim f_{\alpha}\Big(\Big[\sum_{i=1}^{N}(|\phi_{ji}|\gamma_i)^{\alpha} + \gamma_{\eta}^{\alpha}\Big]^{1/\alpha}, 0\Big) \ , \ j = 1, \ldots, M \ , \tag{7}$$

where ϕ_{ji} is the element of $\mathbf{\Phi}$ in row-j and column-i, that is, the CS measurements are jointly $S\alpha S$ with the signal and noise components. In addition, since only the FLOMs are finite for $S\alpha S$ variables, then, if $\vec{X}, \vec{Y}$ are two jointly $S\alpha S$ random vectors we consider the following statistical "inner product":

$$(\vec{X}, \vec{Y}) = \|\vec{Y}\|_{\alpha}^{2-\alpha}\,[\vec{X}, \vec{Y}]_{\alpha} \ . \tag{8}$$

We observe that most of the above quantities associated with a $S\alpha S$ distribution depend on the parameter p, whose value depends on α. By noting that in the subsequent analysis we are restricted in $1 \leq \alpha \leq 2$, we computed the *optimal value of* p as a function of α via a Monte-Carlo simulation resulting in a lookup table, which is then used to find the optimal p for every $1 \leq \alpha \leq 2$ through interpolation (ref. Table 1 in [1]). This table is also employed for the estimation of p in the proposed distributed $S\alpha S$-CS algorithm.

In addition, the efficiency of a CS method is highly affected by the selection of an appropriate measurement matrix $\mathbf{\Phi}$, which embeds the information content of the sparse signal in a low-dimensional vector of CS measurements $\vec{g}$. The disadvantage of the previous CS methods is that they employ measurement matrices which, in general, are not adapted to the true statistics of the sparse signal. However, in the companion paper [1] we introduced a measurement matrix, which is *best adapted* to the underlying heavy-tailed statistics of highly impulsive signal and noise components as expressed by a $S\alpha S$ model. In particular, we showed that the most appropriate measurement matrix $\mathbf{\Phi}$, which will be also used in the subsequent derivations, satisfying the stability property (7), as well as an analogue of the restricted isometry property (RIP), is obtained by drawing independent and identically distributed (i.i.d.) samples from a standard $S\alpha S$ distribution and then normalizing its columns to unit covariation norm. The normalization of a vector $\vec{x} \in \mathbb{R}^M$ to unit covariation norm is performed as follows:

$$\frac{\vec{x}}{\|\vec{x}\|_{\alpha}} \overset{(2),(3)}{=} \frac{\vec{x}}{\big(\frac{1}{M}\sum_{i=1}^{M}|x_i|^p\big)^{1/p}\big(C(p,\alpha)\big)^{-1}} = C(p,\alpha)M^{1/p}\frac{\vec{x}}{\|\vec{x}\|_{\ell_p}} \ . \tag{9}$$

3. DISTRIBUTED $S\alpha S$-CS

In the following, we extend our previous $S\alpha S$-CS algorithm [1] in a distributed fashion in order to deal with the potentially limited resources in a WSN scenario. There are cases in a WSN application where the CS-related information, namely, the measurement matrix $\mathbf{\Phi}$ and the sparse signal $\vec{w}$ may not be available in a single node. In particular, we consider the case of a network, where each sensor has access only to a *portion* of $\mathbf{\Phi}$. In the following, we assume that the columns of $\mathbf{\Phi}$ are distributed across the nodes of the network.

Due to the lack of second-order moments for $S\alpha S$ distributions the Minimum Mean Squared Error (MMSE) criterion is not valid and it should be replaced by the *Minimum Dispersion* (MD) criterion since, unlike the variance, their dispersion is finite and gives a good measure of the spread of estimation errors around zero. We also observe that from (2) the MD criterion can be viewed as a *least ℓ_p-norm estimation error* criterion since the FLOM $\mathbb{E}\{|X|^p\}$ can be estimated as the ℓ_p norm of the vector X. This justifies the use of the following objective function to be optimized:

$$J_p(\vec{w}) = \sum_{i=1}^{N}|w_i|^p \ , \ \vec{w} \in \mathbb{R}^N \ , \ 0 \leq p \leq 1 \ . \tag{10}$$

The problem under consideration is stated as follows: "given K nodes each one storing a subset of columns of $\mathbf{\Phi}$, find appropriate network topologies along with distributed algorithms for solving the following problem (P1)",

PRIMAL (P1): $\min J_p(\vec{w})$ s.t. $\vec{g} = \mathbf{\Phi}\vec{w} + \vec{\eta}, \ -\vec{w} \leq 0$,

where by $-\vec{w} \leq 0$ ($\Leftrightarrow \vec{w} \geq 0$) we mean that each component of $\vec{w}$ should be non-negative. We assume that the columns of $\mathbf{\Phi}$ are distributed among K nodes, such that the k-th node stores the k-th submatrix in the horizontal partition of $\mathbf{\Phi} = [\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_k, \ldots, \mathbf{\Phi}_K]$, where $\mathbf{\Phi}_k \in \mathbb{R}^{M \times n_k}$, and $n_1 + \cdots + n_K = N$. A corresponding partition also holds for the sparse vector, $\vec{w} = [\vec{w}_1, \ldots, \vec{w}_k, \ldots, \vec{w}_K]$, where $\vec{w}_k \in \mathrm{R}^{n_k}$. The proposed method is based on the *Duality Theory* [9] for the solution of the primal problem (P1). Under the appropriate conditions, such as separability of the objective function and the constraints, dual problems can be confronted by distributed methods. Hereafter, we assume that $\mathbf{\Phi}$ has full rank in order to ensure the feasibility of (P1) with high probability.

By construction of the objective function the following *feasibility & boundedness* condition holds: There exists at least one feasible solution for the primal problem (P1) and the objective function is bounded, that is, $-\infty < J_p^* < \infty$, where J_p^* denotes the optimal value of $J_p(\vec{w})$.

In order to enforce a direct applicability in a distributed setting (since the objective function is already separable), we introduce a redundant constraint [9]. In particular, let $U > 0$ be an upper bound of the ℓ_∞ norm of any solution of (P1). Then, a bounded version of (P1) is given by:

***Bounded* PRIMAL (P2):**

$$\min J_p(\vec{w}) \quad \text{s.t. } \vec{g} = \mathbf{\Phi}\vec{w} + \vec{\eta}\ ,\ -\vec{w} \le 0\ ,\ \|\vec{w}\|_\infty^p \le U\ .$$

The following inequalities determine a rule for the selection of U:

$$\|\vec{w}^*\|_\infty^p \le J_p(\vec{w}^*) \le J_p(\vec{w}) = J_p(\mathbf{\Phi}^\dagger(\vec{g}-\vec{\eta})) \le N\|\mathbf{\Phi}^\dagger(\vec{g}-\vec{\eta})\|_\infty^p$$
$$\le N\,\|\mathbf{\Phi}^\dagger\|_{max}\,\|\mathbf{1}(\vec{g}-\vec{\eta})\|_\infty^p < (N\,\|\mathbf{\Phi}^\dagger\|_{max}\,R^p) = U\ ,$$

where $\|\mathbf{\Phi}^\dagger\|_{max} = \max\{|[\mathbf{\Phi}^\dagger]_{nm}|\}_{1\le n\le N, 1\le m\le M}$ ($\mathbf{\Phi}^\dagger$ is the pseudoinverse), $\mathbf{1} \in \mathbb{R}^{N\times M}$ is the matrix with all of its entries being equal to one and R is a positive constant greater than the maximum amplitude component of $\vec{g}-\vec{\eta}$. Since these are $S\alpha S$ random vectors, this maximum is unknown in advance, but it suffices to select an R that satisfies this requirement with "high-probability". In a specific signal processing application, there is usually some prior knowledge about the signal content so that we can achieve an appropriate choice for R by assigning a relatively large value to it in comparison with the entries of the (known) measurement vector $\vec{g}$ and the expected noise amplitude.

3.1 Dualization and distributed solution of P1

We consider the *dual function* $\mathscr{L}(\cdot)$ defined for $\vec{\lambda} \in \mathbb{R}^M$ as follows:

$$\mathscr{L}(\vec{\lambda}) = \inf_{\substack{\vec{w}\in\mathbb{R}^N \\ \|\vec{w}\|_\infty^p \le U}} \mathscr{L}(\vec{w},\vec{\lambda})\ , \tag{11}$$

where $\mathscr{L}(\vec{w},\vec{\lambda})$ is the *Lagrangian function* and $\vec{\lambda}$ is the vector of *Lagrange multipliers*. The *dual problem* is defined by:

DUAL (D1): $\max \mathscr{L}(\vec{\lambda}) \quad \text{s.t. } \vec{\lambda} \ge 0\ .$

Following the standard dualization approach on all constraints except for the redundant ones and exploiting the separability of the objective function, as well as the partition of $\mathbf{\Phi}$ and $\vec{w}$, the Lagrangian function is expressed as follows:

$$\mathscr{L}(\vec{w},\vec{\lambda}) = J_p(\vec{w}) + \vec{\lambda}^T(\vec{g}-\mathbf{\Phi}\vec{w})\ . \tag{12}$$

Notice that although the noise component is not explicitly employed in the above expression, its presence will always result in an approximation $\widehat{\vec{w}}^*$ of the optimal vector $\vec{w}^*$.

Because of the lack of second-order statistics we are interested in developing a distributed $S\alpha S$-CS algorithm based on FLOMs. The standard Lagrangian function (12) employs the usual (Euclidean) inner product, which can be viewed as a measure of variance between the associated vectors, and thus it is not suitable for representing the statistics of a $S\alpha S$ model. For this purpose, we introduce the following Lagrangian function that exploits covariations instead of variances and thus it best adapts to our $S\alpha S$ framework:

$$\begin{aligned}\mathscr{L}^S(\vec{w},\vec{\lambda}) &= J_p(\vec{w}) + (\vec{\lambda},\vec{g}-\mathbf{\Phi}\vec{w}) \\ &\overset{(8)}{=} J_p(\vec{w}) + \underbrace{\|\vec{g}-\mathbf{\Phi}\vec{w}\|_\alpha^{2-\alpha}[\vec{\lambda},\vec{g}-\mathbf{\Phi}\vec{w}]_\alpha}_{\text{Ⓢ}}\ .\end{aligned} \tag{13}$$

For convenience, we will restrict ourselves to the case $1 \le \alpha \le 2$. By noting that $\|\vec{g}-\mathbf{\Phi}\vec{w}\|_\alpha^{2-\alpha} = \|\vec{\eta}\|_\alpha^{2-\alpha} = \gamma_\eta^{2-\alpha}$ (from (3)) and using the pseudo-linearity property (6), the second term of (13) takes the following form:

$$\begin{aligned}\text{Ⓢ} &= \gamma_\eta^{2-\alpha}\left([\vec{\lambda},\vec{g}]_\alpha + (-1)^{<\alpha-1>}[\vec{\lambda},\mathbf{\Phi}\vec{w}]_\alpha\right) \\ &\overset{(2),(4)}{=} \gamma_\eta^{2-\alpha}\left(\frac{\mathbb{E}\{\vec{\lambda}\,.*\,\vec{g}^{<p-1>}\}}{C(p,\alpha)^p\,\gamma_g^{p-\alpha}} - \frac{\mathbb{E}\{\vec{\lambda}\,.*\,(\sum_{k=1}^K \mathbf{\Phi}_k\vec{w}_k)^{<p-1>}\}}{C(p,\alpha)^p\,\gamma_{\mathbf{\Phi}\vec{w}}^{p-\alpha}}\right)\end{aligned} \tag{14}$$

where ".∗" denotes element-by-element multiplication between two vectors. In order to avoid numerical instability caused by the estimation of γ_g and $\gamma_{\mathbf{\Phi}\vec{w}}$, we will consider scenarios where the signal power is greater than the noise power (analogous to a relatively high SNR assumption). In this case $\gamma_g \simeq \gamma_{\mathbf{\Phi}\vec{w}}$. We also note that in (14) the first expectation is taken over $\vec{g}$, while the second expectation is over $\vec{w}_k$. However, the computational implementation proceeds by substituting the expectations with the corresponding arithmetic means (expressed as inner products) and (14) takes the following form:

$$\text{Ⓢ} = \frac{\gamma_\eta^{2-\alpha}\,\gamma_g^{\alpha-p}}{MC(p,\alpha)^p}\left[\vec{\lambda}^T\left(\vec{g}^{<p-1>} - (\sum_{k=1}^K \mathbf{\Phi}_k\vec{w}_k)^{<p-1>}\right)\right]\ . \tag{15}$$

Since γ_η is unknown and also along with γ_g they act as positive scaling factors and thus they do not affect the minimization operator, the final expression of the proposed Lagrangian function is given by:

$$\mathscr{L}^S(\vec{w},\vec{\lambda}) = \sum_{k=1}^K J_p(\vec{w}_k) + \vec{\lambda}^T\left(\vec{g}^{<p-1>} - (\sum_{k=1}^K \mathbf{\Phi}_k\vec{w}_k)^{<p-1>}\right) \tag{16}$$

which is in a separable form and thus amenable to a distributed implementation. In particular, we solve the dual problem (D1) by replacing $\mathscr{L}(\vec{w},\vec{\lambda})$ with the $S\alpha S$-based Lagrangian $\mathscr{L}^S(\vec{w},\vec{\lambda})$. We proceed by employing the method of subgradients [9], that is,

$$\vec{\lambda}^{i+1} = [\vec{\lambda}^i + s^i\vec{d}(\vec{\lambda}^i)]^+\ , \tag{17}$$

where $\vec{\lambda}^i$ is the estimated dual variable in the i-th iteration, $s^i > 0$ is a step-size parameter, $\vec{d}(\vec{\lambda}^i)$ is a supergradient[1] of the dual function $\mathscr{L}^S(\vec{\lambda})$, obtained by substituting (16) in (11), and $[\cdot]^+$ denotes the projection of a vector on the non-negative halfplane (due to the constraint of (D1)). This method guarantees that for a sufficiently small step-size s^i the distance of the current iterate, $\vec{\lambda}^{i+1}$, to the optimal solution is reduced. In practice, the convergence of the subgradient method is ensured using the following step-size:

$$s^i = c^i\left(\widehat{\mathscr{L}^S}(\vec{\lambda}^i) - \mathscr{L}^S(\vec{\lambda}^i)\right)/\|\vec{d}(\vec{\lambda}^i)\|_2\ , \tag{18}$$

where $\widehat{\mathscr{L}^S}$ is an approximation to the (unknown) optimal dual solution, which can be estimated using the best current dual value $\widehat{\mathscr{L}^S}(\vec{\lambda}^i) = \max_{0\le i'\le i}\mathscr{L}^S(\vec{\lambda}^{i'})$. In (18), c^i is a number chosen such that it guarantees a diminishing step-size. This can be achieved by setting $c^i = (1+\beta)/(i+\beta)$, where β is a fixed positive integer.

Turning back into (17), for a given $\vec{\lambda}$ a supergradient $\vec{d}(\vec{\lambda})$ can be obtained by differentiating (16) with respect to $\vec{\lambda}$ as follows:

$$\vec{d}(\vec{\lambda}) = \vec{g}^{<p-1>} - (\sum_{k=1}^K \mathbf{\Phi}_k\vec{w}_k^*(\vec{\lambda}))^{<p-1>}\ , \tag{19}$$

where $\vec{w}_k^*(\vec{\lambda})$ maximizes $\mathscr{L}^S(\vec{\lambda})$. We select the $\{\vec{w}_k^*(\vec{\lambda})\}_{k=1}^K$ by employing a heuristic approach. First, in the current i-th iteration the term $\vec{\lambda}^{iT}\vec{g}^{<p-1>}$ can be considered as a constant and thus it suffices

[1] The vector $\vec{h}$ is a supergradient (resp. subgradient) of a concave (resp. convex) function f at the point $\vec{x}$ if $\forall\vec{y}$, $f(\vec{y}) \le f(\vec{x}) + \vec{h}^T(\vec{y}-\vec{x})$ (resp. $f(\vec{y}) \ge f(\vec{x}) + \vec{h}^T(\vec{y}-\vec{x})$).

to find $\vec{w}_k^*(\vec{\lambda}^i)$ such that the vector $\vec{w}^*(\vec{\lambda}^i) = [\vec{w}_1^*(\vec{\lambda}^i), \ldots, \vec{w}_K^*(\vec{\lambda}^i)]$ satisfies the expression

$$\vec{w}^*(\vec{\lambda}^i) = \arg \inf_{\substack{\vec{w} \in \mathbb{R}_+^N \\ \|\vec{w}\|_\infty^p \le U}} \Big(\sum_{k=1}^K J_p(\vec{w}_k) - \vec{\lambda}^{iT} \big(\sum_{k=1}^K \mathbf{\Phi}_k \vec{w}_k \big)^{<p-1>} \Big) \ . \quad (20)$$

The following relations hold under the consideration $\{\vec{a} \le \vec{b} \Leftrightarrow a_i \le b_i, \forall i\}$ (the same holds for "$\ge$"):

$$\big(\sum_{k=1}^K \mathbf{\Phi}_k \vec{w}_k \big)^{<p-1>} = \big[|v_1|^{p-1} \mathrm{sign}(v_1), \ldots, |v_M|^{p-1} \mathrm{sign}(v_M) \big]^T$$

$$\le_{\mathrm{sign}(\cdot) \le 1} \big[|v_1|^{p-1}, \ldots, |v_M|^{p-1} \big]^T \quad \text{with} \quad v_m = \sum_{k=1}^K \sum_{j=1}^{n_k} [\mathbf{\Phi}_k]_{mj} w_{kj} \ , \quad (21)$$

where $[\mathbf{\Phi}_k]_{mj}$ denotes the (mj)-th element of the submatrix $\mathbf{\Phi}_k$ and w_{kj} is the j-th component of $\vec{w}_k$. Taking the inner products of both sides of the above inequality with $\vec{\lambda}^i$ under the dual constraint $\vec{\lambda}^i \ge 0$ results in the following relations:

$$\begin{aligned} \vec{\lambda}^{iT} \big(\sum_{k=1}^K \mathbf{\Phi}_k \vec{w}_k \big)^{<p-1>} &\le \vec{\lambda}^{iT} \big[|v_1|^{p-1}, \ldots, |v_M|^{p-1} \big]^T \\ = \sum_{m=1}^M \lambda_m^i |v_m|^{p-1} &= \sum_{m=1}^M |(\lambda_m^i)^{\frac{1}{p-1}} v_m|^{p-1} \ . \end{aligned} \quad (22)$$

From (22) we can see that the m-th component of the current Lagrange multiplier λ_m^i, risen to the power of $1/(p-1)$, multiplies the m-th row of each submatrix $\mathbf{\Phi}_k$. We seek for a vector $\vec{w}^*(\vec{\lambda}^i)$ that minimizes (20). By combining with the inequality in (22), whose right-hand side consists of non-negative terms, we suggest that instead of finding a $\vec{w}^*(\vec{\lambda}^i)$ satisfying (20) we relax this requirement by searching for a $\vec{w}^*(\vec{\lambda}^i)$ such that

$$\vec{w}^*(\vec{\lambda}^i) = \arg \inf_{\substack{\vec{w} \in \mathbb{R}_+^N \\ \|\vec{w}\|_\infty^p \le U}} \Big(\sum_{k=1}^K J_p(\vec{w}_k) - \sum_{m=1}^M |(\lambda_m^i)^{\frac{1}{p-1}} v_m|^{p-1} \Big) \ , \quad (23)$$

where $\{v_m\}_{m=1}^M$ (ref. (21)) depend explicitly on $\{\vec{w}_k^*(\vec{\lambda}^i)\}_{k=1}^K$ and the relaxation refers to the fact that the estimated $\vec{w}^*$ does not achieve exactly the infimum of (20) but a lower value with our goal being to make this difference as small as possible. This relaxation has the advantage that we estimate $\vec{w}^*$ without the ambiguity of the $\mathrm{sign}(\cdot)$ function.

Since both terms in the parentheses of (23) are non-negative the infimum of their difference, under a non-negativity constraint for $\vec{w}^*$, will be equal to zero. Notice also that the second term implies that the parts of the partition of $\mathbf{\Phi}$ and $\vec{w}$ corresponding to the k-th sensor are distributed over M (additive) terms in a row-wise way (ref. (21)). Thus, in order to enforce this contribution of the k-th sensor to be close to its associated objective function value, $J_p(\vec{w}_k)$, we keep only these components of $\vec{w}_k$ for which the sum of their coefficients over those M terms is non-negative, that is, for the k-th sensor the set of indices $\mathscr{T}_k^i$ corresponding to the active components in the i-th iteration is given by

$$\mathscr{T}_k^i = \Big\{ j : \sum_{m=1}^M (\lambda_m^i)^{\frac{1}{p-1}} [\mathbf{\Phi}_k]_{mj} \ge 0 \Big\} \ , \ 1 \le j \le n_k \ . \quad (24)$$

Each sensor computes individually its set $\mathscr{T}_k^i$, which is then transmitted to the central node (fusion center), where the single set of the current active components, $\mathscr{T}^i$, is obtained as the union of the K sets,

$$\mathscr{T}^i = \bigcup_{k=1}^K \mathscr{T}_k^i \ . \quad (25)$$

Finally, the n-th component of the current "optimal" vector $\vec{w}^*(\vec{\lambda}^i)$ is formed as follows:

$$[\vec{w}^*(\vec{\lambda}^i)]_n = \begin{cases} 0 & , \text{if } n \notin \mathscr{T}^i \\ U^{\frac{1}{p}} & , \text{if } n \in \mathscr{T}^i \ . \end{cases} \quad (26)$$

The above discussion indicates a natural star-shaped network topology for the distributed implementation of the proposed subgradient method, where each sensor transmits directly to the fusion center only its index set $\mathscr{T}_k^i$.

Through the Karush-Kuhn-Tucker (KKT) conditions and the relaxation represented by (23) it can be seen that for any optimal solution $\vec{\lambda}^*$ of (D1), using $\mathscr{L}^S(\cdot)$ instead of $\mathscr{L}(\cdot)$, we have that $supp(\vec{w}^*) \subset \mathscr{T}^*$, where $supp(\vec{w}^*) = \{n : [\vec{w}^*]_n \ne 0\}$ is the support of the optimal sparse vector satisfying (P2) and $\mathscr{T}^*$ is the final set of active components after the algorithm has converged. In other words, once the central node computes $\vec{\lambda}^*$ then, it obtains an overestimate of the support of a solution of (P2) and thus of (P1), since the two problems are equivalent. This means that at this point the central node could solve a problem (P1) of reduced dimensionality by removing the columns of $\mathbf{\Phi}$, whose indices are not included in $\mathscr{T}^*$ (and consequently setting $[\vec{w}]_{n'} = 0$ for $n' \notin \mathscr{T}^*$).

In particular, the central node estimates the sparse vector $\vec{w}$, which satisfies the observation model $\vec{g} = \mathbf{\Phi}_{\mathscr{T}^*} \vec{w}_{\mathscr{T}^*} + \vec{\eta}$, using the centralized $S\alpha S$-CS algorithm described in [1]. The proposed distributed $S\alpha S$-CS strategy relies on the knowledge of $\vec{\lambda}^*$. In practice, the subgradient method terminates at the central node after a maximum finite number of iterations is reached, or when the relative error of the estimated dual variable falls below a predefined tolerance ε, $\|\vec{\lambda}^{i+1} - \vec{\lambda}^i\|_2 < \varepsilon \cdot \|\vec{\lambda}^{i+1} - \vec{\lambda}^0\|_2$. As a result, the distributed algorithm converges to a suboptimal $\vec{\lambda}_s^*$ and consequently to a suboptimal set $\mathscr{T}_s^*$.

Notice that $\mathbf{\Phi}$ is distributed over the K nodes. However, $\mathbf{\Phi}_{\mathscr{T}_s^*}$ is required to the central node to estimate the sparse vector. This is carried out as follows: after stopping the subgradient method, the central node sends $\vec{\lambda}_s^*$ to the K nodes, which compute their corresponding fragments of $\mathscr{T}_s^*$ in a parallel way and transmit them back to the central node.

4. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed distributed $S\alpha S$-CS algorithm and compare it with its centralized version [1]. We start by noting that the so-called *Fractional-order SNR* (FSNR) is employed as a signal distortion measure as an alternative to the usual signal-to-noise ratio, which is not valid in the $S\alpha S$ case due to the lack of finite second-order statistics. For jointly $S\alpha S$ signal and noise components ($\alpha_g = \alpha_\eta$) the FSNR takes the form

$$\mathrm{FSNR} = 10 \log_{10} \Big(\frac{\mathbb{E}\{|\vec{g}|^p\}}{\mathbb{E}\{|\vec{\eta}|^p\}} \Big) = p \cdot 10 \log_{10} \big(\gamma_g / \gamma_\eta \big) \ , \quad (27)$$

where γ_g, γ_η are the signal and noise dispersions, respectively. The reconstruction quality is measured via the relative reconstruction SNR, $\mathrm{rSNR} = 10 \log_{10} \big(\|\vec{w}\|_{\ell_2}^2 / \|\vec{w} - \widehat{\vec{w}}\|_{\ell_2}^2 \big)$, with $\widehat{\vec{w}}$ denoting the reconstructed sparse vector.

Under the non-negativity assumption for the sparse vector, first we generate vectors $\vec{x} \in \mathbb{R}^N$, $N = 512$, with $L = 10$ non-zero components, whose values are drawn from a $S\alpha S$ distribution, placed in randomly chosen positions. Then, the non-negative vector to be reconstructed is $\vec{w} = abs(\vec{x}) \triangleq (|x_1|, \ldots, |x_N|)$. The value of α varies in $[1.1, 2]$, while the dispersion γ_w is chosen randomly from $[0.1, 1]$. Then, the noise dispersion γ_η is determined via (27) for a given pair (α_w, γ_w) and FSNR value (in dB). The entries of the measurement matrix $\mathbf{\Phi}$ are standard $S\alpha S$ samples, and then its columns are normalized to unit covariation norm. We also note that the subsequent results are represented as an average over 100 Monte-Carlo runs.

First, we validate the efficiency of the proposed FLOM-based Lagrangian function (16) in capturing the significant basis functions

(columns of $\mathbf{\Phi}$) to be activated for the estimation of the sparse vector $\vec{w}$, in contrast to the standard Lagrangian given by (12). We do so using simulated $S\alpha S$ signal ($\vec{w}$) and noise ($\vec{\eta}$) components with $\alpha \in [1.1, 2]$, $\gamma_w = 0.7$, FSNR $\in [5, 15]$ by repeating the process for each triplet $(\alpha, \gamma_w = 0.7, \text{FSNR})$ for 100 Monte-Carlo runs. Then, for each signal $\vec{w}$ the centralized $S\alpha S$-CS algorithm ([1]) is executed to estimate $\hat{\vec{w}}$, as well as the corresponding set of significant basis functions, whose indices are stored in a vector $\mathcal{T}_{\mathcal{R}}$. We proceed by setting $K = 15$, $M = 100$, $\beta = 1$ and $\varepsilon = 10^{-6}$.

The proposed distributed algorithm is executed next using the standard and the FLOM-based Lagrangian function for reconstructing the same sparse vector $\vec{w}$, resulting in the vectors $\mathcal{T}$ and $\mathcal{T}_S$, respectively, containing the corresponding indices of the significant basis functions. We also note that a different partition of $\mathbf{\Phi}$ (and $\vec{w}$) is created in each Monte-Carlo run, by assigning a different number of columns n_k to the k-th sensor ($k = 1, \ldots, K$). However, we take care of generating "balanced" partitions in the sense that all sensors obtain a similar number of columns of $\mathbf{\Phi}$.

Fig. 1 shows the average percentage of successful retrievals of the significant basis functions, as expressed via the cardinalities of the intersections $\mathcal{T}_{\mathcal{R}} \cap \mathcal{T}$ and $\mathcal{T}_{\mathcal{R}} \cap \mathcal{T}_S$ as a function of α and FSNR (in dB). It is clear that, on average, the standard Lagrangian function[2], which is based on second-order statistics, is able to retrieve less than half of the significant basis functions as estimated by the centralized $S\alpha S$-CS method. On the other hand, the distributed $S\alpha S$-CS method combined with the FLOM-based Lagrangian function has an 100% percentage of success in retrieving the significant basis functions given by its centralized implementation.

Fig. 2 shows the relative reconstruction SNR for the proposed distributed $S\alpha S$-CS algorithm as a function of α and FSNR. First, we observed that the reduced dimensionality problem resulting by implementing the distributed $S\alpha S$-CS method, which is then solved at the central node, achieved the same reconstruction performance with its centralized full dimensional counterpart. The reason is that the FLOM-based Lagrangian employed by the distributed $S\alpha S$-CS method is able to capture accurately the significant columns of $\mathbf{\Phi}$. In addition, we can see that the reconstruction performance increases as the values of α and FSNR increase. The decrease of rSNR as $\alpha \to 1$ is related to the increased inaccuracy in estimating the characteristic exponent α using a measurement vector $\vec{g}$ of small size $M = 100$. This problem can be alleviated by increasing M.

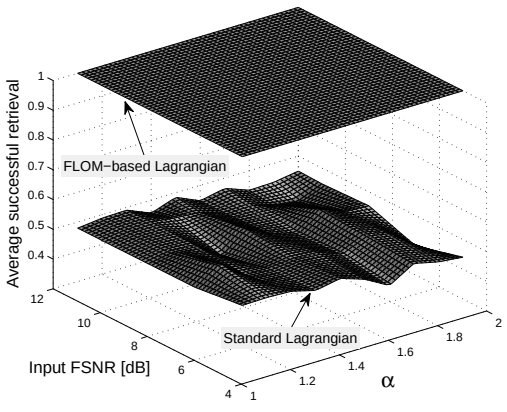

Figure 1: Average percentage of successful retrievals of the significant basis functions for the standard and FLOM-based Lagrangian function, as a function of α and FSNR (in dB).

[2]We mention for clarification that when using the standard Lagrangian the i-th column of $\mathbf{\Phi}$, $\vec{\phi}_i$, is considered to be significant if $|\vec{\phi}_i^T \vec{\lambda}| \geq 1$.

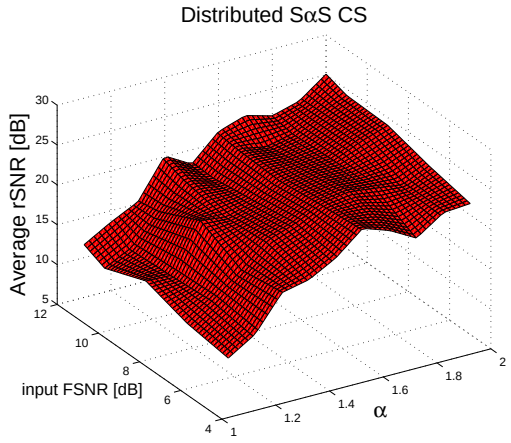

Figure 2: Average reconstruction rSNR (in dB) for the distributed $S\alpha S$-CS method, as a function of α and FSNR.

5. CONCLUSIONS AND FUTURE WORK

In this work, we developed a distributed method for CS reconstruction of highly impulsive signals with non-negative components working in a non-linear programming framework with application in a WSN. The high sparsity of the signal and noise components was modelled directly by using $S\alpha S$ distributions as their priors. The experimental results revealed that the distributed method maintained the increased reconstruction performance of its centralized counterpart, while also reducing significantly the cost for processing and transmitting the data at each sensor meeting the limitations of a WSN. As a future work, we will extend the distributed $S\alpha S$-CS method in more complex network topologies, as well as in the case of disjointly $S\alpha S$ signal and noise components ($\alpha_w \neq \alpha_\eta$).

REFERENCES

[1] G. Tzagkarakis and P. Tsakalides, "Greedy sparse reconstruction of non-negative signals using symmetric alpha-stable distributions," *in Proc. Eur. Sig. Proc. Conf. (EUSIPCO 2010)*, Aug. 23–27, Aalborg, Denmark.

[2] D. Baron *et al*, "Distributed compressed sensing," *submitted to IEEE Trans. Inform. Theory* [Online: `http://dsp.rice.edu/cs/DCS112005.pdf`].

[3] W. Bajwa, J. Haupt, A. Sayeed and R. Nowak, "Compressive wireless sensing," *proc. IEEE Int. Conf. on Inf. Proc. in Sensor Net. (IPSN'06)*, April 19–21, 2006, Nashville, TN, USA.

[4] M. Bouvet and S. Schwartz, "Comparison of adaptive and robust receivers for signal detection in ambient underwater noise," *IEEE Trans. on Acoustic, Speech, Sig. Proc.*, Vol. 37, pp. 621–626, May 1989.

[5] P. Tsakalides, F. Trinci and C. Nikias, "Radar CFAR Thresholding in Heavy-Tailed Clutter Environments, *Proc.* 32nd *Asilomar Conf. on Sig., Systems & Comp.*, Pacific Grove, CA, Nov. 1–4, 1998.

[6] S. Boyd and L. Vandenberghe, "Convex Optimization", Cambridge University Press, 2004.

[7] G. Samorodnitsky, and M. S. Taqqu, "Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance", Chapman & Hall, New York, 1994.

[8] J. Nolan, "Numerical calculation of stable densities and distribution functions," *Commun. Statist.-Stochastic Models*, Vol. 13, pp. 759–774, 1997.

[9] D. Bertsekas, "Nonlinear Programming", Athena Scientific, 2nd Edition, 1999.