GREEDY SPARSE RECONSTRUCTION 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

An accurate representation of the acquired data, while also conserving limited resources, such as power, bandwidth and storage capacity, is a challenging task. Besides, the Gaussian assumption, which plays a predominant role in signal processing being widely used as a signal and noise model, is unrealistic for a wide range of real-world data, which can be highly sparse in appropriate orthonormal bases. In the present work, the inherent property of compressed sensing (CS) working simultaneously as a sensing and compression protocol using a small subset of random projections is exploited to reduce the total amount of data. In particular, we propose a novel iterative algorithm for sparse representation and reconstruction of non-negative signals in highly impulsive background using the family of symmetric alpha-stable distributions. The experimental evaluation shows that our proposed method results in an increased reconstruction performance, while also achieving a higher sparsity when compared with state-of-the-art CS algorithms.

1. INTRODUCTION

Acquisition of high-resolution data with modern digital devices rises a very important issue, that is, how to effectively and precisely describe the information content of a given source signal such that it can be stored, processed or transmitted by taking into consideration the limited power and storage resources. Several studies [1] have shown that many natural signals result in highly sparse representations when they are projected on localized orthonormal bases (e.g., wavelets, sinusoids). Motivated by this, compressed sensing (CS) [2] enables a potentially significant reduction in sampling and computation costs at a sensing system with limited capabilities. In particular, a signal having a sparse representation in a transform domain can be reconstructed from a small set of random incoherent projections onto a suitable measurement basis.

The majority of previous works on CS-based reconstruction of a sparse signal solve constrained optimization problems. Commonly used approaches are typically based on convex relaxation (basis pursuit), non-convex (gradient based) local optimization, or greedy strategies ((Orthogonal) Matching Pursuit ((O)MP) [3, 4]. All these methods are primarily based on a Gaussian assumption for the underlying signal and/or noise processes, which is often violated in several distinct environments, such as in underwater acoustics, in sonar/radar and in medical imaging, where the associate signals take large-amplitude values much more frequently than what a Gaussian model implies. In addition, several studies have shown 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 [5, 6, 7].

Despite its efficiency, the family of $S\alpha S$ distributions has never been exploited in the framework of CS due to the lack of closed-form expressions for the density functions, as well as the lack of second-order statistics. For this purpose, we develop a novel iterative greedy algorithm for CS reconstruction of sparse signals with non-negative components corrupted by additive heavy-tailed noise. In particular, the prior belief for a highly impulsive signal and/or noise is modelled using members from the family of $S\alpha S$ distributions. To the best of our knowledge, this is the first effort to bridge the fields of CS and alpha-stable modelling.

The paper is organized as follows: in Section 2, we briefly review the main properties of $S\alpha S$ distributions exploited in our proposed method. In Section 3, the $S\alpha S$-based CS reconstruction algorithm is presented by considering jointly $S\alpha S$ signal and noise components. In Section 4, we compare the performance of the proposed approach with recent state-of-the-art CS recovery methods, while we conclude in Section 5.

2. STATISTICAL SIGNAL MODEL

In the framework of CS, 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 $N \times N$ matrix, whose columns correspond to 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 $\vec{f}$). Most of the recent literature on CS has concentrated on solving constrained optimization problems, where the unknown vector $\vec{w}$ and/or the noise $\vec{\eta}$ (with unknown variance σ_η^2) are modelled as Gaussian random variables. However, the Gaussian assumption is inadequate for a highly sparse vector $\vec{w}$. For this purpose, recent studies incorporated several non-Gaussian (heavy-tailed) distributions (e.g., Laplace, Cauchy) [8, 9] for modelling the prior belief that the vast majority of $\vec{w}$'s components have negligible amplitude. 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 work, the prior belief that $\vec{w}$ is highly sparse is exploited by using a $S\alpha S$ distribution as its prior, which is heavy-tailed and thus suitable for representing accurately an impulsive behavior. In the following, we consider that the noise $\vec{\eta}$ is also drawn from a $S\alpha S$ distribution. 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

In the following, we introduce briefly the family of univariate $S\alpha S$ distributions, as well as some of their fundamental properties exploited in the proposed CS method. A $S\alpha S$ distribution is best defined by its characteristic function [10]:

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

where α ($0 < \alpha \leq 2$) is the *characteristic exponent*, which is a shape parameter controlling the "thickness" of the tails of the den-

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.

sity 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 (with $\alpha < 2$) 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 [10]:

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

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 [11], which gives reliable estimates and provides the tightest possible confidence intervals. From (2) we get the following expression for the dispersion of X:

$$\gamma_X = \big(\mathbb{E}\{|X|^p\}\big)^{1/p}\big(C(p,\alpha)\big)^{-1} \,. \tag{3}$$

The *covariation norm* of $X \sim f_\alpha(\gamma, 0)$ with $\alpha > 1$ is defined by $\|X\|_\alpha = \gamma_X$, where γ_X is given from (3). This definition is extended to a quasi-norm for $\alpha < 1$, resulting in the following expressions:

$$\|X\|_\alpha = \begin{cases} \gamma_X & \text{, for } 0 < p < \alpha,\ 1 \le \alpha \le 2 \\ \gamma_X^\alpha & \text{, for } 0 < p < \alpha,\ 0 < \alpha < 1 \end{cases} \tag{4}$$

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

$$[X,Y]_\alpha = \frac{\mathbb{E}\{XY^{<p-1>}\}}{\mathbb{E}\{|Y|^p\}}\|Y\|_\alpha^\alpha \,, \tag{5}$$

where for any $z \in \mathbb{R}$ and $a \ge 0$ we use the notation $z^{<a>} = |z|^a sign(z)$, while for a real vector $\vec{z} \in \mathbb{R}^N$ and $a \ge 0$ we write $\vec{z}^{<a>} = [|z_1|^a sign(z_1), \ldots, |z_N|^a sign(z_N)]$. 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:

$$[aX_1 + bX_2, Y]_\alpha = a[X_1,Y]_\alpha + b[X_2,Y]_\alpha \tag{6}$$

$$[Y, aX_1 + bX_2]_\alpha = a^{<\alpha-1>}[Y,X_1]_\alpha + b^{<\alpha-1>}[Y,X_2]_\alpha \,. \tag{7}$$

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

$$g_j \sim f_\alpha\Big(\Big[\sum_{i=1}^N (|\phi_{ji}|\gamma_i)^\alpha + \gamma_\eta^\alpha\Big]^{1/\alpha}, 0\Big) \,, \ j = 1, \ldots, M \,, \tag{8}$$

where ϕ_{ji} is the element of $\mathbf{\Phi}$ in row-j and column-i, that is, each CS measurement follows a $S\alpha S$ distribution with the same α.

2.2 Estimation of p parameter

Most of the quantities involved in the $S\alpha S$-CS algorithm, which we will introduce in Section 3, will depend on the parameter p, whose optimal value depends on α. We also note that in the subsequent analysis we are restricted in $1 \le \alpha \le 2$. The optimal value of p is computed as the one that *minimizes the standard deviation* of the FLOM-based covariation estimator obtained from (5) by replacing the expectations with the arithmetic means. For this purpose, we studied the influence of p on the performance of the estimator via Monte-Carlo runs using two $S\alpha S$ random variables X, Y of length $N = 2048$. We executed 1000 Monte-Carlo runs with $\alpha \in [1 : 0.05 : 2]$ and for dispersions (γ_X, γ_Y) ranging in the interval $[0.01, 5]$ with a denser sampling in the sub-interval $[0.01, 2.5]$.

Table 1 shows the averaged optimal p values as a function of $\alpha \ge 1$, where the average is taken over the corresponding optimal p's for all dispersion pairs (γ_X, γ_Y) and Monte-Carlo runs. This table is then used as a lookup (interpolation) table in order to find the optimal p for every $1 \le \alpha \le 2$. It is important to note that the optimal value of p is close to $\alpha/2$, as observed in the table.

α	1	1.05	1.1	1.15	1.2	1.25	1.3	1.35
Optimal p	0.52	0.54	0.56	0.57	0.58	0.59	0.61	0.62
α	1.4	1.45	1.5	1.55	1.6	1.65	1.7	1.75
Optimal p	0.64	0.66	0.69	0.71	0.72	0.74	0.76	0.79
α	1.8	1.85	1.9	1.95	2			
Optimal p	0.81	0.84	0.88	0.93	0.8			

Table 1: Optimal p parameter as a function of α.

2.3 An adaptive $S\alpha S$ measurement matrix

A disadvantage of the previous CS methods is that they employ measurement matrices $\mathbf{\Phi}$ which, in general, are not adapted to the true statistics of the sparse signal. In this section, we study the performance of existing measurement matrices in the case of alpha-stable statistics and we introduce a new measurement matrix, which is best adapted to the statistics of $S\alpha S$ data. In particular, the desired measurement matrix must: i) approximate the *stability property* as expressed by (8), that is, the output characteristic exponent estimated from $\vec{g}$ is close to the input characteristic exponent of $\vec{w}$, ii) satisfy the *restricted isometry property* (RIP) [2].

First, we test the stability property by carrying out a set of 1000 Monte-Carlo runs for each $\alpha \in [0.9 : 0.05 : 2]$, where in each run we generate distinct vectors $\vec{w} \in \mathbb{R}^{1000}$ and $\vec{\eta} \in \mathbb{R}^M$, for a varying M, containing i.i.d. $S\alpha S$ entries (for clarity we present the results only for $M = 250$, by noting that as M increases the stability property is approximated more closely). Their corresponding dispersions are chosen uniformly at random in the interval $[0.01, 2.5]$. The performance was evaluated for several measurement matrices with i.i.d. entries drawn from distributions satisfying the RIP (e.g., Bernoulli, Gaussian), but we present the results for the following six matrices, which approximated more closely the stability property: 1) $\mathbf{\Phi_1}$ - its entries are i.i.d. Bernoulli samples, 2) $\mathbf{\Phi_2}$ - its entries are i.i.d. standard Gaussian samples ($S\alpha S$ with $\alpha = 2$) and normalized columns to unit ℓ_2-norm, 3) $\mathbf{\Phi_3}$ - its entries are i.i.d. standard $S\alpha S$ samples, 4) $\mathbf{\Phi_4}$ - obtained by normalizing the columns of $\mathbf{\Phi_3}$ to unit ℓ_2-norm, 5) $\mathbf{\Phi_5}$ - obtained by normalizing the columns of $\mathbf{\Phi_3}$ to unit ℓ_α-norm and 6) $\mathbf{\Phi_6}$ - obtained by normalizing the columns of $\mathbf{\Phi_3}$ to unit covariation norm, where the normalization of a vector $\vec{x} \in \mathbb{R}^M$ to unit covariation norm is performed as follows:

$$\frac{\vec{x}}{\|\vec{x}\|_\alpha} \overset{(3),(4)}{=} \frac{\vec{x}}{\left(\frac{1}{M}\sum_{i=1}^M |x_i|^p\right)^{1/p}\big(C(p,\alpha)\big)^{-1}} = C(p,\alpha)M^{1/p}\frac{\vec{x}}{\|\vec{x}\|_{\ell_p}} \,, \tag{9}$$

that is, the normalization to unit covariation norm is a scaled version of the normalization to unit ℓ_p norm with $p < 1$. Fig. 1 shows the output values of α estimated from $\vec{g}$, averaged over all Monte-Carlo runs, as a function of the input α. We observe that $\mathbf{\Phi_3}$ and $\mathbf{\Phi_6}$ are the top candidates, since they approximate more closely the stability property (diagonal line).

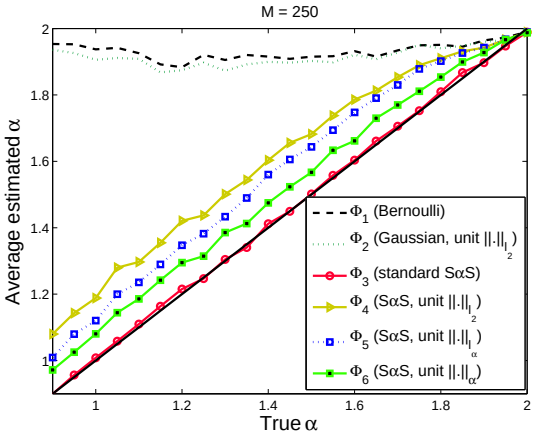

Figure 1: Stability property testing for six measurement matrices $\mathbf{\Phi_i}$ (ref. Sec. 2.3).

The second property to be satisfied by $\mathbf{\Phi}$ is the RIP, which is known to be valid for sub-Gaussian random matrices[1] (e.g., Gaussian, Bernoulli), and random partial bounded orthogonal matrices (e.g., the partial Fourier ensemble) [2]. However, since our method will be based on ℓ_p norms we consider a variant of the RIP, namely, the *restricted p-isometry property* (RpIP). A matrix $\mathbf{\Phi}$ satisfies the RpIP of order L if there exists a $\delta_L \in (0,1)$ such that:

$$(1-\delta_L)\|\vec{w}_I\|_{\ell_2}^p \le \|\mathbf{\Phi}_I \vec{w}_I\|_{\ell_p}^p \le (1+\delta_L)\|\vec{w}_I\|_{\ell_2}^p \qquad (10)$$

for all sets I with $card(I) \le L$ and all the L-sparse vectors $\vec{w} \in \mathbb{R}^N$ ($card(\cdot)$ denotes the cardinality of a set).

A theoretical proof of the RpIP for general random matrices is a difficult problem. To the best of our knowledge, the RIP/RpIP has not been shown for alpha-stable random matrices. Although this is, by all means, of great independent theoretical interest, however, here we give a computational evidence that an RpIP-like condition holds with high probability for $S\alpha S$ matrices. A theoretical proof is by its own a special study, which we are currently pursuing.

In the case of a $S\alpha S$ measurement matrix we exploit a concentration of measure inequality derived for alpha-stable vectors, which states that if f is a Lipschitz function (under the ℓ_2 norm) and $\vec{w}$ is an alpha-stable vector with $\alpha \in (1,2)$, then the deviation probability $\Pr\{|f(\vec{w}) - \mathbb{E}\{f(\vec{w})\}| \ge u\}$ is upper bounded by e^{-cu^α}, for u taking values in a finite range interval, where $c>0$ is a constant depending on α [12]. We begin with the random variable $f(\vec{w}) = \|\mathbf{\Phi}\vec{w}\|_\alpha^p$, where the randomness concerns the selection of $\mathbf{\Phi}$, while the vector $\vec{w}$ is considered to be fixed at a given realization. The expected value of $f(\vec{w})$, is obtained by combining (2),(8) and noting that $\gamma_{\phi_{ji}} = 1$, since the entries of $\mathbf{\Phi}$ are i.i.d standard $S\alpha S$ samples, as follows:

$$\mathbb{E}\{f(\vec{w})\} \overset{(3),(4)}{=} \frac{1}{M\big(C(p,\alpha)\big)^p} \sum_{j=1}^{M} \mathbb{E}\Big\{\big|\sum_{i=1}^{N} \phi_{ji} w_i\big|^p\Big\}$$

$$= \frac{1}{M\big(C(p,\alpha)\big)^p} \sum_{j=1}^{M} \Big(C(p,\alpha)\Big[\sum_{i=1}^{N} |w_i|^\alpha\Big]^{1/\alpha}\Big)^p = \|\vec{w}\|_{\ell_\alpha}^p \ . \qquad (11)$$

Notice that in the last equation, $\|\vec{w}\|_{\ell_\alpha}$ denotes the standard ℓ_α norm and not the covariation norm ($\|\vec{w}\|_\alpha$).

By setting $u = \varepsilon\|\vec{w}\|_{\ell_\alpha}^p$ in the above deviation probability for alpha-stable vectors and combining with (11), the following RpIP-like inequality holds for a $S\alpha S$ measurement matrix with probability exceeding $1 - e^{-c(\varepsilon\|\vec{w}\|_{\ell_\alpha}^p)^\alpha}$ (where $C_{M,\alpha} = M\big(C(p,\alpha)\big)^p$):

$$(1-\varepsilon)C_{M,\alpha}\|\vec{w}\|_{\ell_\alpha}^p \le \|\mathbf{\Phi}\vec{w}\|_{\ell_p}^p \le (1+\varepsilon)C_{M,\alpha}\|\vec{w}\|_{\ell_\alpha}^p \ . \qquad (12)$$

[1] A random matrix $\mathbf{\Phi}$ is sub-Gaussian if its i.i.d. entries are drawn from a sub-Gaussian random variable X with variance 1, that is, whose tail distribution is dominated by that of the standard Gaussian random variable: $\exists c_1, c_2 > 0$ s.t. $\Pr\{|X| > x\} \le c_1 e^{-c_2 x^2}$, for all $x > 0$.

Fig. 2 validates the Lipschitz condition for the candidate measurement matrices $\mathbf{\Phi_3}$ and $\mathbf{\Phi_6}$, by showing the ratios $\frac{\big|\|\mathbf{\Phi}\vec{w}_1\|_{\ell_p}^p - \|\mathbf{\Phi}\vec{w}_2\|_{\ell_p}^p\big|}{M\big(C(p,\alpha)\big)^p \|\vec{w}_1 - \vec{w}_2\|_{\ell_2}}$ averaged over a set of 500 Monte-Carlo runs, where in each run we generate a pair of matrices $(\mathbf{\Phi_3}, \mathbf{\Phi_6})$, with $M = 250$, and two sparse vectors $\vec{w}_1, \vec{w}_2 \in \mathbb{R}^{1000}$ with $L \in \{50,\ldots,250\}$ non-zero components, whose locations are selected uniformly at random and their corresponding values are drawn from a $S\alpha S$ distribution with a dispersion drawn uniformly from the interval $[0.1, 1]$. First we observe that the two matrices satisfy the Lipschitz condition, since the values of the above ratio are less than one. In addition, the values of $\mathbf{\Phi_6}$'s ratio are equal or less than the values of $\mathbf{\Phi_3}$ for the corresponding pairs $(\alpha, L/N)$ and thus the Lipschitz condition is satisfied in a higher extend. This fact, in addition to the observation that $\mathbf{\Phi_6}$ approximates the stability property quite closely, as shown in Fig. 1, motivated the selection of $\mathbf{\Phi_6}$ as the most suitable $S\alpha S$ measurement matrix.

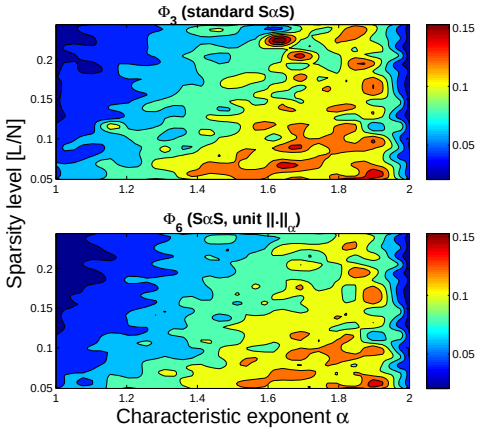

Figure 2: Verification of Lipschitz condition for $f(\vec{w}) = \|\mathbf{\Phi}\vec{w}\|_\alpha^p$ and for each matrix $\mathbf{\Phi_3}$, $\mathbf{\Phi_6}$, as a function of $(\alpha, \frac{L}{N})$ ($\alpha \in (1,2)$, $L \in \{50,\ldots,250\}$, $N = 1000$, $M = 250$).

3. MINIMUM DISPERSION CS INVERSION

The development of the proposed iterative $S\alpha S$-CS method starts by noting that for alpha-stable data, 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, the dispersion is finite and gives a good measure of the spread of estimation errors around zero. This provides a natural justification for the eligibility of ℓ_p norms with $p < 1$ in conjunction with a $S\alpha S$ prior model. From (3) we also observe that 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.

In the subsequent derivations the following conventions are used: i) I^n: set with n indices corresponding to the "active" columns of $\mathbf{\Phi}$ in the current iteration, ii) $\mathbf{\Phi}_{I^n}$: submatrix of $\mathbf{\Phi}$ containing only those columns with indices in I^n (the same notation is used for vectors, e.g., $\vec{w}_{I^n}$), iii) $\vec{w}_{I^n}^{n-1}$: $card(I^n)$-dimensional vector calculated in iteration $n-1$, that is, the elements in $\vec{w}_{I^n}^{n-1}$ which are not in $\vec{w}_{I^{n-1}}^{n-1}$ are set to zero.

Similarly to previous greedy CS algorithms, the proposed approach minimizes an objective function depending on the norm of the approximation error. However, in contrast to the majority of them, which are based on ℓ_2 or ℓ_1 norms of the error, our $S\alpha S$-CS method employs ℓ_p norms ($p < 1$), approximating more closely the ideal ℓ_0 case:

$$J_p(\vec{w}) = \sum_{i=1}^{N} |w_i|^p \ , \ \vec{w} \in \mathbb{R}^N \ , \ 0 \le p \le 1 \ . \qquad (13)$$

In particular, $J_p(\cdot)$ is minimized in terms of the estimation error $r(\vec{w}) = \vec{g} - \mathbf{\Phi}\vec{w}$ (for convenience we will also use the notation $\vec{r}$ instead of $r(\vec{w})$). For this purpose, a hybrid scheme of *directional updates* is employed. Specifically, in the n-th iteration the estimated sparse vector is updated by calculating a direction $\vec{v}^n_{I^n}$ and a step-size μ^n as follows

$$\vec{w}^n_{I^n} = \vec{w}^{n-1}_{I^n} + \mu^n \vec{v}^n_{I^n} \ . \tag{14}$$

Following a conjugate gradient approach, the step-size μ^n and the updated residual are given by

$$\mu^n = \left(\vec{r}^{n^T} \mathbf{\Phi}_{I^n} \vec{v}^n_{I^n}\right) \left(\vec{v}^{n^T}_{I^n} \mathbf{\Phi}^T_{I^n} \mathbf{\Phi}_{I^n} \vec{v}^n_{I^n}\right)^{-1} \tag{15}$$

$$\vec{r}^n = \vec{r}^{n-1} - \mu^n \mathbf{\Phi}_{I^n} \vec{v}^n_{I^n} \ . \tag{16}$$

A more computationally efficient sub-optimal direction is computed by combining the current gradient and the previous direction only

$$\vec{v}^n_{I^n} = [\vec{\nabla} J_p]_{I^n} + b^n \vec{v}^{n-1}_{I^n} \ , \tag{17}$$

where b^n is a step-size parameter and $[\vec{\nabla} J_p]$ is the negative gradient vector of the cost function with respect to $\vec{w}$

$$[\vec{\nabla} J_p] = \mathbf{\Phi}^T |p| \, diag\left(|r_1|^{p-2}, \ldots, |r_M|^{p-2}\right) \vec{r} \ , \tag{18}$$

where $diag\left(|r_1|^{p-2}, \ldots, |r_M|^{p-2}\right)$ is the $M \times M$ diagonal matrix with elements the components of the residual vector $\vec{r} = \vec{g} - \mathbf{\Phi}\vec{w}$.

In order to calculate the step-size b^n, we introduce a statistical pseudo-orthogonality condition between two $S\alpha S$ random vectors. First we note that if $\vec{X}$, $\vec{Y}$ are two jointly Gaussian random vectors, they are considered to be orthogonal if their covariance is equal to zero. 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 them to be orthogonal if the following "inner product" is zero:

$$(\vec{X}, \vec{Y}) = \|\vec{Y}\|_\alpha^{2-\alpha} \, [\vec{X}, \vec{Y}]_\alpha \ . \tag{19}$$

Thus, the step-size b^n is computed by requiring the new direction to be "orthogonal" to the previous one, that is, $(\vec{v}^n_{I^n}, \vec{v}^{n-1}_{I^n}) = 0$, where

$$\begin{aligned} (\vec{v}^n_{I^n}, \vec{v}^{n-1}_{I^n}) &= \|\vec{v}^{n-1}_{I^n}\|_\alpha^{2-\alpha} [\vec{v}^n_{I^n}, \vec{v}^{n-1}_{I^n}]_\alpha \\ &\overset{(17),(6)}{=} \|\vec{v}^{n-1}_{I^n}\|_\alpha^{2-\alpha} \left(\left[[\vec{\nabla} J_p]_{I^n}, \vec{v}^{n-1}_{I^n} \right]_\alpha + b^n \left[\vec{v}^{n-1}_{I^n}, \vec{v}^{n-1}_{I^n} \right]_\alpha \right) . \end{aligned}$$

By equating the last expression with zero and noting that $\|\vec{v}^{n-1}_{I^n}\|_\alpha \neq 0$ and $[\vec{v}^{n-1}_{I^n}, \vec{v}^{n-1}_{I^n}]_\alpha = \gamma^\alpha_{\vec{v}^{n-1}_{I^n}}$, the step-size b^n is given by

$$b^n = -\frac{\mathbb{E}\{[\vec{\nabla} J_p]_{I^n} .* (\vec{v}^{n-1}_{I^n})^{<p-1>}\}}{\mathbb{E}\{|\vec{v}^{n-1}_{I^n}|^p\}} \ , \tag{20}$$

where the expectations are estimated by taking the mean of the corresponding vectors and "$.*$" denotes element-by-element multiplication between two vectors.

3.1 Basis selection rule

The performance of a CS algorithm is affected significantly by the appropriate selection of the sparsest subset of basis vectors (columns of $\mathbf{\Phi}$) that best represents the data $\vec{g}$. For instance, MP selects iteratively the column of $\mathbf{\Phi}$ resulting in the largest (in absolute magnitude) inner product with the current approximation error $\vec{r}^n$. In the proposed $S\alpha S$-CS algorithm, we select the optimal set of basis functions, I^n, by introducing the following "distance measure" between two $S\alpha S$ random vectors $\vec{X}, \vec{Y} \in \mathbb{R}^N$, based on FLOMs (with $0 < p < \alpha$):

$$d_\alpha(\vec{X}, \vec{Y}) = \|\vec{X} - \vec{Y}\|_\alpha = \begin{cases} \frac{\left(\mathbb{E}\{|\vec{X} - \vec{Y}|^p\}\right)^{1/p}}{C(p,\alpha)} & , 1 \leq \alpha \leq 2 \\ \frac{\left(\mathbb{E}\{|\vec{X} - \vec{Y}|^p\}\right)^{\alpha/p}}{C(p,\alpha)} & , 0 < \alpha < 1 \end{cases} \tag{21}$$

where $|\vec{X} - \vec{Y}|^p \triangleq (|x_1 - y_1|^p, \ldots, |x_N - y_N|^p)$. In particular, an index i is added in I^n if the "distance" between the i-th basis vector, $\vec{\phi}_i$, and the current residual, $\vec{r}^n$, is below a certain threshold

$$I^n = I^{n-1} \cup \{i \mid d_\alpha(\vec{\phi}_i, \vec{r}^n) \leq \xi \cdot \min_j (d_\alpha(\vec{\phi}_j, \vec{r}^n))\} \ , \tag{22}$$

with $\xi > 1$ (usually it suffices $\xi \approx (1 + \delta)$ with δ being close to zero). Notice that when $\xi = 1$ the basis selection rule reduces to the completely greedy approach, where a single optimal basis vector is selected in each iteration. The algorithm terminates whether a predefined maximum number of iterations is reached, or when the relative decrease of the residual ℓ_p-norm falls bellow a threshold

$$\frac{\|\vec{r}^n\|^p_{\ell_p} - \|\vec{r}^{n-1}\|^p_{\ell_p}}{\|\vec{r}^{n-1}\|^p_{\ell_p}} < \varepsilon \ , \quad \varepsilon \ll 1 \ .$$

4. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed $S\alpha S$-CS algorithm by comparing it with state-of-the-art norm-based CS methods. Of course there are quite many CS methods in the recent literature with which we could compare, however, the scope of this study is to highlight the advantages of the $S\alpha S$ model in developing CS reconstruction algorithms and to exhibit its superiority against some of the most recent norm-based iterative CS methods. Hence, we compared the performance with several CS techniques, which employ ℓ_2, ℓ_1, or ℓ_p ($p \leq 1$) norms, but hereafter we present the results with respect to the main competitors of the proposed $S\alpha S$-CS method only, namely, the LASSO with a non-negativity constraint (nnLasso), the smoothed ℓ_0 (SL0), and the stagewise weak conjugate gradient pursuit (SWCGP)[2].

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, the so-called *Fractional-order SNR* (FSNR) is employed as a signal distortion measure. When the signal and noise components are jointly $S\alpha S$ ($\alpha_g = \alpha_\eta$) the FSNR takes the following form:

$$\text{FSNR} = 10 \log_{10}\left(\frac{\mathbb{E}\{|\vec{g}|^p\}}{\mathbb{E}\{|\vec{\eta}|^p\}}\right) = p \cdot 10 \log_{10}\left(\gamma_g / \gamma_\eta\right) \ , \tag{23}$$

where γ_g, γ_η are the signal and noise dispersion, respectively.

Under the non-negativity assumption for the sparse vector, first we generate vectors $\vec{x}$ of length $N = 512$ with $L \in \{5, \ldots, 20\}$ 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 from $[0.1, 1]$. According to the results of section 2.1, the entries of the measurement matrix $\mathbf{\Phi}$ are standard $S\alpha S$ samples, and then its columns are normalized to unit covariation norm. The noise dispersion γ_η is determined via (23) for a given pair (α_w, γ_w) and an FSNR value (in dB). We also note that the subsequent results are represented as an average over 100 Monte-Carlo runs. Besides, the parameter ξ involved in the basis selection rule is set to 1.005 in order to accelerate the $S\alpha S$-CS approach by permitting the simultaneous selection of more than one basis vectors in each iteration. The reconstruction quality is measured via the relative reconstruction SNR (rSNR), which is defined as follows: $\text{rSNR} = 10 \log_{10}\left(\|\vec{w}\|^2_{\ell_2} / \|\vec{w} - \widehat{\vec{w}}\|^2_{\ell_2}\right)$, with $\widehat{\vec{w}}$ denoting the reconstructed sparse vector.

In the following, we compare the reconstruction performance and the degree of sparsity achieved by $S\alpha S$-CS, along with LASSO, SL0 and SWCGP, for $M = 120$ and FSNR $= 10$ dB. In particular, for each pair (α, L) we perform a set of 100 Monte-Carlo runs, where

[2] For the implementation of the other CS methods we used the MATLAB codes included in the packages: `http://sparselab.stanford.edu/`, `http://www.acm.caltech.edu/l1magic`, `http://ee.sharif.ir/~SLzero`, `http://www.see.ed.ac.uk/~tblumens/sparsify/sparsify.html`.

in each run the dispersion γ_w is chosen from $[0.1, 1]$ uniformly at random. Fig. 3 shows the corresponding average rSNR values for the four methods. We observe that for highly impulsive signals, that is, for $\alpha \to 1$ the proposed $S\alpha S$-CS method outperforms clearly all the others. In addition, $S\alpha S$-CS outperforms all the other methods for a higher range of α as L increases. This means that the capability of the other CS methods in discriminating an increasing number of non-negative components drawn from a heavy-tailed $S\alpha S$ distribution is reduced.

On the other hand, Fig. 4 shows the corresponding CS ratios. The CS ratio, which is used as an index of sparsity, is defined as follows:

$$\text{CS ratio} = \frac{\text{number of CS measurements } M}{\text{number of non-zero components of } \widehat{\vec{w}}}, \qquad (24)$$

where the number of non-zero components of the estimated $\widehat{\vec{w}}$ (sparsity) depends on the algorithm. The higher the CS ratio the higher the sparsity is for a fixed value of M. We can see that on average the proposed $S\alpha S$-CS method results in much higher CS ratios, or equivalently in much sparser solutions when compared with the other CS techniques. In addition, we note that the slightly better reconstruction performance of the SWCGP method when $\alpha \to 2$ comes at the cost of a significant increase in the number of basis functions (much smaller CS ratio), which must be employed to represent accurately a highly sparse signal. Besides, the CS ratio of the $S\alpha S$-CS method decreases as the number of non-zero components increases and $\alpha \to 2$. In other words, as the underlying statistics tend to a Gaussian distribution, the $S\alpha S$-based algorithm requires more basis functions to capture the inherent information content which is related to second-order moments.

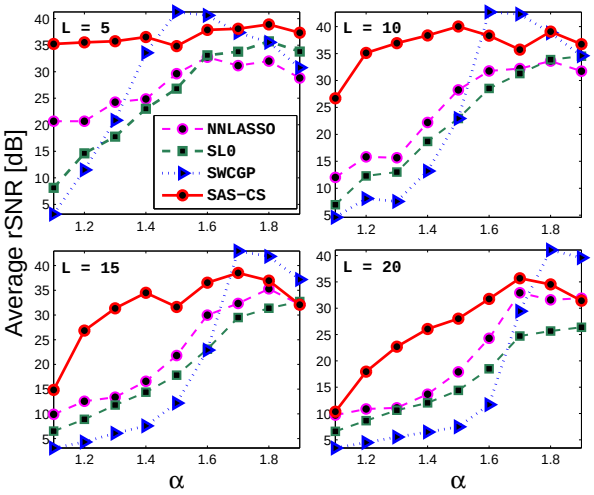

Figure 3: Average rSNR for $S\alpha S$-CS, LASSO, SL0 and SWCGP as a function of α with $L \in \{5, 10, 15, 20\}$ ($N = 512$, $M = 120$, FSNR $= 10$ dB).

5. CONCLUSIONS AND FUTURE WORK

In this work, we described an iterative CS algorithm for the reconstruction of highly sparse non-negative signals corrupted by heavy-tailed noise. The highly sparse behavior was modelled directly with a heavy-tailed distribution selected from the family of $S\alpha S$ distributions. The experimental results revealed an increased reconstruction performance, while also achieving an increased sparsity, when compared with other state-of-the-art iterative greedy CS algorithms.

As a future work, we are interested in exploiting the $S\alpha S$ model for developing a CS algorithm in a purely Bayesian framework. We expect that a probabilistic approach will provide further control on the sparsity of the sparse vector and furthermore it will permit the optimal design of future CS measurements with the goal of reducing the uncertainty of the reconstructed signal, something which is not possible with a norm-based iterative approach. In addition, we will extend the $S\alpha S$-CS method in the case of non-jointly $S\alpha S$ signal and noise components ($\alpha_w \neq \alpha_\eta$).

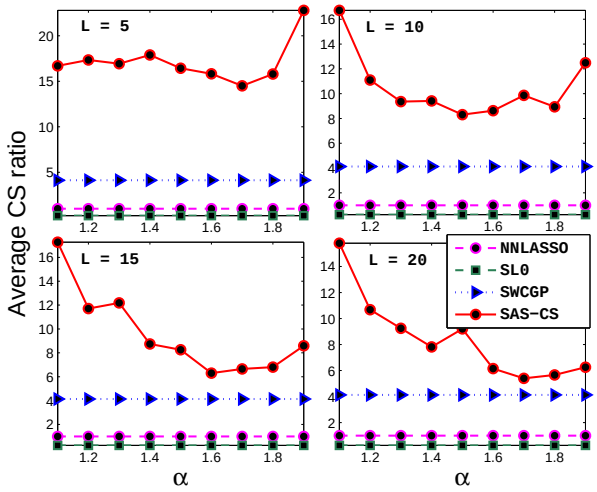

Figure 4: Average CS ratios for $S\alpha S$-CS, LASSO, SL0 and SWCGP as a function of α with $L \in \{5, 10, 15, 20\}$ ($N = 512$, $M = 120$, FSNR $= 10$ dB).

REFERENCES

[1] S. Mallat, "A Wavelet Tour of Signal Processing", 2nd ed., New York: Academic Press, 1998.

[2] D. Donoho, "Compressed Sensing", *IEEE Trans. Inf. Theory*, Vol. 52, No. 4, pp. 1289–1306, Apr. 2006.

[3] S. Chen, D. Donoho and M. Saunders, "Atomic decomposition by Basis Pursuit", *SIAM J. on Sci. Comp.*, Vol. 20, No. 1, pp. 33–61, 1998.

[4] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. on Sig. Proc.*, Vol. 41, No. 12, pp. 3397–3415, 1993.

[5] R. Kapoor, A. Banerjee, G. Tsihrintzis and N. Nandhakumar, "UWB radar detection of targets in foliage using alpha-stable clutter models," *IEEE Trans. on Aerospace & Elec. Sys.*, Vol. 35, No. 3, pp. 819–834, Jul. 1999.

[6] M. Taroudakis, G. Tzagkarakis and P. Tsakalides, "Classification of shallow-water acoustic signals via alpha-Stable modeling of the one-dimensional wavelet coefficients," *J. Acoust. Soc. Am.*, Vol. 119, No. 3, pp. 1396–1405, Mar. 2006.

[7] S. Liao and A. C. Chung, "Feature based nonrigid brain MR image registration with symmetric alpha stable filters," *IEEE Trans. on Med. Img.*, Vol. 29, No. 1, pp. 106–119, Jan. 2010.

[8] S. Babacan, R. Molina and A. Katsaggelos, "Fast Bayesian Compressive Sensing using Laplace Priors," *in Proc. IEEE Int. Conf. on Acoust., Speech & Sig. Proc. (ICASSP'09)*, Taipei, Taiwan, April 2009.

[9] G. Tzagkarakis and P. Tsakalides, "Bayesian Compressed Sensing of a Highly Impulsive Signal in Heavy-Tailed Noise using a Multivariate Cauchy Prior," *in Proc. Europ. Sig. Proc. Conf. (EUSIPCO'09)*, Glasgow, Scotland, Aug. 24–28, 2009.

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

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

[12] J. Breton and C. Houdré, "On finite range stable-type concentration," *Theory Prob. Appl.*, Vol. 52 (4), pp. 543–564, 2008.