# BAYESIAN COMPRESSED SENSING OF A HIGHLY IMPULSIVE SIGNAL IN HEAVY-TAILED NOISE USING A MULTIVARIATE CAUCHY PRIOR

*George Tzagkarakis and Panagiotis Tsakalides*

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

**ABSTRACT**

Recent studies reveal that if a signal is highly compressible in some orthonormal basis, then an accurate reconstruction can be obtained from random projections using a very small subset of the projection coefficients, and thus, reducing the complexity of the sensing system. A Bayesian framework was introduced recently with respect to the reconstruction of the original (noisy) signal, providing some advantages when compared with reconstruction methods, employing norm-based constrained minimization approaches. These Bayesian methods were designed by using mixtures of Gaussians to approximate the sparsity of the prior distribution of the projection coefficients. However, there are cases in which a signal exhibits a highly impulsive behavior, and thus, resulting in an even sparser coefficient vector. In this paper, we develop a Bayesian approach for estimating the original signal based on a set of compressed-sensing measurements corrupted by heavy-tailed noise. The prior belief that the vector of projection coefficients should be sparse is enforced by fitting its prior distribution by means of a heavy-tailed multivariate Cauchy distribution. The experimental results show that our proposed method achieves an improved reconstruction performance, in terms of a smaller reconstruction error, while increasing the sparsity using less basis functions, compared with the recently introduced Gaussian-based Bayesian implementation.

## 1. INTRODUCTION

Sampling is a key concept of signal processing because it allows real-world signals in the continuous-domain to be acquired, represented, and processed in the discrete-domain. In many modern applications, including digital image and video cameras, the classical Shannon/Nyquist sampling rate is so high, resulting in too many samples, and thus, making compression a necessity prior to storage or transmission. Several works [1, 2, 3] have shown that many natural signals result in a highly compact (sparse) representation, when they are projected on orthonormal basis functions (e.g.,wavelets and sinusoids). The traditional approach to compressing such a sparse signal is to compute its transform coefficients and then store or transmit only a small number of large amplitude coefficients.

Compressed Sensing (CS) is a new framework introduced recently for simultaneous sensing and compression [4, 5]. CS enables a potentially significant reduction in the sampling and computation costs at a sensor with limited capabilities. According to the CS framework, a signal having a sparse representation in a transform basis can be reconstructed from a small set of projections onto a second, measurement basis that is incoherent with the first one.

Let $\mathbf{\Psi}$ be a $N \times N$ matrix, whose columns correspond to the transform basis functions. Then, a given signal $\vec{f} \in \mathbb{R}^N$ can be represented as $\vec{f} = \mathbf{\Psi}\vec{w}$, where $\vec{w} \in \mathbb{R}^N$ is the weight vector. As mentioned above, for many signals $\vec{f}$ present in nature, the majority of the components of $\vec{w}$ have negligible amplitude. In particular, a signal $\vec{f}$ is $L$-sparse in basis $\mathbf{\Psi}$ if the corresponding weight vector $\vec{w}$ has $L$ non-zero components ($L \ll N$). In a real-world scenario $\vec{f}$ is not strictly $L$-sparse, but it is said to be *compressible* when the re-ordered components of $\vec{w}$ decay at a power-law.

Consider also an $M \times N$ measurement matrix $\mathbf{\Phi}$, $M < N$, where the rows of $\mathbf{\Phi}$ are incoherent with the columns of $\mathbf{\Psi}$. For example, let $\mathbf{\Phi}$ contain independent and identically distributed (i.i.d.) Gaussian entries. Such a matrix is incoherent with any fixed transform matrix $\mathbf{\Psi}$ with high probability (universality property) [5].

If the signal $\vec{f}$ is compressible in $\mathbf{\Psi}$, then, it is possible to perform directly a set of compressed sensing (CS) measurements $\vec{g}$, through random projections. That is, the $m$-th CS measurement, $\vec{g}_m$, results by projecting $\vec{f}$ onto a random linear combination of the basis functions, $\vec{g}_m = \vec{f}^T(\mathbf{\Psi}\vec{\phi}_m)$ , where $\vec{\phi}_m \in \mathbb{R}^N$ is a random vector with i.i.d. components. The CS measurements can be written in the following compact form, $\vec{g} = \mathbf{\Phi}\mathbf{\Psi}^T\vec{f} = \mathbf{\Phi}\vec{w}$ , where $\mathbf{\Phi} = [\vec{\phi}_1, \ldots, \vec{\phi}_M]^T$. Thus, the problem of reconstructing the original signal $\vec{f}$ from the CS measurements $\vec{g}$, is equivalent to estimating the (sparse) weight vector $\vec{w}$.[1]

Most of the recent literature on CS [6, 7] has concentrated on constrained optimization-based methods for signal reconstruction. For instance, the $\ell_1$-norm minimization approach seeks a sparse weight vector $\vec{w}$ by solving the following linear problem,

$$\hat{\vec{w}} = arg\min_{\vec{w}} \|\vec{w}\|_1 \ , \ s.t. \ \ \vec{g} = \mathbf{\Phi}\vec{w} \ . \tag{1}$$

The main approaches for the solution of such an optimization problem include linear programming [8] and greedy algorithms [9], resulting in a *point estimate* of the weight vector $\vec{w}$.

In recent studies [10, 11, 12, 13], the inversion of CS measurements was considered from a probabilistic/Bayesian perspective. In particular, given a prior belief that the weight vector $\vec{w}$ should be sparse and the set of CS measurements $\vec{g}$ (observables), the objective is to formulate a *posterior probability distribution* for $\vec{w}$. This approach improves the accuracy over point-estimate methods and provides confidence intervals (error bars) in the approximation of $\vec{f}$. For computational purposes, and in order to get closed-form expressions, current Bayesian techniques employ the multivariate Gaussian distribution as the prior probability model.

In the present work, the estimation of $\vec{w}$ is also performed in a Bayesian framework. However, there are many cases in which a signal exhibits a highly impulsive behavior, resulting in an even sparser coefficient vector $\vec{w}$. In contrast to previous studies, our proposed method consists of modeling the prior probability of $\vec{w}$ with a heavy-tailed distribution, which promotes the sparsity of $\vec{w}$. This is motivated by the fact that a heavy-tailed density function is suitable for modeling highly impulsive signals. In particular, in our case, $\vec{w}$ can be considered as a highly impulsive "signal" since it

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

[1] Obviously, $\vec{f}$ and $\vec{w}$ are equivalent representations of the signal, with $\vec{f}$ being in the time (or space) domain and $\vec{w}$ in the (transform) $\mathbf{\Psi}$ domain.

is characterized by a large number of small-amplitude components and a small number of large-amplitude components. For this purpose, a multivariate Cauchy distribution (which is a member of the so-called sub-Gaussian family) is employed to model the heavy-tailed behavior of $\vec{w}$, as well as the noise corrupting the projection coefficients, since it models efficiently highly impulsive environments [14, 15] and also it yields closed form expressions in the subsequent Bayesian inference.

The rest of the paper is organized as follows: in Section 2, the heavy-tailed statistical signal model is described briefly. In Section 3, the CS inversion process for the estimation of the weight vector $\vec{w}$ is analyzed in detail. In Section 4, the performance of the proposed method is compared with the original Bayesian CS method first introduced in [13]. Finally, in Section 5 conclusions are drawn, together with directions for future work.

## 2. STATISTICAL SIGNAL MODEL

As mentioned in the previous section, the representation of the time-domain signal $\vec{f}$, is equivalent to its representation $\vec{w}$ in the frequency-domain (for instance, $\vec{w}$ may contain the wavelet coefficients of $\vec{f}$). Thus, without loss of generality, in the following study we consider that the noisy CS measurements are acquired in the transform domain, using the model:

$$\vec{g} = \mathbf{\Phi}\vec{w} + \vec{\eta} \ , \tag{2}$$

where the measurement matrix $\mathbf{\Phi}$ is described in Section 1, and $\vec{\eta}$ is the associated noise component. Assuming that matrix $\mathbf{\Phi}$ is known, the quantities to be estimated, given the CS measurements $\vec{g}$, are the sparse weight vector $\vec{w}$ and the noise underlying variance $\sigma_\eta^2$. In this work, the assumption that $\vec{w}$ and $\vec{\eta}$ are highly sparse is formalized by modeling their prior distribution using a member of the so-called *sub-Gaussian Symmetric $\alpha$-Stable (S$\alpha$S)* family [16, 17]:

**Definition 1** *A vector $\vec{w}$ is called a sub-Gaussian S$\alpha$S random vector (in $\mathbb{R}^N$), with underlying Gaussian vector $\vec{G}$, iff it can be written as $\vec{w} = A^{1/2}\vec{G}$, where $A$ is a positive $\frac{\alpha}{2}$-stable random variable with parameters $A \sim S_{\alpha/2}\big((\cos\frac{\pi\alpha}{4})^{2/\alpha}, 1, 0\big)$ and $\vec{G} = [G_1, G_2, \ldots, G_N]^T$ is a zero-mean Gaussian random vector, independent of $A$, with covariance matrix $\mathbf{\Sigma}$.*

A multivariate sub-Gaussian distribution, with underlying covariance matrix $\mathbf{\Sigma}$ and location parameter $\vec{\mu}$, is often denoted by $\alpha$-SG($\vec{\mu}, \mathbf{\Sigma}$), where the parameter $\alpha$ is the characteristic exponent, controlling the heaviness of the tails of the marginal sub-Gaussian distributions. The multivariate Cauchy distribution (MvC) is $\alpha$-SG($\vec{\mu}, \mathbf{\Sigma}$) for $\alpha = 1$, and its density function is given by:

$$p(\vec{x}) = \frac{\Gamma\big(\frac{N+1}{2}\big)}{\pi^{(N+1)/2}} |\mathbf{\Sigma}|^{-\frac{1}{2}} \big[1 + (\vec{x}-\vec{\mu})^T \mathbf{\Sigma}^{-1} (\vec{x}-\vec{\mu})\big]^{-\frac{(N+1)}{2}} \ , \tag{3}$$

where $\Gamma(\cdot)$ is the Gamma function and $|\cdot|$ denotes the matrix determinant.

In (2) we consider that $\vec{w} \sim$ 1-SG($\vec{\mu}, \mathbf{\Sigma}$), where $\mathbf{\Sigma} = diag(\sigma_1^2, \ldots, \sigma_N^2)$ and $\vec{\eta} \sim$ 1-SG($\vec{0}, \sigma_\eta^2 \mathbf{I}_{M\times M}$), where $\vec{0}$ is a $M \times 1$ zero vector and $\mathbf{I}_{M\times M}$ the identity matrix. The stability property states that the sum of two sub-Gaussian (and consequently two multivariate Cauchy) distributions with the same $\alpha$, is again sub-Gaussian (multivariate Cauchy). Thus, the measurement vector $\vec{g} \sim$ 1-SG($\vec{\mu}_g = \mathbf{\Phi}\vec{\mu}, \mathbf{R} := \mathbf{\Phi}\mathbf{\Sigma}\mathbf{\Phi}^T + \sigma_\eta^2\mathbf{I}$).

From the above, the multivariate Cauchy distribution can be viewed as a mixture of Gaussians scaled by a S$\alpha$S r.v. $A^{1/2}$, where $\mathbf{\Sigma}$, and thus, $\mathbf{R}$, are determined by a discrete random vector $\vec{\tau} = [\tau_0, \ldots, \tau_{N-1}]^T$ of mixture parameters. In the following, we will denote the matrices by $\mathbf{\Sigma}(\vec{\tau}), \mathbf{R}(\vec{\tau})$. For simplicity (as in [13]), we assume that $\vec{\tau} \sim$Bernoulli($\lambda_1$), that is, $\Pr(\tau_i = 1) = \lambda_1$ and $\Pr(\tau_i = 0) = \lambda_0 = 1 - \lambda_1$. Thus, $\mathbf{\Sigma}(\vec{\tau}) = diag(\sigma_{\tau_1}^2, \ldots, \sigma_{\tau_N}^2)$, with $\sigma_{\tau_i}^2 \neq 0$ or $\sigma_{\tau_i}^2 = 0$ depending on whether the $i$-th component is significant and activated in the mixture or not. In the general case, the Gaussian part of each mixture component may be chosen from a set of $\Omega$ Gaussians, with $\mu_{\tau_i} \in \{\mu_\omega\}_{\omega=1}^{\Omega}$ and $\sigma_{\tau_i}^2 \in \{\sigma_\omega^2\}_{\omega=1}^{\Omega}$. In the present study, we consider for simplicity that $\sigma_{\tau_i}^2 \in \{\sigma_0^2, \sigma_1^2\}$, where $\sigma_0^2 = 0$ to enforce sparsity, and $\mu_{\tau_i} \in \{0, \mu_1\}$.

## 3. ESTIMATE A SPARSE VECTOR VIA AN MVC PRIOR

In this section, we describe the process for reconstructing the sparse vector $\vec{w}$ from the CS measurements $\vec{g}$. Following the above analysis, this process is reduced to finding the (sparse) set of the most probable basis configurations (columns of $\mathbf{\Phi}$) associated with the activated mixture components. Then, their corresponding posterior probabilities are employed to obtain a Minimum Mean Squared Error (MMSE), as well as a Maximum A Posteriori (MAP) estimate of the sparse vector $\vec{w}$.

The posterior probability of a given $\vec{\tau}'$ is given by Bayes' rule:

$$p(\vec{\tau}'|\vec{g}) = \frac{p(\vec{g}|\vec{\tau}')p(\vec{\tau}')}{\sum_{\vec{\tau}\in\mathscr{T}} p(\vec{g}|\vec{\tau})p(\vec{\tau})} \ , \tag{4}$$

where $\mathscr{T} = \{0,1\}^N$ contains the $2^N$ possible basis configurations. Let $\mathscr{T}_f$ be the subset of $\mathscr{T}$ containing the vectors $\vec{\tau}$ with the most *significant* posterior probabilities. We expect that the size of $\mathscr{T}_f$ will be much smaller than the size of $\mathscr{T}$ and thus, the $\{p(\vec{\tau}'|\vec{g})\}_{\vec{\tau}'\in\mathscr{T}_f}$ can be estimated from $\{p(\vec{g}|\vec{\tau}')p(\vec{\tau}')\}_{\vec{\tau}'\in\mathscr{T}_f}$.

The basis selection metric, for the MvC prior model, which is used to decide whether to include a given $\vec{\tau}$ in $\mathscr{T}_f$, or not, is defined as follows:

$$\begin{aligned} \rho(\vec{\tau},\vec{g}) &= \ln[p(\vec{g}|\vec{\tau})p(\vec{\tau})] = \ln[p(\vec{g}|\vec{\tau})] + \ln[p(\vec{\tau})] \\ &= \ln\Big[\frac{\Gamma\big((M+1)/2\big)}{\pi^{(M+1)/2}}\Big] - \frac{1}{2}\ln\big[|\mathbf{R}(\vec{\tau})|\big] \\ &- \frac{M+1}{2}\ln\big[1 + (\vec{g}-\vec{\mu}_g)^T\mathbf{R}(\vec{\tau})^{-1}(\vec{g}-\vec{\mu}_g)\big] + \sum_{i=0}^{N-1}\ln[\lambda_{\tau_i}] \ , \end{aligned} \tag{5}$$

where

$$\sum_{i=0}^{N-1}\ln[\lambda_{\tau_i}] = \|\vec{\tau}\|_0 \ln\big[\lambda_1/\lambda_0\big] + N\ln[\lambda_0] \ ,$$

and $\|\vec{\tau}\|_0$ is equal to the number of non-zero ("activated") components of $\vec{\tau}$.

### 3.1 MMSE and MAP estimate of $\vec{w}$

A computationally feasible approximation of the MMSE estimate of $\vec{w}$, using only the most significant posterior probabilities, is given by:

$$\hat{\vec{w}}_{\text{MMSE}} \triangleq \sum_{\vec{\tau}\in\mathscr{T}_f} p(\vec{\tau}|\vec{g})\mathbb{E}\{\vec{w}|\vec{g},\vec{\tau}\} \ , \tag{6}$$

where for the approximation of $\mathbb{E}\{\vec{w}|\vec{g},\vec{\tau}\}$ we use the underlying Gaussian part of the 1-SG($\vec{\mu}_g, \mathbf{R}$) distribution, resulting in the expression:

$$\mathbb{E}\{\vec{w}|\vec{g},\vec{\tau}\} = \vec{\mu} + \mathbf{\Sigma}(\vec{\tau})\mathbf{\Phi}^T\mathbf{R}(\vec{\tau})^{-1}(\vec{g}-\vec{\mu}_g) \ . \tag{7}$$

On the other hand, the MAP basis configuration is given by $\vec{\tau}_{\text{MAP}} = \arg\max_{\vec{\tau}\in\mathscr{T}_f} p(\vec{\tau}|\vec{g})$, resulting in the following approximation of the MAP estimate of $\vec{w}$:

$$\hat{\vec{w}}_{\text{MAP}} \triangleq \mathbb{E}\{\vec{w}|\vec{g},\vec{\tau}_{\text{MAP}}\} \ . \tag{8}$$

### 3.2 Incremental basis selection via a tree-structure

In this section, we review an incremental tree-structured procedure for selecting the next significant basis configuration, presented in [13], adapted to the MvC prior model. The root of the tree consists of the zero vector $\vec{\tau} = \vec{0}$. At the first level, the set $\mathscr{T}_1$ is formed, which contains the $N$ binary vectors $\vec{\tau}$ generated by "activating" one mixture parameter at a time. Then, the values of $\rho(\vec{\tau},\vec{g})$ are computed for these mixture vectors, and those with the $K$ largest values are stored in $\mathscr{T}_{f,1}$. At the second level, for each element of $\mathscr{T}_{f,1}$, a second mixture parameter is "activated" in all possible locations, once at a time, resulting in $\sum_{k=1}^{K}(N-k)$ binary vectors, which form the set $\mathscr{T}_2$. As before, the values of $\rho(\vec{\tau},\vec{g})$ are computed for these mixture vectors, and those with the $K$ largest values are stored in $\mathscr{T}_{f,2}$. The process is repeated until $\mathscr{T}_{f,l_{max}}$ is computed, where $l_{max}$ is the maximum number of tree levels, chosen such that $\Pr(\|\vec{\tau}\|_0 > l_{max})$ is sufficiently small.

When moving from one level of the tree to the next, the values of the metric $\rho(\cdot)$ must be updated. In particular, the change results from the activation of a single mixture parameter at a time. For this purpose, let $\vec{\tau}_q$ denote the mixture vector which is identical to $\vec{\tau}$, except for the $q$-th component, which is "activated" in $\vec{\tau}_q$, while it is "inactive" in $\vec{\tau}$. We are interested in computing the differences $\Delta_q(\vec{\tau}) = \rho(\vec{\tau}_q,\vec{g}) - \rho(\vec{\tau},\vec{g})$ which are then used to decide which mixture components will be activated. More specifically, the set $\mathscr{T}_{f,l}$, at the $l$-th tree-level is formed by keeping the $K$ binary vectors of the set $\mathscr{T}_l$ that correspond to the $K$ largest values of $\Delta_q(\vec{\tau})$, that is, we maintain only these vectors which achieve the highest increase of the basis selection metric (5). From (5), we can see that the key quantities to be updated are the inverse of $\mathbf{R}(\vec{\tau})$ and its determinant. The update of $\mathbf{R}(\vec{\tau})$ when the $q$-th component is activated, is given by:

$$\mathbf{R}(\vec{\tau}_q) = \mathbf{R}(\vec{\tau}) + \sigma_{\tau_q}^2 \vec{\phi}_q \vec{\phi}_q^T \ , \tag{9}$$

and thus, the matrix inversion lemma results in a simple expression for updating the inverse of $\mathbf{R}(\vec{\tau}_q)$:

$$\mathbf{R}(\vec{\tau}_q)^{-1} = \mathbf{R}(\vec{\tau})^{-1} - \gamma_q \vec{v}_q \vec{v}_q^T \ , \tag{10}$$

where $\gamma_q = \sigma_{\tau_q}^2 (1 + \sigma_{\tau_q}^2 \vec{\phi}_q^T \vec{v}_q)^{-1}$, and $\vec{v}_q = \mathbf{R}(\vec{\tau})^{-1}\vec{\phi}_q$. Besides, from (9) the determinant of $\mathbf{R}(\vec{\tau})$ can be easily updated as follows:

$$|\mathbf{R}(\vec{\tau}_q)| = (1 + \sigma_{\tau_q}^2 \vec{\phi}_q^T \mathbf{R}(\vec{\tau})^{-1} \vec{\phi}_q)|\mathbf{R}(\vec{\tau})| = \frac{\sigma_{\tau_q}^2}{\gamma_q}|\mathbf{R}(\vec{\tau})| \ . \tag{11}$$

Notice also that the updated mean vector $\vec{\mu}(\vec{\tau}_q)$ (and consequently, $\vec{\mu}_g(\vec{\tau}_q)$) is the same as $\vec{\mu}(\vec{\tau})$ ($\vec{\mu}_g(\vec{\tau})$) except for a change of its $q$-th component from $\mu_q = \mu_0 = 0$ to $\mu_q = \mu_1$. Finally, the probability of $\vec{\tau}$ is updated as: $p(\vec{\tau}_q) = \frac{\lambda_1}{\lambda_0} p(\vec{\tau})$. The substitution of this update equations in (5), and after some manipulation, results in the following expression for $\Delta_q(\vec{\tau})$ corresponding to the MvC prior model:

$$\Delta_q(\vec{\tau}) = \frac{1}{2}\ln\Big[\frac{\gamma_q}{\sigma_{\tau_q}^2}\Big] + \ln\Big[\frac{\lambda_1}{\lambda_0}\Big] - \frac{M+1}{2} \cdot$$
$$\cdot\Big(\ln\Big[1 - \frac{\gamma_q |\vec{\zeta}_g(\vec{\tau})^T \vec{v}_q + (\mu_1/\sigma_{\tau_q}^2)|^2 - (\mu_1^2/\sigma_{\tau_q}^2)}{1 + \vec{\zeta}_g(\vec{\tau})^T \mathbf{R}(\vec{\tau})^{-1} \vec{\zeta}_g(\vec{\tau})}\Big]\Big) \tag{12}$$

where $\vec{\zeta}_g(\vec{\tau}) = \vec{g} - \vec{\mu}_g(\vec{\tau})$. Due to our assumption that in the present study the mixture vectors consist of two components, the variances $\sigma_{\tau_q}^2$ in (12) can be substituted by $\sigma_1^2$.

## 4. EXPERIMENTAL RESULTS

In this section, we compare the performance of the proposed reconstruction scheme with the FBMP method proposed in [13][2]. For this purpose, we generate simulated measurement vectors $\vec{g}$ according to (2), where the sparse vectors $\vec{w}$ are drawn from an MvC distribution of length $N = 400$ that contain $L$ spikes, whose locations are chosen at random. We set the sparsity as a function of $\lambda_1$ and $N$, $L = \lceil \lambda_1 \cdot N \rceil$. In the subsequent experiments we choose $\lambda_1 = 0.02$, which results in a highly impulsive (and thus, heavy-tailed) vector $\vec{w}$. The measurement noise is generated by drawing samples from a zero-mean MvC distribution with underlying variance $\sigma_\eta^2$. The $M \times N$ measurement matrix $\mathbf{\Phi}$ is constructed by first drawing i.i.d. samples from a standard Gaussian distribution, and then normalizing its columns to unit magnitude.

The reconstruction performance is tested for two distinct Signal-to-Noise Ratio (SNR) values (SNR=10, 15dB), as well as for a range of measurements ($M \in \{90:1:120\}$). In particular, the noise variance $\sigma_\eta^2$ with the mixture variance $\sigma_1^2$ are related via the expression:

$$SNR = \frac{\sigma_1^2 \lambda_1 N}{\sigma_\eta^2 M} \ . \tag{13}$$

The process is repeated for 100 independent Monte-Carlo realizations, and the results are given by averaging over the 100 runs. The normalized mean-squared error of the MMSE estimated sparse vector, $\hat{\vec{w}}$, is given by

$$\mathrm{NMSE}_{\mathrm{MMSE}} = \frac{1}{100}\sum_{j=1}^{100} \frac{\|\hat{\vec{w}}_{\mathrm{MMSE,j}} - \vec{w}\|_2^2}{\|\vec{w}\|_2^2} \ ,$$

where $\hat{\vec{w}}_{\mathrm{MMSE,j}}$ is the MMSE estimate of $\vec{w}$, given by (6), at the $j$-th Monte-Carlo run. Similarly, the normalized mean-squared error of the MAP estimated sparse vector is given by

$$\mathrm{NMSE}_{\mathrm{MAP}} = \frac{1}{100}\sum_{j=1}^{100} \frac{\|\hat{\vec{w}}_{\mathrm{MAP,j}} - \vec{w}\|_2^2}{\|\vec{w}\|_2^2} \ ,$$

where $\hat{\vec{w}}_{\mathrm{MAP,j}}$ is the MAP estimate of $\vec{w}$, given by (8), at the $j$-th Monte-Carlo run.

Fig. 1 shows the MMSE and MAP reconstruction errors averaged over the 100 runs for SNR=10, 15dB. It is clear that the proposed algorithm, based on a heavy-tailed distribution, achieves a better reconstruction performance, when compared with the FBMP method, which is based on a normality assumption with respect to the prior distributions. For both methods, the corresponding MMSE and MAP estimates are close to each other. Also, the SNR level affects more the FBMP method. Besides, it seems that the number of measurements does not affect the performance in the present highly impulsive scenario.

Fig. 2 shows the average number of non-zero taps contained in the most significant basis configurations for both methods. Our proposed method results in a decreased number of activated taps. On the other hand, Fig. 3 shows the average number of vectors contained in the set $\mathscr{T}_{f,l_{max}}$, consisting of the most significant basis configurations (mixture vectors). Our method exploits a smaller number of such configurations, while for a smaller SNR both methods require more mixture vectors to capture the impulsive behavior. In order to make the comparison of the sparsity performance of the two methods more meaningful, we define the following sparsity ratio

$$\mathrm{SpR} = \frac{(\#\text{ non-zero taps}) \times (\#\text{ significant mixture vectors})}{M} \ . \tag{14}$$

The lower the value of SpR, for a fixed $M$, the sparser the solution of the corresponding method.

Fig. 4 shows the SpR ratio for the two methods and for the two SNR values. As it can be seen, for the same number of CS measurements, the SpR ratio of the proposed method is much smaller than

[2] We used the FBMP package downloaded from `http://www.ece.osu.edu/~zinielj/fbmp`, using the standard implementation without parameter re-estimation.

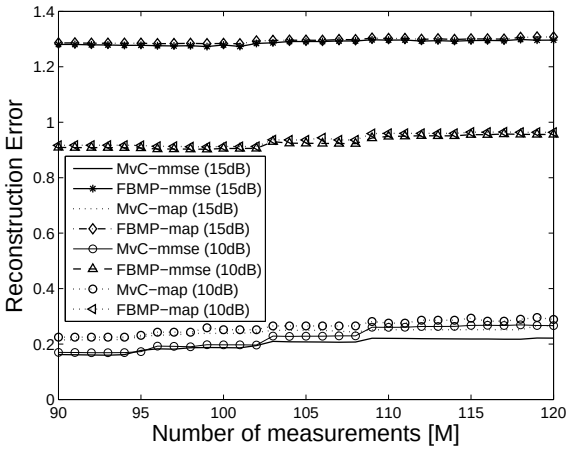


Figure 1: Average MMSE and MAP reconstruction errors, of the FBMP and MvC methods, as a function of M, for SNR=10, 15dB.

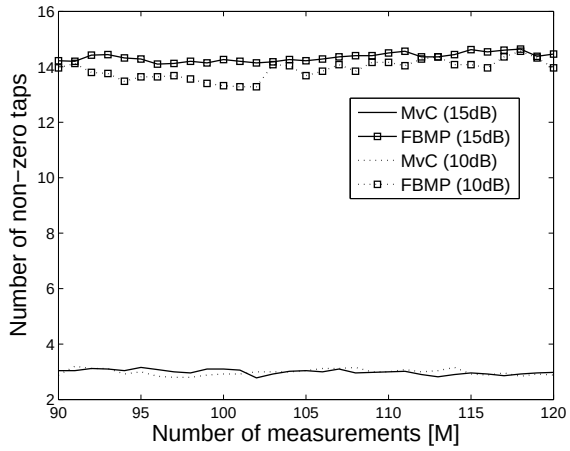


Figure 2: Average number of non-zero taps, as a function of M, of the FBMP and MvC methods, for SNR=10, 15dB.

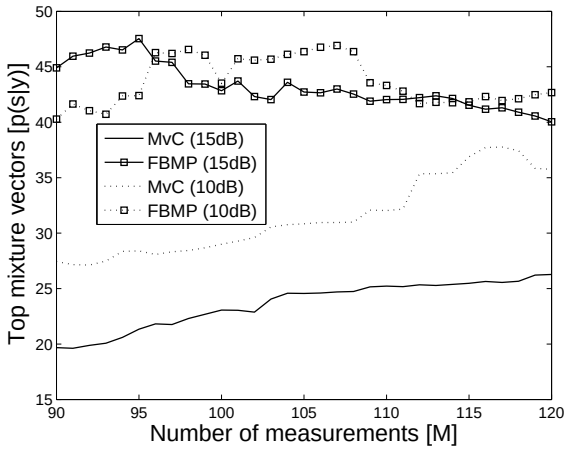


Figure 3: Average number of significant mixture vectors, as a function of M, of the FBMP and MvC methods, for SNR=10, 15dB.

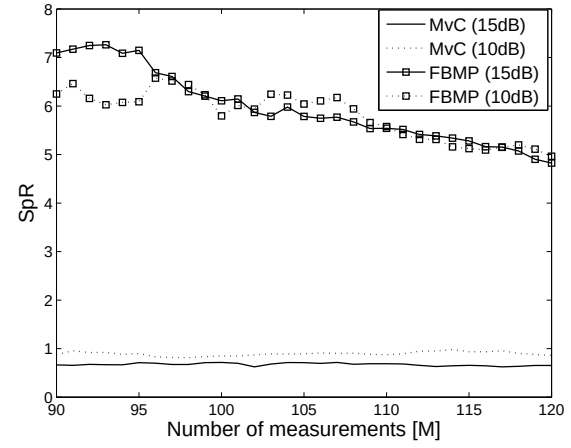


Figure 4: SpR ratio, as a function of M, of the FBMP and MvC methods, for SNR=10, 15dB.

the SpR ratio of the FBMP approach, which means that our proposed method results in a sparser solution. The fact that the value of SpR corresponding to the proposed method is almost constant over the whole range of $M$, may be due to the fact that we do not re-estimate the mixture parameters $(\mu_i, \sigma_i^2)$ during the reconstruction process, as it will be mentioned in section 5. This may affect the sensitivity of the proposed algorithm.

As it was mentioned before, the degree of impulsiveness of the "signal" under consideration is controlled by the value of the characteristic exponent $\alpha$ of the sub-Gaussian $S\alpha S$ distribution. Although in our proposed model we consider the multivariate Cauchy case, with $\alpha = 1$, we are interested in studying the reconstruction performance for other values of $\alpha$ as well, corresponding to different levels of sparseness.

For this purpose, we curry out a second set of Monte-Carlo runs by generating simulated measurement vectors $\vec{g}$ according to (2), where the sparse vectors $\vec{w}$ are drawn from a sub-Gaussian distribution of length $N = 300$ that contain $L$ spikes, whose locations are chosen at random. As before, the sparsity is set as a function of $\lambda_1$ and $N$, $L = \lceil \lambda_1 \cdot N \rceil$, with $\lambda_1 = 0.03$. The measurement noise is generated by drawing samples from a zero-mean sub-Gaussian distribution with underlying standard deviation $\sigma_\eta$. We fix the number of measurements to $M = 90$, and the $M \times N$ measurement matrix $\mathbf{\Phi}$ is constructed by first drawing i.i.d. samples from a standard Gaussian distribution, and then normalizing its columns to unit magnitude.

The reconstruction performance is tested by varying the characteristic exponent, $\alpha$, in the interval $[1,2]$ with a step size of 0.1, and for two SNR values, SNR=8, 10dB. The noise underlying variance $\sigma_\eta^2$ is set in accordance to the mixture variance $\sigma_1^2$ via (13). We also set the values of the mixing parameters $\mu_1$ and $\sigma_1^2$ to 3 and 5, respectively. The process is repeated for 100 independent Monte-Carlo realizations, and the results are given by averaging over the 100 runs.

Fig. 5 shows the $\text{NMSE}_{\text{MMSE}}$ and $\text{NMSE}_{\text{MAP}}$ reconstruction errors, as a function of $\alpha$, corresponding to the FBMP and MvC methods, for the two SNR values. First, we observe that the performance of both methods is improved as the SNR increases. Besides, the MvC approach achieves a smaller reconstruction error in comparison to the FBMP approach, for values of $\alpha$ close to 1, that is, when the actual distribution of the signal is heavy-tailed, while its performance is comparable to the performance of the FBMP as $\alpha$ tends to 2 (that is, to a Gaussian prior model). Thus, when the original signal is highly sparse, the MvC approach should be preferred.

Fig. 6 shows the SpR ratio, as a function of $\alpha$, for the FBMP and the MvC methods and for the two SNR values. As before, for the values of $\alpha$ which are in the vicinity of 1, the SpR ratio

of the proposed method is much smaller than the SpR ratio of the FBMP approach, which means that our proposed method results in a much sparser solution when working in a highly impulsive environment. On the other hand, as the value of $\alpha$ approaches 2 (Gaussian statistics), the SpR ratio of both methods decreases, with the FBMP method, which is based on a Gaussian prior, resulting in a slightly sparser solution.

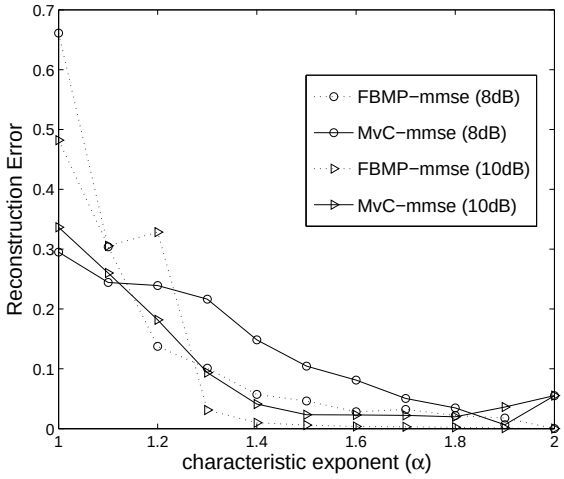


Figure 5: Average MMSE and MAP reconstruction errors, of the FBMP and MvC methods, as a function of $\alpha$, for SNR=8, 10dB.

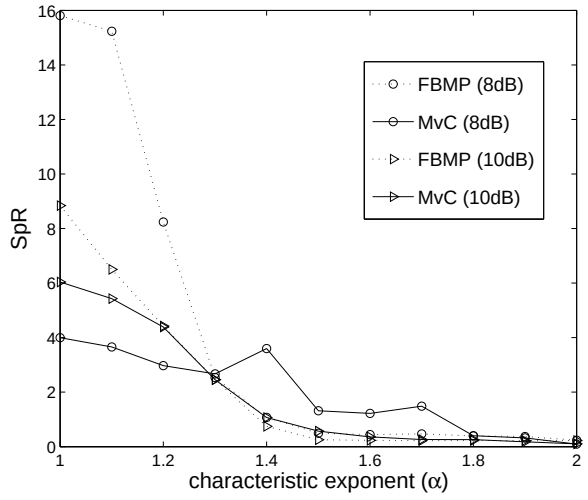


Figure 6: SpR ratio, as a function of $\alpha$, of the FBMP and MvC methods, for SNR=8, 10dB.

## 5. CONCLUSIONS AND FUTURE WORK

In this work, we described a method for CS reconstruction of a highly impulsive vector in heavy-tailed noise, based on a Bayesian framework. We employed a multivariate Cauchy (MvC) distribution as the prior model, and thus, modeling directly the vector $\vec{w}$ with a heavy-tailed distribution that enforces its sparsity. The experimental results revealed an improved performance of the proposed approach when compared with the previous FBMP method. In particular, we showed that the MvC-based implementation achieves a smaller reconstruction error than the FBMP approach when the observed signal is truly sparse (that is, $\alpha \rightarrow 1$), while maintaining a quite low value of the SpR ratio, which is equivalent to an increased sparsity.

In the present work, we made the simplified assumption that the components of a mixture vector $\vec{\tau}$ are chosen from two distributions ("inactive", "active"). Besides, the parameters of these distributions are predetermined and kept fixed during the reconstruction process. As a future work, we are interested in modifying the proposed model so as to permit each mixture component to be chosen from a larger set of candidate mixture distributions. We will also introduce a technique for re-estimating their corresponding parameters $(\mu_i, \sigma_i^2)$ during the reconstruction process.

## REFERENCES


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

[2] I. Daubechies, "Ten lectures on wavelets", SIAM, 1992.

[3] D. S. Taubman, and M. W. Marcellin, "JPEG 2000: Image Compression Fundamentals, Standards and Practice", (Int. Series in Engineering and Computer Science), Norwell, MA: Kluwer, 2002.

[4] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information", *IEEE Trans. Inform. Theory*, Vol. 52, No. 2, pp. 489–509, Feb. 2006.

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

[6] Y. Tsaig, and D. L. Donoho, "Extensions of compressed sensing", *Signal Proc.*, Vol. 86, No. 3, pp. 549-571, Mar. 2006.

[7] J. Haupt, and R. Nowak, "Signal reconstruction from noisy random projections", *IEEE Trans. Inform. Theory*, Vol. 52, No. 9, pp. 4036–4048, Sept. 2006.

[8] S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by Basis Pursuit", *SIAM Journal on Scientific Computing*, Vol. 20, No. 1, pp. 33-61, 1999.

[9] J. Tropp, and A. C. Gilbert, "Signal recovery from random measurements via Orthogonal Matching Pursuit", *IEEE Trans. Information Theory*, Vol. 53, pp. 4655-4666, 2007.

[10] A. Carmi, P. Gurfil, and D. Kanevsky, "A simple method for sparse signal recovery from noisy observations using Kalman filtering", *IBM Res. Report* [RC24709], Dec. 2008.

[11] M. E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine", *Journal of Machine Learning Research*, Vol. 1, pp. 211-244, 2001.

[12] Shihao Ji, Ya Xue, and Lawrence Carin, "Bayesian Compressive Sensing", *IEEE Trans. on Signal Proc.*, Vol. 56, No. 6, pp. 2346–2356, June 2008.

[13] P. Schniter, Lee C. Potter, and J. Ziniel, "Fast Bayesian Matching Pursuit", *Proc. Workshop on Information Theory & Applications*, La Jolla, CA, Jan. 2008.

[14] P. Tsakalides and C. Nikias, "Maximum likelihood localization of sources in noise modeled as a Cauchy process," *Proc. IEEE Milit. Comm. Conf. (MILCOM 1994)*, Fort Monmouth, NJ, Vol. 2, pp. 613–617, Oct. 1994.

[15] R. Raspanti, P. Tsakalides, C. Nikias, and E. Del Re, "Cramer-Rao bounds for target angle and doppler estimation for airborne radar in Cauchy interference," *8th IEEE Sig. Proc. Workshop on Stat. Sig. and Array Proc.*, Corfu, Greece, June 24–26, 1996.

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

[17] J. P. Nolan, "Parameterizations and modes of stable distributions", *Statistics & Probability Letters*, No. 38, pp. 187–195, 1998.

[18] D. Donoho, Y. Tsaig, I. Drori, and J. -L. Starck, "Sparse solution of underdetermined linear equations by Stagewise Orthogonal Matching Pursuit", Tech. Report 06-02, Dept. of Statistics, Stanford Univ., 2006.