

Compressive Sensing Using Symmetric Alpha-Stable Distributions – Part I: Robust Nonlinear Sampling

George Tzagkarakis*, John P. Nolan, and Panagiotis Tsakalides, *Member, IEEE*

Abstract—In the field of compressive sensing (CS), signal and noise models of finite variance are well established due to their analytical tractability and practical appeal. Typically, sampling and reconstruction methods are designed by assuming light-tailed models for the underlying signal and/or noise statistics. However, when we operate in highly impulsive environments, non-Gaussian infinite variance generating processes appear naturally for the signal and/or noise components. Because of this, conventional linear sampling operators, coupled with traditional reconstruction methods, are inefficient to accurately recover the original signal. To address these limitations, recent approaches adopted the use of algebraic-tailed statistical models, such as the Lévy and Cauchy distributions, to construct nonlinear sampling operators and robust reconstruction methods. However, the performance of these methods is degraded when the signal and/or noise statistics deviate significantly from the assumed model. To address this problem, this paper and the companion paper (Part II) propose an innovative CS framework exploiting the power of symmetric alpha-stable distributions in modeling highly impulsive signals. This paper introduces an efficient compressive sampling method that suppresses the effects of impulsive observation noise by designing a robust nonlinear sampling operator based on a generalized alpha-stable matched filter. In the companion paper (Part II), a novel greedy algorithm is introduced for reconstructing sparse signals, which achieves increased robustness to impulsive sampling noise. The theoretical justification along with the experimental evaluation demonstrate the improved performance of the proposed framework when compared against state-of-the-art CS techniques for a broad range of impulsive environments.

Index Terms—Compressive sampling, nonlinear sampling, symmetric alpha-stable distributions, heavy-tailed statistics, fractional lower-order moments, weighted stable matched filter.

I. INTRODUCTION

COMPRESSIVE sensing (CS) has recently been established as an innovative signal acquisition and reconstruction strategy, where additional structure of the signal is exploited to enable sampling rates far below what is dictated by the traditional Nyquist-Shannon sampling theorem. The structure which is primarily associated with CS is that of

signal *sparsity* in a transform domain, or, equivalently, over a sparsifying basis. For instance, at the core of most image compression algorithms is the fact that natural images are sparse, or close to sparse, over sparsifying bases, such as wavelets and cosines [1]. In such a scenario, CS acquisition is realized by taking linear projections of the signal onto a small set of vectors that are incoherent with the sparsity-inducing basis [2], [3]. Given these projections, hereafter called *measurements*, the signal is recovered by searching for the sparsest representation in the transform basis, whilst simultaneously being consistent with the measurements.

The random matrix that generates the undersampled linear measurements must satisfy specific necessary and sufficient conditions (e.g. the null space property and the restricted isometry property for ℓ_1 -norm minimization) to guarantee the successful recovery of sparse signals [4]. An appealing attribute of CS is that the seminal breakthrough was made by employing random vectors and randomly selected vectors from orthonormal matrices [2], [5]. Practical applications, however, often do not allow the use of totally random matrices, due to certain physical constraints on the measurement process. To overcome these limitations, structurally random measurement matrices were introduced enabling the construction of fast and efficient sensing matrices for practical CS [6].

Furthermore, in a real acquisition system, the signal of interest is always corrupted by noise. In this paper, we focus on the design of a robust nonlinear sampling operator, in order to suppress the effects of highly impulsive *observation* noise, while still being able to reconstruct the true sparse signal using a conventional sparse reconstruction algorithm. The accurate reconstruction of sparse signals whose random measurements are corrupted by heavy-tailed *sampling* noise of infinite variance is studied in the companion paper (Part II) [7].

A. Motivation

Most of the well-established sparse reconstruction algorithms provide bounded reconstruction error by assuming bounded or Gaussian noise, as well as light-tailed, finite-variance models for the underlying signal and/or noise statistics, due to their analytical tractability and practical appeal. On the other hand, there is a remarkably wide range of practical applications, where information is recorded in highly *impulsive* environments, giving rise to non-Gaussian, heavy-tailed processes for the representation of the associated signal

This work was partially funded by EONOS Investment Technologies, and the European Union's Horizon 2020 Research and Innovation Programme PHYStIS under grant agreement No. 640174.

G. Tzagkarakis is with EONOS Investment Technologies, Paris, France, e-mail: gtzag@eonos.com.

J. P. Nolan is with the Math/Stat Department, American University, Washington, DC, e-mail: jpnolan@american.edu. Supported by contract W911NF-12-1-0385 from the U.S. Army Research Office.

P. Tsakalides is with the Institute of Computer Science, Foundation for Research and Technology-Hellas (FORTH-ICS) and the Department of Computer Science, University of Crete, Heraklion, Greece, e-mail: tsakalid@ics.forth.gr.

and/or corrupting noise. Examples of such applications can be found in underwater acoustics [8], sonar and radar [9], [10], medical imaging [11], and finance [12], just to name a few.

When we operate in impulsive environments, the true signal is often corrupted by heavy-tailed noise, which occurs in the form of big errors that mask the original information content of the signal. In the framework of CS, the presence of large-amplitude noise samples degrades dramatically the generated linear random measurements, since the large errors are spread throughout the measurements due to the linear sampling process. The uniform spreading of large noise samples, which can be of infinite or very large variance in the case of impulsive environments, over the random measurements, causes traditional CS reconstruction algorithms to fail in recovering a close approximation of the true sparse signal.

Several recent CS techniques tackle the presence of impulsive observation noise either in a completely distribution-agnostic framework [13] or based on the theory of robust statistics [14] and nonlinear signal processing [15]. In contrast to the methods based on robust statistics, which are more generic, the main limitation of the distribution-agnostic approach is that it relies on noise of bounded variance, while the sparse signal is estimated based on a minimum mean squared error criterion, which is not valid for signals characterized by heavy-tailed statistics with infinite variance, due to the lack of second-order moments.

More specifically, the problem of suppressing the effects of impulsive observation noise is addressed in a robust statistical framework by employing heavy-tailed distributions coupled with maximum likelihood (ML) type estimators (also known as M-estimators) as data fidelity functions (ref. [16] for a thorough review). In [17], [18], the reduced set of compressive measurements is constructed using Cauchy random projections, which are not severely degraded if the signal is corrupted by gross errors, and also they are more suitable in applications where ℓ_1 -norm preservation is preferred in the low-dimensional measurements space. A similar approach is proposed in [19], where the generalized Cauchy distribution (GCD) family is employed to design a robust nonlinear measurement operator, based on the weighted myriad estimator [15]. The generated compressive measurements yield an improved reconstruction performance against their linear projections counterparts when combined with traditional CS algorithms (e.g. orthogonal matching pursuit). The interest in Cauchy and GCD random projections mainly arises due to their closed-form expressions, and subsequently their computational tractability in modeling impulsive environments.

Despite the improved performance of the above methods in compressively sampling signals recorded in impulsive environments and corrupted by gross errors, the specific use of Laplace [20], Cauchy [17], or GCD distributions [19] can be restrictive in capturing more generic non-Gaussian, heavy-tailed behaviors. In particular, several studies have demonstrated the power of the *alpha-stable distributions* [21] family, and especially the subclass of *symmetric alpha-stable* (S α S) distributions, in accurately modeling impulsive signal and noise processes [22], [23], [24], [25]. Alpha-stable distributions owe their importance in both theory and practice to

the generalization of the central limit theorem to random variables without second-order moments, since the only possible nontrivial limit of normalized sums of independent identically distributed (i.i.d.) terms, possibly of infinite variance, follows an alpha-stable distribution. However, despite the effectiveness of alpha-stable models in describing a broad range of statistical behaviors, from linear (i.e., Gaussian) to extremely impulsive signals, the lack of closed-form expressions for the density functions of all except for a few stable distributions (Gaussian, Cauchy and Lévy), has been a major drawback to their extensive use by the signal processing community.

B. Main Contributions

The main contributions of this paper are twofold: i) we propose a robust nonlinear sampling operator, based on the *weighted S α S matched filter* (WSMF) [26]. Our generated WSMF compressive measurements exploit the rich class of symmetric alpha-stable distributions. This yields an increased robustness for a broader range of impulsive observation noise behaviors, when compared with the previous methods based on the Cauchy and GCD distributions; ii) we provide an explicit theoretical justification of the improved outlier rejection capability of WSMF measurements, as well as of their asymptotic behavior with respect to the key parameters that control the performance of a weighted S α S matched filter.

C. Paper Organization

The rest of the paper is organized as follows: in Section II, we review the noisy signal model adopted in this study, in conjunction with the main concepts of the S α S distributions family, which is a key ingredient of our proposed method. Section III first points out the limitations of conventional linear compressive sampling in impulsive environments. Then, the weighted stable matched measurements are defined and their properties are proven, as a proper nonlinear sampling operator with increased robustness to impulsive observation noise. An experimental evaluation of the robustness of our proposed sampling operator is presented in Section IV for a variety of impulsive environments, using conventional sparse reconstruction algorithms. Finally, Section V summarizes the main outcomes and gives directions for future research.

D. Notation

In the following, we denote scalars with lower-case letters (e.g. x), column vectors with lower-case boldface letters (e.g. \mathbf{x}), matrices with upper-case boldface letters (e.g. \mathbf{X}), while calligraphic letters are used to denote sets (e.g. \mathcal{S}). We use \mathbf{x}_i to represent the i th column of a matrix \mathbf{X} , x_j to designate the j th element of a vector \mathbf{x} , and \mathcal{S}_i to indicate a subset of a set \mathcal{S} , whose cardinality is denoted by $|\mathcal{S}|$. By $\mathbf{X}_{\mathcal{S}}$ we designate the submatrix formed by the columns $\{\mathbf{x}_i \mid i \in \mathcal{S}\}$, whose indices belong to the set \mathcal{S} . Similarly, by $\mathbf{x}_{\mathcal{S}}$ we denote the subvector formed by the elements $\{x_j \mid j \in \mathcal{S}\}$, whose indices belong to \mathcal{S} . In both cases, the order is preserved among the retained elements. Finally, $\hat{\mathbf{x}}$, \mathbf{x}^T , \mathbf{x}^* , and $\mathbf{x}^{(t)}$ denote the estimate (reconstruction), transpose, optimal solution, and value at t th iteration of a vector \mathbf{x} , respectively. Similar notations are used for the matrices.

II. COMPRESSIVE SAMPLING AND S α S DISTRIBUTIONS

This section describes the typical compressive sampling process for signals corrupted by additive observation noise, along with the basics of S α S distributions, which are at the core of our proposed sampling operator.

A. Compressive Sampling of Noisy Signals

Let $\mathbf{x} = [x_1, x_2, \dots, x_N]^T \in \mathbb{R}^N$ denote an observed discrete-time signal with real-valued elements. In the following, we assume that \mathbf{x} can be either s -sparse ($s \ll N$) by itself, that is, $|\{j \mid x_j \neq 0\}| \leq s$, or sparse in some transform basis (a.k.a. *dictionary*) Ψ , such that $\alpha = \Psi\mathbf{x}$, where $\alpha \in \mathbb{R}^{N'}$ is the s -sparse vector of transform coefficients. Notice that, in general, $N' \geq N$, since Ψ can be overcomplete [1]. In the subsequent analysis, Ψ and Ψ^T denote the analysis (direct) and synthesis (inverse) transforms, respectively.

In practice, the acquired signal is typically corrupted by observation noise, which is defined as a perturbation introduced to the true signal prior to its sampling. In the following, we adopt an *additive* model for the observation noise, that is,

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{e}_o, \quad (1)$$

where $\mathbf{x}_0 \in \mathbb{R}^N$ is the true noiseless signal and $\mathbf{e}_o \in \mathbb{R}^N$ is the observation noise component.

Let $A : \mathbb{R}^N \mapsto \mathbb{R}^M$ with $M < N$ denote a sampling operator that maps a vector of N elements to a lower-dimensional vector of M *measurements*. The compressive sampling of \mathbf{x} is expressed by $\mathbf{y} = A(\mathbf{x})$, where $\mathbf{y} \in \mathbb{R}^M$ is the vector of measurements. In this paper we address only the case of observation noise. The problem of suppressing the effects of sampling noise, corrupting the generated measurements, is the subject of the companion paper [7].

In conventional CS systems the sampling operator $A(\cdot)$ is a linear map. Considering the general case when the true signal is sparse in a transform basis Ψ , $\mathbf{x}_0 = \Psi^T\alpha_0$, the vector of measurements is constructed by taking linear projections onto the rows of a random matrix,

$$\mathbf{y} = \Phi (\Psi^T\alpha_0 + \mathbf{e}_o) = \Phi\Psi^T\alpha_0 + \mathbf{n}, \quad (2)$$

where $\Phi \in \mathbb{R}^{M \times N}$ is a random *measurement matrix* and $\mathbf{n} = \Phi\mathbf{e}_o \in \mathbb{R}^M$ is the projected noise. Φ must satisfy specific conditions (e.g. the null space property and the restricted isometry property for ℓ_1 -norm minimization, and incoherence¹ with Ψ^T) to guarantee the successful reconstruction of a sparse signal. The definition of the restricted isometry property (RIP) and the restricted isometry constant (RIC) is as follows (ref. [4] for more details):

Definition 1 (s-RIC) *The s -restricted isometry constant of Φ is defined as the smallest positive quantity δ_s , such that*

$$(1 - \delta_s)\|\mathbf{v}\|_2^2 \leq \|\Phi\mathbf{v}\|_2^2 \leq (1 + \delta_s)\|\mathbf{v}\|_2^2 \quad (3)$$

holds $\forall \mathbf{v} \in \mathcal{T}_s$, where $\mathcal{T}_s = \{\mathbf{v} \in \mathbb{R}^N \mid \|\mathbf{v}\|_0 \leq s\}$. A matrix Φ is said to satisfy a RIP of order s if $\delta_s \in (0, 1)$.

¹Although incoherence among Φ and Ψ^T is a requirement for guaranteeing accurate sparse reconstruction [2], recent works have proven that for truly redundant dictionaries a no-incoherence restriction on the dictionary can still guarantee accurate sparse recovery [27].

By setting $\mathbf{A} = \Phi\Psi^T$ as the generic linear sampling operator, the true sparse coefficients vector α_0 can be recovered by solving an $\ell_1 - \ell_2$ constrained optimization problem of the form,

$$\min_{\alpha \in \mathbb{R}^{N'}} \|\alpha\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\alpha\|_2 \leq \varepsilon, \quad (4)$$

where $\varepsilon > 0$ is a threshold depending on the noise level. Then, an estimate of the true signal is given by $\hat{\mathbf{x}}_0 = \Psi^T\alpha_0^*$. We emphasize again that the focus of this paper is to develop a novel robust nonlinear sampling operator, which suppresses the effects of highly impulsive observation noise, whereas it achieves superior reconstruction accuracy when combined with conventional sparse reconstruction algorithms. Several efficient optimization formulations have been proposed in the literature; for convenience and without loss of generality we consider the basis pursuit denoising (BPD) formulation in (4), which is solved effectively using, for instance, the orthogonal matching pursuit (OMP) algorithm [28].

B. S α S Modeling of Heavy-Tailed Statistics

In this section, the family of univariate S α S distributions is introduced as a powerful tool for modeling heavy-tailed random variables of infinite variance. The main reason preventing the widespread usage of the S α S family in signal modeling is that, in general, no closed-form expressions exist for most S α S density and distribution functions except for the Gaussian ($\alpha = 2$) and the Cauchy ($\alpha = 1$). Indeed, the probability density function of a general univariate S α S distribution is as follows [29],

$$f_\alpha(x; \gamma, \delta) = \frac{1}{\gamma} q\left(\frac{x - \delta}{\gamma}; \alpha\right), \quad (5)$$

where

$$q(x; \alpha) = \frac{1}{\pi} \int_0^\infty \cos(xt) e^{-t^\alpha} dt. \quad (6)$$

In the above expressions, $\alpha \in (0, 2]$ is the *characteristic exponent*, $\gamma > 0$ is the *dispersion*, and $\delta \in \mathbb{R}$ is the *location parameter* of the distribution. The characteristic exponent is a shape parameter, which controls the thickness of the tails of the density function. The smaller the α , the heavier the tails of the S α S density function. The dispersion parameter determines the spread of the distribution around its location parameter, as the standard deviation does for a Gaussian distribution.

Unlike the Gaussian density, which has exponential tails, all other stable densities have tails following an algebraic rate of decay, that is, $\Pr(X > x) \sim C_f \cdot x^{-\alpha}$, as $x \rightarrow \infty$, where C_f is a constant depending on the model parameters. It follows that S α S random variables with small α values are highly impulsive. Hereafter, the notation $X \sim f_\alpha(\gamma, \delta)$ denotes that a random variable X follows a S α S distribution with parameters α, γ, δ . In the special case where $\gamma = 1$ and $\delta = 0$ the distribution is called *standard* S α S. Furthermore, it holds that, if $X \sim f_\alpha(\gamma_X, 0)$ and $Y \sim f_\alpha(\gamma_Y, 0)$ are two independent jointly S α S (i.e., with equal characteristic exponents) random variables, then,

$$cX + \delta \sim f_\alpha(|c|\gamma_X, \delta) \quad (c \neq 0), \quad (7)$$

$$X + Y \sim f_\alpha\left((\gamma_X^\alpha + \gamma_Y^\alpha)^{1/\alpha}, 0\right). \quad (8)$$

In the following, we assume that the noise statistics is modeled by a $S\alpha S$ distribution located at $\delta = 0$. Nevertheless, the distribution's location can be always shifted to zero via (7) by subtracting δ . In our implementation, the $S\alpha S$ model parameters (α, γ) are estimated from a discrete set $\{x_1, \dots, x_n\}$ of realizations of X using the consistent maximum likelihood (ML) method described in [30], which gives reliable estimates and provides the tightest possible confidence intervals.

III. ROBUST NONLINEAR MEASUREMENT OPERATOR

In this section, we propose a novel nonlinear sampling operator, which extends the state of the art by achieving increased robustness against a broader range of impulsive noise behaviors, from linear (i.e., Gaussian) to extremely impulsive observation noise. First, we demonstrate the inefficiency of linear projections to deal with heavy-tailed observation noise. Then, we define the *weighted $S\alpha S$ matched* measurements as a proper robust nonlinear sampling operator, which generalizes recently introduced nonlinear sampling methods.

A. Effects of Heavy-Tailed Observation Noise on Reconstruction Accuracy

When linear projections are used to generate the measurements, in conjunction with traditional sparse reconstruction methods, the *oracle estimator* achieves the best reconstruction performance when no prior information about the distribution of the true sparse signal is known. The oracle estimator, though, has perfect knowledge of the support of \mathbf{x}_0 (or of α_0 when $\Psi \neq \mathbf{I}$), that is, the set $\mathcal{T} \subset \{1, 2, \dots, N\}$ of indices of the s nonzero elements of the sparse vector. Then, by assuming that the noise is power-limited, or has a finite variance, an estimator of the sparse signal is constructed by taking the least-squares projection onto the subspace which is spanned by the columns of Φ (or of \mathbf{A} in the generic case) with indices in \mathcal{T} .

In [31], the mean squared error (MSE) is derived for the oracle estimator in the case of Gaussian distributed noise. Specifically, let Φ be a measurement matrix with rows of norm $\sqrt{\lambda}$, which satisfies the RIP of order s with constant δ_s , and $\sigma_{e_o}^2$ denotes the variance of the observation noise. When the signal is corrupted by observation noise only, the MSE of the oracle estimator is lower bounded by

$$\mathbb{E} \left\{ \|\mathbf{x}_0 - \hat{\mathbf{x}}\|_2^2 \right\} \geq \frac{s \cdot \lambda \cdot \sigma_{e_o}^2}{1 + \delta_s}, \quad (9)$$

where $\mathbb{E}\{\cdot\}$ denotes the expected value of a random variable.

From (9) it is evident that linear projections do not constitute a satisfactory sampling operator when the original signal is corrupted by heavy-tailed observation noise. Indeed, the expected reconstruction error for the oracle estimator can be very large when $\sigma_{e_o}^2$ is very large, or even infinite, as is the case of heavy-tailed noise.

Fig. 1 demonstrates the inefficiency of traditional $\ell_1 - \ell_2$ CS methods to recover impulsive signals corrupted by gross errors. More specifically, we consider a sparse vector α_0 of length $N = 1024$ with $s = 21$ randomly chosen nonzero elements drawn from the standard Cauchy distribution, as

shown in Fig. 1a. The original signal, shown in Fig. 1b, is given by $\mathbf{x}_0 = \Psi^T \alpha_0$, where Ψ^T is the $N \times N$ discrete cosine transform (DCT) matrix. A vector \mathbf{y}_0 of $M = 256$ linear measurements is generated (Fig. 1c) by projecting \mathbf{x}_0 onto the rows of a matrix $\Phi \in \mathbb{R}^{256 \times 1024}$ whose entries are i.i.d. standard Gaussian samples and its columns are normalized to unit ℓ_2 norm. Then, a single outlier of amplitude $\eta = 10^2$ is added to a randomly chosen element of \mathbf{x}_0 and a new vector \mathbf{y} of $M = 256$ linear measurements is generated using the same matrix Φ (ref. Fig. 1d), with the outlier being spread through all the measurements. Figs. 1e-1f show the reconstruction of the true sparse vector α_0 from \mathbf{y}_0 and \mathbf{y} , respectively, using the NESTA algorithm², which is a fast and robust first-order method that solves Basis Pursuit (BP) problems. Clearly, NESTA recovers perfectly the true sparse vector, and subsequently the original signal \mathbf{x}_0 , in the noiseless case, whereas it fails completely when even a single outlier corrupts the original signal. In particular, the reconstruction in the noiseless case yields a signal-to-error ratio (SER) of 175.32 dB, and a SER of -12.09 dB in the noisy case, where the SER (in dB) is defined as

$$\text{SER}(\mathbf{x}, \hat{\mathbf{x}}) = 10 \log_{10} \left(\frac{\sum_{j=1}^N x_j^2}{\sum_{j=1}^N (x_j - \hat{x}_j)^2} \right). \quad (10)$$

The problem of alleviating the effects of heavy-tailed observation noise in a CS-based signal acquisition system has been addressed recently in the framework of robust statistics [18], [19]. Specifically, the algebraic-tailed Cauchy and generalized Cauchy (GCD) distributions were utilized to derive nonlinear measurement operators based on weighted myriad (WM) estimators. These estimators achieve increased robustness against gross observation errors, whilst the generated random measurements resemble the linear projections approach in the noiseless case.

In the following, we propose a new nonlinear measurement operator, which includes the weighted myriad-based ones as a special case, whilst achieving increased robustness to a broader range of observation noise statistics, from Gaussian to extremely impulsive behaviors. To this end, we construct the compressive *weighted $S\alpha S$ matched* measurements as the output of a weighted matched filter by utilizing the family of $S\alpha S$ distributions, which generalizes the properties of the above models. Also notice that although a GCD-based analysis is carried out in [19], the sparse signal reconstruction relies on weighted myriad measurements and ℓ_1 minimization with a Lorentzian-norm constraint, which are optimal in the special case of a standard Cauchy model.

B. Stable Matched Filter

Under the adverse conditions arising in impulsive environments, the performance of linear adaptive filters, in terms of suppressing the heavy-tailed noise effects, can deteriorate significantly. To address this problem, the use of generalized ML estimators, the so-called M-estimators, has demonstrated

²NESTA toolbox: <http://statweb.stanford.edu/~candes/nesta/nesta.html>.

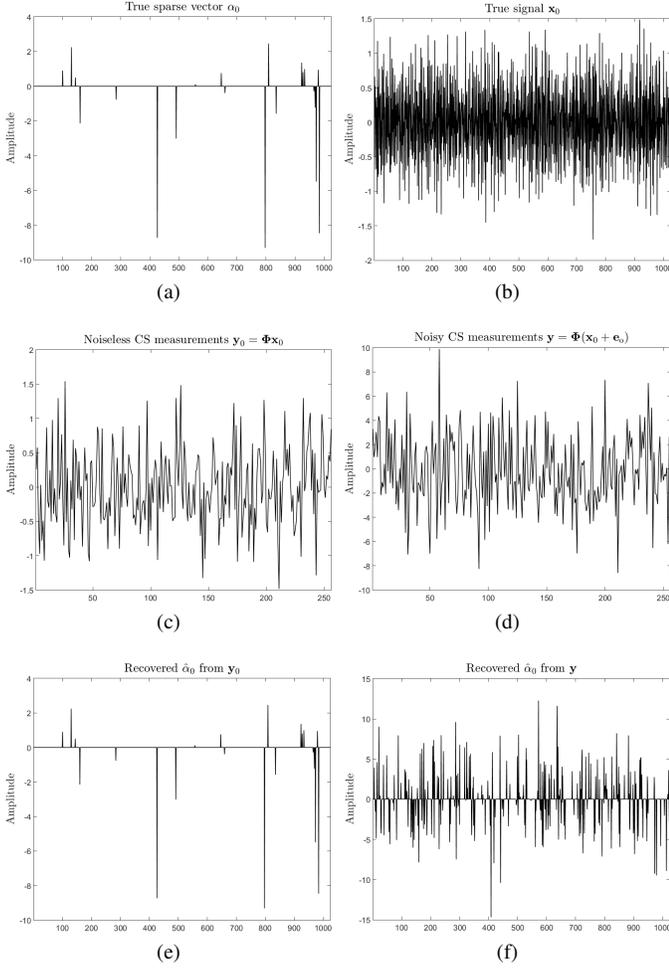


Fig. 1. Reconstruction of an impulsive signal corrupted by a single outlier. NESTA algorithm is used to solve a BP problem. (a) True sparse vector; (b) original DCT synthesized signal; (c) linear measurements in the noiseless case; (d) linear measurements in the noisy case; (e) reconstructed sparse vector in the noiseless case; (f) reconstructed sparse vector in the case of an outlier.

increased robustness to impulsive noise. Motivated by the success of M-estimators, our proposed robust sampling operator is also developed in the framework of robust statistics.

More specifically, an M-estimator built on a S α S distribution yields a robust nonlinear filtering technique, which is optimal for infinite variance noise. Let $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$ be the sampled signal and $\rho(x) = -\log(f_\alpha(x; \gamma, 0))$ be the negative of the log S α S density for $x \in \mathbb{R}$, and define the following cost function,

$$G_{\alpha, \gamma}(\theta; \mathbf{x}) = \sum_{j=1}^N \rho(x_j - \theta) = - \sum_{j=1}^N \log(f_\alpha(x_j - \theta; \gamma, 0)). \quad (11)$$

Then, the *unweighted S α S filter* (USF) is defined as follows:

Definition 2 (Unweighted S α S filter) Given a vector of samples $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$ and the cost function $G_{\alpha, \gamma}(\theta; \mathbf{x})$, the *unweighted S α S filter* is defined as

$$\hat{\theta}_{\alpha, \gamma}(\mathbf{x}) \triangleq \text{USF}(\alpha, \gamma; \mathbf{x}) = \arg \min_{\theta} G_{\alpha, \gamma}(\theta; \mathbf{x}). \quad (12)$$

Since $\rho(x)$ is the negative of the log density, the minimum $\hat{\theta}_{\alpha, \gamma}$ is exactly the M-estimator of location. Note that in the Gaussian case ($\alpha = 2$), the minimum can be found explicitly and is simply the sample mean. However, in the general S α S case ($0 < \alpha < 2$) the filters are nonlinear with no closed-form solution, and the minimum in (12) must be found numerically. We note that all the subsequent numerical calculations and optimizations involving S α S densities are performed using the STABLE toolbox³.

A natural extension of the unweighted S α S filter is obtained by assigning weights to the samples in the ML location estimator. The weights reflect the different levels of reliability of the observed samples. More specifically, consider the samples vector $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$ and a weight vector $\mathbf{w} = [w_1, w_2, \dots, w_N]^T \in \mathbb{R}^N$. A weighted modification of the cost function in (11) is defined as follows,

$$\begin{aligned} G_{\alpha, \gamma}(\theta; \mathbf{w}, \mathbf{x}) &= \sum_{j=1}^N \rho(w_j(x_j - \theta)) \\ &= - \sum_{j=1}^N \log(f_\alpha(w_j(x_j - \theta); \gamma, 0)). \end{aligned} \quad (13)$$

Then, the *weighted S α S matched filter* (WSMF) is defined accordingly:

Definition 3 (Weighted S α S matched filter) Given a vector of samples $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$, a weight vector $\mathbf{w} = [w_1, w_2, \dots, w_N]^T$, and the cost function $G_{\alpha, \gamma}(\theta; \mathbf{w}, \mathbf{x})$, the *weighted S α S matched filter* is defined as

$$\hat{\theta}_{\alpha, \gamma}(\mathbf{w}, \mathbf{x}) \triangleq \text{WSMF}(\alpha, \gamma; \mathbf{w}, \mathbf{x}) = \arg \min_{\theta} G_{\alpha, \gamma}(\theta; \mathbf{w}, \mathbf{x}). \quad (14)$$

The WSMF output is the value of θ at the global minimum of the weighted cost function $G_{\alpha, \gamma}(\theta; \mathbf{w}, \mathbf{x})$. However, reliable and accurate algorithms are required for solving (14) due to the nonconvex behavior of the cost function. A computationally tractable approach for the numerical calculation of the WSMF output is described in [26], [32] and integrated in the STABLE toolbox, which is employed in our implementation. This approach relies on a global minimization scheme based on branch-and-bound to guarantee that the global minimum is found. We also note that it is possible to find sample sets for which the WSMF output is not unique. To protect the formalism of Definition 3, we accept any solution to (14) as a valid calculation of the WSMF output. The degenerate case of getting more than one WSMF values has negligible probability for a large sample size. However, we may get more than one values at the output of a WSMF for small sample sizes. In this case, we choose one of them at random. Nevertheless, the global optimization scheme implemented in the STABLE toolbox guarantees convergence to the global minimum with high probability given a good initial guess.

Most importantly, the *weighted myriad filter* (WMyF), which has been exploited in recent studies [17], [19] for the design of robust measurement operators, constitutes an

³Robust Analysis Inc., STABLE toolbox version 5.3 (<http://www.robustanalysis.com>).

approximation of the WSMF. As such, we expect that our proposed nonlinear sampling operator based on the WSMF will achieve a superior performance, in terms of increased robustness to heavy-tailed observation noise, when compared against its myriad-based counterparts. For completeness of presentation we also cite the definition of the WMyF [19]:

Definition 4 (Weighted myriad filter) *Given a vector of samples $\mathbf{x} = [x_1, x_2, \dots, x_N]^\top$ and a weight vector $\mathbf{w} = [w_1, w_2, \dots, w_N]^\top$, the weighted myriad filter is defined as*

$$\hat{\theta}_k(\mathbf{w}, \mathbf{x}) \triangleq \text{WMyF}(k; \mathbf{w}, \mathbf{x}) \\ = \arg \min_{\theta} \sum_{j=1}^N \log \left(k^2 + |w_j| (\text{sign}(w_j)x_j - \theta)^2 \right) \quad (15)$$

where k is the scale parameter of the (generalized) Cauchy distribution. Notice that, although (15) is the optimal ML estimate of location for the standard Cauchy distribution, it has been also used in the case of the GCD family.

Fig. 2 compares the behavior of $\rho(x)$, which is the main constituent component of the cost functions defined above, for the standard S α S model, $\rho(x) = -\log(f_\alpha(x; 1, 0))$ with $\alpha \in \{0.5, 1, 1.5, 1.9\}$, and the standard Gaussian distribution, $\rho(x) = x^2$. In the standard S α S case, or equivalently, in the ℓ_p case with $p < 2$, the curves indicate the nonconvex nature of $\rho(x)$, which is opposed to the convex behavior of the Gaussian model, or equivalently, the squared ℓ_2 norm. Most importantly, this plot reveals that, in contrast to the Gaussian model, the cost functions based on $\rho(x)$ with $\alpha < 2$ downweight large deviations, thus yielding more robust filters in the presence of outliers. Furthermore, the zoomed inner plot illustrates that, as the characteristic exponent decreases, the function $\rho(x)$ is more robust to outliers, since it increases much more slowly when $|x| \rightarrow \infty$. This behavior also demonstrates the superiority of the S α S matched filter against the previous approaches based on the Cauchy distribution, in terms of better controlling the outlier rejection intensity through an additional degree of freedom (i.e., the parameter α) instead of relying only on the scale parameter.

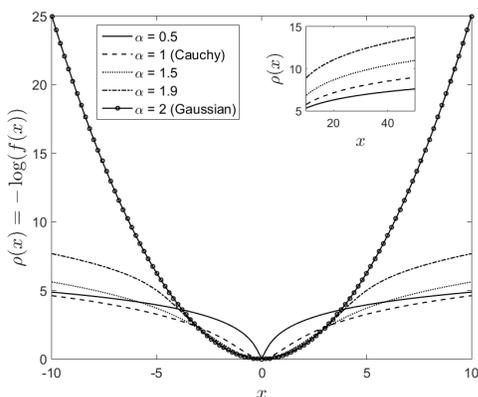


Fig. 2. Comparison of the $\rho(x) = -\log(f_\alpha(x; \gamma, 0))$ function for the standard S α S model ($\gamma = 1$) and the standard Gaussian case $\rho(x) = x^2$.

C. Controlling the Behavior of the S α S Matched Filter

The S α S matched filter enables a rich class of operating modes, which can be controlled by tuning the model parameters (α, γ). In particular, when we operate in light-tailed environments, where the noise statistics is close to a Gaussian, the optimal performance of the filter should be associated with the sample mean. On the other hand, for heavy-tailed noise, the filter should be resistant to large deviations, approximating the behavior of a mode-type estimator.

In Fig. 3, we examine the trade-off between efficiency in the light-tailed case and resistance to heavy-tailed observation noise, by tuning appropriately the values of (α, γ). To this end, we illustrate the behavior of the WSMF output, defined in (14), by varying the model parameters (α, γ), and compare against the performance of the WMyF, defined in (15), which depends on a single parameter k . For convenience, we refer to both γ and k as the scale parameter of the corresponding filter. Fig. 3a shows the histogram of a sample set, which is generated by nonuniformly drawing $N = 64$ random integers in the interval $[-1, 10]$, with an estimated average of 4.94 and a mode at 8. The values of the WSMF and WMyF are calculated for this sample set for varying $\alpha \in \{0.8, 1.2, 1.6, 1.9\}$ and scale parameter in $[10^{-2}, 10^2]$, while setting all the weights $w_j, j = 1, \dots, N$, equal to 1 (the WSMF is reduced to a USF in this case). Fig. 3b shows the value of each filter as a function of the scale parameter, for the various WSMFs and the WMyF. As it can be seen, as the scale parameter increases, the values of the filters tend asymptotically to the sample average, whilst as it decreases all the filters favor the mode 8, which indicates the location where samples are more likely to occur or cluster. The only difference among the filters concerns the regime of the scale parameter, where the filters switch between the mode-mean modes. Specifically, we observe that, as the WSMF tends to the Gaussian case (i.e., $\alpha \rightarrow 2$), the flat regime approximating the sample average covers a larger range of the scale parameter.

The behavior of the stable and the myriad filters to the presence of additive observation noise is illustrated in Figs. 3c-3d. In the first case, the original sample set is corrupted by light-tailed noise with parameters ($\alpha = 1.99, \gamma = 1$). As it can be seen, all the WSMFs and the WMyF present the same robustness against the near-Gaussian additive noise, with their behavior resembling the noiseless case (Fig. 3b). On the contrary, when the corrupting noise is drawn from a highly impulsive S α S distribution with parameters ($\alpha = 0.5, \gamma = 1$), the behavior of the filters becomes very sensitive to large deviations. In particular, we observe that the WMyF output diverges from the sample average for large values of the scale parameter, which is also the case for the WSMF outputs with moderate and large values of α . However, in contrast to the WMyF, whose characteristic exponent is fixed to $\alpha = 1$, the WSMF achieves increased robustness via the double tuning of α and γ . Indeed, as Fig. 3d demonstrates, the outlier rejection capability of the WSMF in the presence of gross noise samples can be enhanced by decreasing the value of α (see USF curve for $\alpha = 0.8$), which is not possible for the myriad filter. On the other hand, for small values of the scale

parameter, the dispersion of the data is assumed to be small, and the corrupting noise samples are considered as outliers, not affecting significantly the output of the WSMF and WMyF.

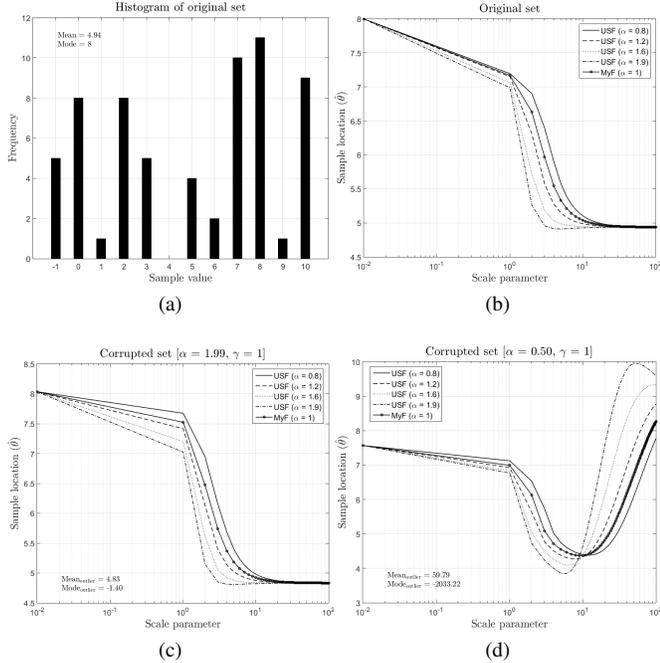


Fig. 3. Behavior of WSMF and WMyF filters as a function of the scale parameter and the characteristic exponent (for the WSMF) (weight vector is fixed to $\mathbf{w} = \mathbf{1}$, thus WSMF=USF). (a) Histogram of original sample set; (b) output of filters applied on original set; (c) output of filters applied on original set corrupted by noise samples drawn from a SaS model with $(\alpha = 1.99, \gamma = 1)$; (d) output of filters applied on original set corrupted by noise samples drawn from a SaS model with $(\alpha = 0.5, \gamma = 1)$.

D. WSMF Projections

Given the additive model for the observation noise as described in (1), our objective is to design a robust sampling operator $A : \mathbb{R}^N \mapsto \mathbb{R}^M$, which maps a vector of N elements to a lower-dimensional vector of M measurements ($M \ll N$). Such an operator must enable the accurate reconstruction of the acquired signal, while being resilient to the presence of gross errors when we operate in impulsive environments.

Let $\mathbf{x}_0 \in \mathbb{R}^N$ be the original noiseless signal, which is sparse in some basis Ψ (for simplicity we assume $\Psi = \mathbf{I}$), and $\mathbf{x} = \mathbf{x}_0 + \mathbf{e}_o$ be the observed noisy signal. Furthermore, let $\phi_i \in \mathbb{R}^N$, $i = 1, \dots, M$, be the measurement kernels that form the measurement matrix $\Phi \in \mathbb{R}^{M \times N}$, that is, $\Phi = [\phi_1 \phi_2 \dots \phi_M]^T$. In traditional CS, the i th measurement is generated as described in (2), that is, by taking the linear projection of \mathbf{x} onto the i th row of Φ , $y_i = \phi_i^T \mathbf{x}$.

In contrast to the conventional linear approach, our proposed compressive sampling operator is defined in a non-linear fashion. In particular, the i th measurement is given by $y_i = h(\phi_i, \mathbf{x})$, where $h : \mathbb{R}^N \times \mathbb{R}^N \mapsto \mathbb{R}$ denotes the nonlinear action of the i th measurement kernel on the signal. Subsequently, the overall nonlinear sampling operator is defined as

$$\mathbf{y} = A_{\Phi}(\mathbf{x}) \triangleq [h(\phi_1, \mathbf{x}), h(\phi_2, \mathbf{x}), \dots, h(\phi_M, \mathbf{x})], \quad (16)$$

where the subscript in A_{Φ} highlights the dependence of the operator on the measurement kernels.

When heavy-tailed observation noise corrupts the signal of interest, our nonlinear sampling operator must satisfy two basic requirements, namely, i) $A_{\Phi}(\mathbf{x})$ must preserve the information content of the original signal \mathbf{x}_0 by suppressing the effects of large noise samples; ii) $A_{\Phi}(\mathbf{x})$ should resemble the linear measurements as much as possible in the noiseless case, that is, $\mathbf{y} = A_{\Phi}(\mathbf{x}) \approx \Phi \mathbf{x}_0$ when $\mathbf{e}_o \rightarrow \mathbf{0}$. The second requirement enables the use of traditional sparse reconstruction algorithms for the recovery of \mathbf{x}_0 from the nonlinear measurements \mathbf{y} .

Given the problems arising in the presence of impulsive observation noise and the desired requirements of a robust sampling operator, we suggest the WSMF as a robust nonlinear sampling method. This is motivated by its increased robustness in a wide range of impulsive noise behaviors, from linear (i.e., Gaussian) to extremely impulsive observation noise, as well as by its asymptotic properties, which enable accurate reconstruction of the noise-free sparse signal using conventional CS reconstruction algorithms as if we operated in a light-tailed environment.

Our proposed WSMF projections for the generation of compressive measurements are defined as follows:

Definition 5 (WSMF projections) Let $\Phi = [\phi_1 \dots \phi_M]^T \in \mathbb{R}^{M \times N}$ ($M \ll N$) be a measurement matrix, whose i th row is the measurement kernel ϕ_i^T , and $\phi_{i,j}$ its (i, j) th element. Then, the WSMF projections are defined as

$$h_{\alpha, \gamma}(\phi_i, \mathbf{x}) \triangleq c_i \cdot \text{WSMF}(\alpha, \gamma; \phi_i, \mathbf{x}), \quad (17)$$

for $i = 1, \dots, M$, where the weighted stable matched filter is defined in (14) and c_i is an appropriate scaling factor used to adjust the magnitude of the measurements (e.g. $c_i = \sum_{j=1}^N |\phi_{i,j}|$).

As mentioned above, a desirable property of a robust sampling operator is the downweighting of outliers, that is, gross errors have negligible influence on the generated WSMF measurements. Indeed, the following property states the outlier rejection capability of our proposed WSMF projections.

Property 1 (Outlier rejection of WSMF projections) Let $\gamma < \infty$, then

$$\lim_{x_N \rightarrow \pm\infty} h_{\alpha, \gamma}(\phi_i, [x_1, x_2, \dots, x_N]) = h_{\alpha, \gamma}(\phi_i, [x_1, x_2, \dots, x_{N-1}]). \quad (18)$$

Proof: The above property results by combining (13), (14) and (17), and noticing that by adding a constant (see third equation below) to the objective function it does not affect the optimization, as follows:

$$\begin{aligned} & \lim_{x_N \rightarrow \pm\infty} h_{\alpha, \gamma}(\phi_i, [x_1, x_2, \dots, x_N]) \\ &= \lim_{x_N \rightarrow \pm\infty} \arg \min_{\theta} - \sum_{j=1}^N \left(\log(f_{\alpha}(\phi_{i,j}(x_j - \theta); \gamma, 0)) \right) \left(\right. \\ &= \lim_{x_N \rightarrow \pm\infty} \arg \min_{\theta} - \sum_{j=1}^{N-1} \left(\log(f_{\alpha}(\phi_{i,j}(x_j - \theta); \gamma, 0)) \right) \end{aligned}$$

$$\begin{aligned}
& -\log(f_\alpha(\phi_{i,N}(x_N - \theta); \gamma, 0)) \Bigg) \Bigg(\\
& = \lim_{x_N \rightarrow \pm\infty} \arg \min_{\theta} - \sum_{j=1}^{N-1} \left(\log(f_\alpha(\phi_{i,j}(x_j - \theta); \gamma, 0)) \right. \\
& \left. - \log(f_\alpha(\phi_{i,N}(x_N - \theta); \gamma, 0)) + \log(f_\alpha(\phi_{i,N}x_N; \gamma, 0)) \right) \Bigg) \Bigg(\\
& = \lim_{x_N \rightarrow \pm\infty} \arg \min_{\theta} - \sum_{j=1}^{N-1} \left(\log(f_\alpha(\phi_{i,j}(x_j - \theta); \gamma, 0)) \right. \\
& \left. - \log \left(\frac{f_\alpha(\phi_{i,N}(x_N - \theta); \gamma, 0)}{f_\alpha(\phi_{i,N}x_N; \gamma, 0)} \right) \right) \Bigg) \Bigg(\quad (19)
\end{aligned}$$

Evaluating the above limit is equivalent to finding the limit of the second logarithmic term in (19). Due to the continuity of the logarithm for all $x > 0$, the lim and log can be interchanged yielding an indeterminate form $0/0$, since $f_\alpha(x; \gamma, 0) \rightarrow 0$ as $x \rightarrow \pm\infty$. By applying L'Hospital's rule and using the expansion of the first derivative of a S α S density $f'_\alpha(x; \gamma, 0)$ as $x \rightarrow \pm\infty$ [33], the previous indeterminate form is evaluated as follows:

$$\begin{aligned}
& \lim_{x_N \rightarrow \pm\infty} \log \left(\frac{f_\alpha(\phi_{i,N}(x_N - \theta); \gamma, 0)}{f_\alpha(\phi_{i,N}x_N; \gamma, 0)} \right) \Bigg(\\
& = \log \left(\lim_{x_N \rightarrow \pm\infty} \frac{f'_\alpha(\phi_{i,N}(x_N - \theta); \gamma, 0)}{f'_\alpha(\phi_{i,N}x_N; \gamma, 0)} \right) \Bigg(\\
& = \log \left(\lim_{x_N \rightarrow \pm\infty} \frac{\frac{1}{\gamma^2 \pi} \sum_{k=1}^{\infty} c_{k,\alpha} \left(\frac{\phi_{i,N}(x_N - \theta)}{\gamma} \right)^{-\alpha k - 2}}{\frac{1}{\gamma^2 \pi} \sum_{k=1}^{\infty} c_{k,\alpha} \left(\frac{\phi_{i,N}x_N}{\gamma} \right)^{-\alpha k - 2}} \right) \Bigg) \Bigg(\\
& = \log(1) = 0, \quad (20)
\end{aligned}$$

where $c_{k,\alpha} = \frac{\Gamma(\alpha k + 2)}{k!} (-1)^k \sin\left(\frac{\pi \alpha k}{2}\right)$. The last equality follows from the fact that the terms of the fraction corresponding to the same power are cancelled out as $x \rightarrow \pm\infty$. Combining (19) and (20) yields

$$\begin{aligned}
& \lim_{x_N \rightarrow \pm\infty} h_{\alpha,\gamma}(\phi_i, [x_1, x_2, \dots, x_N]) \\
& = \arg \min_{\theta} \left(- \sum_{j=1}^{N-1} \left(\log(f_\alpha(\phi_{i,j}(x_j - \theta); \gamma, 0)) \right) \right) \Bigg(\\
& = h_{\alpha,\gamma}(\phi_i, [x_1, x_2, \dots, x_{N-1}]), \quad (21)
\end{aligned}$$

which completes the proof. \blacksquare

E. Asymptotic Behavior of WSMF Projections and Parameter Setting

Regarding the asymptotic behavior of our WSMF projections with respect to the characteristic exponent α , the following property holds as α approaches the critical points 0, 1, and 2.

Property 2 (Asymptotic behavior of WSMF projections with respect to α) Let $\gamma < \infty$, then the asymptotic behavior

of the WSMF projections as α approaches the critical points 0, 1, and 2, is as follows:

$$\lim_{\alpha \rightarrow 1} h_{\alpha,\gamma}(\phi_i, \mathbf{x}) = h_{1,\gamma}(\phi_i, \mathbf{x}) \quad (22)$$

$$\lim_{\alpha \rightarrow 2} h_{\alpha,\gamma}(\phi_i, \mathbf{x}) = h_{2,\gamma}(\phi_i, \mathbf{x}) \quad (23)$$

$$\lim_{\alpha \rightarrow 0} h_{\alpha,\gamma}(\phi_i, \mathbf{x}) = c_i \cdot \min_{j=1, \dots, N} \{G_{\alpha,\gamma}(x_j; \phi_i, \mathbf{x})\}, \quad (24)$$

where c_i is a scaling factor as in Definition 5.

Proof: The first two equalities result directly from the definition of the WSMF projections, and specifically from the smoothness of the cost function $G_{\alpha,\gamma}(\theta; \phi_i, \mathbf{x})$ and the continuity of $\rho(x) = -\log(f_\alpha(x; \gamma, 0))$ with respect to α . On the other hand, as $\alpha \rightarrow 0$, the cost function is highly nonconvex with multiple local minima located exactly at the points x_j , $j = 1, \dots, N$, that is, where $\phi_{i,j}(x_j - \theta) = 0$. If some $\phi_{i,j} = 0$, the corresponding term is ignored, since in this case $\rho(\phi_{i,j}(x_j - \theta)) = \rho(0)$ does not depend on θ . As a result, the overall minimum of the WSMF in (14), and subsequently the limit of $h_{\alpha,\gamma}(\phi_i, \mathbf{x})$ as $\alpha \rightarrow 0$, is attained among the points x_j with $\phi_{i,j} \neq 0$, namely, $\min_{j=1, \dots, N} \{G_{\alpha,\gamma}(x_j; \phi_i, \mathbf{x})\}$. This is exactly a selection filter. \blacksquare

Apart from the outlier rejection property of WSMF projections (see Property 1), the second desirable property is that the nonlinear sampling operator, $\Lambda_{\Phi}(\mathbf{x})$, should resemble the linear measurements as much as possible in the noiseless case. The following property states the asymptotic behavior of WSMF projections as $\gamma \rightarrow \infty$.

Property 3 (Asymptotic behavior of WSMF projections with respect to γ) Fix $\alpha \in (0, 2]$. In the limit as $\gamma \rightarrow \infty$, the WSMF projections reduce to a linear projection onto the elementwise square of ϕ_i , that is,

$$\lim_{\gamma \rightarrow \infty} h_{\alpha,\gamma}(\phi_i, \mathbf{x}) = c_i \cdot \frac{\sum_{j=1}^N \phi_{i,j}^2 x_j}{\sum_{j=1}^N \phi_{i,j}^2}, \quad (25)$$

where c_i is a scaling factor as in Definition 5.

Proof: For a fixed $\alpha \in (0, 2]$, the function $\rho(x) = -\log(f_\alpha(x; \gamma, 0))$ becomes convex on an increasingly larger interval as $\gamma \rightarrow \infty$ [26]. Furthermore, the nonlinear score function $g(x) = \rho'(x) = -\frac{f'_\alpha(x; \gamma, 0)}{f_\alpha(x; \gamma, 0)}$ is approximately linear on any bounded set $|x| \leq B$. Specifically, when $\alpha = 2$ (Gaussian case) $\rho(x) = \frac{x^2}{2}$ and $g(x) = x$, that is, the score function is linear and the WSMF reduces to a linear filter. On the other hand, for $\alpha < 2$ the function $\rho(x)$ becomes flared and wider, yet nonconvex (ref. Fig. 2), whilst $g(x)$ decreases to zero as $|x| \rightarrow \infty$ (ref. Fig. 4a). The practical implication of this behavior is that the WSMF suppresses the significance of extreme values. Given that $g(x)$ is smooth and infinitely differentiable, a linear approximation is obtained locally around $x = 0$ using Taylor's theorem, that is,

$$g(x) = g(0) + g'(0)x + \mathcal{O}(x^2), \quad (26)$$

where $\mathcal{O}(x^2)$ denotes the higher order terms. Focusing on expansions of the first derivative of a S α S density function

as $x \rightarrow 0$ it holds [33],

$$f'_\alpha(x; \gamma, 0) = \frac{1}{\gamma^2 \pi \alpha} \sum_{k=1}^{\infty} \frac{\Gamma(\frac{2k+1}{\alpha})}{(2k-1)!} (-1)^k \left(\frac{x}{\gamma}\right)^{2k-1}. \quad (27)$$

From (27) we get $f'_\alpha(0; \gamma, 0) = 0$ and subsequently $g(0) = -\frac{0}{f'_\alpha(0; \gamma, 0)} = 0$. Regarding the first derivative of the score function we have

$$g'(x) = -\frac{f''_\alpha(x; \gamma, 0) f'_\alpha(x; \gamma, 0) - (f'_\alpha(x; \gamma, 0))^2}{(f'_\alpha(x; \gamma, 0))^2}. \quad (28)$$

Given that $f'_\alpha(0; \gamma, 0) = 0$, we get $g'(0) = -\frac{f''_\alpha(0; \gamma, 0)}{f'_\alpha(0; \gamma, 0)}$. After some algebraic manipulation, this ratio equals $g'(0) = \frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}$. Then, scaling by $\gamma \rightarrow \infty$ expands the scale of the score function, so the linear approximation (26) holds over increasingly larger intervals (ref. Fig. 4b).

From the above, we deduce that the first derivative of the weighted S α S cost function is given by

$$\begin{aligned} G'_{\alpha, \gamma}(\theta; \phi_i, \mathbf{x}) &= -\sum_{j=1}^N \phi_{i,j} \rho'(\phi_{i,j}(x_j - \theta)) \\ &\approx -\sum_{j=1}^N \phi_{i,j} g'(0) (\phi_{i,j}(x_j - \theta)) \\ &= -g'(0) \left(\sum_{j=1}^N \phi_{i,j}^2 x_j - \theta \sum_{j=1}^N \phi_{i,j}^2 \right) \end{aligned} \quad (29)$$

Setting $G'_{\alpha, \gamma}(\theta; \phi_i, \mathbf{x}) = 0$ and solving with respect to θ gives

$$\hat{\theta}_{\alpha, \gamma}(\phi_i, \mathbf{x}) = \frac{\sum_{j=1}^N \phi_{i,j}^2 x_j}{\sum_{j=1}^N \phi_{i,j}^2}. \quad (30)$$

Combining (30) with (17) concludes the proof. \blacksquare

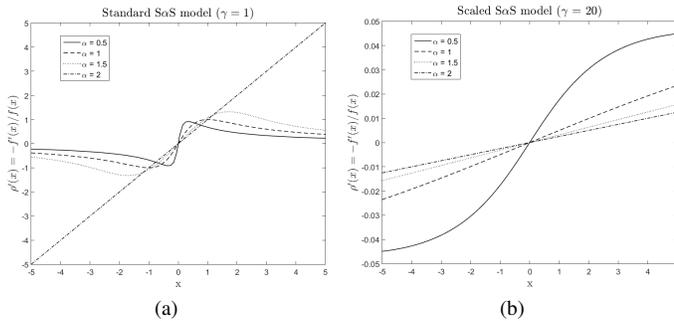


Fig. 4. Comparison of $\rho'(x) = -\frac{f'_\alpha(x; \gamma, 0)}{f'_\alpha(x; \gamma, 0)}$ function with $\alpha \in \{0.5, 1, 1.5, 2\}$ for (a) the standard S α S model ($\gamma = 1$) and (b) a scaled S α S model ($\gamma = 20$).

It is important to emphasize that calibrating the WSMF projections by appropriately setting the model parameters (α, γ) optimizes the shape of the weighted S α S cost function $G_{\alpha, \gamma}(\theta; \phi_i, \mathbf{x})$ for a given set of samples \mathbf{x} and measurement kernels ϕ_i ($i = 1, \dots, M$), and subsequently the performance of the WSMF projections. However, determining the optimal (α, γ) – optimal in the sense that the WSMF projections suppress the effects of large noise samples, whilst they resemble

linear measurements as much as possible in the noiseless case – from the corrupted signal $\mathbf{x} = \mathbf{x}_0 + \mathbf{e}_0$ is a nontrivial task.

To address this issue, we introduce an iterative algorithm, which successively filters the noisy signal \mathbf{x} and updates the S α S model parameters (α, γ) via ML estimation on the residual between the input noisy signal and its current filtered version. The filtering is performed using the weighted S α S matched filter of (14) in overlapping windows of length $l_{\text{win}} = L$ and step size $s_{\text{win}} = 1$, with weights $\mathbf{w} = \mathbf{1} \in \mathbb{R}^L$. The choice of an optimal filter order, L , is still an open question, but our experimental evaluation showed that a good trade-off between execution speed and estimation accuracy is obtained by setting $L = 0.05 \cdot N$. The end extremes of the input signal are treated using constant padding, where the first and the last sample are repeated at the beginning and at the end of the signal, respectively. The algorithm terminates when either a maximum number of iterations, $\text{maxIter}_{\text{filt}}$, is reached, or the relative change between consecutive S α S parameter estimates falls below a threshold tol_{filt} . In our implementation we set $\text{maxIter}_{\text{filt}} = 10$ and $\text{tol}_{\text{filt}} = 0.001$. The algorithm is summarized in Algorithm 1.

Algorithm 1 Calibration of S α S parameters (α, γ) for WSMF projections

Input: \mathbf{x} , L , $\mathbf{w} \in \mathbb{R}^L$, $\text{maxIter}_{\text{filt}}$, tol_{filt}

Initialize:

S α S parameters: $[\alpha^{(0)}, \gamma^{(0)}] = \text{mlfit}(\mathbf{x})$ (*)
 $\text{relChange} = 1000 \cdot \text{tol}_{\text{filt}}$, $t = 0$

- 1: **while** ($\text{relChange} > \text{tol}_{\text{filt}}$ or $t < \text{maxIter}_{\text{filt}}$) **do**
- 2: $\mathbf{P}_{\text{prev}} = [\alpha^{(t)}, \gamma^{(t)}]$
- 3: **Signal filtering:** $\mathbf{x}_f^{(t)} = \text{stablesigfilt}(\mathbf{x}, \mathbf{w}, L, \alpha^{(t)}, \gamma^{(t)})$ (**)
- 4: **Update residual:** $\mathbf{r}^{(t)} = \mathbf{x} - \mathbf{x}_f^{(t)}$
- 5: **Update S α S parameters:** $[\alpha^{(t+1)}, \gamma^{(t+1)}] = \text{mlfit}(\mathbf{r}^{(t)})$
- 6: $\mathbf{P}_{\text{new}} = [\alpha^{(t+1)}, \gamma^{(t+1)}]$
- 7: $\text{relChange} = \|\mathbf{P}_{\text{new}} - \mathbf{P}_{\text{prev}}\|_2 / \|\mathbf{P}_{\text{prev}}\|_2$
- 8: $t = t + 1$
- 9: **end while**

Output: The final S α S model parameters $(\alpha^{(t)}, \gamma^{(t)})$

(*) $\text{mlfit}(\mathbf{x})$ denotes a function that returns the ML estimates of the S α S model parameters for an input signal \mathbf{x} .

(**) $\text{stablesigfilt}(\mathbf{x}, \mathbf{w}, L, \alpha, \gamma)$ denotes a function that implements the WSMF of (14) in overlapping rolling windows of length L and step size equal to 1 using the weights \mathbf{w} .

An interesting observation is that when the original signal \mathbf{x}_0 is sparse by itself and sparse impulsive noise is added directly, then the signal and noise become indistinguishable, except if the noise samples have significantly larger magnitude. However, in practice noise is added in the observation domain, which rarely coincides with the sparsity inducing basis. Furthermore, WSMF projections are more expensive to be generated in terms of computational complexity, since each projection is produced by solving an optimization problem. Thus, WSMF projections should be considered for compressively sampling a signal when the sensing conditions are not ideal or when a robust sensing process is required.

IV. PERFORMANCE EVALUATION

This section evaluates the effectiveness of WSMF projections as a robust compressive sampling technique in impulsive environments. To this end, numerical experiments are performed with synthetic signals, along with a comparison against random linear projections and myriad measurements.

For the generation of synthetic signals, the following settings are employed unless stated otherwise: signal length $N = 1024$; cardinality of the sparse support $s = |\mathcal{S}| = \lceil 2\% \cdot N \rceil$; the nonzero coefficients are drawn from a Student's- t distribution with one degree of freedom and their positions are chosen uniformly at random from the index set $\{1, 2, \dots, N\}$; the DCT matrix is used as the sparsifying dictionary Ψ , that is, the linear measurement operator is given by $\mathbf{A} = \Phi\Psi^T$; the measurement matrix Φ has i.i.d. entries drawn from a standard Gaussian distribution and its columns are normalized to unit ℓ_2 norm; the number of random projections (measurements) is set to $M = \lceil 25\% \cdot N \rceil$ unless otherwise specified. Furthermore, the results of each experiment are averaged over 500 Monte Carlo repetitions with different realizations of the sparse signals, the random measurement matrices, and the additive observation noise component. The reconstruction quality is measured in terms of the signal-to-error ratio (SER) (in dB) defined by (10). In the subsequent experimental evaluation, the orthogonal matching pursuit (OMP) [28]⁴ is used with the prior assumption that the sparsity level s is known.

First, we illustrate experimentally the outlier rejection capability of WSMF projections (Property 1). For this purpose, two impulses are added to the original signal at randomly chosen positions. The amplitude of the first impulse is equal to 10^3 and of the second equals 10^2 . The true sparse signal is reconstructed using OMP for both linear and WSMF projections. The resulting SER is equal to -36.01 dB for the linear projections and 34.72 dB for the WSMF projections, where the filter parameters (α, γ) are estimated directly from the corrupted signal using Algorithm 1. Fig. 5 shows the true sparse signal, the associated linear and WSMF measurements, and the corresponding OMP reconstructions.

Next, we address the more challenging case, where the original signal is corrupted by additive S α S observation noise. Specifically, the noise characteristic exponent α_n takes values in $[0.8, 2]$, whilst its dispersion γ_n varies in $\{0.01, 0.05\}$. This gives a *geometric power* range between -21.25 dB and -12.38 dB. The geometric power (in dB) is a measure of strength for random variables with infinite variance, which is defined as

$$P_g(\alpha, \gamma) = 10 \log_{10} \left(\gamma \left(C_g^{\frac{1}{\alpha}-1} \right) \right), \quad (31)$$

where $C_g = e^{C_e}$ and $C_e = 0.5772\dots$ is the Euler constant.

Fig. 6 compares the reconstruction performance of OMP when using WSMF projections against linear and myriad [19] projections for S α S observation noise with model parameters as mentioned above. Fig. 6a corresponds to a low-dispersion (or, equivalently, low geometric power) noise, while Fig. 6b demonstrates the performance in higher-dispersion (or, equivalently, higher geometric power) noise conditions. In both

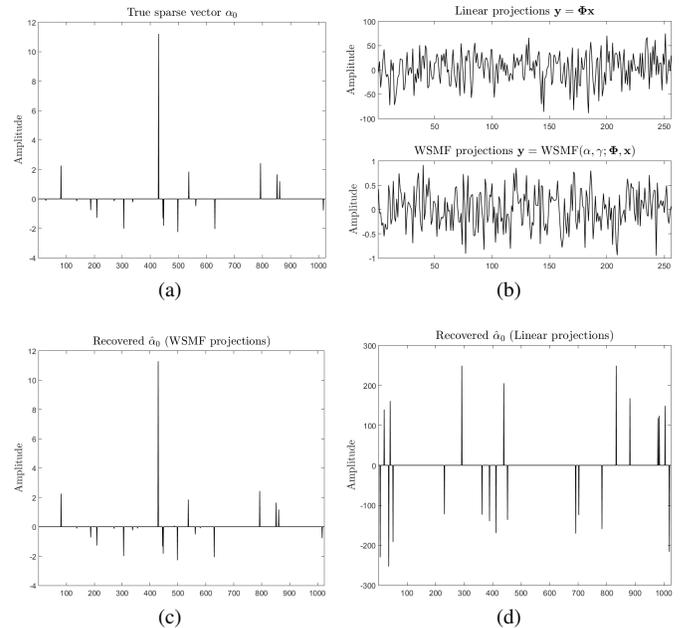


Fig. 5. Outlier rejection capability of WSMF projections. (a) True sparse signal; (b) linear and WSMF measurements for the DCT synthesized signal corrupted by two impulses; (c) OMP reconstructed signal from WSMF projections; (d) OMP reconstructed signal from linear projections.

cases, WSMF projections outperform significantly the myriad projections, especially as the noise characteristic exponent tends to 2, deviating significantly from the Cauchy distribution. This illustrates the increased robustness of WSMF measurements to a broader range of noise behavior, due to increased adaptability to the underlying statistics of the observation noise. Concerning the linear projections, in highly impulsive environments (small α_n values) they totally fail to give a reliable reconstruction of the true sparse signal, even for moderate noise dispersion values. On the other hand, they converge to, or even slightly exceed, the performance of WSMF projections when the noise statistics tends to the Gaussian ($\alpha_n \rightarrow 2$) and the noise dispersion is small. In the higher-dispersion case, OMP yields a deteriorated reconstruction accuracy for all the three types of projections as α_n decreases, nevertheless yielding an increased SER for the WSMF projections. This simulation also reveals the inefficiency of traditional CS reconstruction algorithms to achieve even fair approximations of the true sparse signals when operating in highly impulsive noisy conditions. It is exactly this shortcoming that motivated our study in the companion paper (Part II) [7] towards developing a novel CS reconstruction algorithm, which better adapts to the underlying heavy-tailed statistics of signals that are compressively sampled in impulsive environments.

Finally, we examine the performance of WSMF measurements as their number, M , varies from $2s$ (i.e., twice the cardinality of the sparse support) to $N/2$, for a varying noise impulsiveness, $\alpha_n \in \{1, 1.5, 1.9\}$, and a fixed $\gamma_n = 0.1$. The OMP-based reconstruction using linear projections contaminated by near-Gaussian noise ($\alpha_n = 1.9$) is used as a benchmark. Fig. 7 shows the SER (in dB) averaged over 500 Monte Carlo runs between the original noiseless signal and its

⁴MATLAB code available from <https://goo.gl/VHvyJe>.

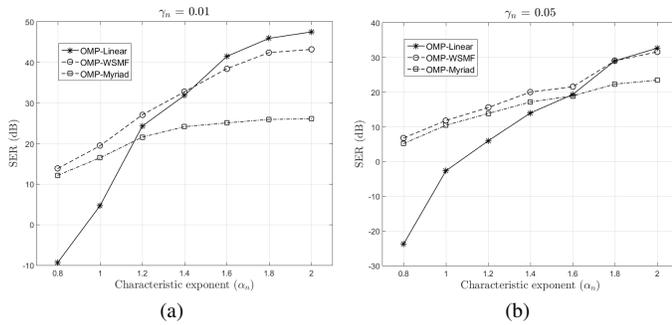


Fig. 6. Comparison of WSMF projections with linear and myriad projections for $S\alpha S$ observation noise with $\alpha_n \in [0.8, 2]$ and $\gamma_n \in \{0.01, 0.05\}$. Average SER is shown for OMP-based reconstruction over 500 Monte Carlo runs.

OMP-based reconstructions, as a function of M , using WSMF, linear, and myriad measurements. First, we observe again the significantly improved reconstruction performance when using WSMF measurements against their myriad counterpart, which becomes more prominent as the noise statistics tends to a Gaussian. Moreover, although WSMF projections introduce nonlinear distortions to the generated measurements, however, their performance coincides with the benchmark in the near-Gaussian case. This is in contrast to the myriad projections, which yield a significantly degraded performance. Another interesting remark is that WSMF projections are more efficient than myriad projections even for very small values of M . This reveals that, although the generation of WSMF measurements relies on the implicit estimation of (α_n, γ_n) , which may not be accurate for very small sample sizes, however, they still outperform the myriad case where α_n is inherently fixed to 1. Finally, the results show that, as expected, more measurements are required to improve the reconstruction quality and compensate for the increased impulsiveness when α_n decreases, for all the three types of projections.

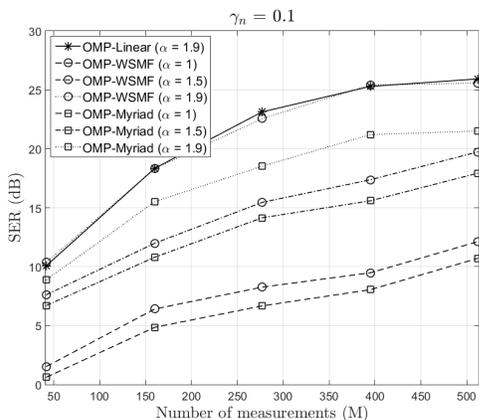


Fig. 7. Comparison of WSMF projections with linear and myriad projections for $S\alpha S$ observation noise with $\alpha_n \in \{1, 1.5, 1.9\}$ and $\gamma_n = 0.1$. Average SER is shown as a function of M for OMP-based reconstruction over 500 Monte Carlo runs.

V. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, a novel method was proposed for robust nonlinear compressive sampling of signals corrupted by heavy-tailed, infinite variance observation noise. Specifically, the noise statistics was modeled via $S\alpha S$ distributions. Subsequently, the weighted $S\alpha S$ matched filter (WSMF) projections were proposed as efficient nonlinear sampling operators, which suppress the effects of impulsive observation noise. The outlier rejection and asymptotic behavior properties of WSMF projections were proven theoretically, and guidelines were given for tuning appropriately their parameters to achieve improved adaptation to the underlying heavy-tailed statistics of the noise. Moreover, it was shown that WSMF measurements can be utilized with traditional sparse reconstruction methods outperforming linear projections, as well as the previously introduced myriad projections that are derived from Cauchy statistics.

However, a theoretical framework for selecting the optimal values of the key parameters for the WSMF projections is still an open task. Furthermore, we are interested in extending the compressive sampling method proposed herein in the case of observation noise modeled by a general (skewed) alpha-stable distribution. We expect that the incorporation of an additional free parameter (i.e., the skewness) will improve the adaptability of the nonlinear sampling operator to the underlying heavy-tailed statistics of infinite variance noise, thus further increasing its robustness against gross outliers.

Our proposed CS framework for signals corrupted by impulsive noise is completed in the companion paper (Part II) [7], which addresses the case of heavy-tailed sampling noise. To this end, a novel iterative hard thresholding method is designed based on a minimum dispersion optimization criterion, which emerges naturally in the case of additive impulsive sampling noise modeled by $S\alpha S$ distributions. We emphasize that the methodologies presented herein (for compressive sampling) and the companion paper (for sparse signal reconstruction) can be used independently of each other, as well as together to form an integrated CS system.

REFERENCES

- [1] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*. 3rd Ed., Academic Press, 2009.
- [2] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [3] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [4] E. Candès, "The restricted isometry property and its implications for compressed sensing," *C. R. Acad. Sci. Paris, Ser. I*, vol. 346, pp. 589–592, 2008.
- [5] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Uniform uncertainty principle for Bernoulli and Subgaussian ensembles," *Constructive Approximation*, vol. 28, no. 3, pp. 277–289, Dec. 2008.
- [6] T. T. Do, L. Gan, N. H. Nguyen, and T. D. Tran, "Fast and efficient compressive sensing using structurally random matrices," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 139–154, Jan. 2012.
- [7] G. Tzagkarakis, J. P. Nolan, and P. Tsakalides, "Compressive sensing using symmetric alpha-stable distributions – Part II: Robust sparse reconstruction," *submitted in IEEE Trans. Signal Process.*, Nov. 2017.
- [8] S. Banerjee and M. Agrawal, "Underwater acoustic communication in the presence of heavy-tailed impulsive noise with bi-parameter Cauchy-Gaussian mixture model," in *Proc. Int. Symp. on Ocean Electron. (SYM-POL)*, Kochi, India, 23–25 Oct. 2013.

- [9] P. Tsakalides, F. Trinci, and C. Nikias, "Radar CFAR thresholding in heavy-tailed clutter and positive alpha-stable measurements," in *Proc. 32nd Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, 1–4 Nov. 1998.
- [10] V. Aalo *et al.*, "Performance of CA-CFAR receivers in alpha-stable clutter," in *Proc. IEEE Int. Symp. on Signal Process. Inf. Technol. (ISSPIT)*, Athens, Greece, 12–15 Dec. 2013.
- [11] A. Achim, A. Bezerianos, and P. Tsakalides, "Novel Bayesian multiscale method for speckle removal in medical ultrasound images," *IEEE Trans. Med. Imag.*, vol. 20, no. 8, pp. 772–783, Aug. 2001.
- [12] P. Glasserman, P. Heidelberger, and P. Shahabuddin, "Portfolio value-at-risk with heavy-tailed risk factors," *Math. Finance*, vol. 12, no. 3, pp. 239–269, July 2002.
- [13] M. Masood and T. Al-Naffouri, "Sparse reconstruction using distribution agnostic Bayesian matching pursuit," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5298–5309, Nov. 2013.
- [14] P. Huber and E. Ronchetti, *Robust Statistics*. 2nd Ed., Wiley, 2009.
- [15] G. Arce, *Nonlinear Signal Processing: A Statistical Approach*. Wiley, 2005.
- [16] R. Carrillo *et al.*, "Robust compressive sensing of sparse signals: a review," *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 1, pp. 108, Oct. 2016.
- [17] A. Ramirez, G. Arce, and B. Sadler, "Fast algorithms for reconstruction of sparse signals from Cauchy random projections," in *Proc. 18th Eur. Signal Process. Conf. (EUSIPCO)*, Aalborg, Denmark, 23–27 Aug. 2010.
- [18] A. Ramirez *et al.*, "Reconstruction of sparse signals from ℓ_1 dimensionality-reduced Cauchy random projections," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5725–5737, Nov. 2012.
- [19] R. Carrillo, K. Barner, and T. Aysal, "Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 392–408, Apr. 2010.
- [20] S. Babacan, R. Molina, and A. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53–63, Jan. 2010.
- [21] G. Samorodnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall, New York, 1994.
- [22] C. Tzagkarakis, A. Mouchtaris, and P. Tsakalides, "Musical genre classification via generalized Gaussian and alpha-stable modeling," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Toulouse, France, 14–19 May, 2006.
- [23] H. E. Ghannudi *et al.*, "Alpha-stable interference modeling and Cauchy receiver for an IR-UWB ad hoc network," *IEEE Trans. Commun.*, vol. 58, no. 6, pp. 1748–1757, June 2010.
- [24] S. Liao and A. C. Chung, "Feature based nonrigid brain MR image registration with symmetric alpha stable filters," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 106–119, Jan. 2010.
- [25] C. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*. Wiley, 1995.
- [26] J. P. Nolan, J. Gonzalez, and R. Núñez, "Stable filters: A robust signal processing framework for heavy-tailed noise," in *Proc. IEEE Int. Radar Conf.*, Arlington, VA, 10–14 May, 2010.
- [27] E. Candès, Y. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmonic Anal.*, vol. 31, no. 1, pp. 59–73, July 2011.
- [28] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [29] J. P. Nolan, "Numerical calculation of stable densities and distribution functions," *Commun. Statist.-Stochastic Models*, vol. 13, no. 4, pp. 759–774, 1997.
- [30] J. P. Nolan, "Maximum likelihood estimation and diagnostics for stable distributions," *Lévy Processes: Theory and Applications*, O. E. Barndorff-Nielsen, T. Mikosch, and S. I. Resnick (Eds.), Birkhäuser, Boston, pp. 379–400, 2001.
- [31] M. Davenport *et al.*, "The pros and cons of compressive sensing for wideband signal acquisition: Noise folding versus dynamic range," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4628–4642, Sept. 2012.
- [32] J. P. Nolan, "Advances in nonlinear signal processing for heavy tailed noise," in *Proc. Int. Workshop on Appl. Prob.*, Compiègne, France, July 2008.
- [33] M. Matsui and A. Takemura, "Some improvements in numerical evaluation of symmetric stable density and its derivatives," *Commun. in Statist.*, vol. 35, no. 1, pp. 149–172, 2006.