Joint Sparsity with Different Measurement Matrices

Reinhard Heckel and Helmut Bölcskei

Dept. of IT & EE, ETH Zurich, Switzerland

{heckel,boelcskei}@nari.ee.ethz.ch

Abstract— We consider a generalization of the multiple measurement vector (MMV) problem, where the measurement matrices are allowed to differ across measurements. This problem arises naturally when multiple measurements are taken over time, e.g., and the measurement modality (matrix) is time-varying. We derive probabilistic recovery guarantees showing that—under certain (mild) conditions on the measurement matrices—ℓ_2/ℓ_1-norm minimization and a variant of orthogonal matching pursuit fail with a probability that decays exponentially in the number of measurements. This allows us to conclude that, perhaps surprisingly, recovery performance does not suffer from the individual measurements being taken through different measurement matrices. What is more, recovery performance typically benefits (significantly) from diversity in the measurement matrices; we specify conditions under which such improvements are obtained. These results continue to hold when the measurements are subject to (bounded) noise.

I. INTRODUCTION

An interesting generalization of the sparse signal recovery problem as studied, e.g., in [1], [2], [3], is the so-called multiple measurement vector (MMV) problem [4], [5], [6], [7]. Application areas of the MMV problem include neuromagnetic imaging, array processing, and nonparametric spectral analysis of time series [4]. The MMV problem is formalized as follows: Given the vectors $\mathbf{x}^{(0)}, ..., \mathbf{x}^{(d-1)}$, that share the sparsity pattern $\mathcal{S}$, i.e., the entries of $\mathbf{x}^{(0)}, ..., \mathbf{x}^{(d-1)}$ are equal to zero on $\bar{\mathcal{S}}$, we want to recover the $\mathbf{x}^{(i)}$ from the noisy measurements $\mathbf{y}^{(i)} = \mathbf{A}\mathbf{x}^{(i)} + \mathbf{e}^{(i)}, i = 0, ..., d-1$, where the $\mathbf{e}^{(i)}$ are noise vectors and the measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is assumed known. For the noiseless case, i.e., $\mathbf{e}^{(i)} = \mathbf{0}$, for all i, it was shown in [5], [8] that the program

$$\text{(P0-MMV)} \begin{cases} \text{minimize} & |\mathcal{S}| \\ \text{subject to} & \mathbf{y}^{(i)} = \mathbf{A}\mathbf{x}^{(i)}, \quad i = 0, .., d-1 \end{cases}$$

recovers all $\mathbf{x}^{(0)}, ..., \mathbf{x}^{(d-1)}$ with $\text{rank}[\mathbf{x}^{(0)} ... \mathbf{x}^{(d-1)}] = K$ if and only if

$$|\mathcal{S}| < \frac{\text{spark}(\mathbf{A}) - 1 + K}{2} \tag{1}$$

where $\text{spark}(\mathbf{A})$ is the cardinality of the smallest set of linearly dependent columns of $\mathbf{A}$ [2]. The threshold (1) constitutes a potentially significant improvement over the well-known $\text{spark}(\mathbf{A})/2$-threshold [2] for the single measurement vector (SMV) case, i.e., for $d = 1$. Necessity of the threshold (1) shows that when asking for recovery of *all* sets of vectors $\mathbf{x}^{(0)}, ..., \mathbf{x}^{(d-1)}$, including linearly dependent collections, multiple measurements do not result in an improvement in the recovery threshold over the SMV case. It is therefore sensible to ask whether performance improvements can be

expected for "typical" $\mathbf{x}^{(0)}, ..., \mathbf{x}^{(d-1)}$. Since (P0-MMV) is NP-hard [9], this question is usually posed with the proviso that computationally efficient algorithms such as ℓ_2/ℓ_1-norm minimization or a variant of orthogonal matching pursuit (OMP) [5], [6], [7] should be used for recovery. Indeed, a corresponding probabilistic performance analysis carried out in [10], [11] shows that multiple measurements yield significant improvements in recovery performance over the SMV case.

In practical applications the measurement matrix (modality) often changes across measurements, e.g., when measurements are taken over time and the underlying measurement modality exhibits characteristics that vary over time. It is therefore natural to ask whether improvements thanks to multiple measurements depend critically on the measurements all being taken through the same measurement matrix $\mathbf{A}$. We answer this question by considering the following modification of the MMV problem, termed generalized MMV (GMMV) problem henceforth: Given the vectors $\mathbf{x}^{(0)}, ..., \mathbf{x}^{(d-1)}$, that share the sparsity pattern $\mathcal{S}$, recover the $\mathbf{x}^{(i)}$ from the (possibly noisy) measurements

$$\mathbf{y}^{(i)} = \mathbf{A}^{(i)}\mathbf{x}^{(i)} + \mathbf{e}^{(i)}, \quad i = 0, ..., d-1 \tag{2}$$

assuming knowledge of the measurement matrices $\mathbf{A}^{(i)} \in \mathbb{R}^{m \times n}$. Here, the $\mathbf{e}^{(i)}$ are noise vectors.

The GMMV problem also occurs in the recovery of sparse signals that lie in the union of shift-invariant subspaces [12], [13], as detailed in an extended version of this paper [14].

As the MMV problem is a special case of the GMMV problem, obtained by setting $\mathbf{A}^{(i)} = \mathbf{A}$, for all i, it follows immediately that, for general $\mathbf{A}^{(i)}$, a worst-case (with respect to the $\mathbf{x}^{(i)}$) analysis reveals no improvements resulting from multiple measurements.

Contributions: The main theme of this paper is a probabilistic (with respect to the $\mathbf{x}^{(i)}$) performance analysis of an ℓ_2/ℓ_1-norm based recovery algorithm, called LOPT, and a variant of OMP, called MOMP, for deterministic measurement matrices $\mathbf{A}^{(i)}$. For the noiseless case, under very general conditions on the $\mathbf{A}^{(i)}$, we find that the failure probability of LOPT and MOMP decays exponentially in the number of measurements d. We show that, perhaps surprisingly, having different measurement matrices $\mathbf{A}^{(i)}$ can lead to (substantial) performance improvements over the MMV case $\mathbf{A}^{(0)} = ... = \mathbf{A}^{(d-1)}$. What is more, these improvements are obtained under very mild "isometry" conditions on the $\mathbf{A}^{(i)}$. Furthermore, we show that our results continue to hold when the measurements are subject to bounded noise.

The probabilistic model on the $\mathbf{x}^{(i)}$ we use is more general than that employed in [10], [11] for the MMV case. Particularizing our results to the MMV case therefore yields generalizations of the main results in [10], [11]. For the noisy case our result for LOPT is new, even in the MMV case.

We note that the GMMV problem can be cast as a block-sparse problem [15], which in turn is contained in the model-based [16] setting. However, formulating the GMMV problem as a block-sparse (or model-based) problem, and applying the corresponding recovery results available in the literature yields worst-case recovery conditions only.

In terms of mathematical tools, we note that the proofs of our main results, provided in [14], consist of two steps. First, we derive conditions for LOPT and MOMP to succeed and then we use concentration of measure results to show that these conditions are satisfied with high probability, provided that mild conditions on the $\mathbf{A}^{(i)}$ are satisfied. While the proofs in [10], [11] follow these two general steps as well, the technical specifics are quite different. Concretely, the more general probabilistic model for the $\mathbf{x}^{(i)}$ requires the use of concentration of measure results that are more general than those employed in [10], [11]. In addition, our recovery conditions are new, and, in particular in the noisy case, non-trivial to derive.

Notation: We use lowercase boldface letters to denote column vectors, e.g., $\mathbf{x}$, and uppercase boldface letters to designate matrices, e.g., $\mathbf{A}$. For a vector $\mathbf{x}$, $[\mathbf{x}]_q$ and x_q denote the qth entry. For the matrix $\mathbf{A}$, $\mathbf{A}^\dagger$ is its pseudo-inverse and $\|\mathbf{A}\|_{2\to 2} := \max_{\|\mathbf{v}\|_2=1} \|\mathbf{A}\mathbf{v}\|_2$ its spectral norm. The superscript ^{H} stands for Hermitian transposition. For the set $\mathcal{S}$, $|\mathcal{S}|$ is its cardinality and $\bar{\mathcal{S}}$ stands for its complement in $\{0,...,n-1\}$. We say that a random variable x is standard Gaussian, if it is of zero-mean and unit variance; x is standard complex Gaussian if $x = x_R + jx_I$, where x_R, x_I are i.i.d. Gaussian with mean zero and variance $1/2$.

II. PROBLEM FORMULATION

The formal statement of the problem we consider is as follows. Suppose we observe the m-dimensional vectors

$$\mathbf{y}^{(i)} = \mathbf{A}^{(i)}\mathbf{x}^{(i)} + \mathbf{e}^{(i)}, \quad i = 0,...,d-1 \qquad (3)$$

where the $\mathbf{e}^{(i)} \in \mathbb{R}^m$ account for (unknown) noise, the $\mathbf{x}^{(i)} \in \mathbb{R}^n$, $n > m$, share the sparsity pattern $\mathcal{S} \subsetneq \{0,...,n-1\}$, i.e., for each $\mathbf{x}^{(i)}$ the entries with index in $\bar{\mathcal{S}}$ are equal to zero, and the measurement matrices $\mathbf{A}^{(0)},...,\mathbf{A}^{(d-1)} \in \mathbb{R}^{m\times n}$ are known. We want to recover $\mathbf{x}^{(0)},...,\mathbf{x}^{(d-1)}$ from the $\mathbf{y}^{(0)},...,\mathbf{y}^{(d-1)}$.

We first consider the noiseless case, i.e., $\mathbf{e}^{(i)} = \mathbf{0}$, for all i. Recovery can be accomplished by solving

$$(\text{P0-GMMV}) \begin{cases} \text{minimize} & |\mathcal{S}| \\ \text{subject to} & \mathbf{y}^{(i)} = \mathbf{A}^{(i)}\mathbf{x}^{(i)}, \ i = 0,...,d-1 \end{cases}$$

which is, however, NP-hard [9]. Computationally efficient alternative recovery algorithms, with, however, weaker recovery guarantees, are specified next. A convex relaxation of P0-GMMV is given by

$$(\text{LOPT}) \begin{cases} \text{minimize} \ \sum_{l=0}^{n-1}\left(\sum_{i=0}^{d-1}\left|x_l^{(i)}\right|^2\right)^{1/2} \\ \text{subject to} \ \mathbf{y}^{(i)} = \mathbf{A}^{(i)}\mathbf{x}^{(i)}, \ i = 0,...,d-1. \end{cases}$$

Another alternative, which is an adaptation of OMP, and will be called MOMP, is defined as follows. MOMP iteratively builds up the joint support set of $\mathbf{x}^{(0)},...,\mathbf{x}^{(d-1)}$. The algorithm is initialized by choosing the residuals in iteration 0 as $\mathbf{r}_0^{(i)} = \mathbf{y}^{(i)}, i = 0,...,d-1$, and the set of selected indices as $\mathcal{S}_0 = \emptyset$. In the pth iteration ($p \geq 1$) we find the index

$$l_p = \arg\max_l \sum_{i=0}^{d-1}\left|(\mathbf{a}_l^{(i)})^H \mathbf{r}_{p-1}^{(i)}\right|^2$$

and update the set of selected indices by setting $\mathcal{S}_p = \mathcal{S}_{p-1}\cup \{l_p\}$. The residuals are updated according to

$$\mathbf{r}_p^{(i)} = \mathbf{y}^{(i)} - \mathbf{A}_{\mathcal{S}_p}^{(i)}\mathbf{x}_{\mathcal{S}_p}^{(i)} = (\mathbf{I} - \mathbf{P}_{\mathcal{S}_p}^{(i)})\mathbf{y}^{(i)}, \quad i = 0,...,d-1$$

where $\mathbf{A}_{\mathcal{S}_p}^{(i)}$ is the matrix obtained from $\mathbf{A}^{(i)}$ by selecting the columns with indices in $\mathcal{S}_p$ and $\mathbf{P}_{\mathcal{S}_p}^{(i)} := \mathbf{A}_{\mathcal{S}_p}^{(i)}(\mathbf{A}_{\mathcal{S}_p}^{(i)})^\dagger$ is the orthogonal projector onto the span of the columns in $\mathbf{A}_{\mathcal{S}_p}^{(i)}$. Both LOPT and MOMP are trivial generalizations of corresponding algorithms for the MMV case [4], [5], [6], [7].

Proceeding to the noisy case, we assume that noise is bounded in the sense of

$$\sum_{i=0}^{d-1}\left\|\mathbf{e}^{(i)}\right\|_2^2 \leq \epsilon^2. \qquad (4)$$

As exact recovery of the $\mathbf{x}^{(i)}$ will, in general, no longer be possible, we will be content with ensuring that the estimates of the $\mathbf{x}^{(i)}$ are "close" to the true $\mathbf{x}^{(i)}$. The recovery algorithms we analyze in the noisy case are MOMP and a convex program closely related to LOPT, namely

$$(\text{POPT}) \ \text{minimize} \ \frac{1}{2}\sum_{i=0}^{d-1}\left\|\mathbf{y}^{(i)} - \mathbf{A}^{(i)}\mathbf{x}^{(i)}\right\|_2^2$$
$$+ \gamma\sum_{l=0}^{n-1}\left(\sum_{i=0}^{d-1}\left|x_l^{(i)}\right|^2\right)^{1/2}$$

which, for $d = 1$, is known as *the lasso* [17] in the statistics literature, and for $d > 1$, is a particular variant of *the group lasso* [18]. The first term in the cost function of POPT accounts for the recovery error and the second term enforces sparsity; the parameter $\gamma > 0$ controls the tradeoff between these two terms.

III. REVIEW OF WORST-CASE RECOVERY RESULTS

We briefly discuss worst-case recovery results for the GMMV problem. Formulating the GMMV problem as a block-sparse recovery problem and evaluating the corresponding recovery conditions in [15] yields the following proposition.

Proposition 1: Let $\mathcal{S}$ be the sparsity pattern of $\mathbf{x}^{(0)}, ..., \mathbf{x}^{(d-1)}$ and assume that

$$\max_{l \notin \mathcal{S}} \sum_{q} \max_{i=0,...,d-1} \left| [(\mathbf{A}_{\mathcal{S}}^{(i)})^{\dagger} \mathbf{a}_l^{(i)}]_q \right| < 1. \tag{5}$$

Then, LOPT and MOMP recover $\mathbf{x}^{(0)}, ..., \mathbf{x}^{(d-1)}$ exactly from $\mathbf{y}^{(i)} = \mathbf{A}^{(i)} \mathbf{x}^{(i)}, i = 0, ..., d-1$.

For the MMV case, Proposition 1 reduces to [5, Th. 3.1]. Condition (5) can be viewed as the GMMV-equivalent of the SMV-exact recovery condition, a standard recovery condition for ℓ_1-minimization and OMP [19].

An alternative recovery condition can be obtained by viewing the GMMV problem as separate SMV problems and requiring exact recovery for each of the resulting SMV problems. Following this route, based on the SMV exact recovery condition [19, Th. A], we get that ℓ_1-minimization and OMP applied individually to $\mathbf{y}^{(i)} = \mathbf{A}^{(i)} \mathbf{x}^{(i)}$ recover $\mathbf{x}^{(0)}, ..., \mathbf{x}^{(d-1)}$ correctly if

$$\max_{l \notin \mathcal{S}} \max_{i=0,...,d-1} \left\| (\mathbf{A}_{\mathcal{S}}^{(i)})^{\dagger} \mathbf{a}_l^{(i)} \right\|_1 < 1. \tag{6}$$

This is a slightly weaker condition than (5). Hence Proposition 1 does not predict any improvement of using LOPT or MOMP over treating the recovery problem as individual SMV problems (solved through ℓ_1-minimization and/or OMP).

IV. MAIN RESULTS

We discuss the noiseless and the noisy case separately.

A. Recovery in the noiseless case

For the noiseless case the probabilistic model on the $\mathbf{x}^{(i)}$ is as follows: For a given support set $\mathcal{S} \subseteq \{0, ..., n-1\}$, we take the entries of the vectors $\mathbf{x}_{\mathcal{S}}^{(0)}, ..., \mathbf{x}_{\mathcal{S}}^{(d-1)}$ to be independent sub-Gaussian [20].

Definition 1: A zero-mean random variable is ρ-sub-Gaussian[1], with $\rho > 0$, if its moment generating function satisfies

$$\mathbb{E}\left[e^{tx} \right] \leq e^{\rho t^2}. \tag{7}$$

Sub-Gaussian random variables contain Gaussian and all bounded[2] random variables as special cases. We start with our main result for LOPT in the noiseless case.

Theorem 1: Fix $\mathcal{S} \subseteq \{0, ..., n-1\}$ with cardinality $s := |\mathcal{S}|$, and take the entries of $\mathbf{x}_{\mathcal{S}}^{(0)}, ..., \mathbf{x}_{\mathcal{S}}^{(d-1)} \in \mathbb{R}^s$ to be i.i.d. zero-mean ρ-sub-Gaussian with unit variance[3]. Assume that the measurement matrices $\mathbf{A}^{(0)}, ..., \mathbf{A}^{(d-1)} \in \mathbb{R}^{m \times n}$ satisfy

$$\left(\frac{1}{d} \sum_{i=0}^{d-1} \left\| (\mathbf{A}_{\mathcal{S}}^{(i)})^{\dagger} \mathbf{a}_l^{(i)} \right\|_2^2 \right)^{1/2} \leq \alpha < 1, \ \text{for all } l \notin \mathcal{S} \tag{8}$$

[1]Sub-Gaussian random variables are often equivalently defined through tail bounds or through moment bounds, see e.g. [20]. The definition we chose is the most convenient for our purposes.
[2]The random variable x is bounded if there exists an $M \geq 0$ such that $\mathrm{P}[|x| \leq M] = 1$.
[3]This is w.l.o.g. as the entries of the $\mathbf{x}^{(i)}$ can be scaled to account for non-unit variance.

and

$$\max_{i} \left\| (\mathbf{A}_{\mathcal{S}}^{(i)})^{\dagger} \mathbf{a}_l^{(i)} \right\|_2 \leq \gamma, \ \text{for all } l \notin \mathcal{S} \tag{9}$$

for some $\gamma > 0$, where $\mathbf{a}_l^{(i)}$ denotes the lth column of $\mathbf{A}^{(i)}$. Then, for every $\xi > 0$ satisfying $\max\{1 - 32e\rho, \alpha^2\} < \xi^2 \leq \alpha^2 (1 + 32e\rho)$, with probability at least

$$1 - (n-s) \exp\left(-d \frac{(\xi^2 - \alpha^2)^2}{2^{11} e^2 \rho^2 \gamma^2 \alpha^2} \right) - s \exp\left(-d \frac{(1 - \xi^2)^2}{2^{11} e^2 \rho^2} \right) \tag{10}$$

LOPT applied to $\mathbf{y}^{(i)} = \mathbf{A}^{(i)} \mathbf{x}^{(i)}, i = 0, ..., d-1$, recovers the correct solution $\mathbf{x}^{(0)}, ..., \mathbf{x}^{(d-1)}$.

The main implication of Theorem 1 is that, provided (8) (and (9)) is satisfied, the probability that LOPT fails decays exponentially in the number of measurements d. This has been shown before for the MMV case under the assumption of i.i.d. Gaussian $\mathbf{x}_{\mathcal{S}}^{(i)}$ [10, Th. 4.4].

The constants in the exponents of (10) can be improved (significantly) for certain distributions. For example, when the entries of the $\mathbf{x}_{\mathcal{S}}^{(i)}$ are i.i.d. standard Gaussian (note that a standard Gaussian is sub-Gaussian with $\rho = 1/2$), the recovery probability is at least [14]

$$1 - (n-s) \exp\left(-d \frac{(\xi - \alpha)^2}{2\gamma^2} \right) - s \exp\left(-d \frac{(1 - \xi^2)^2}{4} \right). \tag{11}$$

Improvements over worst-case results: First note that γ in (9) can be chosen arbitrarily, hence (9) is not restrictive. To see that the recovery condition (8) is weaker than the worst-case recovery condition (6) (recall that (6) implies (5)), we simply note that

$$\left(\frac{1}{d} \sum_{i=0}^{d-1} \left\| (\mathbf{A}_{\mathcal{S}}^{(i)})^{\dagger} \mathbf{a}_l^{(i)} \right\|_2^2 \right)^{1/2} \leq \left(\frac{1}{d} \sum_{i=0}^{d-1} \left\| (\mathbf{A}_{\mathcal{S}}^{(i)})^{\dagger} \mathbf{a}_l^{(i)} \right\|_1^2 \right)^{1/2}$$
$$\leq \max_{i=0,...,d-1} \left\| (\mathbf{A}_{\mathcal{S}}^{(i)})^{\dagger} \mathbf{a}_l^{(i)} \right\|_1.$$

Improvements due to different measurement matrices: Evaluating (8) for the MMV case yields

$$\left\| (\mathbf{A}_{\mathcal{S}})^{\dagger} \mathbf{a}_l \right\|_2 \leq \alpha < 1, \ \text{for all } l \notin \mathcal{S}. \tag{12}$$

Note that (12) is the recovery condition stated in [10, Th. 4.4] and applying to the case where the entries of the $\mathbf{x}_{\mathcal{S}}^{(i)}$ are i.i.d. Gaussian. Comparing (12) to (8), we see that in the GMMV case the measurement matrices have to satisfy $\left\| (\mathbf{A}_{\mathcal{S}}^{(i)})^{\dagger} \mathbf{a}_l^{(i)} \right\|_2^2 \leq \alpha^2$ only on average (i.e., across i). This essentially says that having different measurement matrices allows for some of them to be "bad" as long as the collection $\{\mathbf{A}^{(0)}, ..., \mathbf{A}^{(d-1)}\}$ is good enough on average. In contrast, in the MMV case, the single measurement matrix $\mathbf{A}$ has to be "good" in the sense of (12).

This can be nicely illustrated by way of an example. Suppose we are given a measurement matrix $\mathbf{A}$ which does not satisfy (12) for all $\mathcal{S} \subseteq \{0, ..., n-1\}$ with $|\mathcal{S}| \leq k$, for a given k, but does so on average over those $\mathcal{S}$. Now, take the matrices $\mathbf{A}^{(0)}, ..., \mathbf{A}^{(d-1)}$ to be obtained independently

by permuting the columns of $\mathbf{A}$. Then, if d is sufficiently large, with high probability (8) will be satisfied for all $\mathcal{S}$ with $|\mathcal{S}| \leq k$.

We next state our recovery results for MOMP and start by defining the following quantities, which are used to formulate "local" (i.e., pertaining to the (given) set $\mathcal{S}$) isometry conditions. These quantities were also used in [10], [11] in the performance analysis of MOMP for the MMV case.

For a given set $\mathcal{S} \subseteq \{0, ..., n-1\}$, let

$$\delta_i(\mathcal{S}) = \left\| (\mathbf{A}_{\mathcal{S}}^{(i)})^H \mathbf{A}_{\mathcal{S}}^{(i)} - \mathbf{I} \right\|_{2 \to 2}.$$

Observe that

$$(1 - \delta_i(\mathcal{S})) \|\mathbf{x}_{\mathcal{S}}\|_2^2 \leq \left\| \mathbf{A}_{\mathcal{S}}^{(i)} \mathbf{x}_{\mathcal{S}} \right\|_2^2 \leq (1 + \delta_i(\mathcal{S})) \|\mathbf{x}_{\mathcal{S}}\|_2^2$$

for all $\mathbf{x}_{\mathcal{S}} \in \mathbb{R}^s$. Define

$$\mu_i(\mathcal{S}) = \max \left\{ \max_{l \notin \mathcal{S}} \left\| (\mathbf{A}_{\mathcal{S}}^{(i)})^H \mathbf{a}_l^{(i)} \right\|_2, \max_{l \in \mathcal{S}} \left\| (\mathbf{A}_{\mathcal{S} \setminus l}^{(i)})^H \mathbf{a}_l^{(i)} \right\|_2 \right\}$$

and let $\delta_{\max}(\mathcal{S}) = \max_i \delta_i(\mathcal{S})$ and $\mu_{\max}(\mathcal{S}) = \max_i \mu_i(\mathcal{S})$.

Theorem 2: Fix $\mathcal{S} \subseteq \{0, ..., n-1\}$ with cardinality $s := |\mathcal{S}|$, let the measurement matrices $\mathbf{A}^{(0)}, ..., \mathbf{A}^{(d-1)} \in \mathbb{R}^{m \times n}$ have unit norm columns with $\mu_{\max}(\mathcal{S}) < 1$ and $\delta_{\max}(\mathcal{S}) < 1$, and let the entries of $\mathbf{x}_{\mathcal{S}}^{(0)}, ..., \mathbf{x}_{\mathcal{S}}^{(d-1)} \in \mathbb{R}^s$ be i.i.d. zero-mean ρ-sub-Gaussian with unit variance. If

$$\frac{\sum_{i=0}^{d-1} \left(\frac{\mu_i(\mathcal{S})}{1 - \delta_i(\mathcal{S})} \right)^2}{\sum_{i=0}^{d-1} \left(1 - \frac{\mu_i^2(\mathcal{S})}{1 - \delta_i(\mathcal{S})} \right)^2} \leq \frac{(1 - \beta)}{(1 + \beta)} \tag{13}$$

for β with $0 < \beta \leq 32e\rho$, then MOMP applied to $\mathbf{y}^{(i)} = \mathbf{A}^{(i)} \mathbf{x}^{(i)}, i = 0, ..., d-1$, recovers the correct solution $\mathbf{x}^{(0)}, ..., \mathbf{x}^{(d-1)}$ with probability at least

$$1 - 2^s (n + 1 - s) \exp \left(-d\beta^2 \frac{c(\mathcal{S}, \mathbf{A})}{2^{11} e^2 \rho^2} \right) \tag{14}$$

where $c(\mathcal{S}, \mathbf{A})$ is a constant that depends on the $\mathbf{A}^{(i)}$, but is independent of d.

Remark: The constant $c(\mathcal{S}, \mathbf{A})$ can be lower-bounded in terms of the $\mu_i(\mathcal{S})$ and $\delta_i(\mathcal{S})$, see [14].

The main implication of Theorem 2 is that, provided (13) is satisfied, the probability that MOMP fails decays exponentially in the number of measurements d. This has been shown before in [10], [11] for the MMV case, under the assumption of i.i.d. Gaussian $\mathbf{x}_{\mathcal{S}}^{(i)}$. The implications of Theorem 2 concerning improvements over the worst-case results and over the MMV case are as discussed above, for LOPT. Furthermore, as in the case of LOPT, Theorem 2 can be strengthened for certain distributions. For example, when the entries of $\mathbf{x}_{\mathcal{S}}^{(0)}, ..., \mathbf{x}_{\mathcal{S}}^{(d-1)}$ are i.i.d. standard Gaussian, Theorem 2 holds with Condition (13) replaced by

$$\frac{\sum_{i=0}^{d-1} \left(\frac{\mu_i(\mathcal{S})}{1 - \delta_i(\mathcal{S})} \right)^2}{\sum_{i=0}^{d-1} \left(1 - \frac{\mu_i^2(\mathcal{S})}{1 - \delta_i(\mathcal{S})} \right)^2} \leq \frac{(1 - \beta)^2 \varsigma^2}{(1 + \beta)^2} \tag{15}$$

for $\beta > 0$, where $\varsigma > 1$ is a constant that tends to 1 as d grows, and (14) replaced by

$$1 - 2^s \left((n - s) \exp \left(-d\beta^2 c(\mathcal{S}, \mathbf{A}) \right) + \exp \left(-d\beta^2 \varsigma^2 c(\mathcal{S}, \mathbf{A}) \right) \right). \tag{16}$$

We finally note that condition (13) is slightly stronger than condition (8) pertaining to LOPT [14].

B. The noisy GMMV problem

We next present our results for the noisy GMMV problem and start with the probabilistic analysis of POPT. For the following result, we assume that the entries of $\mathbf{x}_{\mathcal{S}}^{(0)}, ..., \mathbf{x}_{\mathcal{S}}^{(d-1)}$ are i.i.d. Rademacher random variables, i.e., they take on the values $+1$ and -1 with equal probability. We chose this model for convenience and note that similar results can be obtained for the sub-Gaussian case. The corresponding analysis is, however, much more cumbersome and does not yield additional insights.

Theorem 3: Fix $\mathcal{S} \subseteq \{0, ..., n-1\}$, with cardinality $s := |\mathcal{S}|$, and take the entries of $\mathbf{x}_{\mathcal{S}}^{(0)}, ..., \mathbf{x}_{\mathcal{S}}^{(d-1)} \in \mathbb{R}^s$ to be i.i.d. Rademacher. Suppose the measurement matrices $\mathbf{A}^{(0)}, ..., \mathbf{A}^{(d-1)} \in \mathbb{R}^{m \times n}$ satisfy conditions (8) and (9) for $\alpha < 1$ and some $\gamma > 0$. Suppose the noise level ϵ in (4) and γ satisfy

$$\left(c_3 \epsilon + \gamma c_4 \sqrt{|\mathcal{S}|} \right) \left(2c_2 + 1 - \frac{\epsilon}{\gamma} c_1 - \beta \right)$$
$$< \sqrt{d} \left(1 - \frac{\epsilon}{\gamma} c_1 - \xi \right) \tag{17}$$

where $c_1, c_2, c_3,$ and c_4 are constants depending on $\delta_{\max}(\mathcal{S})$ and $\mu_{\max}(\mathcal{S})$ only. Then, for $\xi > 0$ such that $\max\{1 - 16e, \alpha^2\} < \xi^2 \leq \alpha^2 (1 + 16e)$, with probability at least

$$1 - \exp \left(-d \frac{(\xi^2 - \alpha^2)^2}{512 e^2 \gamma^2 \alpha^2} \right) \tag{18}$$

the solution to POPT applied to $\mathbf{y}^{(i)} = \mathbf{A}^{(i)} \mathbf{x}^{(i)}, i = 0, ..., d-1$, and denoted by $\tilde{\mathbf{x}}^{(0)}, ..., \tilde{\mathbf{x}}^{(d-1)}$, is supported on $\mathcal{S}$ and satisfies

$$\left(\sum_{i=0}^{d-1} \left\| \tilde{\mathbf{x}}^{(i)} - \mathbf{x}^{(i)} \right\|_2^2 \right)^{1/2} \leq c_3 \epsilon + \gamma c_4 \sqrt{|\mathcal{S}|}. \tag{19}$$

The main implication of Theorem 3 is that, under certain conditions on the $\mathbf{A}^{(i)}$ and for the noise level ϵ sufficiently small, the probability that POPT produces a solution with correct support set that is "close" in ℓ_2-norm to the true $\mathbf{x}^{(i)}$, tends to 1 exponentially fast in d. This result is also new for the MMV case. Condition (17) ensures that the noise level ϵ is sufficiently small. Note that (17) depends on the "worst" measurement matrix through $\delta_{\max}(\mathcal{S})$ and $\mu_{\max}(\mathcal{S})$. This is sensible as noise has the largest effect on the measurement $\mathbf{y}^{(i)}$ taken through the "worst" measurement matrix.

We finally turn to the performance of MOMP. This result will be stated for i.i.d. sub-Gaussian $\mathbf{x}_{\mathcal{S}}^{(i)}$.

Theorem 4: Fix $\mathcal{S} \subseteq \{0, ..., n-1\}$ with cardinality $s := |\mathcal{S}|$, and let the measurement matrices $\mathbf{A}^{(0)}, ..., \mathbf{A}^{(d-1)} \in \mathbb{R}^{m \times n}$ have unit norm columns with $\mu_{\max}(\mathcal{S}) < 1$ and

$\delta_{\max}(\mathcal{S}) < 1$. Let the entries of $\mathbf{x}_{\mathcal{S}}^{(0)}, ..., \mathbf{x}_{\mathcal{S}}^{(d-1)} \in \mathbb{R}^s$ be i.i.d. zero-mean ρ-sub-Gaussian with unit variance. Suppose that

$$\epsilon \leq \frac{1 - \delta_{\max}(\mathcal{S})}{1 - \delta_{\max}(\mathcal{S}) + (1 - \delta_{\max}(\mathcal{S}))\mu_{\max}(\mathcal{S})}\varkappa \qquad (20)$$

for some $\varkappa \geq 0$. If

$$\sqrt{1 - \beta} \left(\frac{1}{d} \sum_{i=0}^{d-1} \left(1 - \frac{\mu_i^2(\mathcal{S})}{1 - \delta_i(\mathcal{S})} \right)^2 \right)^{1/2}$$
$$- \sqrt{1 + \beta} \left(\frac{1}{d} \sum_{i=0}^{d-1} \left(\frac{\mu_i(\mathcal{S})}{1 - \delta_i(\mathcal{S})} \right)^2 \right)^{1/2} \geq \varkappa \quad (21)$$

for β satisfying $0 < \beta \leq 32e\rho$, then with probability at least (14), MOMP applied to $\mathbf{y}^{(i)} = \mathbf{A}^{(i)}\mathbf{x}^{(i)}, i = 0, ..., d - 1$, yields an estimate of the $\mathbf{x}^{(i)}$, denoted by $\tilde{\mathbf{x}}^{(i)}$, that is supported on $\mathcal{S}$ and satisfies

$$\left(\sum_{i=0}^{d-1} \left\| \tilde{\mathbf{x}}^{(i)} - \mathbf{x}^{(i)} \right\|_2^2 \right)^{1/2} \leq \frac{1 + \delta_{\max}(\mathcal{S})}{1 - \delta_{\max}(\mathcal{S})}\epsilon. \qquad (22)$$

Again, the main implication of Theorem 4 is that, under certain mild conditions on the $\mathbf{A}^{(i)}$ and for the noise level ϵ sufficiently small, the probability that MOMP produces a solution with correct support set that is "close" to the true $\mathbf{x}^{(i)}$, tends to 1 exponentially fast in d. This was shown in [11] for the MMV case and for i.i.d. Gaussian $\mathbf{x}_{\mathcal{S}}^{(i)}$. Note that for $\epsilon = 0$, i.e., in the noiseless case, (21) reduces to (13) and Theorem 4 reduces to Theorem 2. For $\epsilon > 0$, and hence $\varkappa > 0$, (21) is more restrictive than Condition (13). Condition (20) depends on the "worst" measurement matrix, and ensures that the noise level ϵ is sufficiently small. The constants in Theorem (4) can be improved for i.i.d. Gaussian $\mathbf{x}_{\mathcal{S}}^{(i)}$ [14].

We conclude by noting that the results in this paper extend straightforwardly to the case of complex $\mathbf{A}^{(i)}$ and $\mathbf{x}^{(i)}$.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[2] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proc. Natl. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[3] M. Elad, *Sparse and redundant representations: From theory to applications in signal and image processing.* Springer, 2010.

[4] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, 2005.

[5] 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.

[6] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Process.*, vol. 86, no. 3, pp. 589–602, Mar. 2006.

[7] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, 2006.

[8] M. E. Davies and Y. C. Eldar, "Rank awareness in joint sparse recovery," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1135–1146, 2012.

[9] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constr. Approx.*, vol. 13, no. 1, pp. 57–98, Mar. 1997.

[10] 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.

[11] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, "Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms," *J. Fourier Anal. Appl*, vol. 14, no. 5, pp. 655–687, 2008.

[12] Y. M. Lu and M. N. Do, "A theory for sampling signals from a union of subspaces," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2334–2345, 2008.

[13] Y. C. Eldar, "Compressed sensing of analog signals in shift-invariant spaces," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 2986–2997, 2009.

[14] R. Heckel and H. Bölcskei, "The generalized multiple measurement vector problem," *in preparation*, 2012.

[15] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, 2010.

[16] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[17] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 58, no. 1, pp. 267–288, Jan. 1996.

[18] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 68, no. 1, pp. 49–67, 2006.

[19] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[20] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," in *Compressed Sensing Theory and Applications*, Y. C. Eldar and G. Kutyniok, Eds. Cambridge Univ. Press, 2012, pp. 210–268.