RANDOMIZED SIMULTANEOUS ORTHOGONAL MATCHING PURSUIT

Aqib Ejaz, Esa Ollila and Visa Koivunen*

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

ABSTRACT

In this paper, we develop randomized simultaneous orthogonal matching pursuit (RandSOMP) algorithm which computes an approximation of the Bayesian minimum mean-squared error (MMSE) estimate of an unknown rowsparse signal matrix. The approximation is based on greedy iterations, as in SOMP, and it elegantly incorporates the prior knowledge of the probability distribution of the signal and noise matrices into the estimation process. Unlike the exact MMSE estimator which is computationally intractable to solve, the Bayesian greedy pursuit approach offers a computationally feasible way to approximate the MMSE estimate. Our simulations illustrate that the proposed RandSOMP algorithm outperforms SOMP both in terms of mean-squared error and probability of exact support recovery. The benefits of RandSOMP are further illustrated in direction-of-arrival estimation with sensor arrays and image denoising.

***Index Terms*—** Bayes, minimum mean-squared error (MMSE), multichannel sparse recovery, compressed sensing.

1. INTRODUCTION

Multichannel sparse signal recovery, an extension of compressed sensing or single measurement vector (SMV) model, has found applications in many areas of signal processing [1–3]. A number of greedy pursuit algorithms (SOMP, SIHT, etc) have been proposed to recover the multichannel sparse signals in a non-Bayesian framework [4–6]. In cases where some prior knowledge about the distribution of the signals is available, Bayesian methods often provide a more effective approach as shown in [7, 8]. In this paper, we develop an algorithm for multichannel sparse signals that takes the prior statistical knowledge of the signal and noise distributions into account and thereby produces more accurate estimates than the conventional greedy pursuit methods such as simultaneous orthogonal matching pursuit (SOMP) [4].

We consider the multiple measurement vectors (MMV) model [2] in which the goal is to recover a set of Q unknown complex-valued signal vectors $\mathbf{x}_i \in \mathbb{C}^N$ from a set of Q measurement vectors $\mathbf{y}_i \in \mathbb{C}^M$. The vectors $\mathbf{x}_i$ are assumed to be *sparse*, i.e., only a small fraction of their entries are nonzero. The sparse vectors are measured according to the model, $\mathbf{y}_i = \mathbf{\Phi}\mathbf{x}_i + \mathbf{e}_i, i = 1, 2, \ldots, Q$ where $\mathbf{\Phi} \in \mathbb{C}^{M\times N}$ is a known *measurement matrix* and $\mathbf{e}_i \in \mathbb{C}^M$ denote the unobservable measurement *noise*. Typically, $M < N$, so the linear model is underdetermined. In matrix form, the model can be written as

$$\mathbf{Y} = \mathbf{\Phi}\mathbf{X} + \mathbf{E}, \tag{1}$$

where $\mathbf{Y} \in \mathbb{C}^{M\times Q}$ with $\mathbf{y}_i$ as its ith column vector, $\mathbf{X} \in \mathbb{C}^{N\times Q}$ contains the unknown sparse vectors $\mathbf{x}_i$ while $\mathbf{E} \in \mathbb{C}^{M\times Q}$ is the noise matrix.

If the unknown sparse vectors were incoherent then there is no obvious gain in trying to recover the sparse vectors simultaneously (jointly). The benefits of simultaneous sparse recovery arise when sparse signal vectors are all predominantly supported on a common support set. In other words, the signal matrix $\mathbf{X}$ is assumed to be K*-rowsparse*, i.e., at most K rows of $\mathbf{X}$ contain non-zero entries. This means that the *row support* of $\mathbf{X}$, defined as the index set of rows containing non-zero elements, $\mathrm{rsupp}(\mathbf{X}) = \{i \in \{1, \ldots, N\} : x_{ij} \neq 0 \text{ for some } j\}$, has cardinality less than or equal to K. The objective of multichannel sparse recovery is to recover the unknown signal matrix $\mathbf{X}$ by knowing only $\mathbf{\Phi}$ and $\mathbf{Y}$.

In this paper, we derive the Bayesian MMSE estimator of $\mathbf{X}$. Since computing the exact MMSE estimate is infeasible, we show how the greedy pursuit approach of SOMP can be utilized in finding an approximate MMSE estimator. The method is a generalization of randomized orthogonal matching pursuit (RandOMP) [7] from SMV model to complex-valued MMV model. The algorithm can also be used for real-valued measurements with obvious modifications.

The paper is organized as follows. In Section 2 we develop the Bayesian MMSE estimator. In Section 3 we develop the RandSOMP algorithm for approximating the MMSE estimator. Section 4 provides simulation results illustrating the effectiveness of the proposed RandSOMP algorithm. In Section 5, we demonstrate the usage of RandSOMP for direction-of-arrival (DOA) estimation using sensor arrays and image denoising. We conclude our paper in Section 6.

2. THE MMSE ESTIMATOR

2.1. Assumptions

The elements of the noise matrix $\mathbf{E}$ are assumed to be independent and identically distributed (i.i.d.) complex normal

*The author is currently affiliated with Nokia Technologies, Finland.

random variables with zero mean and known variance σ_e^2, so $\mathbf{E}$ has *matrix variate complex normal (MCN) distribution*, denoted as $\mathbf{E} \sim \mathcal{MCN}_{M,Q}\left(\mathbf{0}, \sigma_e^2 \mathbf{I}_M, \mathbf{I}_Q\right)$, where $\mathbf{I}_M$ denotes the $M \times M$ identity matrix. In the Bayesian framework we treat $\mathbf{X}$ and its row support $\Gamma = \mathrm{rsupp}(\mathbf{X})$ as random variables with known prior probability distributions. The row support Γ has known fixed cardinality K (i.e. $|\Gamma| = K$) and a uniform prior distribution (i.e., all row supports are equiprobable), $p(\Gamma) = 1/|\Omega|$ for $\Gamma \in \Omega$, where Ω denotes the set of all row supports that have cardinality K and $|\Omega| = \binom{N}{K}$. For a given row support Γ, let $\mathbf{X}_{(\Gamma)}$ denote the $K \times Q$ matrix restricted to those K rows of $\mathbf{X}$ that are indexed by the support set Γ. Then $\mathbf{X}_{(\overline{\Gamma})} = \mathbf{0}$ by definition, where $\overline{\Gamma}$ is the complement of Γ. Furthermore, let each element of $\mathbf{X}_{(\Gamma)}$ be an i.i.d. complex normal random variable with zero mean and known variance σ_x^2, so $\mathbf{X}_{(\Gamma)} \sim \mathcal{MCN}_{K,Q}\left(\mathbf{0}, \sigma_x^2 \mathbf{I}_K, \mathbf{I}_Q\right)$, with p.d.f.

$$p(\mathbf{X}_{(\Gamma)} \,|\, \Gamma) = \frac{1}{(\pi\sigma_x^2)^{KQ}} \exp\left(-\frac{1}{\sigma_x^2}\left\|\mathbf{X}_{(\Gamma)}\right\|_F^2\right), \quad (2)$$

where $\|\cdot\|_F$ denotes the Frobenius norm, i.e., $\|\mathbf{X}\|_F = \sqrt{\mathrm{Tr}(\mathbf{X}^{\mathrm{H}}\mathbf{X})} = \|\mathrm{vec}(\mathbf{X})\|$, where $\mathrm{vec}(\mathbf{X})$ is a vector formed by stacking the columns of $\mathbf{X}$ on top of each other and $\|\cdot\|$ denotes the usual Euclidean norm. We define the signal-to-noise ratio (SNR) as $\gamma = \sigma_x^2/\sigma_e^2$.

2.2. Derivation of the MMSE estimator

The MMSE estimate of $\mathbf{X}$ is obtained as the conditional mean, i.e., the mean of the posterior distribution of $\mathbf{X}$,

$$\widehat{\mathbf{X}}_{\text{MMSE}} = \mathbb{E}\left[\mathbf{X} \,|\, \mathbf{Y}\right] = \sum_{\Gamma \in \Omega} p(\Gamma|\mathbf{Y}) \mathbb{E}\left[\mathbf{X} \,|\mathbf{Y}, \Gamma\right], \quad (3)$$

where $\mathbb{E}[\mathbf{X}|\mathbf{Y},\Gamma]$ is K-rowsparse since $\mathbb{E}\left[\mathbf{X}_{(\overline{\Gamma})} \,|\, \mathbf{Y}, \Gamma\right] = \mathbf{0}$ for the complement of Γ. In order to evaluate the non-zero sub-matrix $\mathbb{E}[\mathbf{X}_{(\Gamma)} \,|\, \mathbf{Y}, \Gamma]$ one can use Bayes' rule as follows,

$$p(\mathbf{X}_{(\Gamma)}|\mathbf{Y},\Gamma) = \frac{p\big(\mathbf{X}_{(\Gamma)} \,|\, \Gamma\big)\, p\big(\mathbf{Y} \,|\, \mathbf{X}_{(\Gamma)}, \Gamma\big)}{p(\mathbf{Y}|\Gamma)}, \quad (4)$$

where $p(\mathbf{Y}|\Gamma)$ is a normalizing constant for fixed $\mathbf{Y}$ and Γ. Moreover,

$$\mathbf{Y} \,|\, \mathbf{X}_{(\Gamma)}, \Gamma = \mathbf{\Phi}_\Gamma \mathbf{X}_{(\Gamma)} + \mathbf{E},$$

where $\mathbf{\Phi}_\Gamma$ is an $M \times K$ matrix restricted to those columns of $\mathbf{\Phi}$ that are indexed by Γ. For a given $\mathbf{X}_{(\Gamma)}$, we have that $\mathbf{Y} \,|\, \mathbf{X}_{(\Gamma)}, \Gamma \sim \mathcal{MCN}_{M,Q}\left(\mathbf{\Phi}_\Gamma \mathbf{X}_{(\Gamma)}, \sigma_e^2 \mathbf{I}_M, \mathbf{I}_Q\right)$, so

$$p\big(\mathbf{Y} \,|\, \mathbf{X}_{(\Gamma)}, \Gamma\big) = \frac{1}{(\pi\sigma_e^2)^{MQ}} \exp\bigg(-\frac{1}{\sigma_e^2}\left\|\mathbf{Y} - \mathbf{\Phi}_\Gamma \mathbf{X}_{(\Gamma)}\right\|_F^2\bigg). \quad (5)$$

Ignoring the normalizing constant $p(\mathbf{Y}|\Gamma)$ and using the closed-form expressions of $p(\mathbf{X}_{(\Gamma)}|\Gamma)$ and $p(\mathbf{Y} \,|\, \mathbf{X}_{(\Gamma)}, \Gamma)$ from (2) and (5), we can rewrite (4) as

$$p(\mathbf{X}_{(\Gamma)}|\mathbf{Y},\Gamma) \propto \exp\bigg(-\frac{\left\|\mathbf{X}_{(\Gamma)}\right\|_F^2}{\sigma_x^2} - \frac{\left\|\mathbf{Y} - \mathbf{\Phi}_\Gamma \mathbf{X}_{(\Gamma)}\right\|_F^2}{\sigma_e^2}\bigg).$$

Since the prior $p(\mathbf{X}_{(\Gamma)}|\Gamma)$ and the likelihood $p(\mathbf{Y}|\mathbf{X}_{(\Gamma)},\Gamma)$ are matrix variate complex normal with known variance, the posterior $p(\mathbf{X}_{(\Gamma)}|\mathbf{Y},\Gamma)$ will also be matrix variate complex normal which is a symmetric unimodal distribution. Therefore, the mean of the posterior will be equal to its mode,

$$\mathbb{E}\big[\mathbf{X}_{(\Gamma)}|\mathbf{Y},\Gamma\big] = \underset{\mathbf{X}_{(\Gamma)} \in \mathbb{C}^{K \times Q}}{\arg\max} \; \log p(\mathbf{X}_{(\Gamma)} \,|\, \mathbf{Y}, \Gamma).$$

The above convex optimization problem can be solved by setting the matrix gradient of the objective function to $\mathbf{0}$ and solving the resulting set of equations. This results in the following closed-form solution:

$$\mathbb{E}[\mathbf{X}_{(\Gamma)}|\mathbf{Y},\Gamma] = \left(\mathbf{\Phi}_\Gamma^{\mathrm{H}}\mathbf{\Phi}_\Gamma + \frac{1}{\gamma}\mathbf{I}_K\right)^{-1} \mathbf{\Phi}_\Gamma^{\mathrm{H}}\mathbf{Y}, \quad (6)$$

where $\mathbf{\Phi}_\Gamma^{\mathrm{H}}$ denotes the conjugate transpose of $\mathbf{\Phi}_\Gamma$.

For a fixed $\mathbf{Y}$ and all $\Gamma \in \Omega$, both $p(\mathbf{Y})$ and $p(\Gamma)$ are constant. Therefore, from Bayes' rule we can conclude that $p(\Gamma|\mathbf{Y}) \propto p(\mathbf{Y}|\Gamma)$. Moreover,

$$p(\mathbf{Y}|\Gamma) = \int_{\mathbf{X}_{(\Gamma)} \in \mathbb{C}^{K \times Q}} p\Big(\mathbf{Y}, \mathbf{X}_{(\Gamma)}|\Gamma\Big) \mathrm{d}\mathbf{X}_{(\Gamma)} \propto$$
$$\int \exp\bigg\{-\frac{\left\|\mathrm{vec}(\mathbf{X}_{(\Gamma)})\right\|^2}{\sigma_x^2} - \frac{\left\|\mathrm{vec}(\mathbf{Y} - \mathbf{\Phi}_\Gamma \mathbf{X}_{(\Gamma)})\right\|^2}{\sigma_e^2}\bigg\} \mathrm{d}\,\mathbf{X}_{(\Gamma)} \quad (7)$$

where the integration is over $\mathbf{X}_{(\Gamma)} \in \mathbb{C}^{K \times Q}$. Since $\mathrm{vec}(\mathbf{Y} - \mathbf{\Phi}_\Gamma \mathbf{X}_{(\Gamma)}) = \mathrm{vec}(\mathbf{Y}) - (\mathbf{I}_Q \otimes \mathbf{\Phi}_\Gamma)\mathrm{vec}(\mathbf{X}_{(\Gamma)})$, where $\otimes$ denotes the Kronecker product, the integral in (7) simplifies to

$$p(\mathbf{Y}|\Gamma) \propto w_\Gamma = \exp\Bigg(\frac{\mathrm{vec}(\mathbf{\Phi}_\Gamma^{\mathrm{H}}\mathbf{Y})^{\mathrm{H}}\mathbf{P}_\Gamma^{-1}\mathrm{vec}(\mathbf{\Phi}_\Gamma^{\mathrm{H}}\mathbf{Y})}{\sigma_e^4} + \log\left(\det(\mathbf{P}_\Gamma^{-1})\right)\Bigg),$$

where

$$\mathbf{P}_\Gamma = \frac{1}{\sigma_e^2}\mathbf{I}_Q \otimes \mathbf{\Phi}_\Gamma^{\mathrm{H}}\mathbf{\Phi}_\Gamma + \frac{1}{\sigma_x^2}\mathbf{I}_{KQ}.$$

The above result is derived using similar arguments as those used in [9, p. 214, 215] for the real-valued SMV model ($Q = 1$). Finally, the posterior $p(\Gamma|\mathbf{Y})$ of Γ, which is proportional to its likelihood $p(\mathbf{Y}|\Gamma)$, is obtained by normalizing w_Γ, i.e.,

$$p(\Gamma|\mathbf{Y}) = \frac{w_\Gamma}{\sum_{\widetilde{\Gamma} \in \Omega} w_{\widetilde{\Gamma}}}. \quad (8)$$

3. RANDOMIZED SIMULTANEOUS OMP

Now we develop RandSOMP algorithm for approximating the MMSE estimator. The MMSE estimate of $\mathbf{X}$ given in (3) can be expressed in closed-form using (6) and (8). But the summation in (3) needs to be evaluated for all possible $\Gamma \in \Omega$, which is computationally infeasible due to the large size of Ω ($|\Omega| = \binom{N}{K}$). One could approximate the MMSE estimate by randomly drawing a small number of sample row supports according to $p(\Gamma|\mathbf{Y})$ and evaluating the summation in (3) over these row supports. Again, due to the large size of the sample space Ω, this approach is not feasible. Next we describe how one can overcome these limitations via SOMP-like greedy pursuit. We propose to draw row supports iteratively by sampling one element at a time from a much smaller space. This results in row supports that are drawn from an approximation of $p(\Gamma|\mathbf{Y})$.

Suppose the rowsparsity of $\mathbf{X}$ is one, i.e., $|\Gamma| = K = 1$. This implies that $\Omega = \{1, 2, \ldots, N\}$, and

$$\mathbf{P}_{\Gamma=\{j\}} = (c_j/\sigma_x^2) \cdot \mathbf{I}_Q \quad \text{for} \quad j \in \Omega,$$

where $c_j = 1 + \gamma \cdot \|\boldsymbol{\phi}_j\|^2$ and $\boldsymbol{\phi}_j$ denotes the jth column of $\mathbf{\Phi}$. Furthermore, $p(\Gamma = \{j\}|\mathbf{Y}) \propto w_{\Gamma=\{j\}} = w_j$ is given by

$$w_j = \exp\left\{\frac{\gamma}{\sigma_e^2} \cdot \frac{\|\mathbf{Y}^{\mathrm{H}}\boldsymbol{\phi}_j\|^2}{c_j} - Q \log c_j\right\}. \tag{9}$$

Since the size of the sample space $|\Omega| = N$, drawing a random row support Γ becomes trivial. For the case when $|\Gamma| = K > 1$, we propose an iterative greedy procedure that looks similar to SOMP and builds up the random row support iteratively one element at a time (see Algorithm 1). In the first iteration, we randomly draw an element j_1 of Γ with probability proportional to the weights given in (9). We then compute the MMSE estimate $\widetilde{\mathbf{X}}$ of $\mathbf{X}$ assuming $\Gamma = \{j_1\}$ and the residual error matrix $\mathbf{R} = \mathbf{Y} - \mathbf{\Phi}\widetilde{\mathbf{X}}$. In the second iteration we modify the weights in (9) by substituting the matrix $\mathbf{Y}$ by $\mathbf{R}$. We randomly draw another element j_2 of the row support using the modified weights and compute the MMSE estimate assuming $\Gamma = \{j_1, j_2\}$. This process continues for K iterations. After K-th iteration we have a randomly drawn row support $\Gamma = \{j_1, j_2, \ldots, j_K\}$ from a distribution that approximates $p(\Gamma|\mathbf{Y})$.

To reduce the estimation error, we run the above greedy procedure L number of times and average the results. Let $\Omega_\star$ denote the set of L row supports obtained from L independent runs. The *approximate MMSE estimate* of $\mathbf{X}$ is then

$$\widehat{\mathbf{X}}_{\text{AMMSE}} = \frac{1}{L}\sum_{\Gamma\in\Omega_\star} \mathbb{E}\big[\mathbf{X} \,|\, \mathbf{Y}, \Gamma\big]. \tag{10}$$

The posterior probability mass $p(\Gamma|\mathbf{Y})$ of each row support does not appear explicitly in the above formula. This is because the row supports having high probability are more likely to be selected in the sampling process than the row supports with low probability, so the row supports are represented in $\Omega_\star$ in (approximate) proportion to their probability masses. Hence, to approximate (3) simple averaging suffices.

In general, the approximate MMSE estimate of $\mathbf{X}$ given in (10) will not be rowsparse. In order to obtain rowsparse

Algorithm 1: RandSOMP algorithm

Input: $\mathbf{\Phi} \in \mathbb{C}^{M\times N}$, $\mathbf{Y} \in \mathbb{C}^{M\times Q}$, K, L, γ, σ_e^2
for $l \leftarrow 1$ **to** L **do**
  $\Gamma^{(0)} \leftarrow \varnothing$, $\widetilde{\mathbf{X}}^{(0)} \leftarrow \mathbf{0}$
  for $i \leftarrow 1$ **to** K **do**
    $\mathbf{R} \leftarrow \mathbf{Y} - \mathbf{\Phi}\widetilde{\mathbf{X}}^{(i-1)}$
    Draw an integer j randomly with probability proportional to
    $$w_j = \exp\left\{\frac{\gamma}{\sigma_e^2} \cdot \frac{\|\mathbf{R}^{\mathrm{H}}\boldsymbol{\phi}_j\|^2}{c_j} - Q \log c_j\right\}$$
    $\Gamma^{(i)} \leftarrow \Gamma^{(i-1)} \cup \{j\}$
    $\widetilde{\mathbf{X}}^{(i)}_{(\Gamma^{(i)})} \leftarrow \left(\mathbf{\Phi}^{\mathrm{H}}_{\Gamma^{(i)}}\mathbf{\Phi}_{\Gamma^{(i)}} + \frac{1}{\gamma}\cdot\mathbf{I}_i\right)^{-1}\mathbf{\Phi}^{\mathrm{H}}_{\Gamma^{(i)}}\mathbf{Y}$
    $\widetilde{\mathbf{X}}^{(i)}_{(\overline{\Gamma}^{(i)})} \leftarrow \mathbf{0}$
  end
  $\widehat{\mathbf{X}}^{(l)} \leftarrow \widetilde{\mathbf{X}}^{(K)}$
end
Output: $\widehat{\mathbf{X}}_{\text{AMMSE}} \leftarrow \frac{1}{L}\sum_{l=1}^{L}\widehat{\mathbf{X}}^{(l)}$

approximate MMSE estimate, we use an approach similar to the one used in RandOMP [7]. Let $H_K(\widehat{\mathbf{X}}_{\text{AMMSE}})$ denote the matrix obtained by setting all but the K largest (in terms of their ℓ_2-norm) rows of $\widehat{\mathbf{X}}_{\text{AMMSE}}$ to $\mathbf{0}$. Let $\widetilde{\Gamma}$ be the row support of $H_K(\widehat{\mathbf{X}}_{\text{AMMSE}})$, i.e., $\widetilde{\Gamma} = \operatorname{rsupp}\big(H_K(\widehat{\mathbf{X}}_{\text{AMMSE}})\big)$. A K-rowsparse estimate of $\mathbf{X}$ is then obtained as

$$\widehat{\mathbf{X}}_{\text{rowsparse}} = \mathbb{E}\big[\mathbf{X} \,|\, \mathbf{Y}, \widetilde{\Gamma}\,\big]. \tag{11}$$

4. SIMULATION RESULTS

Next we provide simulation results comparing the performance of SOMP and RandSOMP. In order to illustrate the fact that the joint recovery of the rowsparse matrix in the MMV model is a much more effective approach, we include the results of the case where RandOMP is used to recover the individual sparse vectors one by one. It is demonstrated that RandSOMP performs better than RandOMP both in terms of normalized mean-squared error (MSE) and probability of exact support recovery. RandSOMP also outperforms SOMP in terms of normalized MSE, which is expected since RandSOMP approximates the MMSE estimate. More importantly, RandSOMP also outperforms SOMP in terms of probability of exact support recovery.

The simulation set-up is as follows. We generate $Q = 30$ sparse vectors each of length $N = 300$ which share the same randomly chosen support of cardinality $K = 20$. All non-zero entries of the sparse vectors are independently drawn from the standard complex normal distribution (i.e., $\sigma_x^2 = 1$). The elements of the measurement matrix $\mathbf{\Phi} \in \mathbb{C}^{150\times 300}$ (i.e., $M = 150$) are independently drawn from the standard com-

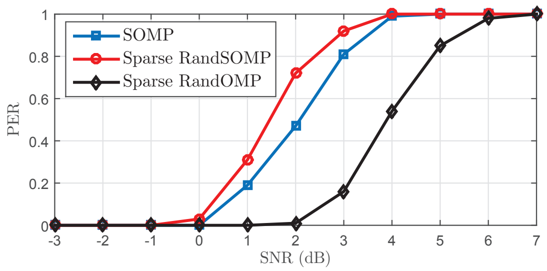

Fig. 1. PER rates vs SNR. $M = 150$, $N = 300$, $K = 20$, $L = 15$, $Q = 30$. Sparse RandSOMP has higher PER rates for all the given SNR levels.

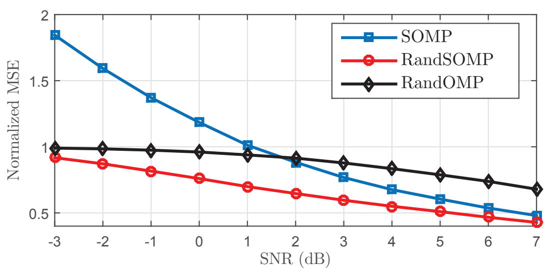

Fig. 2. Normalized MSE vs SNR. $M = 150$, $N = 300$, $K = 20$, $L = 15$, $Q = 30$. RandSOMP has lower normalized MSE for all the given SNR levels.

plex normal distribution and each column is normalized to have unit norm. The measurements are contaminated by zero mean additive complex Gaussian noise with variance σ_e^2. The SNR is defined as $\gamma = \sigma_x^2/\sigma_e^2$. The number of randomly drawn row supports in both RandSOMP and RandOMP is $L = 15$. The experiments are averaged over 100 realizations.

First, the recovery of the correct signal support is considered. Figure 1 depicts the empirical *probability of exact (support) recovery (PER)* rates as a function of SNR. Since RandSOMP and RandOMP produce estimates that are not necessarily rowsparse, PER rates are computed for the sparsified estimates given in (11). (In addition to ensuring sparsity of the estimates, this approach also brings the benefits of multiple measurement vectors to RandOMP.) As shown in Figure 1, the proposed RandSOMP algorithm outperforms both SOMP and RandOMP. Figure 2 depicts the empirical normalized MSE ($\|\widehat{\mathbf{X}} - \mathbf{X}\|_F^2 / \|\mathbf{X}\|_F^2$). Again RandSOMP has the best performance at all SNR levels. One can also observe that in low SNR regime, the randomized algorithms are performing much better than the non-randomized SOMP. With increasing SNR, the performance of RandOMP improves much more slowly in comparison to RandSOMP. The performance of SOMP improves much more sharply, and at 2 dB SNR SOMP performs better than RandOMP. As expected, RandSOMP has the best performance since it is an approximate MMSE estimate.

Next we fix the SNR value at 3 dB and vary the number of measurement vectors Q. Figure 3 depicts the PER as a function of Q. When $Q = 1$, the MMV model is reduced

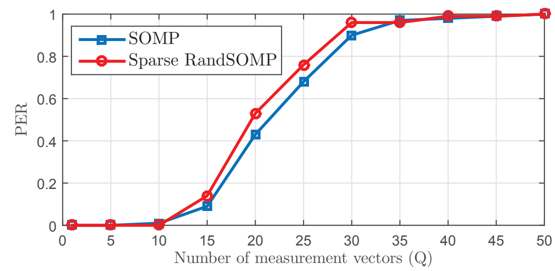

Fig. 3. PER rates vs Q. $M = 150$, $N = 300$, $K = 20$, $L = 15$, SNR $= 3$ dB. Sparse RandSOMP has higher PER rates for all the given values of Q.

to the SMV model and thus SOMP and RandSOMP become equivalent to OMP [10] and RandOMP respectively. As Q increases we expect to gain some improvement through joint processing of multiple measurement vectors. This is evident from Figure 3. When $Q = 1$, the PER rate is near 0 for both SOMP and RandSOMP. With increasing Q the PER rate increases and reaches full PER ($= 1$) at $Q = 50$. Again observe that for almost all the employed values of Q, the proposed RandSOMP algorithm performs better than SOMP.

5. APPLICATIONS

5.1. DOA estimation with sensor arrays

We consider a DOA estimation problem using a uniform linear array (ULA) of M sensors that receives signals from K narrowband incoherent farfield plane-wave sources ($M > K$). The output of the array at time t is modeled as $\mathbf{y}(t) = \mathbf{\Phi}(\boldsymbol{\theta})\mathbf{x}_{\boldsymbol{\theta}}(t) + \mathbf{e}(t)$, where $\mathbf{x}_{\boldsymbol{\theta}}(t) \in \mathbb{C}^K$ represents the source signals, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^{\mathrm{T}}$ contains the directions-of-arrival (DOAs) of the K sources, $\mathbf{\Phi}(\boldsymbol{\theta}) \in \mathbb{C}^{M \times K}$ is the measurement matrix whose columns consist of steering vectors and $\mathbf{e}(t) \in \mathbb{C}^M$ is the additive noise vector. For the ULA with half-wavelength sensor spacing and ideal omnidirectional sensors, the steering vectors are $\mathbf{\Phi}_i(\theta_i) = (1 \;\; e^{-j\pi \sin(\theta_i)} \;\ldots\; e^{-j\pi(M-1)\sin(\theta_i)})^{\mathrm{T}}$. Since $\boldsymbol{\theta}$ is unknown, the matrix $\mathbf{\Phi}(\boldsymbol{\theta})$ is also unknown. The goal is to estimate $\boldsymbol{\theta}$ and $\mathbf{x}_{\boldsymbol{\theta}}(t)$ from $\mathbf{y}(t)$.

In [3], DOA estimation was modeled as a sparse recovery problem using steering vectors for a pre-determined set of N different DOA values. If the set contains the true DOAs of the source signals, then the array output at time t is $\mathbf{y}(t) = \mathbf{\Phi}\mathbf{x}(t) + \mathbf{e}(t)$, where $\mathbf{\Phi} \in \mathbb{C}^{M \times N}$ is the matrix of the N *known* steering vectors and $\mathbf{x}(t) \in \mathbb{C}^N$ is a K-sparse vector whose non-zeros entries consists of source signal $\mathbf{x}_{\boldsymbol{\theta}}(t)$. Given that we have multiple measurements (snapshots) at time instants $t = t_1, t_2, \ldots, t_Q$, the array output can be modeled by the MMV model, $\mathbf{Y} = \mathbf{\Phi}\mathbf{X} + \mathbf{E}$, where each column of $\mathbf{Y}$ represents array's output at one time instant. Therefore, recovering the row support of the K-sparse matrix $\mathbf{X}$ is equivalent to finding the DOAs of the K sources.

We investigate the performance of RandSOMP for DOA

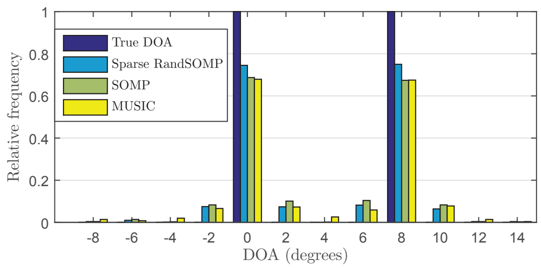

Fig. 4. Relative frequency vs DOA (degrees). $M = 20$, $N = 90$, $Q = 50$, $K = 2$, $L = 10$, SNR $= -15$ dB.

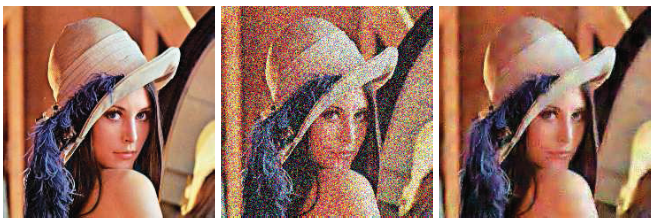

Fig. 5. Image denoising with RandSOMP ($L = 10$, $C = 0.99$). (a) Original image; (b) Noisy image with PSNR $= 12$ dB; (c) Denoised image with PSNR $= 22.34$ dB.

estimation under the following simulation settings. The array consists of $M = 20$ sensors which receive signals from $K = 2$ sources located at $\theta_1 = 0°$ and $\theta_2 = 8°$. The source signals are (spatially and temporally) independent standard complex Gaussian random variables. The set of $N = 90$ predetermined DOA values are taken in the interval $[-90°, 88°]$ with $2°$ step size. We take $Q = 50$ time measurements that are corrupted by a zero mean additive white Gaussian noise with SNR $= -15$ dB. The number of randomly drawn supports in RandSOMP is $L = 10$. The DOAs of the two sources are estimated using RandSOMP, SOMP and MUSIC and the experiments are repeated over 1000 trials. Figure 4 shows the relative frequencies of estimated DOAs of the three algorithms. At the given challenging SNR scenario (-15 dB) and rather low sample size ($Q = 50$), none of the considered methods is able to offer perfect recovery (i.e., mass 1 at true DOA's). Nevertheless, RandSOMP is able to provide accurate estimation more frequently than both SOMP and MUSIC.

5.2. Image denoising

Next we apply RandSOMP for denoising RGB color images. We process the noisy Lena RGB image of size 256×256 in small overlapping patches of size 8×8. The patches are modeled as $\mathbf{Y}_i = \mathbf{U}_i + \mathbf{E}_i$, where $\mathbf{U}_i \in \mathbb{R}^{64\times 3}$ represents the original noise-free patches and $\mathbf{E}_i$ represents additive white Gaussian noise. The original patches $\mathbf{U}_i$ are modeled to have rowsparse representations $\mathbf{X}_i \in \mathbb{R}^{192\times 3}$ in an overcomplete dictionary $\mathbf{\Phi} \in \mathbb{R}^{64\times 192}$ (i.e., $\mathbf{U}_i = \mathbf{\Phi}\mathbf{X}_i$). The dictionary $\mathbf{\Phi}$ is formed from the concatenation of the following three unitary subdictionaries: symlet-4, coiflet-4, and the discrete cosine transform (DCT-II). The aim in image denoising is to recover the noise-free patches $\mathbf{U}_i$ by finding their rowsparse representations $\mathbf{X}_i$ from the MMV model, $\mathbf{Y}_i = \mathbf{\Phi}\mathbf{X}_i + \mathbf{E}_i$. In image denoising, the stopping criterion of RandSOMP is based on the norm of the residual error, i.e., RandSOMP stops when $\|\mathbf{R}\|_F < \sqrt{64 \times 3} \cdot C\sigma_e$, where C is a tuning parameter. Figure 5 shows the denoised Lena image using RandSOMP with $L = 10$ and $C = 0.99$. The peak signal-to-noise ratio (PSNR) in the noisy image is set at 12 dB. RandSOMP produces more than 10 dB improvement in the PSNR value.

6. CONCLUSION

In this paper, we developed an algorithm called RandSOMP that approximates the Bayesian MMSE estimator of the rowsparse signal in the MMV model. In comparison to SOMP, the non-Bayesian greedy method, RandSOMP achieves higher PER and lower MSE when adequate prior statistical knowledge of the signal distribution is available. The benefits of using RandSOMP for DOA estimation and image denoising were demonstrated by practical examples.

REFERENCES

[1] 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, Jan 2010.

[2] 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, Sept 2011.

[3] D. Malioutov, M. Cetin, and A.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.

[4] Joel A. Tropp, Anna C. Gilbert, and Martin J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Processing*, vol. 86, pp. 572–588, Mar. 2006.

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

[6] M. Moussallam, A. Gramfort, L. Daudet, and G. Richard, "Blind denoising with random greedy pursuits," *Signal Processing Letters, IEEE*, vol. 21, no. 11, pp. 1341–1345, 2014.

[7] M. Elad and I. Yavneh, "A plurality of sparse representations is better than the sparsest one alone," *Information Theory, IEEE Transactions on*, vol. 55, pp. 4701–4714, Oct. 2009.

[8] P. Schniter, L.C. Potter, and J. Ziniel, "Fast bayesian matching pursuit," in *Proc. IEEE Information Theory and Applications Workshop*, San Diego, CA, Jan 27-Feb 1 2008, pp. 326–333.

[9] Michael Elad, *Sparse and Redundant Representations*, Springer, 2010.

[10] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," *Proc. Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 40–44, 1993.