Signal Processing 182 (2021) 107944

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

# Robust nonlinear compressive sampling using symmetric alpha-stable distributions

George Tzagkarakis[a,*], John P. Nolan[b], Panagiotis Tsakalides[a,c]

[a] Institute of Computer Science, Foundation for Research and Technology Hellas, Greece
[b] Math/Stat Department, American University, Washington, DC, USA
[c] Department of Computer Science, University of Crete, Heraklion, Greece

ARTICLE INFO

*Article history:*
Received 15 March 2020
Revised 29 October 2020
Accepted 13 December 2020
Available online 16 December 2020

*Keywords:*
Nonlinear compressive sampling
Alpha-stable distributions
Heavy-tailed statistics
Stable matched filter

ABSTRACT

Conventional compressive sampling (CS) primarily assumes light-tailed models for the underlying signal and/or noise statistics. Nevertheless, this assumption is abolished when operating in impulsive environments, where non-Gaussian infinite-variance processes arise for the signal and/or noise components. This drives traditional linear sampling operators to failure, since the gross observation errors are spread uniformly over the generated compressed measurements, whilst masking the critical information content of the observed signal. To address this problem, this paper exploits the power of symmetric alpha-stable ($S\alpha S$) distributions to design a robust nonlinear compressive sampling operator capable of suppressing the effects of infinite-variance additive observation noise. Specifically, a generalized alpha-stable matched filter is introduced for generating compressed measurements in a nonlinear fashion, which achieves increased robustness to impulsive observation noise, thus subsequently improving the accuracy of traditional sparse reconstruction algorithms. This filter emerges naturally in the case of additive observation noise modeled by $S\alpha S$ distributions, as an effective mechanism for downweighting gross outliers in the noisy signal. The theoretical justification along with the experimental evaluation demonstrate the improved performance of our nonlinear CS framework when compared against state-of-the-art CS techniques for a broad range of impulsive environments.



## 1. Introduction

Compressive sampling (CS) is established as a powerful signal acquisition and reconstruction strategy, which enables a potentially large reduction in the sampling and computation costs for capturing signals at sampling rates far below what is dictated by the traditional Nyquist-Shannon's sampling theorem. This reduction is feasible for signals that admit *sparse* or *compressible* representations in appropriate transform domains, or over suitable sparsifying bases.

Conventional CS acquisition is realized by taking linear projections of the observed signal onto a small set of vectors that are incoherent with the sparsity-inducing basis[1] [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 conditions (e.g. the null space property and the restricted isometry property for $\ell_1$-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]. Nevertheless, practical applications 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 [6–8].

In real acquisition systems, the observed signal is often corrupted by additive *observation* noise, which is spread uniformly over the compressed measurements due to the linear sampling operation. Subsequently, this impacts dramatically the accuracy of sparse reconstruction algorithms. The majority of previous CS-based methods is primarily based on light-tailed, finite-variance

* Corresponding author at: FORTH-ICS, N. Plastira 100, Vassilika Vouton, GR-70013 Heraklion, Crete, Greece.
*E-mail addresses:* gtzag@ics.forth.gr (G. Tzagkarakis), jpnolan@american.edu (J.P. Nolan), tsakalid@ics.forth.gr (P. Tsakalides).

[1] Although incoherence among the sampling vectors and the sparsifying basis (a.k.a dictionary) is a requirement for guaranteeing accurate sparse reconstruction, recent works have proven that for truly redundant dictionaries a no-incoherence restriction on the dictionary can still guarantee accurate sparse recovery [1].

https://doi.org/10.1016/j.sigpro.2020.107944
0165-1684/

assumptions for the statistics of the signal and/or noise generating processes. Despite the analytic tractability and practical appeal, these assumptions may yield a dramatic degradation of the reconstruction quality when we operate in highly impulsive environments, which give rise to *heavy-tailed* processes with infinite variance. To alleviate the effects of gross errors that mask the information conveyed by the compressed measurements, recent state-of-the-art methods rely on robust statistics [9] and algebraic-tailed models, such as the Cauchy and generalized Cauchy (GCD) distributions [10], to design nonlinear sampling operators.

Notice that, from an engineering viewpoint, processes with infinite variance may sound counter-intuitive, since they give raise to infinite power that does not really exist in real world. Nevertheless, from a probabilistic perspective, variance is just a measure of how spread out a distribution is. Distributions with infinite variance present fat upper tails that decrease at an extremely slow rate. Intuitively, this means that the distribution will take a very long time to vanish. Actually, in theory, it never vanishes, and this is precisely the reason we say that the upper tail is "unbounded". The slow decay of probability in this area increases the odds of extreme values (outliers), "bumps" in the distribution, and other surprising last-minute events at some point in the future. Although the model has infinite variance, this does not imply that the real-world phenomenon being modeled also extends to infinity. It just means that the model is a "good enough" fit to describe the behavior of the phenomenon under study. On the other hand, notice that with such reasoning (that infinite variance is inappropriate in a practical signal processing context), the Gaussian distribution would not be a physically appropriate assumption because it is unbounded. Gaussian models are adequate for modeling many types of observed data. In some cases, especially when outliers or heavy tails appear in the observed data, an infinite-variance stable model may have to be used. Next section gives examples in several real application domains where infinite-variance models provide an excellent fit of the associated recorded signals.

On the other hand, *alpha-stable* distributions [11] have been proven very powerful in accurately modeling impulsive phenomena. However, their intractability due to the lack of closed-form expressions for the density functions of all except for a few stable distributions (Gaussian, Cauchy and Lévy) has prevented their exploitation in the framework of CS. To address this problem, whilst also revealing the advantages of alpha-stable models in designing efficient compressive sampling operators, this paper proposes the design of a robust nonlinear sampling operator under heavy-tailed observation noise, by modeling the noise statistics via *symmetric alpha-stable* (S$\alpha$S) distributions. We emphasize that the proposed method aims at generating a set of nonlinear compressed measurements, with the goal of improving the accuracy of conventional sparse reconstruction algorithms. To the best of our knowledge, this is the first thorough study that bridges the fields of nonlinear compressive sampling and S$\alpha$S models for the design of a CS acquisition framework with increased robustness to heavy-tailed infinite-variance observation noise.

### 1.1. Motivation

In practical CS acquisition systems, either the observed signal or the generated random measurements are often corrupted by noise. In a recent study [12], we addressed the latter problem of suppressing the effects of additive impulsive sampling noise. In this work, we focus on the former case, that is, on diminishing the effects of additive *observation* noise. There is a remarkably wide range of applications, where information is recorded in highly impulsive environments, giving rise to non-Gaussian, heavy-tailed processes for the representation of the observed signal and/or corrupting noise, including i) underwater acoustics [13,14], ii) sonar and radar systems [15,16], iii) medical imaging [17,18], iv) finance [19,20], and v) local damage detection in vibrating systems [21], just to name a few. The presence of large-amplitude noise in the signal domain masks the original information content of the signal, thus degrading dramatically the quality of the generated linear random measurements. The uniform spread of large noise samples, which can be of infinite or very large variance, over the linear random measurements, causes traditional CS reconstruction algorithms to fail in recovering a close approximation of the true sparse signal.

The presence of impulsive observation noise has been previously tackled either in the framework of robust statistics [9] and nonlinear signal processing [10], or in a completely distribution-agnostic framework [22]. In contrast to the former methods, which are more generic, the main limitation of the distribution-agnostic techniques is that they rely on an assumption of bounded-variance noise, whilst the sparse signal is estimated based on a minimum mean squared error criterion. However, this assumption is not valid for environments characterized by heavy-tailed distributions, whose second-order moments are infinite or even undefined.

Recently, the problem of suppressing the effects of impulsive observation noise during compressive sampling has been addressed efficiently by employing heavy-tailed distributions coupled with maximum likelihood (ML) type estimators (a.k.a M-estimators) as data fidelity functions [23]. In [24,25], the reduced set of compressed measurements is generated using Cauchy random projections, which are not severely degraded by gross observation errors. Furthermore, such measurements are better suited in applications where $\ell_1$-norm preservation is preferred in the low-dimensional measurements space. A similar approach is proposed in [26], where the generalized Cauchy distributions (GCD) family is employed to design a robust nonlinear sampling operator based on the weighted myriad estimator [10]. The generated compressed 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 is mainly motivated by their closed-form expressions, and subsequently their computational tractability in modeling impulsive environments.

Despite the increased robustness of the above methods in compressively sampling signals corrupted by gross additive errors, the specific use of Cauchy [24], GCD [26] or Laplace distributions [27] can be restrictive in capturing more generic non-Gaussian heavy-tailed behaviors of the observation noise. Instead, several studies have demonstrated the power of *symmetric alpha-stable* (S$\alpha$S) distributions in accurately modeling impulsive signal and noise processes [28–31]. Motivated by the limitations of previous CS methods, first we model the statistics of impulsive observation noise, with possibly infinite variance, by members of the S$\alpha$S family [11]. Then, we propose a novel nonlinear compressive sampling operator based on alpha-stable filtering of the observed signal. Specifically, the set of random measurements is generated by applying a weighted stable matched filter on the signal, with appropriate measurement kernels as the weight vectors. As it will become clear in the subsequent analysis, the filter's output is interpreted naturally as the estimated location of random variables modeled by S$\alpha$S distributions, as is the case with the infinite variance observation noise adopted herein. Nevertheless, we emphasize that the subsequent analysis focuses on the compressive sampling part, while considering conventional algorithms for the CS reconstruction of sparse signals.

### 1.2. Main contributions

The major contribution of this paper is twofold: i) we propose a novel robust nonlinear sampling operator based on the

*weighted SαS matched filter* (WSMF). The generated WSMF-based compressed measurements exploit the rich class of SαS distributions. This yields an increased robustness for a broader range of impulsive observation noise behaviors, when compared with the previous methods based on Cauchy and GCD distributions; ii) we provide an explicit theoretical justification of the improved outlier rejection capability of WSMF-based random measurements, as well as of their asymptotic behavior with respect to the key parameters that control the performance of the filter.

Furthermore, the use of WSMF provides additional degrees of freedom, namely, the parameters of a SαS distribution, yielding increased robustness against gross observation errors. The proposed compressive sampling framework resembles a non-universal approach in the sense that, given an ensemble of random measurement kernels and a sparse signal $\mathbf{x}$, the compressed measurements generated by applying our nonlinear sampling operator will be robust to heavy-tailed observation noise for *this specific* $\mathbf{x}$.

### 1.3. Paper organization

The rest of the paper is organized as follows: Section 2 overviews the signal model adopted in this study, along with the main concepts of the SαS family as a proper model for heavy-tailed infinite-variance observation noise. Section 3 analyzes the design and implementation of our proposed nonlinear compressive sampling operator, whilst also providing theoretical proofs for its outlier rejection performance and asymptotic behavior. An experimental evaluation of the robustness of our sampling operator is presented in Section 4 for a variety of impulsive environments, compared against state-of-the-art sampling operators tailored to impulsive observation noise. Finally, Section 5 summarizes the main outcomes and gives directions for future work.

### 1.4. Notation

In the following, scalars are denoted by lower-case letters (e.g. $x$), column vectors by lower-case boldface letters (e.g. $\mathbf{x}$), and matrices by upper-case boldface letters (e.g. $\mathbf{X}$). The $i$th column of a matrix $\mathbf{X}$ is denoted by $\mathbf{x}_i$, whereas $x_j$ indicates the $j$th element of a vector $\mathbf{x}$. Finally, we use $\hat{\mathbf{x}}$, $\mathbf{x}^\mathsf{T}$, $\mathbf{x}^*$, and $\mathbf{x}^{(t)}$ to 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.

## 2. Linear sampling and SαS distributions

Let $\mathbf{x} = [x_1, x_2, \ldots, x_N]^\mathsf{T} \in \mathbb{R}^N$ be a real discrete-time signal. In the general case, we assume that $\mathbf{x}$ can be sparsified over a, possibly overcomplete, transform basis $\boldsymbol{\Psi} \in \mathbb{R}^{N' \times N}$ with $N' \geq N$, such that $\boldsymbol{\alpha} = \boldsymbol{\Psi}\mathbf{x} \in \mathbb{R}^{N'}$ is an $s$-sparse vector of transform coefficients. $\boldsymbol{\Psi}$ and $\boldsymbol{\Psi}^\mathsf{T}$ denote the analysis (direct) and synthesis (inverse) transforms, respectively. In practice, the observed signal is typically corrupted by additive noise, which is defined as a perturbation introduced to the true signal prior to its sampling,

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{e}_\mathrm{o} \tag{1}$$

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

Let $\mathrm{A} : \mathbb{R}^N \longmapsto \mathbb{R}^M$ with $M < N$ denote a sampling operator that maps a vector $\mathbf{x} \in \mathbb{R}^N$ to a lower-dimensional vector $\mathbf{y} \in \mathbb{R}^M$ of *measurements*. This compressive sampling process is expressed by $\mathbf{y} = \mathrm{A}(\mathbf{x})$. In conventional CS systems, the sampling operator A is a linear map. More specifically, given a random measurement matrix $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$ ($M < N$), which satisfies all the necessary and sufficient conditions for accurate sparse reconstruction, the traditional linear sampling model is as follows,

$$\mathbf{y} = \boldsymbol{\Phi}\left(\boldsymbol{\Psi}^\mathsf{T}\boldsymbol{\alpha}_0 + \mathbf{e}_\mathrm{o}\right) = \boldsymbol{\Phi}\boldsymbol{\Psi}^\mathsf{T}\boldsymbol{\alpha}_0 + \mathbf{n} \tag{2}$$

where $\boldsymbol{\alpha}_0 \in \mathbb{R}^{N'}$ is the true sparse signal to be reconstructed, and $\mathbf{n} = \boldsymbol{\Phi}\mathbf{e}_\mathrm{o} \in \mathbb{R}^M$ is the projected observation noise.

By setting $\mathbf{A} = \boldsymbol{\Phi}\boldsymbol{\Psi}^\mathsf{T}$ as the generic linear sampling operator, the sparse signal is recovered by solving, among others, an $\ell_1 - \ell_2$ constrained optimization problem,

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^{N'}} \|\boldsymbol{\alpha}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\boldsymbol{\alpha}\|_2 \leq \varepsilon \tag{3}$$

where $\varepsilon > 0$ is a threshold depending on the noise level. Having obtained an estimate of the optimal sparse coefficients vector $\boldsymbol{\alpha}_0^*$, the true signal is given by $\hat{\mathbf{x}}_0 = \boldsymbol{\Psi}^\mathsf{T}\boldsymbol{\alpha}_0^*$. Without loss of generality, in the following, we consider the basis pursuit denoising (BPD) formulation in (3), which is solved effectively using, for instance, the orthogonal matching pursuit (OMP) algorithm [32].

At the core of our proposed nonlinear compressive sampling operator is the use of SαS distributions for modeling the statistics of impulsive observation noise. In particular, we assume that the random noise $\mathbf{e}_\mathrm{o}$ consists of independent and identically distributed (i.i.d.) components, $e_{\mathrm{o},i} \in \mathbb{R}$, $i = 1, \ldots, N$, that follow a univariate SαS distribution. The probability density function of a general univariate SαS distribution is as follows [33],

$$f_\alpha(x; \gamma, \delta) = \frac{1}{\gamma}\, q\left(\frac{x - \delta}{\gamma}; \alpha\right) \tag{4}$$

where

$$q(x; \alpha) = \frac{1}{\pi}\int_0^\infty \cos(xt)e^{-t^\alpha}\, \mathrm{d}t. \tag{5}$$

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 $\alpha$, the heavier the tails of the SαS density function. The dispersion parameter determines the spread of the distribution around its location, as the standard deviation does for a Gaussian distribution. Hereafter, the notation $X \sim f_\alpha(\gamma, \delta)$ denotes that a random variable $X$ follows a SαS distribution with parameters $\alpha$, $\gamma$, $\delta$. In the special case where $\gamma = 1$ and $\delta = 0$ the distribution is called *standard* SαS. Furthermore, it holds that, if $X \sim f_\alpha(\gamma_X, 0)$ and $Y \sim f_\alpha(\gamma_Y, 0)$ are two independent jointly SαS (i.e., with equal characteristic exponents) random variables, then,

$$cX + \delta \sim f_\alpha(|c|\gamma_X, \delta)\ (c \neq 0), \tag{6}$$

$$X + Y \sim f_\alpha\left((\gamma_X^\alpha + \gamma_Y^\alpha)^{1/\alpha}, 0\right). \tag{7}$$

In the following, we assume that the noise statistics is modeled by a SαS distribution located at $\delta = 0$. Nevertheless, the distribution's location can be always shifted to zero via (6) by subtracting $\delta$. In our implementation, the SαS model parameters $(\alpha, \gamma)$ are estimated from a discrete set $\{x_1, \ldots, x_N\}$ of $X$'s realizations using the consistent maximum likelihood (ML) method described in [34], which gives reliable estimates and provides the tightest possible confidence intervals.

## 3. Robust nonlinear compressive sampling operator

In this section, we propose a novel nonlinear compressive sampling operator, which suppresses efficiently the effects of heavy-tailed observation noise of infinite variance, whilst achieving increased robustness to a broader range of impulsive noise behaviors, from near linear (i.e., $\alpha \to 2$) to extremely impulsive observation noise (i.e., $\alpha \to 0$). To this end, we analyze the design and implementation of our proposed sampling operator, which exploits a

weighted stable matched filter as a robust outlier rejection mechanism.

### 3.1. Stable matched filter

Under the adverse conditions arising in impulsive environments, the performance of linear adaptive filters, in terms of suppressing the effects of heavy-tailed observation noise, may deteriorate dramatically. To address this problem, the use of generalized ML estimators, the so-called M-estimators, has demonstrated increased robustness to impulsive noise. Motivated by the success of M-estimators, our proposed compressive sampling operator is developed in the framework of robust statistics.

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

$$G_{\alpha,\gamma}(\theta; \mathbf{w}, \mathbf{x}) = \sum_{j=1}^{N} \rho(w_j(x_j - \theta)) = -\sum_{j=1}^{N} \log\left(f_\alpha(w_j(x_j - \theta); \gamma, 0)\right) \tag{8}$$

where the weight vector $\mathbf{w} = [w_1, w_2, \ldots, w_N]^{\mathrm{T}} \in \mathbb{R}^N$ reflects the different levels of reliability of the observed samples. Then, the *weighted S$\alpha$S matched filter* (WSMF) is defined as follows:

**Definition 1** (**Weighted S$\alpha$S matched filter**). Given the observed samples $\mathbf{x} = [x_1, x_2, \ldots, x_N]^{\mathrm{T}}$, a weight vector $\mathbf{w} = [w_1, w_2, \ldots, w_N]^{\mathrm{T}}$, and the cost function $G_{\alpha,\gamma}(\theta; \mathbf{w}, \mathbf{x})$, the weighted S$\alpha$S matched filter is defined by

$$\mathrm{WSMF}(\alpha, \gamma; \mathbf{w}, \mathbf{x}) \doteq \hat{\theta}_{\alpha,\gamma}(\mathbf{w}, \mathbf{x}) = \arg\min_{\theta} G_{\alpha,\gamma}(\theta; \mathbf{w}, \mathbf{x}). \tag{9}$$

The WSMF output is the value of $\theta$ at the global minimum of the weighted cost function $G_{\alpha,\gamma}(\theta; \mathbf{w}, \mathbf{x})$. For its computation, reliable and accurate algorithms are required for solving (9) due to the nonconvex behavior of the cost function. A computationally tractable approach for the numerical calculation of the WSMF output is described in [35,36]. This approach relies on a global minimization scheme based on a branch-and-bound method, which does not rely on any starting point, to guarantee that the global minimum is found. Instead, it starts with an interval which is subdivided, ruling out pieces until a global minimum is found. The branch-and-bound method uses an interval determined by the data values (e.g. in the unweighted case it uses the extremes of the input data). Besides, this method always finds a global minimum, it just may not be unique. 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 1, we accept any solution to (9) 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 described in the above papers guarantees convergence to the global minimum.

Furthermore, since $\rho(x)$ is the negative of a 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$\alpha$S case ($0 < \alpha < 2$) the stable matched filter is nonlinear with no closed-form solution, and the minimum in (9) must be found numerically. We note that all the subsequent numerical calculations and optimizations involving S$\alpha$S densities are performed using the STABLE toolbox.[2]

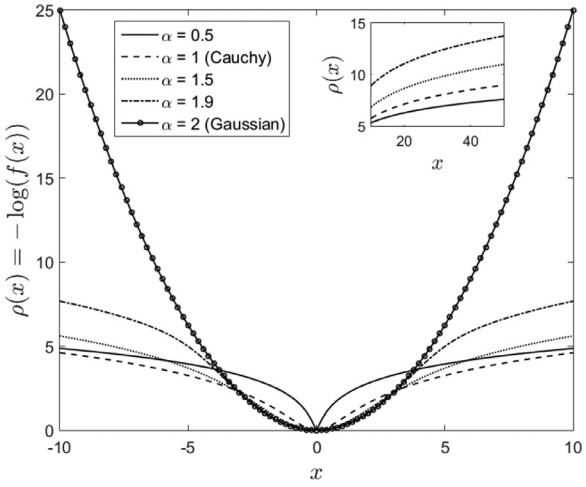


**Fig. 1.** Comparison of the $\rho(x) = -\log\left(f_\alpha(x; \gamma, 0)\right)$ function for the standard S$\alpha$S model ($\gamma = 1$) and the standard Gaussian case $\rho(x) = x^2$.

It must be noted that the weighted myriad filter (WMyF), which has been exploited by [24,26] for the design of sampling operators robust to gross errors, constitutes an approximation of the WSMF tailored to Cauchy statistics. As such, we expect that our WSMF-based nonlinear sampling operator will achieve a superior performance, in terms of increased robustness to heavy-tailed observation noise, when compared against its myriad-based counterpart. For completeness of presentation we cite the definition of the WMyF [26]:

**Definition 2** (**Weighted myriad filter**). Given the observed samples $\mathbf{x} = [x_1, x_2, \ldots, x_N]^{\mathrm{T}}$ and a weight vector $\mathbf{w} = [w_1, w_2, \ldots, w_N]^{\mathrm{T}}$, the weighted myriad filter is defined by

$$\mathrm{WMyF}(k; \mathbf{w}, \mathbf{x}) \doteq \hat{\theta}_k(\mathbf{w}, \mathbf{x}) = \arg\min_{\theta} \sum_{j=1}^{N} \log\left(k^2 + |w_j|\left(\mathrm{sign}(w_j)x_j - \theta\right)^2\right) \tag{10}$$

where $k$ is the scale parameter of the (generalized) Cauchy distribution. Notice that, although (10) 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. 1 compares the behavior of $\rho(x)$, which is the main component of the cost functions defined above, for the standard S$\alpha$S model, $\rho(x) = -\log\left(f_\alpha(x; 1, 0)\right)$ with $\alpha \in \{0.5, 1, 1.5, 1.9\}$, and the standard Gaussian distribution, $\rho(x) = x^2$. In the standard S$\alpha$S case, or equivalently, in the $\ell_p$ 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 $\ell_2$ 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 gross outliers. Furthermore, the zoomed inner plot illustrates that, as the characteristic exponent decreases, $\rho(x)$ is more robust to outliers, since it increases much more slowly when $|x| \to \infty$. This behavior also demonstrates the superiority of the S$\alpha$S matched filter against the previous approaches based on the Cauchy distribution, in terms of better controlling the outlier re-

[2] Robust Analysis Inc., STABLE toolbox version 5.3 (http://www.robustanalysis.com).

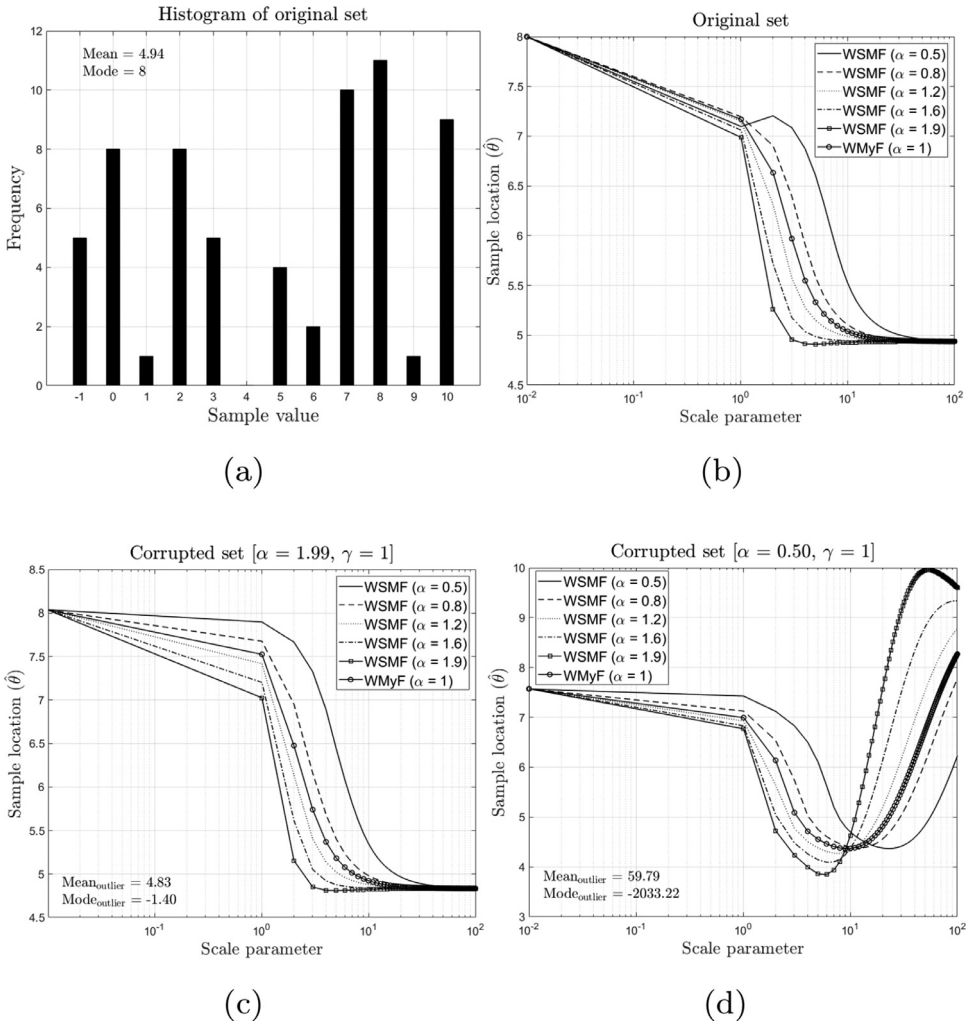


**Fig. 2.** Behavior of WSMF and WMyF filters as a function of the scale parameter and the characteristic exponent (for the WSMF) (weight vector is fixed to $\mathbf{w} = \mathbf{1}$). (a) Histogram of original sample set; (b) output of filters applied on original set; (c) output of filters applied on original set corrupted by noise drawn from a S$\alpha$S model with ($\alpha = 1.99$, $\gamma = 1$); (d) output of filters applied on original set corrupted by noise drawn from a S$\alpha$S model with ($\alpha = 0.5$, $\gamma = 1$).

jection intensity through an additional degree of freedom (i.e., the parameter $\alpha$) instead of relying only on the scale parameter.

### 3.2. Controlling the behavior of S$\alpha$S matched filters

The S$\alpha$S matched filter enables a rich class of operating modes, which can be controlled by tuning the model parameters $(\alpha, \gamma)$. 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. 2 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 $(\alpha, \gamma)$. To this end, we illustrate the behavior of the WSMF output defined in (9) by varying the model parameters $(\alpha, \gamma)$, and compare against the performance of the WMyF defined in (10), which depends on a single parameter $k$. For convenience, we refer to both $\gamma$ and $k$ as the scale parameter of the corresponding filter. Fig. 2 a shows the histogram of a sample set, which is generated by drawing non-uniformly $N = 64$ 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 a varying $\alpha \in \{0.5, 0.8, 1.2, 1.6, 1.9\}$ for WSMF and $\alpha = 1$ for WMyF, and scale parameter in $[10^{-2}, 10^{2}]$, while setting all the weights $w_j$, $j = 1, \ldots, N$, equal to 1. Fig. 2 b shows the value of each filter as a function of the scale parameter, for the various WSMFs and the WMyF. We observe that, 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., $\alpha \to 2$), 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. 2 c-2 d. 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 achieve the same robustness against the near-Gaussian additive noise, with their behavior resembling the noiseless case (Fig. 2 b). As before, their difference lies in the different regimes of the associated scale parameter, where the mean-mode behavior switches. Specifically, we observe that as the characteristic exponent of the WSMF decreases, the mode of the data

set is better approximated for a wider range of the scale parameter. On the contrary, when the corrupting noise is drawn from a highly impulsive S$\alpha$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 $\alpha$. However, in contrast to the WMyF, whose characteristic exponent is fixed to $\alpha = 1$, the WSMF achieves an increased robustness via the double tuning of $\alpha$ and $\gamma$. Indeed, as Fig. 2 d demonstrates, the outlier rejection capability of the WSMF in the presence of gross noise samples can be enhanced by decreasing the value of $\alpha$ (see WSMF curve for $\alpha = 0.5$), yielding a better approximation of the mode for a wider range of scale parameter values, 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.

### 3.3. WSMF-based compressed random measurements

Our proposed compressive sampling operator aims at enabling accurate reconstruction of the observed signal, while being resilient to the presence of gross errors when we operate in impulsive environments. Let $\mathbf{x} \in \mathbb{R}^N$ be the observed noisy signal given by (1), and $\boldsymbol{\phi}_i \in \mathbb{R}^N$, $i = 1, \ldots, M$, be the measurement kernels that form the measurement matrix $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$, that is, $\boldsymbol{\Phi} = \left[\boldsymbol{\phi}_1 \, \boldsymbol{\phi}_2 \, \cdots \, \boldsymbol{\phi}_M\right]^{\mathrm{T}}$. According to (2), in traditional CS, the random measurements are generated by taking linear projections of $\mathbf{x}$ onto the rows of $\boldsymbol{\Phi}$.

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

$$\mathbf{y} = \mathrm{A}_{\Phi}(\mathbf{x}) \doteq \left[h\left(\boldsymbol{\phi}_1, \mathbf{x}\right), h\left(\boldsymbol{\phi}_2, \mathbf{x}\right), \ldots, h\left(\boldsymbol{\phi}_M, \mathbf{x}\right)\right] \tag{11}$$

where the subscript in $\mathrm{A}_{\boldsymbol{\Phi}}$ 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) $\mathrm{A}_{\boldsymbol{\Phi}}(\mathbf{x})$ must preserve the information content of the true signal $\mathbf{x}_0$ by suppressing the effects of large noise samples; ii) $\mathrm{A}_{\boldsymbol{\Phi}}(\mathbf{x})$ should approximate linear sampling for $\mathbf{e}_o \to \mathbf{0}$. The second requirement enables the use of traditional sparse reconstruction algorithms for recovering $\mathbf{x}_0$ from the nonlinear measurements $\mathbf{y}$.

Given the problems arising in the presence of infinite-variance observation noise and the above two requirements of a robust sampling operator, we propose the use of WSMF-based projections to design a robust nonlinear sampling method. This is motivated by the increased robustness of WSMFs in a wide range of noise statistics, as well as by its asymptotic properties, to be proven next. Our proposed WSMF-based compressive sampling operator for the generation of a reduced set of random measurements is defined as follows:

**Definition 3** (**WSMF-based compressive sampling operator**). Let $\boldsymbol{\Phi} = \left[\boldsymbol{\phi}_1 \, \cdots \, \boldsymbol{\phi}_M\right]^{\mathrm{T}} \in \mathbb{R}^{M \times N}$ $(M \ll N)$ be a measurement matrix, whose $i$th row is the measurement kernel $\boldsymbol{\phi}_i^{\mathrm{T}}$, and $\phi_{i,j}$ its $(i, j)$th element. Then, the $i$th WSMF-based random measurement (or projection) is defined by

$$h_{\alpha,\gamma}(\boldsymbol{\phi}_i, \mathbf{x}) \doteq c_i \cdot \mathrm{WSMF}\left(\alpha, \gamma; \boldsymbol{\phi}_i, \mathbf{x}\right) \tag{12}$$

for $i = 1, \ldots, M$, where the WSMF is defined by (9) 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 before, a robust compressive sampling operator should downweight the potential outliers, such that gross errors have negligible influence on the generated random measurements. The following property states the outlier rejection capability of our proposed WSMF-based sampling operator.

**Property 1** (**Outlier rejection capability of WSMF-based compressive sampling operator**). *Let* $\gamma < \infty$, *then*

$$\lim_{x_N \to \pm\infty} h_{\alpha,\gamma}(\boldsymbol{\phi}, [x_1, x_2, \ldots, x_N]) = h_{\alpha,\gamma}(\boldsymbol{\phi}, [x_1, x_2, \ldots, x_{N-1}]). \tag{13}$$

**Proof.** This property results by combining (8), (9) and (12), and noting that by adding a constant to the objective function (see third equation below) does not affect the optimization, as follows:

$$\begin{aligned}
&\lim_{x_N \to \pm\infty} h_{\alpha,\gamma}(\boldsymbol{\phi}, [x_1, x_2, \ldots, x_N]) \\
&= \lim_{x_N \to \pm\infty} \arg\min_{\theta} \left( - \sum_{j=1}^{N} \log\left(f_\alpha(\phi_j(x_j - \theta); \gamma, 0)\right)\right) \\
&= \lim_{x_N \to \pm\infty} \arg\min_{\theta} \left( - \sum_{j=1}^{N-1} \log\left(f_\alpha(\phi_j(x_j - \theta); \gamma, 0)\right) \right. \\
&\quad \left. - \log\left(f_\alpha(\phi_N(x_N - \theta); \gamma, 0)\right)\right) \\
&= \lim_{x_N \to \pm\infty} \arg\min_{\theta} \left( - \sum_{j=1}^{N-1} \log\left(f_\alpha(\phi_j(x_j - \theta); \gamma, 0)\right) \right. \\
&\quad - \log\left(f_\alpha(\phi_N(x_N - \theta); \gamma, 0)\right) \\
&\quad \left. + \log\left(f_\alpha(\phi_N x_N; \gamma, 0)\right)\right) = \lim_{x_N \to \pm\infty} \\
&\quad \arg\min_{\theta} \left( - \sum_{j=1}^{N-1} \log\left(f_\alpha(\phi_j(x_j - \theta); \gamma, 0)\right) \right. \\
&\quad \left. - \log\left(\frac{f_\alpha(\phi_N(x_N - \theta); \gamma, 0)}{f_\alpha(\phi_N x_N; \gamma, 0)}\right)\right).
\end{aligned} \tag{14}$$

Evaluating the above limit is equivalent to finding the limit of the second logarithmic term in (14). 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 $f_\alpha(x; \gamma, 0) \to 0$ as $x \to \pm\infty$. By applying L'Hospital's rule and using the expansion of the first derivative of a S$\alpha$S density $f'_\alpha(x; \gamma, 0)$ as $x \to \pm\infty$ [37], the previous indeterminate form is evaluated as follows:

$$\begin{aligned}
&\lim_{x_N \to \pm\infty} \log\left(\frac{f_\alpha(\phi_N(x_N - \theta); \gamma, 0)}{f_\alpha(\phi_N x_N; \gamma, 0)}\right) \\
&= \log\left(\lim_{x_N \to \pm\infty} \frac{f'_\alpha(\phi_N(x_N - \theta); \gamma, 0)}{f'_\alpha(\phi_N x_N; \gamma, 0)}\right) \\
&= \log\left(\lim_{x_N \to \pm\infty} \frac{\frac{1}{\gamma^2\pi} \sum_{k=1}^{\infty} c_{k,\alpha} \left(\frac{\phi_N(x_N - \theta)}{\gamma}\right)^{-\alpha k - 2}}{\frac{1}{\gamma^2\pi} \sum_{k=1}^{\infty} c_{k,\alpha} \left(\frac{\phi_N x_N}{\gamma}\right)^{-\alpha k - 2}}\right) = \log(1) = 0
\end{aligned} \tag{15}$$

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 $x \to \pm\infty$. Combining (14) and (15) yields

$$\lim_{x_N \to \pm\infty} h_{\alpha,\gamma}(\boldsymbol{\phi}, [x_1, x_2, \ldots, x_N])$$

$$= \arg\min_{\theta} \left( -\sum_{j=1}^{N-1} \log\left(f_\alpha(\phi_j(x_j-\theta);\gamma,0)\right)\right)$$
$$= h_{\alpha,\gamma}(\boldsymbol{\phi},[x_1,\, x_2,\, \ldots,\, x_{N-1}]) \tag{16}$$

which completes the proof. □

### 3.4. Asymptotic behavior of WSMF-based random projections and parameter setting

In this section, we examine the asymptotic behavior of our WSMF-based random projections with respect to the SαS model parameters $(\alpha,\gamma)$. Regarding the characteristic exponent, the following property holds as $\alpha$ approaches the critical points 0, 1, and 2. We note that the values 0 and 2 are critical points as bounds of the feasible range of $\alpha$, whereas 1 is also characterized a critical point, in order to compare with the Cauchy-based counterparts.

**Property 2** (**Asymptotic behavior of WSMF-based random projections with respect to** $\alpha$)**.** *Let $\gamma<\infty$, then, the asymptotic behavior of the WSMF-based random projections as $\alpha$ approaches the critical points 0, 1, and 2, is expressed by the following limits:*

$$\lim_{\alpha\to 0} h_{\alpha,\gamma}(\boldsymbol{\phi},\mathbf{x}) = c\cdot \min_{j=1,\ldots,N}\left\{G_{\alpha,\gamma}\left(x_j;\boldsymbol{\phi},\mathbf{x}\right)\right\} \tag{17}$$

$$\lim_{\alpha\to 1} h_{\alpha,\gamma}(\boldsymbol{\phi},\mathbf{x}) = h_{1,\gamma}(\boldsymbol{\phi},\mathbf{x}) \tag{18}$$

$$\lim_{\alpha\to 2} h_{\alpha,\gamma}(\boldsymbol{\phi},\mathbf{x}) = h_{2,\gamma}(\boldsymbol{\phi},\mathbf{x}) \tag{19}$$

*where $h_{\alpha,\gamma}(\cdot,\cdot)$ is defined by (12), $G_{\alpha,\gamma}(\cdot;\cdot,\cdot)$ is given by (8), and $c$ is a scaling factor as in Definition 3.*

**Proof.** The last two equalities result directly from the definition of the WSMF-based random projections, and specifically from the smoothness of the cost function $G_{\alpha,\gamma}(\theta;\boldsymbol{\phi},\mathbf{x})$ and the continuity of $\rho(x)=-\log(f_\alpha(x;\gamma,0))$ with respect to $\alpha$. Notice also that (18) and (19) are equal to the weighted location of a Cauchy and the weighted mean of a Gaussian, respectively, estimated directly from $\mathbf{x}$ weighted by $\boldsymbol{\phi}$. On the other hand, as $\alpha\to 0$, the cost function is highly nonconvex with multiple local minima located exactly at the points $x_j,\ j=1,\ldots,N$, that is, where $\phi_j(x_j-\theta)=0$. If some $\phi_j=0$, the corresponding term is ignored, since in this case $\rho(\phi_j(x_j-\theta))=\rho(0)$ does not depend on $\theta$. As a result, the overall minimum of the WSMF in (9), and subsequently the limit of $h_{\alpha,\gamma}(\boldsymbol{\phi},\mathbf{x})$ as $\alpha\to 0$, is attained among the points $x_j$ with $\phi_j\neq 0$, namely, $\min_{j=1,\ldots,N}\left\{G_{\alpha,\gamma}\left(x_j;\boldsymbol{\phi},\mathbf{x}\right)\right\}$. This is exactly a selection filter. □

Apart from the outlier rejection property of WSMF-based projections (ref. Property 1), the second desirable property is that the nonlinear sampling operator, $\mathrm{A}_{\boldsymbol{\Phi}}(\mathbf{x})$, should approximate a linear sampling scheme as much as possible in the noiseless case. The following property states the asymptotic behavior of WSMF-based random projections as $\gamma\to\infty$.

**Property 3** (**Asymptotic behavior of WSMF-based random projections with respect to** $\gamma$)**.** *Fix $\alpha\in(0,2]$. In the limit as $\gamma\to\infty$, the WSMF-based random projections reduce to a linear projection onto the elementwise square of the measurement kernel $\boldsymbol{\phi}$, that is,*

$$\lim_{\gamma\to\infty} h_{\alpha,\gamma}(\boldsymbol{\phi},\mathbf{x}) = c\cdot\frac{\sum_{j=1}^{N}\phi_j^2 x_j}{\sum_{j=1}^{N}\phi_j^2} \tag{20}$$

*where $c$ is a scaling factor as in Definition 3.*

**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 $\gamma\to\infty$ [35]. 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 $|x|\le B$, for some $B>0$. 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. 1), whilst $g(x)$ decreases to zero as $|x|\to\infty$ (ref. Fig. 3a). 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,

$$g(x) = g(0) + g'(0)x + \mathcal{O}(x^2) \tag{21}$$

where $\mathcal{O}(x^2)$ denotes the higher order terms. Focusing on expansions of the first derivative of a SαS density function as $x\to 0$ it holds [37],

$$f'_\alpha(x;\gamma,0) = \frac{1}{\gamma^2\pi\alpha}\sum_{k=1}^{\infty}\frac{\Gamma\left(\frac{2k+1}{\alpha}\right)}{(2k-1)!}(-1)^k\left(\frac{x}{\gamma}\right)^{2k-1}. \tag{22}$$

From (22) 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

$$g'(x) = -\frac{f''_\alpha(x;\gamma,0)f_\alpha(x;\gamma,0)-(f'_\alpha(x;\gamma,0))^2}{(f_\alpha(x;\gamma,0))^2}. \tag{23}$$

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 $\gamma\to\infty$ expands the scale of the score function, thus the linear approximation (21) holds over increasingly larger intervals (ref. Fig. 3b).

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

$$G'_{\alpha,\gamma}(\theta;\boldsymbol{\phi},\mathbf{x}) = -\sum_{j=1}^{N}\phi_j\,\rho'(\phi_j(x_j-\theta)) \approx -\sum_{j=1}^{N}\phi_j\,g'(0)\,(\phi_j(x_j-\theta))$$
$$= -g'(0)\left(\sum_{j=1}^{N}\phi_j^2 x_j - \theta\sum_{j=1}^{N}\phi_j^2\right). \tag{24}$$

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

$$\hat{\theta}_{\alpha,\gamma}(\boldsymbol{\phi},\mathbf{x}) = \frac{\sum_{j=1}^{N}\phi_j^2 x_j}{\sum_{j=1}^{N}\phi_j^2}. \tag{25}$$

Combining (25) with (12) concludes the proof. □

It is important to emphasize that calibrating the WSMF-based random projections by appropriately setting the model parameters $(\alpha,\gamma)$ optimizes the shape of the cost function $G_{\alpha,\gamma}(\theta;\boldsymbol{\phi},\mathbf{x})$ for an observed signal $\mathbf{x}$ and measurement kernels $\boldsymbol{\phi}_i,\ i=1,\ldots,M$, and subsequently the robustness of the compressed measurements to impulsive observation noise. Although this calibration step may seem contradictory to the universality of CS, as we highlighted in Section 1.2, our sampling framework resembles a non-universal approach [38] in the sense that, given an ensemble of measurement kernels and a signal $\mathbf{x}$, the compressed measurements generated by our nonlinear sampling operator will be robust to heavy-tailed observation noise for this specific $\mathbf{x}$.

Nevertheless, determining the optimal $(\alpha,\gamma)$ – optimal in the sense that the WSMF-based random projections suppress the effects of large noise samples, whilst they approximate a linear sampling scheme in the noiseless case – from the corrupted signal $\mathbf{x}$ is a nontrivial task.

To address this problem, we design an iterative algorithm, which successively filters the noisy signal $\mathbf{x}$ and updates the SαS

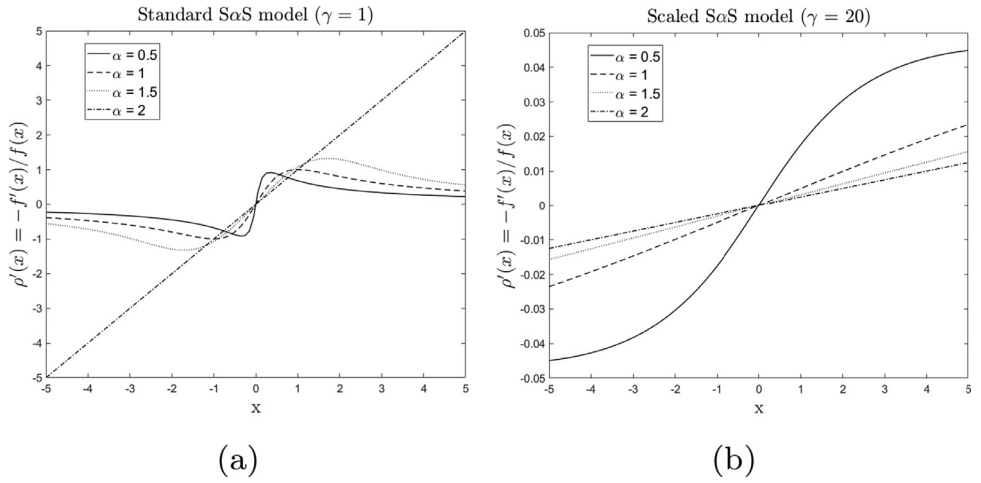


**Fig. 3.** Comparison of $\rho'(x) = -\frac{f'_\alpha(x;\gamma,0)}{f_\alpha(x;\gamma,0)}$ with $\alpha \in \{0.5,\ 1,\ 1.5,\ 2\}$ for (a) the standard S$\alpha$S model ($\gamma = 1$) and (b) a scaled S$\alpha$S model ($\gamma = 20$).

model parameters $(\alpha, \gamma)$ via ML estimation on the residual between the input noisy signal and its current filtered version. The filtering is performed using the weighted S$\alpha$S matched filter of (9) 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$\alpha$S parameter estimates falls below a threshold $\text{tol}_{\text{filt}}$. In our implementation, we set empirically the values of $\text{maxIter}_{\text{filt}} = 10$ and $\text{tol}_{\text{filt}} = 0.001$. The above iterative process is summarized in Algorithm 1. We emphasize that a systematic setting of an optimal termination criterion and/or threshold value is by no means a very important, yet non-trivial problem. Nevertheless, this depends on the given input data, the weights $\mathbf{w}$, the random measurement matrix, the filter's order $L$, and the noise impulsiveness, thus making the automatic selection of $\text{maxIter}_{\text{filt}}$ and $\text{tol}_{\text{filt}}$ a highly challenging task.

**Algorithm 1** Calibration of S$\alpha$S parameters $(\alpha, \gamma)$ for WSMF-based random projections.

**Input:** $\mathbf{x}$, $L$,$\mathbf{w} \in \mathbb{R}^L$,$\text{maxIter}_{\text{filt}}$,$\text{tol}_{\text{filt}}$
**Initialize:**
S$\alpha$S parameters: $[\alpha^{(0)}, \gamma^{(0)}] = \text{mlfit}(\mathbf{x})^{(*)}$
$\text{relChange} = 1000 \cdot \text{tol}_{\text{filt}}$, $t = 0$
**while** (relChange $>$ $\text{tol}_{\text{filt}}$and$t <$ $\text{maxIter}_{\text{filt}}$) **do**
  $\mathbf{p}_{\text{prev}} = [\alpha^{(t)}, \gamma^{(t)}]$
  Signal filtering: $\mathbf{x}_f^{(t)} = \text{stablesigfilt}(\mathbf{x}, \mathbf{w}, L, \alpha^{(t)}, \gamma^{(t)})^{(**)}$
  Update residual: $\mathbf{r}^{(t)} = \mathbf{x} - \mathbf{x}_f^{(t)}$
  Update S$\alpha$S parameters: $[\alpha^{(t+1)}, \gamma^{(t+1)}] = \text{mlfit}(\mathbf{r}^{(t)})$
  $\mathbf{p}_{\text{new}} = [\alpha^{(t+1)}, \gamma^{(t+1)}]$
  $\text{relChange} = \|\mathbf{p}_{\text{new}} - \mathbf{p}_{\text{prev}}\|_2 / \|\mathbf{p}_{\text{prev}}\|_2$
  $t = t + 1$
**end while**
**Output:** The final S$\alpha$S model parameters$(\alpha^{(t)}, \gamma^{(t)})$

$(*)$ mlfit($\mathbf{x}$) denotes a functionthat returns the ML estimates of the S$\alpha$S model parametersfor an input signal $\mathbf{x}$.
$(**)$ stablesigfilt($\mathbf{x}$, $\mathbf{w}$, $L$, $\alpha$, $\gamma$)denotes a function that implements the WSMFof (**??**) in overlapping rolling windows oflength $L$ and step size equal to 1 using the weights$\mathbf{w}$.

An interesting observation is that when the true (noiseless) 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, the WSMF-based random measurements are more expensive to be generated in terms of computational complexity, since each measurement is produced by solving an optimization problem. Thus, our proposed nonlinear sampling method should be considered for compressively sampling a signal when the sensing conditions are not ideal or when a robust sensing process is required.

## 4. Performance evaluation

This section evaluates the robustness of our WSMF-based compressive sampling operator in effectively suppressing the effects of impulsive observation noise, towards increasing the reconstruction accuracy of conventional sparse reconstruction algorithms. To this end, numerical experiments are performed with synthetic signals, as well as with real data. A performance comparison is also carried out against traditional linear random projections and state-of-the-art nonlinear sampling methods tailored to Cauchy and GCD distributed noise. The orthogonal matching pursuit (OMP) [32][3] algorithm is used for sparse signal reconstruction, by assuming that the sparsity level $s$ and noise tolerance $\eta$ are known and used as stopping criteria. We emphasize again that the objective of this work is to focus on the sampling part of a CS system, and, as such, our proposed method can be coupled with any of the conventional sparse reconstruction methods.

### 4.1. Experiments with synthetic signals

The synthetic signals are generated using the following settings, unless stated otherwise: signal length $N = 1024$; cardinality of the sparse support $s = \lceil 2\%N \rceil$; the nonzero coefficients are drawn from a Student's-$t$ distribution with one degree of freedom

[3] MATLAB code available from https://goo.gl/VHvyJe.

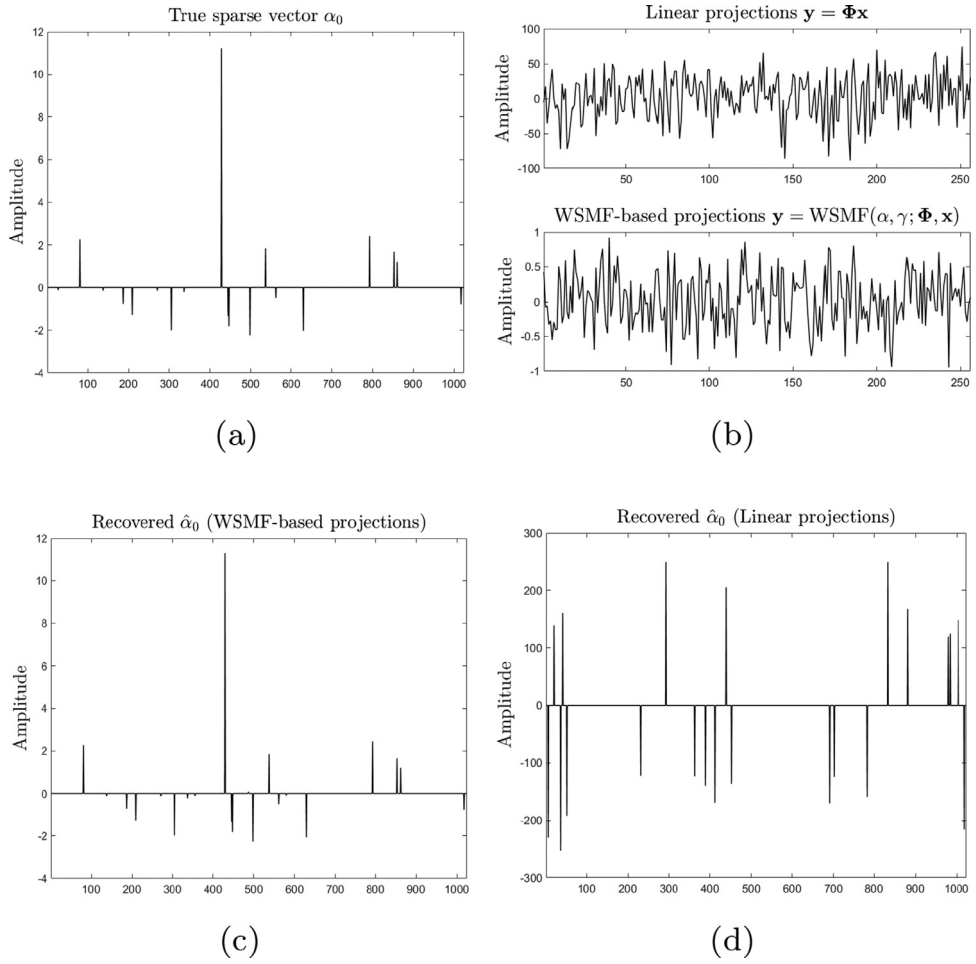


**Fig. 4.** Outlier rejection capability of WSMF-based projections. (a) True sparse signal; (b) linear and WSMF-based measurements for the DCT synthesized signal corrupted by two impulses; (c) OMP reconstructed signal from WSMF-based measurements; (d) OMP reconstructed signal from linear measurements.

and their positions are chosen uniformly at random from the index set $\{1, 2, \ldots, N\}$; the DCT matrix is used as the sparsifying dictionary $\mathbf{\Psi}$, that is, the linear measurement operator is given by $\mathbf{A} = \mathbf{\Phi}\mathbf{\Psi}^{\mathsf{T}}$; the measurement matrix $\mathbf{\Phi}$ has i.i.d. entries $\{-1, 1\}$ drawn from a Bernoulli distribution with equal probability; the number of generated random measurements is set to $M = \lceil 25\%N \rceil$ 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 measurement kernels, and the additive observation noise term. The reconstruction quality is measured in terms of the signal-to-error ratio (SER) (in dB) defined by

$$\mathrm{SER}(\mathbf{x}, \hat{\mathbf{x}}) = 10 \log_{10} \left( \frac{\sum_{j=1}^{N} x_j^2}{\sum_{j=1}^{N} (x_j - \hat{x}_j)^2} \right) \tag{26}$$

where $\mathbf{x}$ and $\hat{\mathbf{x}}$ correspond to the original and reconstructed signals, respectively.

First, we illustrate experimentally the outlier rejection capability of our WSMF-based random measurements (Property 1). For this purpose, two impulses with amplitudes equal to $10^3$ and $10^2$ are added to the original signal at randomly chosen positions. The true sparse signal is reconstructed using OMP for both linear and WSMF-based random projections. The resulting SER equals $-36.55$ dB for the linear case and 31.07 dB for our WSMF-based method, where the filter parameters $(\alpha, \gamma)$ are estimated directly from the corrupted signal using Algorithm 1. Fig. 4 shows the true sparse signal, the associated linear and WSMF-based measurements, and the corresponding OMP reconstructions.

Next, we address the more challenging case, where the original signal is corrupted by additive S$\alpha$S observation noise. Specifically, the noise characteristic exponent $\alpha_n$ takes values in $[0.5, 2]$, whilst its dispersion $\gamma_n$ varies in $\{0.1, 1\}$. This gives a *geometric power* range between $-11.19$ dB and 2.51 dB. The geometric power (in dB) is a measure of strength for random variables with infinite variance, which is defined by

$$\mathrm{P_g}(\alpha, \gamma) = 10 \log_{10} \left( \gamma C_g^{\frac{1}{\alpha}-1} \right) \tag{27}$$

where $C_g = e^{C_e}$ and $C_e = 0.5772\ldots$ is the Euler constant. Notice that, although $\alpha \in (0, 2]$ in theory, however, in the subsequent evaluations with synthetic signals we vary $\alpha$ in $[0.5, 2]$. This is because, as described in [34] and the STABLE toolbox, the utilized algorithms for the numerical calculations and optimizations involving S$\alpha$S densities are robust for $\alpha \geq 0.5$. Nevertheless, in practical applications (ref. Section 1) $\alpha$ is typically greater than 1, which also explains why we set $\alpha \in [1, 2]$ in the case of real EEG signals employed in Section 4.2.

Fig. 5 compares the reconstruction performance of OMP when using WSMF projections against linear and myriad [26] projections for S$\alpha$S observation noise with model parameters as mentioned above. Fig. 5a corresponds to a low-dispersion (or, equivalently, low geometric power) noise, while Fig. 5b demonstrates the performance in higher-dispersion (or, equivalently, higher ge-

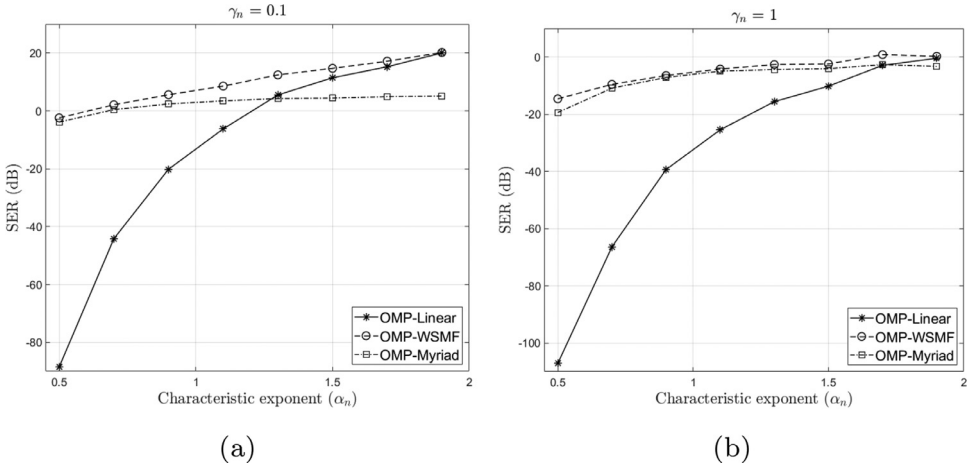


**Fig. 5.** Comparison of WSMF projections with linear and myriad projections for S$\alpha$S observation noise with $\alpha_n \in [0.5, 2]$ and $\gamma_n \in \{0.1, 1\}$. Average SER is shown for OMP-based reconstruction over 500 Monte Carlo runs.

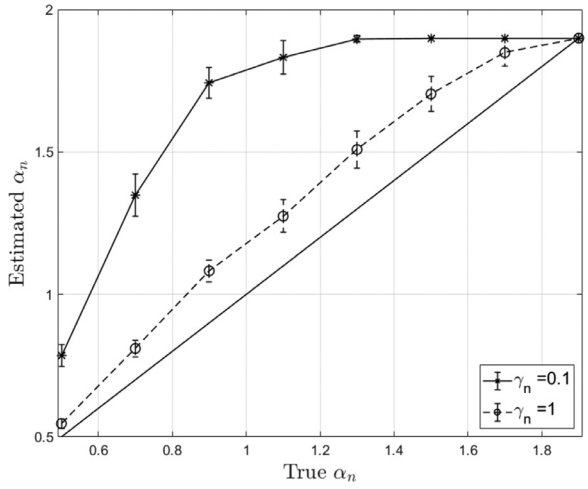


**Fig. 6.** True vs. estimated (using Algorithm 1) noise characteristic exponent for $\alpha_n \in [0.5, 2]$ and $\gamma_n \in \{0.1, 1\}$ (the straight diagonal corresponds to the ideal case, where the estimated equals the true value). The average and standard deviation of the estimated $\alpha_n$ are calculated over 500 Monte Carlo runs.

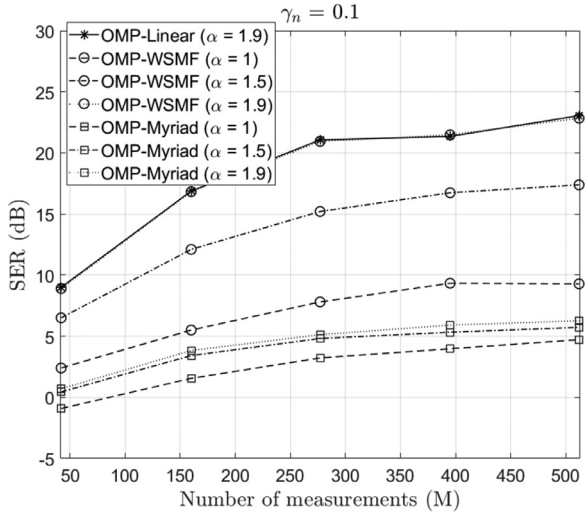


**Fig. 7.** Comparison of WSMF projections with linear and myriad projections for S$\alpha$S observation noise with $\alpha_n \in \{1, 1.5, 1.9\}$ and $\gamma_n = 0.1$. Average SER is shown as a function of $M$ for OMP-based reconstruction over 500 Monte Carlo runs.

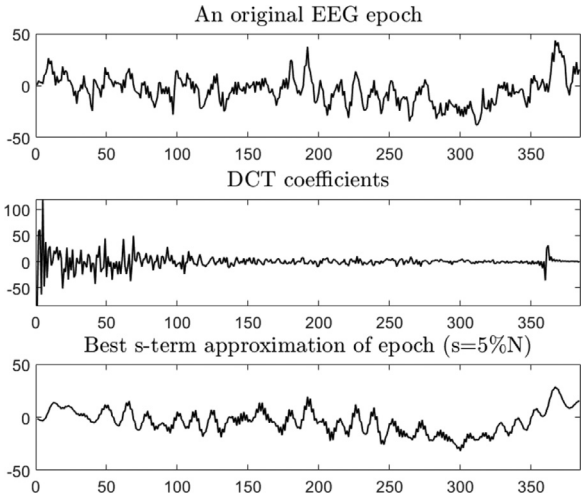


**Fig. 8.** Example of an original EEG epoch, its DCT coefficients and best $s$-term approximation for $s = \lceil 5\%N \rceil$.

ometric power) noise conditions. In both 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. Indeed, as shown in Fig. 6, the noise characteristic exponent, $\alpha_n$, estimated by Algorithm 1 approximates closely its true value, especially as the noise dispersion, $\gamma_n$, increases. This is justified by the fact that, as the dispersion increases the impulsive heavy-tailed noise dominates the underlying signal, which allows the parameters calibration algorithm to better discriminate between the noise and signal statistics. Nevertheless, even for small noise dispersion values, for which the estimation of $\alpha_n$ is not very accurate, our proposed compressive sampling operator still achieves to sufficiently suppress the impulsive observation noise, yielding an accurate reconstruction of the original noiseless signals.

Concerning the linear projections, in highly impulsive environments (small $\alpha_n$ 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 the performance of WSMF projections when the noise statistics tends to the

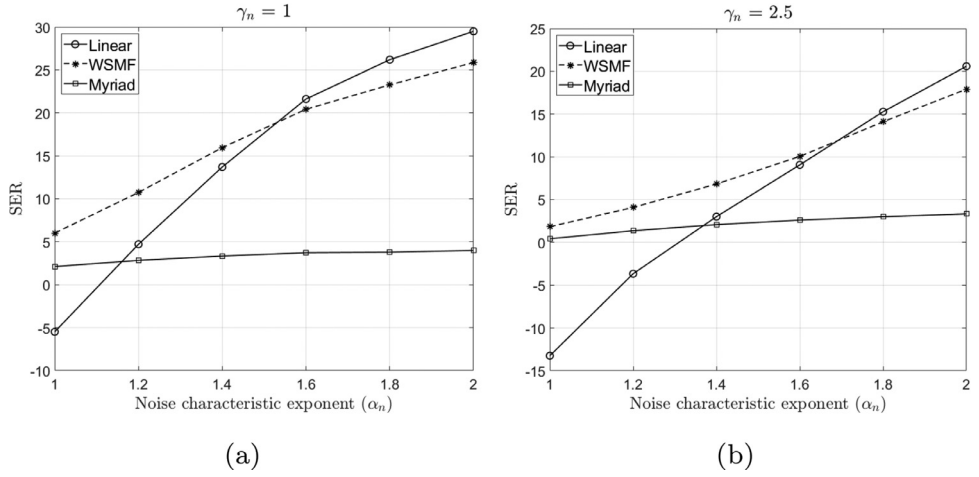


**Fig. 9.** Reconstruction performance for the EEG dataset using WSMF, myriad and linear projections. S$\alpha$S observation noise is added with $\alpha_n \in \{1 : 0.2 : 2\}$ and $\gamma_n \in \{1, 2.5\}$. Average SER is shown for OMP-based reconstruction over 500 Monte Carlo runs.

Gaussian ($\alpha_n \to 2$) 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 $\alpha_n$ 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.

As a last experiment, we examine how the number, $M$, of WSMF projections affects the reconstruction accuracy. For this, $M$ 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 average SER (in dB) between the original noiseless signal and its OMP-based reconstructions, as a function of $M$, using WSMF, linear, and myriad projections. First, we observe the significantly improved reconstruction performance in the WSMF case against the 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 $(\alpha_n, \gamma_n)$, which may not be accurate for very small sample sizes, however, they still outperform the myriad case where $\alpha_n$ is inherently fixed to 1. Finally, as expected, the results show that additional measurements are required to improve the reconstruction quality and compensate for the increased impulsiveness when $\alpha_n$ decreases, for all the three types of projections.

### 4.2. Experiments with EEG data

The following experiments evaluate and compare the performance of our proposed nonlinear compressive sampling method on real data [39]. Specifically, the utilized dataset contains electroencephalography (EEG) signals of 32 channels with sequence length of 30,720 data points. Each channel consists of 80 epochs, each one containing $N = 384$ points. Artifacts caused by muscle movement also occur in the signals. The EEG signals are compressively sampled in an epoch-by-epoch fashion using a Bernoulli matrix $\mathbf{\Phi}$, whereas the $384 \times 384$ DCT matrix is used as the sparsifying dictionary $\mathbf{\Psi}$. The sparsity level is fixed at $s = \lceil 5\%N \rceil$, whilst the number of random measurements is set to $M = \lceil N/3 \rceil$. The reconstruction quality is measured in terms of the achieved SER averaged over all the EEG epochs and Monte Carlo runs. Fig. 8 shows an original EEG epoch, along with the corresponding DCT coefficients and the best $s$-term approximation for $s = \lceil 5\%N \rceil$. Clearly, the signal is nonsparse both in time and frequency. Nevertheless, for the reconstruction we consider that the targeted sparsity level of the DCT coefficients is upper bounded by $s = \lceil 5\%N \rceil$.

In order to investigate the effects of noise impulsiveness strength on the outlier rejection capability of the generated random measurements, and subsequently on the reconstruction quality, we vary $\alpha_n$ in $\{1 : 0.2 : 2\}$ and $\gamma_n$ in $\{1, 2.5\}$. As before, the number of random measurements is set to $M = \lceil N/3 \rceil$ and the noise tolerance is fixed to $M\gamma_n^2$ for the OMP algorithm. Fig. 9 compares the performance, in terms of the achieved SER averaged over all the EEG epochs and Monte Carlo runs, of an OMP-based reconstruction by employing our WSMF projections, along with the myriad and linear counterparts. Clearly, as the noise impulsiveness increases ($\alpha_n \to 1$), the linear measurements fail to achieve a fair reconstruction of the epochs for both noise dispersion values. On the other hand, the WSMF projections are highly robust over the whole range of $\alpha_n$ values, outperforming significantly their myriad counterpart. Interestingly, our sampling operator performs better than the myriad-based one even for Cauchy observation noise ($\alpha_n = 1$). This is justified by the fact that, although the noise part follows a Cauchy distribution, however, the distribution of the noisy signal the myriad filter (WMyF) is applied to deviates significantly from a Cauchy. In particular, the estimated characteristic exponent averaged over all the noisy epochs and Monte Carlo runs is equal to 1.79 and 1.61 for $\gamma_n = 1$ and $\gamma_n = 2.5$, respectively. On the contrary, as the noise statistics tend to a Gaussian, the performance of linear measurements improves, exceeding the two nonlinear sampling operators. However, the larger the noise dispersion, the closer to the Gaussian the noise statistics must be, in order for the linear projections to compete the WSMF-based ones. Nevertheless, this behavior can be also attributed to the fact that the OMP algorithm is especially tailored to linear random projections.

## 5. Conclusions and future work

In this paper, a robust method was proposed for nonlinear compressive sampling of signals corrupted by impulsive observation noise. Specifically, the heavy-tailed noise statistics, with possibly infinite variance, was modeled by means of S$\alpha$S distributions. Sub-

sequently, the weighted S$\alpha$S matched filter (WSMF) projections were proposed as an efficient nonlinear sampling operator, which suppresses 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 robustness against the underlying heavy-tailed noise statistics. Experimental evaluations with synthetic and real data revealed that, when coupled with traditional sparse reconstruction methods, our WSMF-based sampling mechanism outperforms significantly the linear and myriad counterparts in the case of highly impulsive observation noise, whilst converging to the performance of linear projections for light-tailed (near-Gaussian) noise statistics.

However, a theoretical framework for optimally setting the key parameters of our WSMF-based operator 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 increase the robustness of our nonlinear sampling operator against gross outliers due to the underlying heavy-tailed statistics of infinite variance noise. Besides, although the robust compressive sampling operator proposed herein is tailored to impulsive noise modelled by alpha-stable distributions, we are highly interested in testing its performance in practical scenarios with alternative (heavy-tailed) noise models (e.g. Laplace, Log-normal, Middleton class A/B). In addition, our current study focused on multivariate observation noise with i.i.d. components. As a further extension, the more generic case of observation noise with dependent, jointly S$\alpha$S components will also be investigated. Finally, motivated by the work in [21], we are interested in testing the performance of our robust compressive sampling operator, in conjunction with a local anomaly detector for structural health monitoring in impulsive environments. Specifically, two research directions will be studied, namely, i) the efficiency of WSMF-based random measurements at improving the anomaly detector's accuracy by sufficiently suppressing the impulsive observation noise at the reconstructed signal, and ii) the design of a local anomaly detector which operates directly on the WSMF compressive samples without necessitating signal reconstruction.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## CRediT authorship contribution statement

**George Tzagkarakis:** Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing, Funding acquisition. **John P. Nolan:** Methodology, Software, Writing - original draft, Funding acquisition. **Panagiotis Tsakalides:** Writing - original draft, Validation, Funding acquisition.

## Acknowledgements

This work was partially funded by the Interreg V-A Greece-Cyprus 2014–2020 programme, co-financed by the European Union (ERDF) and National Funds of Greece and Cyprus, under the project SmartWater2020, and the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT) under grant agreements No. 2285 (neuronXnet) and No. 1725 (V4-ICARUS). J. P. Nolan is supported by contract W911NF-12-1-0385 from the U.S. Army Research Office.

## References

[1] E. Candés, Y. Eldar, D. Needell, P. Randall, Compressed sensing with coherent and redundant dictionaries, Appl. Comput. Harmonic Anal. 31 (1) (2011) 59–73, doi:10.1016/j.acha.2010.10.002.
[2] E. Candés, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inf. Theory 52 (2) (2006) 489–509, doi:10.1109/TIT.2005.862083.
[3] D. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (4) (2006) 1289–1306, doi:10.1109/TIT.2006.871582.
[4] E. Candés, The restricted isometry property and its implications for compressed sensing, C. R. Acad. Sci. Paris, Ser. I 346 (2008) 589–592, doi:10.1016/j.crma.2008.03.014.
[5] S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, Uniform uncertainty principle for bernoulli and subgaussian ensembles, Constructive Approx. 28 (3) (2008) 277–289, doi:10.1007/s00365-007-9005-8.
[6] T.T. Do, L. Gan, N.H. Nguyen, T.D. Tran, Fast and efficient compressive sensing using structurally random matrices, IEEE Trans. Signal Process. 60 (1) (2012) 139–154, doi:10.1109/TSP.2011.2170977.
[7] R.R. Naidu, P. Jampana, C.S. Sastry, Deterministic compressed sensing matrices: construction via Euler squares and applications, IEEE Trans. Signal Process. 64 (14) (2016) 3566–3575, doi:10.1109/TSP.2016.2550020.
[8] P. Sasmal, S.S. Thoota, C.R. Murthy, Disjunct matrices for compressed sensing, in: Proc. of IEEE Intl' Conf. Acoust., Speech, Signal Process. (ICASSP), 2019. Brighton, UK.
[9] P.J. Huber, E.M. Ronchetti, Robust statistics, 2nd ed., Wiley, 2009.
[10] G. Arce, Nonlinear signal processing: A Statistical approach, Wiley, 2005.
[11] G. Samorodnitsky, M. Taqqu, Stable non-Gaussian random processes: Stochastic models with infinite variance, Chapman & Hall, New York, 1994.
[12] G. Tzagkarakis, J.P. Nolan, P. Tsakalides, Compressive sensing using symmetric alpha-stable distributions for robust sparse signal reconstruction, IEEE Trans. Signal Process. 67 (3) (2019) 808–820, doi:10.1109/TSP.2018.2887400.
[13] S. Banerjee, M. Agrawal, Underwater acoustic communication in the presence of heavy-tailed impulsive noise with bi-parameter Cauchy-Gaussian mixture model, in: Proc. of Intl' Symp. on Ocean Electron. (SYMPOL), 2013. Kochi, India
[14] K. Pelekanakis, M. Chitre, Adaptive sparse channel estimation under symmetric alpha-stable noise, IEEE Trans. Wireless Commun. 13 (6) (2014) 3183–3195, doi:10.1109/TWC.2014.042314.131432.
[15] P. Tsakalides, F. Trinci, C. Nikias, Radar CFAR thresholding in heavy-tailed clutter and positive alpha-stable measurements, in: Proc. of 32nd Asilomar Conf. Signals, Syst. Comput., 1998. Pacific Grove, CA.
[16] V.A. Aalo, K.P. Peppas, G. Efthymoglou, M. Alwakeel, S. Alwakeel, Performance of CA-CFAR receivers in alpha-stable clutter, in: Proc. of IEEE Intl' Symp. on Signal Process. Inf. Technol. (ISSPIT), 2013. Athens, Greece.
[17] A. Achim, A. Bezerianos, P. Tsakalides, Novel Bayesian multiscale method for speckle removal in medical ultrasound images, IEEE Trans. Med. Imag. 20 (8) (2001) 772–783, doi:10.1109/42.938245.
[18] A. Achim, A. Basarab, G. Tzagkarakis, P. Tsakalides, D. Kouamé, Reconstruction of ultrasound RF echoes modeled as stable random variables, IEEE Trans. Comput. Imag. 1 (2) (2015) 86–95, doi:10.1109/TCI.2015.2463257.
[19] J.H. McCulloch, Financial applications of stable distributions, Handbook of Statistics 14 (1996) 393–425, doi:10.1016/S0169-7161(96)14015-3.
[20] P. Glasserman, P. Heidelberger, P. Shahabuddin, Portfolio value-at-risk with heavy-tailed risk factors, Math. Finance 12 (3) (2002) 239–269, doi:10.1111/1467-9965.00141.
[21] J. Hebda-Sobkowicz, R. Zimroz, M. Pitera, A. Wyłomańska, Informative frequency band selection in the presence of non-gaussian noise a novel approach based on the conditional variance statistic with application to bearing fault diagnosis, Mech. Syst. and Signal Process. 145 (2020) 106971, doi:10.1016/j.ymssp.2020.106971.
[22] M. Masood, T.Y. Al-Naffouri, Sparse reconstruction using distribution agnostic Bayesian matching pursuit, IEEE Trans. Signal Process. 61 (21) (2013) 5298–5309, doi:10.1109/TSP.2013.2278814.
[23] R. Carrillo, A. Ramirez, G. Arce, K. Barner, B. Sadler, Robust compressive sensing of sparse signals: a review, EURASIP J. Adv. Signal Process. 2016 (1) (2016) 108, doi:10.1186/s13634-016-0404-5.
[24] A. Ramirez, G. Arce, B. Sadler, Fast algorithms for reconstruction of sparse signals from Cauchy random projections, in: Proc. of 18th Eur. Signal Process. Conf. (EUSIPCO), 2010. Aalborg, Denmark.
[25] A. Ramirez, G. Arce, D. Otero, J.-L. Paredes, B. Sadler, Reconstruction of sparse signals from $\ell_1$ dimensionality-reduced Cauchy random projections, IEEE Trans. Signal Process. 60 (11) (2012) 5725–5737, doi:10.1109/TSP.2012.2208954.
[26] R. Carrillo, K. Barner, T. Aysal, Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise, IEEE J. Sel. Topics Signal Process. 4 (2) (2010) 392–408, doi:10.1109/JSTSP.2009.2039177.
[27] S.D. Babacan, R. Molina, A. Katsaggelos, Bayesian compressive sensing using Laplace priors, IEEE Trans. Image Process. 19 (1) (2010) 53–63, doi:10.1109/TIP.2009.2032894.
[28] C. Tzagkarakis, A. Mouchtaris, P. Tsakalides, Musical genre classification via generalized Gaussian and alpha-stable modeling, in: Proc. of IEEE Intl' Conf. Acoust., Speech, Signal Process. (ICASSP), 2006. Toulouse, France
[29] H.E. Ghannudi, L. Clavier, N. Azzaoui, F. Septier, P. alain Rolland, Alpha-stable interference modeling and Cauchy receiver for an IR-UWB ad hoc network, IEEE Trans. Commun. 58 (6) (2010) 1748–1757, doi:10.1109/TCOMM.2010.06.090074.

[30] S. Liao, A.C.S. Chung, Feature based nonrigid brain MR image registration with symmetric alpha stable filters, IEEE Trans. Med. Imag. 29 (1) (2010) 106–119, doi:10.1109/TMI.2009.2028078.
[31] C. Nikias, M. Shao, Signal processing with alpha-stable distributions and applications, Wiley, NY, USA, 1995.
[32] J. Tropp, A. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inf. Theory 53 (12) (2007) 4655–4666, doi:10.1109/TIT.2007.909108.
[33] J.P. Nolan, Numerical calculation of stable densities and distribution functions, Commun. Statist.-Stochastic Models 13 (4) (1997) 759–774, doi:10.1080/15326349708807450.
[34] 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 (2001) 379–400.
[35] J.P. Nolan, J. Gonzalez, R. Núnez, Stable filters: A robust signal processing framework for heavy-tailed noise, in: Proc. of IEEE Intl' Radar Conf., 2010. Washington, DC, USA
[36] J.P. Nolan, Advances in nonlinear signal processing for heavy tailed noise, in: Proc. of Intl' Workshop on Appl. Prob., 2008. Compiegne, France
[37] M. Matsui, A. Takemura, Some improvements in numerical evaluation of symmetric stable density and its derivatives, Commun. in Statist. 35 (1) (2006) 149–172, doi:10.1080/03610920500439729.
[38] E. Candés, Y. Plan, A probabilistic and RIPless theory of compressed sensing, IEEE Trans. Inf. Theory 57 (11) (2011) 7235–7254, doi:10.1109/TIT.2011.2161794.
[39] A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, J. of Neurosc. Methods 134 (1) (2004) 9–21, doi:10.1016/j.jneumeth.2003.10.009.