NONPARAMETRIC SIMULTANEOUS SPARSE RECOVERY: AN APPLICATION TO SOURCE LOCALIZATION

Esa Ollila

Aalto University, Dept. of Signal Processing and Acoustics, P.O.Box 13000, FI-00076 Aalto, Finland

ABSTRACT

We consider multichannel sparse recovery problem where the objective is to find good recovery of jointly sparse unknown signal vectors from the given multiple measurement vectors which are different linear combinations of the same known elementary vectors. Many popular greedy or convex algorithms perform poorly under non-Gaussian heavy-tailed noise conditions or in the face of outliers. In this paper, we propose the usage of mixed $\ell_{p,q}$ norms on data fidelity (residual matrix) term and the conventional $\ell_{0,2}$-norm constraint on the signal matrix to promote row-sparsity. We devise a greedy pursuit algorithm based on simultaneous normalized iterative hard thresholding (SNIHT) algorithm. Simulation studies highlight the effectiveness of the proposed approaches to cope with different noise environments (i.i.d., row i.i.d, etc) and outliers. Usefulness of the methods are illustrated in source localization application with sensor arrays.

Index Terms— multichannel sparse recovery, compressed sensing, robustness, iterative hard thresholding

1. INTRODUCTION

In the *multiple measurement vector (MMV) model*, a single measurement matrix is utilized to obtain multiple measurement vectors, i.e., $\mathbf{y}_i = \boldsymbol{\Phi}\mathbf{x}_i + \mathbf{e}_i$, $i = 1, \ldots, Q$ where $\boldsymbol{\Phi}$ is the $M \times N$ known *measurement matrix* and $\mathbf{e}_i$ are the (unobserved) random noise vectors. Typically there are more column vectors $\boldsymbol{\phi}_i$ than row vectors $\boldsymbol{\phi}_{(j)}$, i.e., $M < N$ (underdetermined linear model). It is still possible to recover the unknown signal vectors $\mathbf{x}_i$, $i = 1, \ldots, Q$ by assuming that signals are sparse, i.e., some of the elements are zero. In matrix form, the MMV model reads $\mathbf{Y} = \boldsymbol{\Phi}\mathbf{X} + \mathbf{E}$, where $\mathbf{Y} = (\mathbf{y}_1 \cdots \mathbf{y}_Q) \in \mathbb{C}^{M\times Q}$, $\mathbf{X} = (\mathbf{x}_1 \cdots \mathbf{x}_Q) \in \mathbb{C}^{N\times Q}$ and $\mathbf{E} = (\mathbf{e}_1 \cdots \mathbf{e}_Q) \in \mathbb{C}^{M\times Q}$ collect the measurement, the signal and the error vectors, respectively. When $Q = 1$, the model reduces to standard *compressed sensing (CS) model* [1]. Then, rather than recovering the sparse/compressible target signals $\mathbf{x}_i$ separately using standard CS reconstruction algorithms, one attempts to simultaneously (jointly) recover all signals. The key assumption is that locations of nonzero values primarily coincide, i.e., signal matrix $\mathbf{X}$ is K rowsparse. Joint estimation can lead both to computational advantages and increased reconstruction accuracy [1–6]. The objective of multichannel sparse recovery problem is finding a row sparse approximation of the signal matrix $\mathbf{X}$ based on knowledge of $\mathbf{Y}$, the measurement matrix $\boldsymbol{\Phi}$ and the sparsity level K. Applications include EEG/MEG [1] and direction-of-arrival (DOA) estimation of sources in array processing [7].

Most greedy CS reconstruction algorithms have been extended for solving MMV problems. These methods, such as simultaneous normalized iterative hard thresholding (SNIHT) algorithm [6] are guaranteed to perform very well provided that suitable conditions (e.g., incoherence of $\boldsymbol{\Phi}$ and non impulsive noise conditions) are met. The derived (worst case) recovery bounds depend linearly on $\|\mathbf{E}\|_2$, so the methods are not guaranteed to provide accurate reconstruction/approximation under heavy-tailed non-Gaussian noise. In this paper, we consider different $\ell_{p,q}$ mixed norms on data fidelity (residual matrix) and devise a greedy SNIHT algorithm for obtaining a sparse solution. We focus on mixed ℓ_1 norms as they can provide robust solutions. As will be shown in the sequel, these methods are then based on *spatial signs* [8] of the residuals and therefore are nonparametric in nature. For an alternative robust approach, see [9].

The paper is organized as follows. In Section 2 we formulate a mixed-norm constrained objective function for the MMV problem and motivate the usage of ℓ_1-norm or the mixed $\ell_{2,1}$- and $\ell_{1,2}$-norms. In Section 3 we formulate the greedy SNIHT algorithm whereas Section 4 provides simulation examples illustrating the improved accuracy of the proposed methods in various noise conditions and signal to noise ratio (SNR) settings. Finally, effectiveness of the methods are illustrated in source localization application with sensor arrays in Section 5.

Notations. Let $[n]$ denote the set $\{1, \ldots, n\}$ for $n \in \mathbb{N}^+$. For a matrix $\mathbf{A} \in \mathbb{C}^{M\times N}$ and an index set Γ of cardinality $|\Gamma| = K$, we denote by $\mathbf{A}_\Gamma$ (resp. $\mathbf{A}_{(\Gamma)}$) the $M \times K$ (resp. $K \times N$) matrix restricted to the columns (resp. rows) of $\mathbf{A}$ indexed by the set Γ. The ith column vector of $\mathbf{A}$ is denoted by $\mathbf{a}_i$ and the hermitian transpose of the ith row vector of $\mathbf{A}$ by $\mathbf{a}_{(i)}$, $\mathbf{A} = (\mathbf{a}_1 \cdots \mathbf{a}_N) = (\mathbf{a}_{(1)} \cdots \mathbf{a}_{(M)})^{\mathrm{H}}$. The *row-support* of $\mathbf{X} \in \mathbb{C}^{N\times Q}$ is the index set of rows containing non-zero elements: $\mathrm{rsupp}(\mathbf{X}) = \{i \in [N] : x_{ij} \neq 0 \text{ for some } j\}$. For $p, q \in [1, \infty)$, the *mixed $\ell_{p,q}$ norm* [10] of

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

$\mathbf{X} \in \mathbb{C}^{N\times Q}$ is defined as

$$\|\mathbf{X}\|_{p,q} = \left(\sum_i \left(\sum_j |x_{ij}|^p\right)^{q/p}\right)^{1/q} = \left(\sum_i \|\mathbf{x}_{(i)}\|_p^q\right)^{1/q}.$$

The mixed norms generalize the usual matrix p-norms: if $p = q$, then $\|\mathbf{X}\|_{p,p} = \|\mathbf{X}\|_p$. The ℓ_2-norm $\|\cdot\|_2$ is called the Frobenius norm and will be denoted shortly as $\|\cdot\|$. In the same spirit, the usual Euclidean norm on vectors is denoted shortly as $\|\cdot\|$. The row-ℓ_0 quasi-norm of a signal matrix $\mathbf{X}$ is the number of nonzero rows, i.e., $\|\mathbf{X}\|_0 = |\operatorname{rsupp}(\mathbf{X})|$. The matrix $\mathbf{X}$ is then said to be K-*rowsparse* if $\|\mathbf{X}\|_0 \leq K$. We use $H_K(\cdot)$ to denote the *hard thresholding operator*: for a matrix $\mathbf{X} \in \mathbb{C}^{N\times Q}$, $H_K(\mathbf{X})$ retains the elements of the K rows of $\mathbf{X}$ that possess largest ℓ_2-norms and set elements of the other rows to zero. Notation $\mathbf{X}|_\Gamma$ refers to a sparsified version of $\mathbf{X}$ such that the entries in the rows indexed by set Γ remain unchanged while all other rows have all entries set to 0.

2. ROBUST MIXED NORM MINIMIZATION

Our objective is to recover K-rowsparse $\mathbf{X}$ in the MMV model. For this purpose, we consider the following constrained optimization problem:

$$\min_{\mathbf{X}} c_{p,q}\|\mathbf{Y} - \mathbf{\Phi X}\|_{p,q}^q \quad \text{subject to} \quad \|\mathbf{X}\|_0 \leq K, \quad (P_{p,q})$$

where $c_{p,q}$ is an irrelevant constant used for making notations compact. For $p = q$, the problem reduces to conventional ℓ_p-norm minimization of the *residual matrix* $\mathbf{R} = \mathbf{Y} - \mathbf{\Phi X} \in \mathbb{C}^{M\times Q}$ under rowsparsity constraint on $\mathbf{X}$. The well-known problem with ℓ_2-norm minimization is that it gives a very small weight on small residuals and a strong weight on large residuals, implying that even a single large outlier can have a large influence on the obtained solution. For robustness, one should utilize ℓ_1 in mixed norms since it gives larger weights on small residuals and less weight on large residuals. In this paper we consider $(P_{p,q})$ in the cases that $p, q \in \{1, 2\}$. The problem $(P_{p,q})$ is combinatorial (NP-hard). Hence suboptimal reduced complexity reconstruction algorithms have been proposed. These can be roughly divided into two classes: convex-relaxation algorithms (e.g., [3,7,10]) and greedy pursuit (e.g., [2,6]) algorithms. In this paper, we devise a greedy simultaneous NIHT (SNIHT) algorithm for the problems $(P_{1,1})$ and $(P_{2,1})$. The case $(P_{1,2})$ is excluded due to the lack of space, but our approach and discussion straightforwardly extends for this mixed ℓ_1 norm as well.

In $(P_{1,1})$ problem, one aims to minimize $\|\mathbf{Y} - \mathbf{\Phi X}\|_1 = \sum_i \sum_j |y_{ij} - \boldsymbol{\phi}_{(i)}^{\mathrm{H}}\mathbf{x}_j|$ under sparsity constraint, so the solution can be viewed as a sparse multivariate least absolute deviation (LAD) regression estimator. The LAD regression (in the real-valued overdetermined linear regression) is well-known to offer robust solution with bounded influence function. In the complex case, this approach can be considered optimal when the error terms e_{ij} are i.i.d. with (circular) complex generalized Gaussian (GG) distribution [11, Example 4] with exponent $s = 1/2$. It is important to realize that minimization of ℓ_p-norms in $(P_{p,p})$ implicitly assumes i.i.d.'ness of the error terms. Since the measurement matrix $\mathbf{Y}$ is in many applications a *space* $\times$ *time* matrix as in medical imaging or sensor array applications, the i.i.d. assumption of the error terms in time/space is often not valid. The benefit of mixed ℓ_1-norms, such as $\ell_{2,1}$ and $\ell_{1,2}$ considered here is that they introduce couplings [10] between the coefficients and offer robustness in case of dependent heavy-tailed errors or outliers. When the errors terms have dependencies in time and/or space, then $\ell_{2,1}$ and $\ell_{1,2}$ minimization can offer advantages over ℓ_1 or ℓ_2 norm approaches. As will be shown later, the usage of ℓ_1-norm or the mixed ℓ_1-norms lead to *non-parametric* approaches that are based on the concept of *spatial sign function* [8] which in the scalar case ($x \in \mathbb{C}$) is defined as

$$\operatorname{sign}(x) = \begin{cases} x/|x|, & \text{for } x \neq 0 \\ 0, & \text{for } x = 0 \end{cases}. \quad (1)$$

In the vector case, $\mathbf{sign}(\mathbf{x}) = \|\mathbf{x}\|^{-1}\mathbf{x}, = 0$ for $\mathbf{x} \neq \mathbf{0}, = \mathbf{0}$.

3. MIXED NORM SNIHT ALGORITHM

Iterative hard thresholding is a *projected gradient descent* method that is known to offer efficient and scalable solution for K-sparse approximation problem [12]. The normalized IHT (NIHT) method updates the estimate of $\mathbf{X}$ by taking steps towards the direction of the negative gradient followed by projection onto the constrained space. In our multichannel sparse recovery problem, at $(n+1)$th iteration the SNIHT update is

$$\mathbf{X}^{n+1} = H_K\big(\mathbf{X}^n + \mu^{n+1}\mathbf{\Phi}^{\mathrm{H}}\psi_{p,q}(\mathbf{Y} - \mathbf{\Phi X}^n)\big)$$

where $\psi_{p,q}(\mathbf{R}) = \nabla_{\mathbf{R}^*}\|\mathbf{R}\|_{p,q}^q$ is the complex matrix derivative [13] with respect to (w.r.t.) $\mathbf{R}^*$, $\mu^{n+1} > 0$ is the stepsize for the current iteration and $p, q \in \{1, 2\}$. For ℓ_2- and ℓ_1-norms the derivatives are easily shown to be

$$\psi_{2,2}(\mathbf{R}) = \mathbf{R} \quad \text{and} \quad \psi_{1,1}(\mathbf{R}) = \operatorname{sign}(\mathbf{R})$$

respectively, where notation $\operatorname{sign}(\mathbf{R})$ refers to element-wise application of the spatial sign function (1), i.e., $[\operatorname{sign}(\mathbf{R})]_{ij} = \operatorname{sign}(r_{ij})$. For $(2, 1)$ mixed norm, we obtain

$$\psi_{2,1}(\mathbf{R}) = \big(\mathbf{sign}(\mathbf{r}_{(1)}) \quad \cdots \quad \mathbf{sign}(\mathbf{r}_{(M)})\big)^{\mathrm{H}},$$

that is, the vector spatial sign function is applied row-wise to the residual matrix $\mathbf{R} = (\mathbf{r}_{(1)} \ \cdots \ \mathbf{r}_{(M)})^{\mathrm{H}}$. Table 1 provides the pseudo-code of the greedy SNIHT algorithm for the problem $(P_{p,q})$, which we call SNIHT(p, q) algorithm for short. Note that SNIHT$(2, 2)$ corresponds to the conventional SNIHT studied in [6] and in [12] for $Q = 1$ case.

Algorithm 1: SNIHT(p,q) algorithm

input : $\mathbf{Y}$, $\mathbf{\Phi}$, sparsity K, mixed norm indices (p,q)
output : $(\mathbf{X}^{n+1}, \Gamma^{n+1})$ estimates of $\mathbf{X}$ and rsupp$(\mathbf{X})$
initialize: $\mathbf{X}^0 = \mathbf{0}, \mu^0 = 0, \Gamma^0 = \emptyset, n = 0$.

1 $\Gamma^0 = \text{rsupp}\big(H_K(\mathbf{\Phi}^{\text{H}}\psi_{p,q}(\mathbf{Y}))\big)$
while *halting criterion false* **do**
2 $\quad \mathbf{R}_\psi^n = \psi_{p,q}(\mathbf{Y} - \mathbf{\Phi}\mathbf{X}^n)$
3 $\quad \mathbf{G}^n = \mathbf{\Phi}^{\text{H}}\mathbf{R}_\psi^n$
4 $\quad \mu^{n+1} = \texttt{CompStepsize}(\Phi, \mathbf{G}^n, \Gamma^n, \mu^n, p, q)$
5 $\quad \mathbf{X}^{n+1} = H_K(\mathbf{X}^n + \mu^{n+1}\mathbf{G}^n)$
6 $\quad \Gamma^{n+1} = \text{rsupp}(\mathbf{X}^{n+1})$
7 $\quad n = n + 1$
end

We now describe the `CompStepsize` function which computes the stepsize update μ^{n+1} in Step 4. Following the approach in [12], assuming that we have identified the correct support at nth iteration, then we may look for a stepsize update μ^{n+1} as the minimizer of $\|\mathbf{Y} - \mathbf{\Phi}\mathbf{X}\|_{p,q}^q$ for the gradient ascent direction $\mathbf{X}^n + \mu\mathbf{G}^n|_{\Gamma^n}$. Thus we find $\mu > 0$ as the minimizer of the convex function

$$\big\|\mathbf{Y} - \mathbf{\Phi}\big(\mathbf{X}^n + \mu\mathbf{G}^n|_{\Gamma^n}\big)\big\|_{p,q}^q = \|\mathbf{R}^n - \mu\mathbf{B}^n\|_{p,q}^q \qquad (2)$$

where $\mathbf{R}^n = \mathbf{Y} - \mathbf{\Phi}\mathbf{X}^n$ and $\mathbf{B}^n = \mathbf{\Phi}_{\Gamma^n}\mathbf{G}_{(\Gamma^n)}$ When $p = q$ this reduces to minimizing a simple linear regression estimation problem, $\min_\mu \|\mathbf{r} - \mu\mathbf{b}\|_p^p$, where the response is $\mathbf{r} = \text{vec}(\mathbf{R}^n)$ and the predictor is $\mathbf{b} = \text{vec}(\mathbf{B}^n)$. Thus when using $p = q = 2$ as in conventional SNIHT [6], the minimizer of (2) is easily found to be $\mu^{n+1} = \|\mathbf{G}_{(\Gamma^n)}^n\|^2 / \|\mathbf{\Phi}_{\Gamma^n}\mathbf{G}_{(\Gamma^n)}^n\|^2$. However, for the robust estimators that we are interested in, i.e., when using $(p,q) = (1,1)$ and $(p,q) = (2,1)$, a minimizer of (2) can not be found in closed-form. In the $(p,q) = (1,1)$ case, it is easy to show that the solution μ verfies the following fixed point (FP) equation $\mu = H(\mu)$, where

$$H(\mu) = \Big(\sum_{i,j} |\tilde{r}_{ij}|^{-1}|b_{ij}|^2\Big)^{-1} \sum_{i,j} |\tilde{r}_{ij}|^{-1}\text{Re}(b_{ij}^* r_{ij}),$$

and $\tilde{\mathbf{R}} = \mathbf{R}^n - \mu\mathbf{B}^n = (\tilde{r}_{ij})$ depends also on the unknown μ. Then, instead of choosing the next update μ^{n+1} as the minimizer of (2) which could be found by running the FP iterations $\mu_i = H(\mu_{i+1})$ for $i = 0, 1, \ldots$ until convergence (with initial value $\mu_0 > 0$), we use a 1-step FP iterate which corresponds to a single iteration with initial value of iteration given by the previous stepsize μ^n. In other words, in Step 4, we set $\mu^{n+1} = H(\mu^n)$. In our simulation studies we noticed that this 1-step FP iterate often gave a very good approximation of the true solution (within 3 decimal accuracy). In case we use $(p,q) = (2,1)$, it is easy to show that the solution μ verifies the FP equation $\mu = H^\star(\mu)$, where

$$H^\star(\mu) = \Big(\sum_i \|\tilde{\mathbf{r}}_{(i)}\|^{-1}\|\mathbf{b}_{(i)}\|^2\Big)^{-1} \sum_i \|\tilde{\mathbf{r}}_{(i)}\|^{-1}\text{Re}(\mathbf{b}_{(i)}^{\text{H}}\mathbf{r}_{(i)})$$

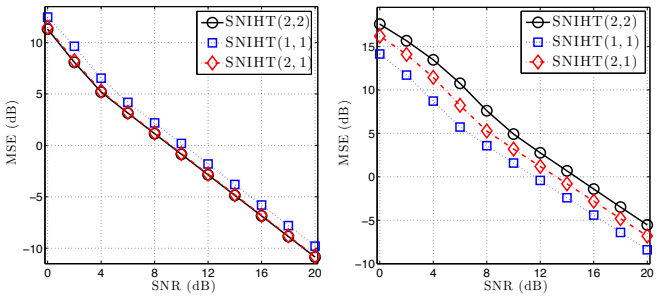


Fig. 1. Average MSE of SNIHT(p,q) methods as a function of SNR in (a) $\mathbb{C}\mathcal{N}(0,\sigma^2)$ noise and (b) $\mathbb{C}t_3(0,\sigma^2)$ noise.

and the same approach, i.e., $\mu^{n+1} = H^\star(\mu^n)$, is used for computing the stepsize update.

4. SIMULATION STUDIES

Next we illustrate the usefulness of the methods in a variety of noise environments and SNR levels. Also, the effect of number of measurement vectors Q on recovery probability will be illustrated. The elements of $\mathbf{\Phi}$ are drawn from $\mathbb{C}\mathcal{N}(0,1)$ distribution and the columns are unit-norm normalized. The coefficients of K non-zero row vectors of $\mathbf{X}$ have equal amplitudes $\sigma_x = |x_{ij}| = 1\ \forall i \in \Gamma, j = 1, \ldots, Q$ and uniform phases, i.e., $\text{Arg}(x_{ij}) \sim Unif(0, 2\pi)$. The support $\Gamma = \text{supp}(\mathbf{X})$ is randomly chosen from $\{1, \ldots, N\}$ without replacement for each trial. We define the (generalized) *signal to noise ratio (SNR)* as $\text{SNR}(\sigma) = 10\log_{10}(\sigma_x^2/\sigma^2) = -20\log_{10}\sigma$ which depends on the scale parameter σ of the error distribution. For robustness purposes, we will study the performance in i.i.d. complex circular t-distributed noise with ν degrees of freedom (d.o.f.), $e_{ij} \sim \mathbb{C}t_\nu(0,\sigma^2)$, when $\nu \leq 5$ and the scale parameter is $\sigma^2 = \text{Med}_{F_e}(|e_{ij}|^2)$. This is an example of a heavy-tailed distribution with $\nu = 1$ corresponding to Cauchy distribution. Also note that at the limit $\nu \to \infty$ one obtains the complex Gaussian distribution.

As performance measures of sparse signal recovery, we use both the (observed) *mean squared error* $\text{MSE}(\hat{\mathbf{X}}) = \frac{1}{LQ}\sum_{\ell=1}^L \big\|\hat{\mathbf{X}}^{[\ell]} - \mathbf{X}^{[\ell]}\big\|^2$ and the empirical *probability of exact recovery*, $\text{PER} \triangleq \frac{1}{L}\sum_{\ell=1}^L \text{I}\big(\hat{\Gamma}^{[\ell]} = \Gamma^{[\ell]}\big)$, where $\text{I}(\cdot)$ denotes the indicator function, $\hat{\mathbf{X}}^{[\ell]}$ and $\hat{\Gamma}^{[\ell]} = \text{rsupp}(\hat{\mathbf{X}}^{[\ell]})$ denote the estimate of the K-sparse signal $\mathbf{X}^{[\ell]}$ and the signal support $\Gamma^{[\ell]}$ for the ℓth Monte Carlo (MC) trial, respectively. In all simulation settings described below, all the reporeted figures are averages over $L = 2000$ MC trials, the length of the signal is $N = 512$, the number of measurements is $M = 256$, and the sparsity level is $K = 8$. The number of measurement vectors is $Q = 16$ unless otherwise stated.

Figure 1(a) depicts the MSE as a function of SNR in i.i.d. circular Gaussian noise, $e_{ij} \sim \mathbb{C}\mathcal{N}(0,\sigma^2)$, where

	SNR (dB)							
	2	4	6	8	10	12	14	16
SNIHT$(2,2)$	0	0	.6	.61	.94	.99	.99	1.0
SNIHT$(1,1)$	0	.25	.91	1.0	1.0	1.0	1.0	1.0
SNIHT$(2,1)$	0	.02	.38	.96	1.0	1.0	1.0	1.0

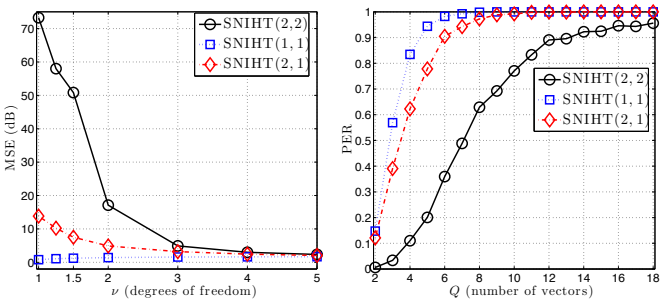
Table 1. PER rates in $\mathbb{C}t_3(0,\sigma)$ distributed noise as a function of SNR (dB). System parameters were $(M,N,K,Q) = (256,512,8,16)$.

$\sigma^2 = \mathbb{E}[|e_{ij}|^2]$. As expected, the conventional SNIHT$(2,2)$ has the best performance, but SNIHT$(2,1)$ suffers a negligible 0.07 dB loss, whereas SNIHT$(1,1)$ attain 1.07 dB performance loss. Note that SNR $= 6$ dB is the cutline for which all methods had full PER rate $(= 1)$. From 4 dB the PER rate declines and reaches 0 at SNR $= 0$ dB for all of the methods.

Next we study the performance in t-distributed noise with $\nu = 3$ d.o.f. Note that $\mathbb{C}t_3(0,\sigma)$ distribution has a finite variance so we can expect that also SNIHT$(2,2)$ can still work reliably in this setting. Figure 1(b) which depict the MSE vs SNR illustrates severe degradation in reconstruction performance for the SNIHT$(2,2)$. This is further illustrated in Table 1 which provides the PER rates for the considered SNIHT(p,q) methods. Note that the decline of PER rate starts much earlier for the conventional SNIHT than for the robust methods.

Figure 2(a) depicts the MSE of the methods in t-distributed noise of SNR$(\sigma) = 10$ dB and d.o.f. ν varying in $\nu \in [1,5]$. We observe that SNIHT$(1,1)$ has the best performance as it retains low MSE for all values of ν. This is in deep contrast to SNIHT$(2,2)$ which starts an exponential increase at $\nu \leq 3$, reaching sky-high MSE levels in Cauchy noise ($\nu = 1$). The PER rates in Table 2 further illustrates the remarkable performance of the robust methods. Note that SNIHT$(1,1)$ is able to maintain full PER rates for all values of ν, whereas SNIHT$(2,2)$ fails completely for $\nu < 3$.

The usefulness of joint recovery becomes more pronounced at low SNR's, where multiple measurements can dramatically improve on the recovery by exploiting the joint information. This is illustrated in our next simulation set up, where d.o.f. ν of the t-distributed noise is fixed at $\nu = 3$ and the SNR is 10 dB. Figure 2(b) depicts the PER rates for increasing number of measurement vectors Q. As can be seen, the PER rate increases as a function of Q from poor 14% (when $Q = 2$) to near full 100% recovery (when $Q = 6$) when using SNIHT$(1,1)$ method. Again, SNIHT$(2,1)$ is slightly behind in performance to SNIHT$(1,1)$. Conventional SNIHT$(2,2)$ is drastically behind the robust methods, reaching highest 96.6% rate when $Q = 18$. This is again in deep contrast with near 100% PER obtained by SNIHT$(1,1)$ method only with $Q = 6$ samples.



(a) (b)

Fig. 2. (a) MSE of SNIHT(p,q) methods in $\mathbb{C}t_\nu(0,\sigma^2)$ noise as a function of ν; (b) Empirical PER rates of SNIHT(p,q) methods as a function of Q in $\mathbb{C}t_3(0,\sigma^2)$ noise. In both setting, the SNR was SNR$(\sigma) = 10$ dB.

SNIHT	Degrees of freedom ν							
(p,q)	1	1.25	1.5	1.75	2	3	4	5
$(2,2)$	0	0	0	0	.04	.94	.99	1.0
$(2,1)$	0	.07	.55	.90	.98	1.0	1.0	1.0
$(1,1)$	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Table 2. PER rates in i.i.d. $\mathbb{C}t_\nu(0,\sigma^2)$ noise for different d.o.f. ν and SNR$(\sigma) = 10$ dB. System parameters were $(M,N,K,Q) = (256,512,8,16)$.

5. APPLICATIONS TO SOURCE LOCALIZATION

Consider a sensor array consisting of M sensors that receives K narrowband incoherent farfield plane-wave sources from a point source $(M > K)$. At discrete time t, the *array output* (snapshot) $\mathbf{y}(t) \in \mathbb{C}^m$ is a weighted linear combination of the signal waveforms $\mathbf{x}(t) = (x_1(t),\ldots,x_K(t))^\top$ corrupted by additive noise $\mathbf{e}(t) \in \mathbb{C}^M$, $\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(t) + \mathbf{e}(t)$, where $\mathbf{A} = \mathbf{A}(\boldsymbol{\theta})$ is the $M \times K$ *steering matrix* parametrized by the vector $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_K)^\top$ of (distinct) unknown DOA's of the sources. Each column vector $\mathbf{a}(\theta_i)$, called the *steering vector*, represents a point in known array manifold $\mathbf{a}(\theta)$. We assume that the number of sources K is known.

As in [7], we cast the source localization problem as a multichannel sparse recovery problem as follows. We construct an overcomplete $M \times N$ steering matrix $\mathbf{A}(\tilde{\boldsymbol{\theta}})$, where $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1,\ldots,\tilde{\theta}_N)^\top$ represents a sampling grid of all source locations of interest. Suppose that $\tilde{\boldsymbol{\theta}}$ contains the true DOA's θ_i, $i = 1,\ldots,K$. In this case the measurement matrix $\mathbf{Y} = \begin{pmatrix}\mathbf{y}(t_1) & \cdots & \mathbf{y}(t_Q)\end{pmatrix} \in \mathbb{C}^{M\times Q}$ consisting of snapshots at time instants $t_1,\ldots,t_Q$ can be *exactly* modelled as MMV model in which the signal matrix $\mathbf{X} \in \mathbb{C}^{N\times Q}$ is K-rowsparse matrix, whose K non-zero row vectors correspond to source signal sequences. Thus finding the DOA's of the sources is equivalent to identifying the support $\Gamma = \mathrm{supp}(\mathbf{X})$. Since the steering matrix $\mathbf{A}(\tilde{\boldsymbol{\theta}})$ is known, we can use SNIHT methods to identify the support.

We assume that $K = 2$ independent (spatially and temporally) complex circular Gaussian source signals of equal power σ_x^2 arrive on uniform linear array (ULA) of $M = 20$ sensors with half a wavelength inter-element spacing from DOA's $\theta_1 = 0^o$ and $\theta_2 = 8^o$. In this case, the array manifold is $\mathbf{a}(\theta) = (1, e^{-\jmath\pi\sin(\theta)}, \cdots, e^{-\jmath\pi(M-1)\sin(\theta)})^\top$. The noise matrix $\mathbf{E} \in \mathbb{C}^{M\times Q}$ has i.i.d. row vectors, each row vector $\mathbf{e}_{(i)}$ having complex Q-variate inverse Gaussian compound Gaussian (IG-CG) distribution [14] with shape parameter $\lambda = 0.1$ and covariance matrix $\mathrm{Cov}(\mathbf{e}_{(i)}) = \mathbf{I}_Q$. Note that the covariance of the snapshot is $\mathrm{Cov}(\mathbf{y}(t_i)) = \sigma_x^2\mathbf{A}(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta})^{\mathrm{H}} + \mathbf{I}_M$, so we may use the popular MUSIC method to localize the sources. In other words, we search for the $K = 2$ peaks of the MUSIC pseudospectrum in the grid. We use a uniform grid $\tilde{\boldsymbol{\theta}}$ on $[-90, 90]$ with 2^o degree spacing, thus containing the true DOA's. In Step 1 of SNIHT(p, q) algorithm, we locate the K largest peaks of rownorms of $\boldsymbol{\Phi}^{\mathrm{H}}\psi_{p,q}(\mathbf{Y})$ instead of taking Γ^0 as indices of K largest rownorms of $\boldsymbol{\Phi}^{\mathrm{H}}\psi_{p,q}(\mathbf{Y})$.

We then identify the support (which gives the DOA estimates) for all the methods over 1000 MC trials and compute the PER rates and the relative frequency of DOA estimates in the grid. Full PER rate $= 1$ implies that the support Γ correctly identified the true DOA's in all MC trials. Such a case is shown in upper plot of Figure 3 for the SNIHT$(1, 1)$ and SNIHT$(2, 1)$ when the number of snapshots is $Q = 50$ and the SNR is -10 dB. The PER rates of SNIHT$(2, 2)$ and MUSIC were considerably lower, 0.81 and 0.73, respectively. Next we keep other parameters fixed, but decrease the SNR to -20 dB. In this case, the MUSIC method fails completely and provides nearly a uniform frequency on the grid. This is illustrated in lower plot of Figure 3. Note that the proposed robust methods, SNIHT$(1, 1)$ and SNIHT$(2, 1)$, provide high peaks on the correct DOA's. The PER rates of SNIHT$(2, 1)$, SNIHT$(1, 1)$, SNIHT$(2, 2)$ and MUSIC were $0.70, 0.64, 0.11$ and 0.05, respectively. Hence the mixed ℓ_1-norm method SNIHT$(2, 1)$ has the best recovery performance. In conclustion, robust sparse recovery methods can offer considerable improvements in performance when the measurement environment is challenging (low SNR, small Q)

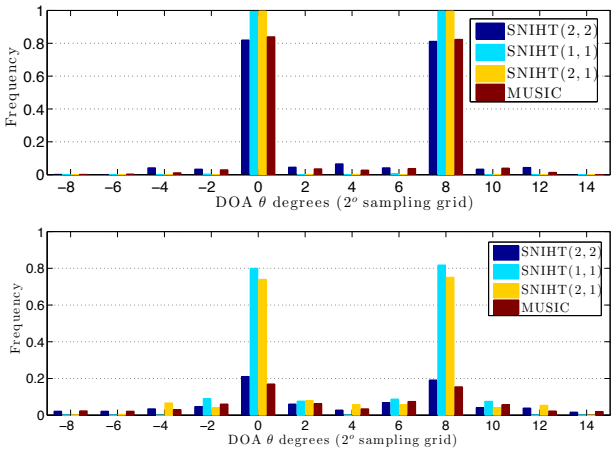


Fig. 3. Relative frequency of DOA estimates. Two equal power Gaussian sources arrive from DOA 0^o and 8^o and the noise has i.i.d. row vectors following IG-CG distribution with covariance matrix $\mathbf{I}$ and shape $\lambda = 0.1$. $\mathrm{SNR}(\sigma) = -10$ dB (upper plot) and $\mathrm{SNR}(\sigma) = -20$ dB (lower plot).

REFERENCES

[1] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, 2011.

[2] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: greedy pursuit," *Signal Processing*, vol. 86, pp. 572–588, 2006.

[3] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: convex relaxation," *Signal Processing*, vol. 86, pp. 589–602, 2006.

[4] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, 2006.

[5] Y. C. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 505–519, 2010.

[6] J. D. Blanchard, M. Cermak, D. Hanle, and Y. Jin, "Greedy algorithms for joint sparse recovery," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1694 – 1704, 2014.

[7] Dmitry Malioutov, Müjdat Çetin, and Alan S Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, 2005.

[8] J. Möttönen and H. Oja, "Multivariate spatial sign and rank methods," *J. Nonparametr. Statist.*, vol. 5, pp. 201–213, 1995.

[9] E. Ollila, "Multichannel sparse recovery of complex-valued signals using Huber's criterion," in *Proc. Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa'15)*, Pisa, Italy, June 16 – 19, 2015.

[10] Matthieu Kowalski, "Sparse regression using mixed norms," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 303–324, 2009.

[11] E. Ollila, J. Eriksson, and V. Koivunen, "Complex elliptically symmetric random variables – generation, characterization, and circularity tests," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 58–69, 2011.

[12] T. Blumensath and M. E. Davies, "Normalized iterative hard thresholding: guaranteed stability and performance," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 298–309, 2010.

[13] A. Hjorungnes and D. Gesbert, "Complex-valued matrix differentiation: techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, pp. 2740–2746, 2007.

[14] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Compound-gaussian clutter modelling with an inverse gaussian texture distribution," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 876–879, 2012.