Rank Awareness in Joint Sparse Recovery

Rank Awareness in Joint Sparse Recovery

Mike E. Davies, *Member IEEE* and Yonina C. Eldar, *Senior Member, IEEE*

Abstract

This paper revisits the sparse multiple measurement vector (MMV) problem, where the aim is to recover a set of jointly sparse multichannel vectors from incomplete measurements. This problem is an extension of single channel sparse recovery, which lies at the heart of compressed sensing. Inspired by the links to array signal processing, a new family of MMV algorithms is considered that highlight the role of rank in determining the difficulty of the MMV recovery problem. The simplest such method is a discrete version of MUSIC which is guaranteed to recover the sparse vectors in the full rank MMV setting, under mild conditions. This idea is extended to a *rank aware* pursuit algorithm that naturally reduces to Order Recursive Matching Pursuit (ORMP) in the single measurement case while also providing guaranteed recovery in the full rank setting. In contrast, popular MMV methods such as Simultaneous Orthogonal Matching Pursuit (SOMP) and mixed norm minimization techniques are shown to be *rank blind* in terms of worst case analysis. Numerical simulations demonstrate that the rank aware techniques are significantly better than existing methods in dealing with multiple measurements.

Index Terms

Compressed sensing, multiple measurement vectors, rank, sparse representations

I. Introduction

Sparse signal representations provide a general signal model that represents or approximates a signal using a linear combination of a small number of elementary waveforms (called atoms) selected from a large collection (the dictionary). Such models make it possible to solve many ill-posed problems such as source separation, denoising and most recently compressed sensing [1], [2] by exploiting the additional sparsity constraint. The key point is that when the signal, $\mathbf{x}$, is sufficiently sparse it can still be uniquely determined from an underdetermined set of measurements $\mathbf{y} = \mathbf{\Phi x}$, where $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ and $m < n$.

The problem of finding the sparsest $\mathbf{x}$ consistent with a given observation vector $\mathbf{y}$ is known to be NP-hard in general [3], [4] and therefore is presumed to not be solvable in polynomial time. Instead, various suboptimal

M. E. Davies is with the Institute for Digital Communication, Edinburgh University, Edinburgh EH9 3JL, U.K. (e-mail: mike.davies@ed.ac.uk).
MED acknowledges support of his position from the Scottish Funding Council and their support of the Joint Research Institute in Signal and Image Processing with the Heriot-Watt University as a component of the Edinburgh Research Partnership. This work was supported in part by EPSRC grant EP/F039697/1 and the European Commission through the SMALL project under FET-Open, grant number 225913.

Y. C. Eldar is with the Technion—Israel Institute of Technology, Haifa 32000, Israel. Phone: +972-4-8293256, fax: +972-4-8295757, E-mail: yonina@ee.technion.ac.il. She is currently on leave at Stanford, USA. Her work was supported in part by the Israel Science Foundation under Grant no. 170/10 and by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM++ (contract no. 216715).

strategies have been proposed and have been demonstrated, both empirically and theoretically, to have good reconstruction performance in a range of settings. Commonly used strategies are typically based on convex relaxation [5], non-convex local optimisation [6] or greedy search strategies [3], [4], [7], [8], [9].

Encouraged by the potential power of sparse representations, researchers have begun to consider a number of extensions to the basic sparse representation model. These include the multiple measurement vector (MMV) problem [10], [11], [12], [13], [14], as well as other union of subspace models [15], [16], [17] such as block sparsity or tree structured sparsity [17], [18], [19] and blind compressed sensing [20]. These ideas have also been recently expanded to include sub-Nyquist sampling of structured analog signals [21]. As with the single measurement vector problem (SMV) several suboptimal methods for finding a sparse matrix solution to the MMV problem have been proposed, that have polynomial complexity [11], [22], [23], [24], [10], [12], [13], [14]. These approaches are generally straightforward extensions of existing SMV solutions and can be roughly divided into greedy methods, and algorithms based on mixed norm optimization. We will discuss these two classes in Section V. One exception to this is the approach in [13] which reduces the MMV problem to a single channel recovery via a random projection that preserves the sparsity pattern.

A variety of different equivalence results between finding the sparsest solution, the so-called ℓ_0-problem, and the output of the proposed efficient algorithms have also been derived. In [12] an equivalence result was obtained for a mixed $\ell_{p,1}$ program in which the objective is to minimize the sum of the ℓ_p-norms of the rows of the estimated matrix whose columns are the unknown vectors. The condition is based on mutual coherence, and turns out to be the same as that obtained from a single measurement problem, so that the joint sparsity pattern does not lead to improved recovery capabilities as judged by this condition. Recovery results for the more general problem of block-sparsity were developed in [19], [17] based on the RIP, and in [18] based on mutual coherence. However reducing these results to the MMV setting leads again to conditions that are the same as in the single measurement case. An exception is the work in [25], [26], [14] which considers average case performance assuming that $\mathbf{X}$ is generated at random from an appropriate distribution. Under a mild condition on the sparsity and on the matrix $\mathbf{\Phi}$, the probability of reconstruction failure decays exponentially with the number of channels l. However, to date, all worst case recovery results have not shown any advantage to the MMV setting over the SMV case. The reason is that in the worst case, the matrix $\mathbf{X}$ may be comprised of a single repeated vector $\mathbf{x}$, in which case effectively the MMV and SMV problems becomes identical, hence the worst case results are not capable of improving the recovery guarantees.

As noted above, one approach to demonstrate the advantage of the MMV formulation is to use an average case analysis where $\mathbf{X}$ is generated at random. Since repeated columns are not likely to occur, this strategy allows for improved recovery guarantees. In this paper, we concentrate on worst-case performance, as in the bulk of prior work on SMV problems. We show that we can break the worst-case analysis bottleneck by exploiting the rank of the matrix $\mathbf{X}$. Although in the case of a rank-one matrix we cannot do better than in SMV recovery, this is no longer true when $\mathbf{X}$ has higher rank. In particular, when the rank of $\mathbf{X}$ is equal to k, we highlight the fact that it can be

recovered exactly from the measurements $\mathbf{Y}$ under a mild condition[1] on $\mathbf{\Phi}$ using polynomial time algorithms from only $m = k + 1$ measurements per signal [27] based on the MUSIC algorithm popular in array signal processing [28]. Clearly this is a big advantage over the SMV problem which cannot guarantee perfect recovery for general $\mathbf{\Phi}$ and all $\mathbf{x}$ with such few measurements. Even using combinatorial algorithms, recovery of all $\mathbf{x}$ is possible only if the number of measurements is at least $2k$.

Interestingly, the links between sparse representations and array signal processing were brought forward early on in the sparse reconstruction literature [6] and the MMV problem in particular has origins in the field of array signal processing. However, the algorithmic connections and particularly the role of the rank of $\mathbf{X}$, appear to have been missed.

The main contribution of this paper is to demonstrate how the rank of $\mathbf{X}$ can be exploited in order to improve MMV recovery results in the worst-case setting. We begin by examining the conditions for which the equation $\mathbf{Y} = \mathbf{\Phi X}$ has a single sparse solution. In Section III, we derive a necessary and sufficient condition for uniqueness and show that it depends directly on the rank of $\mathbf{X}$: the larger the rank the less sparse $\mathbf{X}$ needs to be to still ensure uniqueness. Similarly, in Section IV, we show that the computational effort required to find the unique sparse solution through a combinatorial search is also dependent on the rank of $\mathbf{X}$, and is reduced when the rank increases. In Sections V and VI we turn to discuss polynomial time recovery algorithms. We begin by showing that common MMV methods are not rank aware, namely, they do not efficiently exploit the rank of $\mathbf{X}$ to improve recovery results. In particular, in the full rank case in which the rank of $\mathbf{X}$ is equal to k (which is the largest it can be) we show that most prevalent MMV algorithms do not provide recovery for the worst-case choice of $\mathbf{X}$ from $k+1$ measurements. Moreover, independent of the rank of $\mathbf{X}$, one can find examples that perform almost as badly as worst case SMV problems. We proceed to propose some *rank aware* (RA) algorithms inspired by the popular MUSIC technique, whose behavior improves with increasing rank of $\mathbf{X}$, and are proven to provide exact recovery for all choices of $\mathbf{X}$ when the rank is equal to k, and the number of measurements is $k+1$ (under mild conditions on $\mathbf{\Phi}$). Finally, in Section VII, we present several simulations demonstrating the behavior of these different methods.

Connection with other works

Upon completion of this paper (preprint available on ArXiv [29]) we became aware of recent and independent work by other researchers who, in a similar spirit to this work, have noticed and sought to exploit the link between array signal processing and the MMV sparse problem [30], [31], [32].

In [30], [31] the authors also noted that there is a gap between classical MMV solvers and the guarantees provided by the MUSIC algorithm. Furthermore both papers point out that the MUSIC algorithm could still be applied even when $\mathrm{rank}(\mathbf{X}) < k$ if a partial support set is known. This idea is similar to the reduced combinatorial search that is presented in Section IV. However, the ingenious step noted in [30] and [31] is that the partial support can be estimated using a standard greedy algorithm.

[1]The condition is mild in that it is satistied for almost all $\mathbf{\Phi}$. However verifying such a condition is in practice impossible.

The greedy search proposed in [30], called subspace simultaneous orthogonal matching pursuit (SM-OMP), is similar to Rank Aware OMP (RA-OMP) presented in Section VI. A subtle difference is that SM-OMP invokes a single orthogonalization at the initialization of the algorithm while RA-OMP orthogonalizes the residual matrix at each step. In [30] the authors also propose an additional iterative correction to the augmented MUSIC algorithm based loosely on the ideas of the SMV Subspace Pursuit algorithm. However, the augmented MUSIC requires an estimate for the joint sparsity level of $\mathbf{X}$ in order to know how much of the support set to select in a greedy manner before applying MUSIC. This difficulty is avoided in our proposed RA-ORMP.

In [31] it is noted that any greedy algorithm can in principle be used to select the partial support. The authors then concentrate on the standard simultaneous OMP (SOMP) and a simple thresholding algorithm. We show in Section V that such algorithms are rank blind and therefore expect that they will not perform as well as the RA-ORMP proposed in Section VI or the augmented MUSIC algorithm proposed in [30].

Both [30] and [31] provide performance analysis with a compressed sensing flavour. This is particularly appropriate for when the matrix $\mathbf{\Phi}$ is random. In [30] the performance of SM-OMP is quantified in terms of the restricted isometry property (RIP) of $\mathbf{\Phi}$ and it is shown that the performance bound improves with increasing rank, although it does not reach the performance of MUSIC in the full rank case. This is broadly in line with our own analysis for RA-OMP via a different route. In [31] the authors present performance results for the augmented MUSIC in terms of RIP that also considers noise. They adopt a large scale system model, assuming that $\mathbf{\Phi}$ is Gaussian and that n and m tend to infinity.[2] In [32] Fannjiang explores the performance of MUSIC in the inverse scattering problem with sparse objects. Here again the RIP is applied to a simplified discretization of the original problem. Although no direct reference to the sparse MMV problem is made the derived sensitivity analysis is of interest for compressed sensing matrices $\mathbf{\Phi}$ in the noisy setting.

In contrast to the compressed sensing analsysis in [30], [31], [32], our algorithm analysis in Sections V and VI focusses directly on the worst case scenario with no restriction on the nature of the measurement matrix $\mathbf{\Phi}$. The price for this generality is that our results are more qualitative in nature. In this sense the different analyses are very complementary.

Finally we mention that there has been significant recent work on rank minimization, e.g. [34]. While on the surface, there might appear to be similarities with the MMV problem studied here, the problems are significantly different. Specifically we are generically able to observe the rank of $\mathbf{X}$ through the observation matrix $\mathbf{Y}$ and we can use this to find the unknown sparse support set. In contrast in rank minimization problems it is the unknown rank that is being sought and which therefore plays an equivalent role to sparsity in the MMV problem.

[2]Currently there appear to be some difficulties in a key proof in [31] and it is not clear whether the large scale system results are valid [33].

II. Notation and Problem Formulation

A. Notation

Throughout this paper a coefficient vector $\mathbf{x}$ is said to be k-sparse if the size of the support of $\mathbf{x}$ is no larger than k: $|\operatorname{supp}(\mathbf{x})| \leq k$. We define the support of a collection of vectors $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_l]$ as the union over all the individual supports:

$$\operatorname{supp}(\mathbf{X}) := \bigcup_i \operatorname{supp}(\mathbf{x}_i). \tag{1}$$

A matrix $\mathbf{X}$ is called k joint sparse if $|\operatorname{supp}(\mathbf{X})| \leq k$. In other words, there are at most k rows in $\mathbf{X}$ that contain nonzero elements. The set of k joint sparse matrices of size $n \times l$ is defined as

$$\mathcal{J}_k := \{\mathbf{X} \in \mathbb{R}^{n \times l} : |\operatorname{supp}(\mathbf{X})| \leq k\}. \tag{2}$$

We make use of the subscript notation $\mathbf{x}_\Omega$ to denote a vector that is equal to some $\mathbf{x}$ on the index set Ω and zero everywhere else. Denoting by $|\Omega|$ the cardinality of Ω, the vector $\mathbf{x}_\Omega$ is $|\Omega|$-sparse. For index sets that are sequences of indices, e.g. $1, \ldots, k$ we use the Matlab style notation $1 : k$.

We say that the support of $\mathbf{x}$ lies within Ω whenever $\mathbf{x}_\Omega = \mathbf{x}$. For matrices the subscript notation $\mathbf{\Phi}_\Omega$ will denote a submatrix composed of the columns of $\mathbf{\Phi}$ that are indexed in the set Ω, while the notation $\mathbf{X}_{\Omega,:}$ denotes a row-wise submatrix composed of the rows of $\mathbf{X}$ indexed by Ω. We denote the ith column of a matrix, $\mathbf{\Phi}$, by $\boldsymbol{\phi}_i$, and use $\mathcal{N}(\mathbf{\Phi})$ for the null space and $\mathcal{R}(\mathbf{\Phi})$ for the range of the matrix $\mathbf{\Phi}$.

Throughout the paper, we also require a variety of different norms. We denote by $\|\mathbf{x}\|_p$, $p \geq 1$ the usual ℓ^p norms, and by $\|\mathbf{x}\|_0$ the ℓ^0 quasi-norm that counts the number of non-zero elements of $\mathbf{x}$ so that $\|\mathbf{x}\|_0 = |\operatorname{supp}(\mathbf{x})|$. For matrices we define the $\ell_{p,q}$ norms as:

$$\|\mathbf{X}\|_{p,q} := \left(\sum_i \|\mathbf{X}_{i,:}\|_q^p \right)^{1/p} \tag{3}$$

where, with slight abuse of notation, we also consider the quasi-norms with $p = 0$ such that $\|\mathbf{X}\|_{0,q} = |\operatorname{supp}(\mathbf{X})|$ for any q.

B. MMV sparse recovery problem

We are interested in solving the sparse recovery problem associated with MMV, in which the goal is to recover a jointly sparse matrix $\mathbf{X}$ of size $n \times l$ from $m < n$ measurements per channel. Here l denotes the number of channels, or signals. This is a generalization of the standard SMV problem that has been examined in detail, e.g. [35], [7]. Formally, the problem is defined as follows:

Definition 1 (MMV sparse recovery problem)**.** *Given $\mathbf{Y} \in \mathbb{R}^{m \times l}$ and $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ with $m < n$ find:*

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmin}} |\operatorname{supp}(\mathbf{X})| \text{ s.t. } \mathbf{\Phi}\mathbf{X} = \mathbf{Y}. \tag{4}$$

The SMV problem can be recovered as a special case with $l = 1$.

Throughout the paper we assume that the dictionary $\mathbf{\Phi}$ has unit norm columns, $\|\boldsymbol{\phi}_i\|_2 = 1$ and focus on the ideal model in which the measurements $\mathbf{Y}$ are noiseless, and $\mathbf{X}$ is strictly sparse. For the important issue of sensitivity to noise we direct the reader to [31], [32].

III. MMV Uniqueness

While the MMV problem reduces to the SMV one when each observation vector $\mathbf{Y}_{i,:}$ is colinear, in general, the additional measurement vectors should provide further information and make it easier to determine the support set. The question is: by how much? In this section we focus on uniqueness conditions and show how the rank of $\mathbf{X}$, or alternatively that of $\mathbf{Y}$, can be used to improve the uniqueness conditions. In later sections we will use the rank to develop more efficient recovery algorithms with improved worst-case guarantees over the SMV setting.

In the SMV setting it is well known that a necessary and sufficient condition for the measurements $\mathbf{y} = \mathbf{\Phi x}$ to uniquely determine each k-sparse vector $\mathbf{x}$ is given by

$$k < \frac{\operatorname{spark}(\mathbf{\Phi})}{2} \tag{5}$$

where the *spark* of $\mathbf{\Phi}$ is defined as the smallest number of columns of $\mathbf{\Phi}$ that are linearly dependent. Since $\operatorname{spark}(\mathbf{\Phi}) \leq m + 1$, we have immediately that $m \geq 2k$, namely, at least $2k$ measurements are needed to ensure uniqueness in the SMV case for all possible choices of $\mathbf{x}$. This also defines a necessary and sufficient condition for general MMV sparse recovery, since one instance of this problem is any SMV problem replicated l times [12].

Chen and Huo [12] showed that when $\operatorname{rank}(\mathbf{Y}) > 1$ the sufficient condition for uniqueness in the MMV sparse recovery problem can be relaxed by exploiting the rank of $\mathbf{Y}$, as incorporated in the following theorem.

Theorem 1 (Chen and Huo [12])**.** *A sufficient condition from the measurements* $\mathbf{Y} = \mathbf{\Phi X}$*,* $|\operatorname{supp}(\mathbf{X})| = k$*, to uniquely determine the jointly sparse matrix* $\mathbf{X} \in \mathcal{J}_k$ *is*

$$|\operatorname{supp}(\mathbf{X})| < \frac{\operatorname{spark}(\mathbf{\Phi}) - 1 + \operatorname{rank}(\mathbf{Y})}{2}. \tag{6}$$

This condition was further shown in [13] to hold even in the case where there are infinitely many vectors $\mathbf{y}_i$. A direct consequence of Theorem 1 is that matrices $\mathbf{X}$ which result in matrices $\mathbf{Y}$ with larger rank, can be recovered from fewer measurements. Alternatively, matrices $\mathbf{X}$ with larger support can be recovered from the same number of measurements. Since $\operatorname{rank}(\mathbf{X}) \leq k$, it is obvious that $\operatorname{rank}(\mathbf{Y}) \leq k$. When $\operatorname{rank}(\mathbf{Y}) = k$ and $\operatorname{spark}(\mathbf{\Phi})$ takes on its largest value of $m + 1$, condition (6) becomes $m \geq k + 1$. Therefore, in this best-case scenario, only $k + 1$ measurements per signal are needed to ensure uniqueness. This is much lower than the value of $2k$ obtained in the SMV setting.

Chen and Huo note that it would also be interesting to bound k in terms of $\operatorname{rank}(\mathbf{X})$ instead of $\operatorname{rank}(\mathbf{Y})$. Naturally we have that $\operatorname{rank}(\mathbf{Y}) \leq \operatorname{rank}(\mathbf{X})$. The next lemma shows that we can actually replace $\operatorname{rank}(\mathbf{Y})$ by $\operatorname{rank}(\mathbf{X})$ in condition (6):

Lemma 1. *The sufficient condition of* (6) *is equivalent to*

$$|\operatorname{supp}(\mathbf{X})| < \frac{\operatorname{spark}(\mathbf{\Phi}) - 1 + \operatorname{rank}(\mathbf{X})}{2}. \tag{7}$$

Proof: Since $\operatorname{rank}(\mathbf{Y}) \leq \operatorname{rank}(\mathbf{X})$, (6) automatically implies (7). For the reverse direction, suppose that (7) holds. Since $\operatorname{rank}(\mathbf{X}) \leq |\operatorname{supp}(\mathbf{X})|$, we have from (7) that

$$|\operatorname{supp}(\mathbf{X})| < \operatorname{spark}(\mathbf{\Phi}) - 1. \tag{8}$$

Therefore $\mathbf{\Phi}_\Omega$, $\Omega = \operatorname{supp}(\mathbf{X})$, must be full rank. This implies that $\operatorname{rank}(\mathbf{X}) = \operatorname{rank}(\mathbf{Y})$ and consequently (7) implies (6). ■

We now complete Theorem 1 by showing that both (7) and (6) are necessary and sufficient for uniqueness in the MMV problem.

Theorem 2. *Condition* (7)*, or equivalently* (6)*, is a necessary and sufficient condition for the measurements* $\mathbf{Y} = \mathbf{\Phi X}$ *to uniquely determine the jointly sparse matrix* $\mathbf{X} \in \mathcal{J}_k$.

An immediate consequence of the theorem is that in order to have a unique $\mathbf{X}$ with support set of size k when $\mathbf{X}$ has full rank (i.e. $\operatorname{rank}(\mathbf{X}) = k$), it is enough to take $m = k + 1$ measurements, as long as $\operatorname{spark}(\Phi) \geq k + 2$, namely, that every set of $k + 1$ columns of $\mathbf{\Phi}$ are linearly independent.

Interestingly the $m \geq k + 1$ bound also occurs in the SMV case, where for almost all dictionaries $\mathbf{\Phi}$ almost all k sparse vectors $\mathbf{x}$ are uniquely determined by $\mathbf{y} = \mathbf{\Phi x}$ if $m \geq k + 1$ (for apropriately defined measures) - see [11], [16]. The key difference in the MMV scenario is that the condition on $\mathbf{X}$ that guarantees recovery using only $k + 1$ measurements is readily testable from the observed data, namely that $\operatorname{rank}(\mathbf{Y}) = k$.

Proof: Sufficiency follows immediately from Lemma 1 and Theorem 1.

To show necessity we will show that $2k \geq \operatorname{spark}(\mathbf{\Phi}) - 1 + \tau$ implies that there exists an $\mathbf{X} \in \mathbb{R}^{n \times l}$ with $\operatorname{rank}(\mathbf{X}) = \tau$ that is not uniquely determined by $\mathbf{Y} = \mathbf{\Phi X}$.

Suppose that $2k \geq \operatorname{spark}(\mathbf{\Phi}) - 1 + \tau$. Then there exists a support set T, $|T| = 2k - \tau + 1$ and a vector $\mathbf{v} \neq 0$ such that $\mathbf{\Phi}_T \mathbf{v} = 0$. Let $\mathbf{V} = [\mathbf{v}, \ldots, \mathbf{v}]$ be the matrix consisting of l replications of $\mathbf{v}$. We can now construct an $\mathbf{X}$ with $\operatorname{supp}(\mathbf{X}) \subset T$, as follows:

$$\mathbf{X}_{T,:} = \begin{bmatrix} \mathbf{V}_{\{1:k-\tau+1,:\}} \\ \mathbf{I}_{\tau-1} \quad \mathbf{0} \\ \mathbf{0}_{\{k-\tau+1 \times l\}} \end{bmatrix} \tag{9}$$

where $\mathbf{I}_{\tau-1}$ represents the identity matrix of size $\tau - 1$. Without loss of generality we can assume that $\mathbf{v}_{1:k-\tau+1} \neq 0$ by re-ordering the indices if necessary, then, by construction, $\mathbf{X}$ will k-joint sparse and $\operatorname{rank}(\mathbf{X}) = \tau$. However

Algorithm 1 discrete MUSIC

1: Calculate $\mathbf{U} = \text{orth}(\mathbf{Y})$ an orthonormal basis for $\mathcal{R}(\mathbf{Y})$;

2: Calculate $\Omega = \{i : \|\boldsymbol{\phi}_i^T \mathbf{U}\|_2 \geq \theta_k\}$ where the threshold, θ_k, is set to select the k largest values;

3: $\hat{\mathbf{X}} = \boldsymbol{\Phi}_\Omega^\dagger \mathbf{Y}$.

we can also define the matrix $\tilde{\mathbf{X}}$, with $\text{supp}(\tilde{\mathbf{X}}) \subset T$, by:

$$\tilde{\mathbf{X}}_{T,:} = \mathbf{X}_{T,:} - \mathbf{V}. \tag{10}$$

By construction, $\tilde{\mathbf{X}}$ is also k-joint sparse, and $\boldsymbol{\Phi}\mathbf{X} = \boldsymbol{\Phi}\tilde{\mathbf{X}}$. Therefore, it follows that (7) is also a necessary condition for uniqueness. ■

Theorem 2 shows that $\text{rank}(\mathbf{X})$ plays an important role in uniqueness of MMV problems, since the rank defines the dimension of the subspace of sparse coefficients that generate observation vectors. In the ensuing sections we will further show that $\text{rank}(\mathbf{X})$ also plays an important role in the performance of joint sparse recovery algorithms.

IV. MMV Recovery: Exhaustive Search

Our goal now is to show how the rank of $\mathbf{X}$ can be used to improve MMV recovery. We begin by considering the (generally impractical) exhaustive search approach for $\mathbf{X}$ using combinatorial optimization. That is we seek to minimize the number of nonzero rows of $\mathbf{X}$ subject to the constraint $\boldsymbol{\Phi}\mathbf{X} = \mathbf{Y}$. A solution can be found from an exhaustive search through all $\binom{n}{k}$ support sets Ω with $|\Omega| = k$. However this search can be significantly reduced when we have additional rank information. We begin by considering the full rank case.

A. Full-Rank MMV: MUSIC

Surprisingly, in the case in which $\mathbf{Y}$ has rank equal to k, an exact solution can be found using a very simple linear time search algorithm. This method incorporates the rank information of $\mathbf{Y}$ to efficiently determine $\mathbf{X}$. We refer to this scenario as the full rank case, since, we always have that $\text{rank}(\mathbf{Y}) \leq \text{rank}(\mathbf{X}) \leq k$. The last inequality is a result of the fact that $\mathbf{X}$ is k-sparse.

A rank-aware algorithm for the case in which $\text{rank}(\mathbf{Y}) = k$, can be obtained by using a discrete version of the MUSIC algorithm [28], popular in array signal processing. This technique was first proposed for solving a discrete sparse MMV problem by Feng and Bresler [27], [36] in the context of multiband sampling.

Since, by assumption, $\text{rank}(\mathbf{Y}) = k$, it follows immediately that $\mathcal{R}(\mathbf{Y}) = \mathcal{R}(\boldsymbol{\Phi}_\Omega)$ and hence every correct column $\boldsymbol{\phi}_i$ of $\boldsymbol{\Phi}$, lies within the range of $\mathbf{Y}$. The discrete MUSIC algorithm [27], [36] exploits this property by first calculating an orthonormal basis $\mathbf{U} = \text{orth}(\mathbf{Y})$ for $\mathcal{R}(\mathbf{Y})$ and then selecting the k columns of $\boldsymbol{\Phi}$ that minimize $\|(\mathbf{I} - \mathbf{U}\mathbf{U}^T)\boldsymbol{\phi}_i\|_2$. The algorithm is summarized in Algorithm 1.

If we further assume that the uniqueness condition of Theorem 2 is satisfied (i.e. $\text{spark}(\boldsymbol{\Phi}) > k + 1$), then only

the correct columns lie in $\mathcal{R}(\mathbf{Y})$ and are therefore orthogonal to the null space of $\mathbf{Y}$.[3] That is:

$$\|(\mathbf{I} - \mathbf{U}\mathbf{U}^T)\boldsymbol{\phi}_i\|_2 = 0, \text{ if and only if } i \in \Omega. \tag{11}$$

We summarize this result in the following theorem.

Theorem 3 (Feng [27])**.** *Let* $\mathbf{Y} = \mathbf{\Phi}\mathbf{X}$ *with* $|\operatorname{supp}(\mathbf{X})| = k$*,* $\operatorname{rank}(\mathbf{X}) = k$ *and* $k < \operatorname{spark}(\mathbf{\Phi}) - 1$*. Then* discrete MUSIC *is guaranteed to recover* $\mathbf{X}$ *(i.e.* $\hat{\mathbf{X}} = \mathbf{X}$*).*

A direct consequence of the theorem is that $m = k + 1$ measurements are sufficient to recover $\mathbf{X}$, as long as $\mathbf{\Phi}$ has maximal spark, namely, that all sets of $k + 1$ columns are linearly independent.

As noted in [27], [36], due to the stability of invariant subspaces [37] MUSIC based recovery is also robust to noise and other perturbations. However, this requires that Algorithm 1 be modified to make a practical estimate of the rank of the signal subspace and to separate the signal and noise subspaces. This is typically done through an eigenvalue decomposition of the covariance matrix $\mathbf{Y}\mathbf{Y}^T$ where the eigenvalues are used to distinguish between the signal and noise subspaces [38].

We can still apply the discrete MUSIC algorithm when $\operatorname{rank}(\mathbf{Y}) < k$, however we are no longer guaranteed recovery from Theorem 3. In the reduced rank case, as in the noisy case, care must be taken in determining the best choice of the threshold. This will typically depend on the conditioning of the submatrices of $\mathbf{\Phi}$ as well as the dynamic range of the sparse components: see [31], [32] for details.

B. Reduced-Rank MMV

When $\tau = \operatorname{rank}(\mathbf{Y}) < k$ guaranteed recovery is possible through a reduced combinatorial search. If $\tau < k$ then $\mathcal{R}(\mathbf{Y}) \subset \mathcal{R}(\mathbf{\Phi}_\Omega)$ and there must exist at least one subset $\gamma \subset \Omega$ such that $|\gamma| = k - \tau$ and $\boldsymbol{\phi}_i \notin \mathcal{R}(\mathbf{Y})$. In fact typically *all* $k - \tau$ sized support sets $\gamma \subset \Omega$ will satisfy this. Assuming that the identifiability condition of Theorem 2 is again met, we then have that $\mathcal{R}([\mathbf{\Phi}_\gamma, \mathbf{Y}]) = \mathcal{R}(\mathbf{\Phi}_\Omega)$. Since $|\gamma| = k - \tau$ we only need to search over subsets of size $k - \tau$. Of course there are 'only' $\binom{n}{k-\tau}$ such support sets, so this search still has exponential complexity unless $k \approx \tau$.

Specifically let $Q(\gamma) = \operatorname{orth}([\mathbf{\Phi}_\gamma, \mathbf{Y}])$. Then an optimal γ can be found by solving:

$$\hat{\gamma} = \underset{\gamma, |\gamma| = k-\tau}{\operatorname{argmin}} \|\mathbf{\Phi}_{\gamma^c}^T(\mathbf{I} - Q(\gamma)Q(\gamma)^T)\|_{0,q} \tag{12}$$

for an arbitrary q. This counts the nonzero columns of $\mathbf{\Phi}_{\gamma^c}$ projected onto the null space of $Q(\gamma)$ and will achieve a minimum of $n-k$ when $Q(\gamma)$ spans the range of $\mathbf{\Phi}_\Omega$. The correct support set can then be recovered by considering the full rank problem associated with the augmented measurement matrix $[\mathbf{\Phi}_{\hat{\gamma}}, \mathbf{Y}]$. Generally the optimal solution to (12) will not be unique and there will typically be multiple solutions associated with the $\binom{k}{k-\tau}$ subsets of the

[3]Traditionally MUSIC projects the columns into the null space of $\mathbf{Y}$, however, since we assume that $\|\phi_i\|_2 = 1$ we can equally project into the range of $\mathbf{Y}$ as in Algorithm 1.

true support. This presence of multiple equivalent minima here suggests that such a problem might be difficult to convexify.

As noted in the introduction, independent work [31], [30] recently proposed an ingenious hybrid MMV recovery scheme using similar ideas where the reduced combinatorial search is replaced by a greedy selection strategy in order to find the partial support set γ until it is possible to apply an augmented MUSIC. In Section VI we propose a different solution. First, however we consider the effect of rank on some of the popular MMV sparse recovery algorithms.

V. Rank-Blind MMV Algorithms

Despite the fact that the rank of $\mathbf{X}$ (and therefore that of $\mathbf{Y}$) plays an important role in the MMV problem, we next show that some of the most popular algorithms are effectively *rank blind*. Namely, they do not allow for perfect recovery in the full rank case, and furthermore the worst case behaviour of such algorithms approaches that of the SMV problem. In practice, as we will see in Section VII, such algorithms often exhibit improved (average case) performance when there are multiple measurements, however, they suffer from not properly exploiting the rank information.

A. Greedy Methods

Several greedy algorithms have been proposed to treat the MMV problem. Two examples are extensions of thresholding and OMP to the multiple measurement case [23], [39], [25], [14], [18]. For $1 \leq q \leq \infty$ they produce a k-sparse signal $\hat{\mathbf{X}}$ from measurements $\mathbf{Y} = \mathbf{\Phi X}$ using a greedy search.

In q-thresholding, we select a set Ω of k indices whose q-correlation with $\mathbf{Y}$ are among the k largest. Let θ_k be the kth largest q-correlation of any $\boldsymbol{\phi}_i$ with $\mathbf{Y}$. Then:

$$\Omega = \{i : \|\boldsymbol{\phi}_i^T \mathbf{Y}\|_q \geq \theta_k\}. \tag{13}$$

After the support Ω is determined, the non-zero coefficients of $\hat{\mathbf{X}}$ are computed via an orthogonal projection: $\hat{\mathbf{X}}_\Omega = \mathbf{\Phi}_\Omega^\dagger \mathbf{Y}$.

The q-simultaneous OMP (SOMP) algorithm is an iterative procedure where in each iteration, an atom is selected, and a residual is updated. The next selected atom is the one which maximizes the q-correlation with the current residual. A pseudocode for SOMP is summarized in Algorithm 2.

Different authors have advocated the use of different values for q in step 3. However, it has been shown [12] that, independent of q, SOMP will recover a joint sparse representation with joint support Ω whenever the Exact Recovery Condition (ERC) [7] is met:

$$\max_{j \notin \Omega} ||\mathbf{\Phi}_\Omega^\dagger \boldsymbol{\phi}_j||_1 < 1. \tag{14}$$

For OMP, in the SMV problem, Tropp showed that the ERC is also necessary to guarantee recovery for all vectors with support Ω [7, Theorem 3.10]. It turns out that the necessary condition for SOMP is also the ERC, independent

Algorithm 2 Simultaneous Orthogonal Matching Pursuit (SOMP)

Initialization: $\mathbf{R}^{(0)} = \mathbf{Y}, \mathbf{X}^{(0)} = \mathbf{0}, \Omega^0 = \emptyset$

for $n = 1; n := n + 1$ **until stopping criterion do**

  $i^n = \mathrm{argmax}_i \, \|\boldsymbol{\phi}_i^T \mathbf{R}^{(n-1)}\|_q$

  $\Omega^n = \Omega^{n-1} \cup i^n$

  $\mathbf{X}^{(n)}_{\Omega^n,:} = \boldsymbol{\Phi}^\dagger_{\Omega^n} \mathbf{Y}$

  $\mathbf{R}^{(n)} = \mathbf{Y} - \boldsymbol{\Phi}\mathbf{X}^{(n)}$

end for

of the rank of $\mathbf{X}$. This implies that the SOMP algorithm is not able to exploit rank information in order to improve the recovery ability in the worst-case, as we show in the following theorem.

Theorem 4 (SOMP is not rank aware)**.** *Let τ be given such that $1 \leq \tau \leq k$ and suppose that*

$$\max_{j \notin \Omega} ||\boldsymbol{\Phi}^\dagger_\Omega \boldsymbol{\phi}_j||_1 > 1 \tag{15}$$

for some support Ω, $|\Omega| = k$. Then there exists an $\mathbf{X}$ with $\mathrm{supp}(\mathbf{X}) = \Omega$ and $\mathrm{rank}(\mathbf{X}) = \tau$ that SOMP cannot recover.

Proof: Since the ERC does not hold, we know that for the SMV problem, there exists a vector $\mathbf{x}$ for which OMP will fail (specifically we can set $\mathbf{x} = \mathrm{sign}(\boldsymbol{\Phi}^\dagger_\Omega \boldsymbol{\phi}_j)$ the j that maximises (15), see [7]). Let $\mathbf{x}$ be a vector with $\mathrm{supp}(\mathbf{x}) = \Omega$ such that OMP incorrectly selects atom $j^\star \notin \Omega$ at the first step with

$$|\boldsymbol{\phi}_{j^\star}^T \boldsymbol{\Phi}\mathbf{x}| > \max_{i \in \Omega} |\boldsymbol{\phi}_i^T \boldsymbol{\Phi}\mathbf{x}| + \epsilon \tag{16}$$

for some $\epsilon > 0$. Let $\mathbf{X} := [\mathbf{x}, \mathbf{x}, \ldots, \mathbf{x}]$ be the rank 1 matrix associated with $\mathbf{x}$ that cannot be recovered from $\mathbf{Y} = \boldsymbol{\Phi}\mathbf{X}$ by SOMP using any q. We can now perturb $\mathbf{X}$. Let $\mathbf{E}$ be any rank $\tau - 1$ perturbation matrix such that: $\mathrm{supp}(\mathbf{E}) = \Omega$, $\max_j \|\boldsymbol{\phi}_j^T \boldsymbol{\Phi}\mathbf{E}\|_q \leq l^{1/q}\epsilon/2$ and whose row and column spaces are disjoint with those of $\mathbf{X}$. Then $\tilde{\mathbf{X}} = \mathbf{X} + \mathbf{E}$ has rank τ.

If we now define $\tilde{\mathbf{Y}} = \boldsymbol{\Phi}\tilde{\mathbf{X}}$ we have:

$$\begin{aligned}
\|\boldsymbol{\phi}_{j^\star}^T \tilde{\mathbf{Y}}\|_q &\geq \|\boldsymbol{\phi}_{j^\star}^T \boldsymbol{\Phi}\mathbf{X}\|_q - \|\boldsymbol{\phi}_{j^\star}^T \boldsymbol{\Phi}\mathbf{E}\|_q \\
&\geq l^{1/q} |\boldsymbol{\phi}_{j^\star}^T \boldsymbol{\Phi}\mathbf{x}| - l^{1/q}\epsilon/2 \\
&> l^{1/q} \max_{i \in \Omega} |\boldsymbol{\phi}_i^T \boldsymbol{\Phi}\mathbf{x}| + l^{1/q}\epsilon/2 \\
&= \max_{i \in \Omega} \|\boldsymbol{\phi}_i^T \boldsymbol{\Phi}\mathbf{X}\|_q + l^{1/q}\epsilon/2 \\
&\geq \max_{i \in \Omega} \left\{ \|\boldsymbol{\phi}_i^T \boldsymbol{\Phi}\mathbf{X}\|_q + \|\boldsymbol{\phi}_i^T \boldsymbol{\Phi}\mathbf{E}\|_q \right\} \\
&\geq \max_{i \in \Omega} \|\boldsymbol{\phi}_i^T \tilde{\mathbf{Y}}\|_q.
\end{aligned} \tag{17}$$

Line 1 in (17) is the reverse triangle inequality, line 2 follows from the definition of the q norm and the fact that $\mathbf{X}$ is rank 1, and line 3 follows from (16). Lines $4-6$ are a result of reversing these arguments. Equation (17) therefore shows that no correct atom will be selected at the first step in the perturbed problem and $\tilde{\mathbf{X}}$ will not be correctly recovered. ■

We conclude from Theorem 4 that SOMP is effectively blind to the rank of $\mathbf{X}$. An identical argument can also be used to show that an MMV version of any similar Matching Pursuit type algorithm (e.g. M-MP, M-ORMP [10]) will also be rank blind, including identification of the support set by thresholding $||\boldsymbol{\phi}_j^T \mathbf{Y}||_q$ as proposed in [25].

B. Mixed ℓ^1/ℓ^q minimization

Another popular joint sparse recovery algorithm is to perform mixed norm minimization:

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmin}} ||\mathbf{X}||_{1,q} \text{ s.t. } \boldsymbol{\Phi}\mathbf{X} = \mathbf{Y} \tag{18}$$

for some $q \geq 1$ (values of $q = 1, 2$ and ∞ have been advocated). This is a simple extension of the ℓ^1 minimization used to solve SMV problems. In SMV the necessary and sufficient condition for the recovery of vectors $\mathbf{x}$ with support Ω is given by the *Null Space Property* (see [35], [40], [41]):

$$||z_\Omega||_1 < ||z_{\Omega^c}||_1, \quad \forall z \in \mathcal{N}(\boldsymbol{\Phi}). \tag{19}$$

Here Ω^c is the complement of the set Ω.

As with SOMP we can leverage the SMV conditions for recovery to show that mixed norm methods are not rank aware:

Theorem 5 (ℓ^1/ℓ^q minimization is not rank aware)**.** *Let τ be given such that $1 \leq \tau \leq k$ and suppose that there exists a $z \in \mathcal{N}(\boldsymbol{\Phi})$ such that*

$$||z_\Omega||_1 > ||z_{\Omega^c}||_1 \tag{20}$$

for some support Ω, $|\Omega| = k$. Then there exists an $\mathbf{X}$ with $\operatorname{supp}(\mathbf{X}) = \Omega$, $\operatorname{rank}(\mathbf{X}) = \tau$ that (18) *cannot recover.*

Proof: The proof follows along the same lines as Theorem 4. Let $\mathbf{X} = [\mathbf{x}, \mathbf{x}, \ldots, \mathbf{x}]$ be the rank 1 matrix such that ℓ^1 minimization fails to recover x. Denote the ℓ^1 minimum solution by $\hat{\mathbf{x}}$ and define $\hat{\mathbf{X}} = [\hat{\mathbf{x}}, \hat{\mathbf{x}}, \ldots, \hat{\mathbf{x}}]$. We know that

$$||\mathbf{X}||_{1,q} > ||\hat{\mathbf{X}}||_{1,q} + \epsilon \tag{21}$$

for some $\epsilon > 0$. We now perturb $\mathbf{X}$ by a matrix $\mathbf{E}$, $\tilde{\mathbf{X}} = \mathbf{X} + \mathbf{E}$, such that $\operatorname{supp}(\mathbf{E}) = \Omega$, $\operatorname{rank}(\tilde{\mathbf{X}}) = \tau$ and $||\mathbf{E}||_{1,q} \leq \epsilon/2$. Therefore, the solution $\hat{\tilde{\mathbf{X}}}$ of (18) for the perturbed problem must satisfy:

$$||\hat{\tilde{\mathbf{X}}}||_{1,q} \leq ||\hat{\mathbf{X}}||_{1,q} + ||\mathbf{E}||_{1,q} \leq ||\hat{\mathbf{X}}||_{1,q} + \epsilon/2 < ||\mathbf{X}||_{1,q} - \epsilon/2. \tag{22}$$

On the other hand, by the triangle inequality,

$$||\tilde{\mathbf{X}}||_{1,q} \geq ||\mathbf{X}||_{1,q} - ||\mathbf{E}||_{1,q} \geq ||\mathbf{X}||_{1,q} - \epsilon/2. \tag{23}$$

Therefore, $||\tilde{\mathbf{X}}||_{1,q} > ||\hat{\tilde{\mathbf{X}}}||_{1,q}$ so (18) fails to recover the correct solution. ■

The argument in the proof of the theorem can be applied to any joint sparse recovery that uses $||\cdot||_{p,q}$ for $p < 1$, using the appropriate Null Space Property associated with the p-quasi-norm [40]. Thus the result also applies to the M-FOCUSS family of algorithms [10] that define a specific means of (locally) minimizing the $\{p, q\}$-quasi-norms. Interestingly the above argument does not apply to the ReMBo algorithm [13] which uses a random projection to transform the MMV problem to an SMV counterpart, and then solves the SMV problem using, for example, Basis Pursuit. A partial analysis of ReMBo can be found in [42].

In the next section we develop a new algorithm that is rank aware: in the full rank case it guarantees recovery from $k + 1$ measurements, and in the non full rank case, its performance degrades gracefully with the rank.

VI. Rank Aware Pursuits

In the last section we noted that two classes of popular techniques for joint sparse recovery were effectively rank blind. We now examine methods that can be shown to be rank aware. In particular, we consider two algorithms, both of which can be categorized as "greedy pursuits" and are based upon the discrete MUSIC algorithm.

We have already seen in Section IV-A that the MUSIC algorithm can be viewed as a rank aware form of thresholding. However it is also well known that thresholding techniques can generally be refined through the use of pursuit based methods. These involve iteratively selecting a single atom at a time, calculating an approximate solution for $\mathbf{X}$ and the residual $\mathbf{R}$, and then selecting a new atom. This is repeated until completion. The exact form of the pursuit algorithm depends on the selection strategy and the refinement step. Here we introduce a selection strategy based upon the MUSIC thresholding described above.

A. Rank Aware Selection

Given a residual $\mathbf{R}^{(n-1)}$ at the $(n-1)$th iteration we define the following selection strategy at the nth iteration:

$$\Omega^{(n)} = \Omega^{(n-1)} \cup \underset{i}{\operatorname{argmax}} ||\boldsymbol{\phi}_i^T \mathbf{U}^{(n-1)}||_2, \tag{24}$$

where $\mathbf{U}^{(n-1)} = \text{orth}(\mathbf{R}^{(n-1)})$. We restrict our attention to the 2-norm since this norm is invariant to the orthonormal basis $\mathbf{U}$.

The main difference between rank aware selection and standard OMP type algorithms, is that the criterion is based on the inner products with $\mathbf{U}^{(n)}$, rather than $\mathbf{R}^{(n)}$. This is the same idea that is used in discrete MUSIC: instead of comparing with the given measurements $\mathbf{Y}$ or the resulting residuals $\mathbf{R}^{(n)}$, we compare with an orthonormal basis for the span. This simple difference, which is easy to implement in practice, is enough to allow for rank awareness. Furthermore, as we will see in the simulations section, it results in enhanced performance.

Below, we introduce two algorithms that use the rank aware selection principle. The first is a natural generalization of SOMP, where we simply adopt the rank aware selection step. Although we will see that this substitution improves performance considerably, it does not result in full rank awareness. In the full rank case, perfect recovery from $k+1$ measurements under the spark condition is not guaranteed (or typically observed as we will see in Section VII).

As we show, this is a result of rank degeneration of the residual: the rank of the residual generally decreases in each iteration. To rectify this behaviour, we propose below the use of a different algorithm, a Rank Aware Order Recursive Matching Pursuit (RA-ORMP), with a modified selection step which forces the sparsity of the residual to decrease along with its rank. This ensures that the residual has full rank at each iteration, and results in a fully rank aware OMP-type algorithm. In simulations, we will see that this approach tends to have the best performance for all choices of the rank of $\mathbf{X}$.

Similar to other MP techniques, successful selection of an atom $\boldsymbol{\phi}_i$ from the correct support set $i \in \Omega$ requires the following condition:

$$\frac{\max_{j \notin \Omega} \|\boldsymbol{\phi}_j^T \mathbf{U}\|_2}{\max_{i \in \Omega} \|\boldsymbol{\phi}_i^T \mathbf{U}\|_2} < 1. \tag{25}$$

We begin by noting that for successful selection, i.e. for (25) to hold, the ERC is a sufficient but not necessary. Specifically we have the following proposition.

Proposition 1. *Let* $\mathbf{Y} = \boldsymbol{\Phi}\mathbf{X}$, $\operatorname{supp}(\mathbf{X}) = \Omega$, $|\Omega| = k$ *and* $\operatorname{rank}(\mathbf{Y}) = \tau$. *Then:*

1) $\max_{j \notin \Omega} \|\boldsymbol{\Phi}_\Omega^\dagger \boldsymbol{\phi}_j\|_1 < 1$ *is a sufficient condition for Rank Aware selection to correctly choose an atom* $i \in \Omega$.
2) *If* $\tau > k + 1 - |\operatorname{supp}(\boldsymbol{\Phi}_\Omega^\dagger \boldsymbol{\phi}_{j^\star})|$ *for all* $j^\star := \operatorname{argmax}_{j \notin \Omega} \|\boldsymbol{\Phi}_\Omega^\dagger \boldsymbol{\phi}_j\|_1$ *then*

$$\max_{j \notin \Omega} ||\boldsymbol{\Phi}_\Omega^\dagger \boldsymbol{\phi}_j||_1 < 1 \tag{26}$$

   is not a necessary condition for correct atom selection.
3) *When* $\tau = k$, *then* $k < \operatorname{spark}(\boldsymbol{\Phi}) - 1$ *is necessary and sufficient for rank aware selection to correctly select an atom with index* $i \in \Omega$.

Proof: The proof of part 1 that ERC is sufficient is identical to that for SOMP, given in [12] and is based upon standard norm inequalities. Let $j^\star := \operatorname{argmax}_{j \notin \Omega} \|\boldsymbol{\Phi}_\Omega^\dagger \boldsymbol{\phi}_j\|_1$. We can then bound the norm $\max_{j \notin \Omega} \|\boldsymbol{\phi}_j^T \mathbf{U}\|_2$ for any $\mathbf{U}$ as follows:

$$\begin{aligned}
\max_{j \notin \Omega} \|\boldsymbol{\phi}_j^T \mathbf{U}\|_2 &= \max_{\mathbf{x} \neq 0} \frac{|(\boldsymbol{\Phi}_\Omega^\dagger \boldsymbol{\phi}_{j^\star})^T \boldsymbol{\Phi}_\Omega^T \mathbf{U}\mathbf{x}|}{\|\mathbf{x}\|_2} \\
&= \max_{x \neq 0} \frac{|(\boldsymbol{\Phi}_\Omega^\dagger \boldsymbol{\phi}_{j^\star})^T \boldsymbol{\Phi}_\Omega^T \mathbf{U}\mathbf{x}|}{\|\boldsymbol{\Phi}_\Omega^T \mathbf{U}\mathbf{x}\|_\infty} \frac{\|\boldsymbol{\Phi}_\Omega^T \mathbf{U}\mathbf{x}\|_\infty}{\|\mathbf{x}\|_2} \\
&\leq \|\boldsymbol{\Phi}_\Omega^\dagger \boldsymbol{\phi}_{j^\star}\|_1 \max_{\mathbf{x} \neq 0} \frac{\|\boldsymbol{\Phi}_\Omega^T \mathbf{U}\mathbf{x}\|_\infty}{\|\mathbf{x}\|_2} \\
&= \|\boldsymbol{\Phi}_\Omega^\dagger \boldsymbol{\phi}_{j^\star}\|_1 \max_{i \in \Omega} \|\boldsymbol{\phi}_i^T \mathbf{U}\|_2.
\end{aligned} \tag{27}$$

Hence ERC implies (25) and is sufficient for recovery. However, unlike SOMP, the norm equalities cannot necessarily be approached since $\mathbf{U}$ is constrained to be column orthonormal.

If the ERC is also a necessary condition then we must be able to find a $\mathbf{U}$ that achieves equality in (27). This in turn implies that there must be a single $\mathbf{x}^\star \neq 0$ that simultaneously satisfies:

$$\mathbf{x}^\star = \operatorname*{argmax}_{\mathbf{x}} \frac{|(\boldsymbol{\Phi}_\Omega^\dagger \boldsymbol{\phi}_{j^\star})^T \boldsymbol{\Phi}_\Omega^T \mathbf{U}\mathbf{x}|}{\|\boldsymbol{\Phi}_\Omega^T \mathbf{U}\mathbf{x}\|_\infty} \tag{28}$$

and

$$\mathbf{x}^\star = \underset{\mathbf{x}}{\operatorname{argmax}} \frac{\|\mathbf{\Phi}_\Omega^T \mathbf{U}\mathbf{x}\|_\infty}{\|\mathbf{x}\|_2}. \tag{29}$$

Now, (28) implies that:

$$\mathbf{\Phi}_\Omega^T \mathbf{U}\mathbf{x}^\star \propto \operatorname{sgn}(\mathbf{\Phi}_\Omega^\dagger \boldsymbol{\phi}_{j^\star}), \tag{30}$$

where we use the convention that $\operatorname{sgn}(0)$ can take any value in $[-1, 1]$. Therefore

$$|\langle \phi_i, \mathbf{U}\mathbf{x}^\star \rangle| = |\langle \phi_j, \mathbf{U}\mathbf{x}^\star \rangle| \geq |\langle \phi_k, \mathbf{U}\mathbf{x}^\star \rangle| \tag{31}$$

for all $i, j, k \in \Omega$, $i, j \in \operatorname{supp}(\mathbf{\Phi}_\Omega^\dagger \boldsymbol{\phi}_{j^\star})$ and $k \notin \operatorname{supp}(\mathbf{\Phi}_\Omega^\dagger \boldsymbol{\phi}_{j^\star})$. On the other hand (29) implies that $\mathbf{x}^\star \propto \mathbf{U}^T \boldsymbol{\phi}_{i^\star}$ for some $\boldsymbol{\phi}_{i^\star} = \operatorname{argmax}_{\boldsymbol{\phi}_i} \|\boldsymbol{\phi}_i^T \mathbf{U}\|_2, i \in \Omega$. That is $\mathbf{x}$ must be proportional to the projection of $\boldsymbol{\phi}_{i^\star}$ by $\mathbf{U}$. In fact $\mathbf{x}$ must also be the common projection of all the columns $\phi_i, i \in \operatorname{supp}(\mathbf{\Phi}_\Omega^\dagger \boldsymbol{\phi}_{j^\star})$. Indeed from (31),

$$\|\mathbf{U}^T \boldsymbol{\phi}_{i^\star}\|_2 = \frac{|\langle \phi_{i^\star}, \mathbf{U}\mathbf{x}^\star \rangle|}{\|\mathbf{x}^\star\|_2} = \frac{|\langle \phi_i, \mathbf{U}\mathbf{x}^\star \rangle|}{\|\mathbf{x}^\star\|_2} \leq \|U^T \boldsymbol{\phi}_i\|_2. \tag{32}$$

Therefore we can define the vector $\mathbf{v} := [\mathbf{\Phi}_\Omega^T]^\dagger \operatorname{sgn}(\mathbf{\Phi}_\Omega^\dagger \boldsymbol{\phi}_{j^\star})$ such that:

$$\mathbf{U}\mathbf{U}^T \phi_i \propto \mathbf{v}, \quad \forall i \in \operatorname{supp}(\mathbf{\Phi}_\Omega^\dagger \boldsymbol{\phi}_{j^\star}) \tag{33}$$

Now define the subspace $\mathcal{W}$ as:

$$\mathcal{W} := \{\mathbf{w} : \mathbf{w} = \sum_i \alpha_i \langle \boldsymbol{\phi}_i, \mathbf{v} \rangle \boldsymbol{\phi}_i, i \in \operatorname{supp}(\mathbf{\Phi}_\Omega^\dagger \boldsymbol{\phi}_{j^\star}), \sum_i \alpha_i = 0\}. \tag{34}$$

Note that by definition $\mathbf{U}^T \mathbf{w} = 0$. However we also have $\dim(\mathcal{W}) = |\operatorname{supp}(\mathbf{\Phi}_\Omega^\dagger \boldsymbol{\phi}_{j^\star})| - 1$. Since $\dim(\mathbf{U}) = \tau$ this is only possible if $\tau \leq k + 1 - |\operatorname{supp}(\mathbf{\Phi}_\Omega^\dagger \boldsymbol{\phi}_{j^\star})|$. Thus when $\tau > k - 1 + |\operatorname{supp}(\mathbf{\Phi}_\Omega^\dagger \boldsymbol{\phi}_{j^\star})|$ equality in (27) cannot be achieved. Since the set of column orthonormal matrices $\mathbf{U}$ is compact a maximum must exist:

$$\max_{\mathbf{U}} \frac{\max_{j \notin \Omega} \|\boldsymbol{\phi}_j^T \mathbf{U}\|_2}{\max_{i \in \Omega} \|\boldsymbol{\phi}_i^T \mathbf{U}\|_2} = c < \max_{j \notin \Omega} ||\mathbf{\Phi}_\Omega^\dagger \boldsymbol{\phi}_j||_1. \tag{35}$$

Therefore, (26) is not a necessary condition for correct atom selection, which proves part 2.

Finally when $\tau = k$ the necessary and sufficient conditions in part 3 for correct atom selection follow from the rank aware thresholding result, Theorem 3, and the identifiability conditions in Theorem 2. ■

Proposition 1 shows that in the worst case scenario we lose nothing by incorporating the orthogonalization within the selection step. Furthermore part 2 shows that in general the selection is more effective than that of SOMP, although we have not quantified by how much as this seems difficult to estimate for general τ. The exception is, of course, when $\tau = k$. In this case, RA-selection inherits the desirable property from the discrete MUSIC algorithm that when identifiability conditions are satisfied correct detection is guaranteed.

B. Rank Aware Greedy Algorithms

We now proceed to develop two specific greedy algorithms that exploit the RA-selection.

1) (Partially) Rank Aware OMP: One possible approach to developing a Rank Aware Pursuit would be to substitute (24) as the selection step within SOMP (step 3). We call this algorithm RA-OMP. In Section VII we will see that the incorporation of the rank aware selection step substantially improves the average recovery performance over SOMP. However, curiously even when $\mathrm{rank}(\mathbf{Y}) = k$, RA-OMP is not guaranteed to exactly recover $\mathbf{X}$. That is, it does not achieve the performance of MUSIC in the full rank case.

This can be explained by the fact that the rank of the residual deteriorates at each step of RA-OMP, a process we call *rank degeneration*. When selecting the first atom we have $\mathbf{R}^{(0)} = \mathbf{Y}$ and from Proposition 1, assuming that $\mathrm{rank}(\mathbf{Y}) = k$, we are guaranteed to select an atom $\boldsymbol{\phi}_i$ such that $i \in \Omega$. The updated residual is $\mathbf{R}^{(1)} = (I - \boldsymbol{\phi}_i \boldsymbol{\phi}_i^T)\mathbf{Y}$. Since $\mathcal{R}(\mathbf{Y}) = \mathcal{R}(\boldsymbol{\Phi}_\Omega)$, the rank will be reduced by one such that $\mathrm{rank}(\mathbf{R}^{(1)}) = k - 1$. With a little manipulation we can write $\mathbf{R}^{(1)}$ as follows:

$$\mathbf{R}^{(1)} = \sum_{j \in \Omega \setminus i} \boldsymbol{\phi}_j \mathbf{X}_{j,:} - \boldsymbol{\phi}_i \big(\sum_{j \in \Omega \setminus i} (\boldsymbol{\phi}_j^T \boldsymbol{\phi}_i) \mathbf{X}_{j,:} \big). \tag{36}$$

The matrix $\mathbf{R}^{(1)}$ will therefore still be k-joint sparse unless $\sum_{j \in \Omega \setminus i} (\boldsymbol{\phi}_j^T \boldsymbol{\phi}_i) \mathbf{X}_{j,:} = 0$. However since $\mathbf{X}$ is assumed to be rank k the set of row vectors $\mathbf{X}_{j,:}$, $j \in \Omega \setminus i$ must be linearly independent. Therefore the joint sparsity will decrease only if $\boldsymbol{\phi}_j^T \boldsymbol{\phi}_i = 0, \forall j \in \Omega \setminus i$, i.e. the columns are orthogonal. If the columns are not orthogonal then $\mathbf{R}^{(1)}$ will not have maximal rank and the rank aware selection is no longer guaranteed to make a correct selection. The rank degeneration can continue in subsequent iterations.

2) Rank Aware Order Recursive Matching Pursuit: We can rectify the *rank degeneration* problem using a modified selection step or equivalently modifying the dictionary at each step. The idea is to force the sparsity of the residual to decrease along with its rank, so that Lemma 1 can still be applied. The mechanism we exploit has already been used in the SMV problem and goes by various names including: Orthogonal Least Squares [43] (since it selects the atom that minimizes the residual in the least squares sense at each step), and Order Recursive Matching Pursuit (ORMP) [3].[4] While an MMV extension of ORMP was presented in [10], it is based on similar norm extensions to those used in SOMP and therefore is similarly rank blind. Here we present a Rank Aware Order Recursive Matching Pursuit (RA-ORMP), and show that it has guaranteed recovery in the full rank case, as well as empirically exhibiting improved performance for all values of $\mathrm{rank}(\mathbf{X})$. The pseudocode for RA-ORMP is given in Algorithm 3.

In practice, we do not calculate the projections as detailed above and instead use a Gram-Schmidt orthogonalization procedure as in the standard implementation of ORMP [44]. Furthermore, a practical implementation of RA-ORMP would also need to incorporate an estimate of the signal subspace in step 3 of Algorithm 3 along the same lines as that proposed in Section IV-A.

Note that the key difference between RA-OMP and RA-ORMP is that we not only project the observed data $\mathbf{Y}$ orthogonal to the selected atoms to calculate the residual, we also project the remaining dictionary atoms (step 9) and then, crucially renormalize them (step 10), so that all atoms are again unit norm.

[4]Note that historically OMP and ORMP have been repeatedly confused for each other. For a potted history of the subject see [44].

Algorithm 3 Rank Aware Order Recursive Matching Pursuit (RA-ORMP)

Initialize $\mathbf{R}^{(0)} = \mathbf{Y}, \mathbf{X}^{(0)} = \mathbf{0}, \Omega^0 = \emptyset$ and $\tilde{\boldsymbol{\phi}}_i = \boldsymbol{\phi}_i$ for all i.
for $n = 1; n := n + 1$ **until stopping criterion do**
  Calculate orthonormal basis for residual: $\mathbf{U}^{(n-1)} = \text{Orth}(\mathbf{R}^{(n-1)})$
  $i^n = \text{argmax}_{i \notin \Omega^{(n-1)}} \|\tilde{\boldsymbol{\phi}}_i^T \mathbf{U}^{(n-1)}\|_2$
  $\Omega^n = \Omega^{n-1} \cup i^n$
  $\mathbf{X}^{(n)}_{\Omega^n,:} = \boldsymbol{\Phi}^\dagger_{\Omega^n} \mathbf{Y}$
  Calculate orthogonal projector: $P^\perp_{\Omega^{(n)}} := (\mathbf{I} - \boldsymbol{\Phi}_{\Omega^{(n)}} \boldsymbol{\Phi}^\dagger_{\Omega^{(n)}})$
  $\mathbf{R}^{(n)} = P^\perp_{\Omega^{(n)}} \mathbf{Y}$
  $\boldsymbol{\Phi}' = P^\perp_{\Omega^{(n)}} \boldsymbol{\Phi}$
  Renormalize $\tilde{\boldsymbol{\phi}}_i = \boldsymbol{\phi}'_i / \|\boldsymbol{\phi}'_i\|_2$, for $i \notin \Omega^{(n-1)}$.
end for

We next show that, like MUSIC, RA-ORMP will exactly recover $\mathbf{X}$ in the full rank scenario:

Theorem 6 (RA-ORMP, full rank case)**.** *Let* $\mathbf{Y} = \boldsymbol{\Phi}\mathbf{X}$ *with* $|\operatorname{supp}(\mathbf{X})| = k$, $\operatorname{rank}(\mathbf{X}) = k$ *and* $k < \operatorname{spark}(\boldsymbol{\Phi}) - 1$. *Then RA-ORMP is guaranteed to recover* $\mathbf{X}$ *(i.e.* $\hat{\mathbf{X}} = \mathbf{X}$*).*

Proof: From the spark condition we know that $\boldsymbol{\phi}_j \notin \mathcal{R}(\boldsymbol{\Phi}_\Omega)$ for $j \notin \Omega$ and thus $\mathbf{X}$ is identifiable. From Proposition 1 and the rank assumption, the selection at the first step is successful. It is therefore sufficient to show that the updated residual provides another full rank identifiable problem.

Suppose that we select index $i \in \Omega$ at the first step. The new residual is then:

$$\mathbf{R}^{(1)} = P^\perp_{\Omega^{(n)}} \mathbf{Y}. \tag{37}$$

Therefore $\operatorname{rank}(\mathbf{R}^{(1)}) = k - 1$. If we now expand $\mathbf{R}^{(1)}$ in terms of the atoms $\boldsymbol{\phi}_j$ we get:

$$\begin{aligned} \mathbf{R}^{(1)} &= \sum_{j \in \Omega \setminus i} P^\perp_{\Omega^{(n)}} \boldsymbol{\phi}_j \mathbf{X}_{j,:} \\ &= \sum_{j \in \Omega \setminus i} \tilde{\boldsymbol{\phi}}_j \tilde{\mathbf{X}}_{j,:} \end{aligned} \tag{38}$$

where $\tilde{\mathbf{X}}_{j,:}$ is the jth row of $\mathbf{X}$ rescaled by $\|P^\perp_{\Omega^{(n)}} \boldsymbol{\phi}_j\|_2^{-1}$. Thus $\mathbf{R}^{(1)}$ has a $(k-1)$ sparse representation within the modified dictionary $\tilde{\boldsymbol{\Phi}}$.

Finally we need to show that $\tilde{\mathbf{X}}$ is also identifiable. Since $\boldsymbol{\phi}_j \notin \mathcal{R}(\boldsymbol{\Phi}_\Omega)$ for $j \notin \Omega$, we know that $\|P^\perp_\Omega \boldsymbol{\phi}_j\|_2 > 0$. However this property is preserved under the projection of the dictionary since $P^\perp_\Omega P^\perp_i = P^\perp_\Omega$ for all $i \in \Omega$, so that $\tilde{\mathbf{X}}$ is also identifiable. Recursively applying the above arguments until $\mathbf{R}^{(k)} = 0$ completes the proof. ■

We see that RA-ORMP does not suffer the same rank degeneration that RA-OMP does and provides a natural rank aware algorithm that reduces to ORMP in the SMV case while achieving guaranteed recovery in the full rank MMV setting when $\operatorname{rank}(\mathbf{Y}) = k$.

ORMP is usually championed as superior to OMP, since at each step it selects the atom that most decreases the size of the residual. Furthermore, empirical evidence suggests that ORMP generally outperforms OMP (slightly) but at the expense of additional computation. However, to our knowledge, there is no study of the necessary and sufficient recovery conditions for ORMP. It is easy to see that the ERC provides a necessary condition since in the first selection step OMP and ORMP are identical. Curiously though, it is not clear whether the ERC condition holding for $\mathbf{\Phi}$ implies that it will also hold for the modified dictionary $\tilde{\mathbf{\Phi}}$.

C. Link with sequential MUSIC techniques

Within the array processing literature a number of sequential variants of MUSIC have previously been proposed [45], [46], [47], which relate to our derivations here. In [45] a sequential MUSIC algorithm is introduced which can be thought of as a continuous parameter version of a Rank Aware Matching Pursuit, i.e. RA-OMP, but without the orthogonalization step. In [47] an algorithm called *Recursively Applied and Projected (RAP) MUSIC* was introduced. This is formally equivalent to RA-ORMP proposed here. However, because in array processing the aim is to estimate a continuous parameter vector (even when this is done through discretization) associated with the directions of arrival for multiple sources the type of analysis performed on these algorithms is very different than the exact recovery results presented here for the discrete sparsity model.

VII. Numerical Experiments

In this section we explore the empirical performance of MUSIC, RA-OMP and RA-ORMP. We contrast these with results for the rank blind recovery algorithm SOMP (using $p = 2$). For comparisons with mixed ℓ^1/ℓ^q norm minimization we point the reader to the comparisons performed in [14], where empirically SOMP generally exhibited superior performance.

To test the four algorithms we consider the family of random matrices $\mathbf{\Phi}$ whose elements were drawn independently from a normal distribution: $\mathbf{\Phi}_{i,j} \sim \mathcal{N}(0, 1)$. The columns of $\mathbf{\Phi}$ are then normalized so that $\|\boldsymbol{\phi}_i\|_2 = 1$. The dimensions of $\mathbf{\Phi}$ are fixed to $n = 256$ and $m = 32$, while the number of measurement vectors, l, is varied between 1 and 32. Finally the non-zero entries of $\mathbf{X}$ were also draw independently from a unit variance normal distribution, implying that, with probability one, $\operatorname{rank}(\mathbf{Y}) = l$.

Plots of the empirical probability of recovery for the four algorithms are given in Figure 1. It is clear from these plots that while RA-OMP appears to be superior to SOMP for moderate numbers of measurement vectors ($l = 2, 4$) both SOMP and RA-OMP appear to stall at around the same sparsity level and fail to recover vectors with $k > 11$.

In contrast, both RA-ORMP and MUSIC demonstrate full recovery in the full rank case, including when $l = 32$ up to a sparsity level of $k = 31$ as predicted by the theory. However MUSIC does not appear to provide recovery much beyond the full rank condition. That is, recovery drops immediately once $k > l$. As with SMV thresholding, better performance can be achieved if we restrict the dynamic range of the nonzero coefficients.

The performance of RA-ORMP clearly provides the best recovery. It uniquely amongst the four appears to be able to achieve OMP type performance in the SMV case and consistently provides recovery well beyond the full

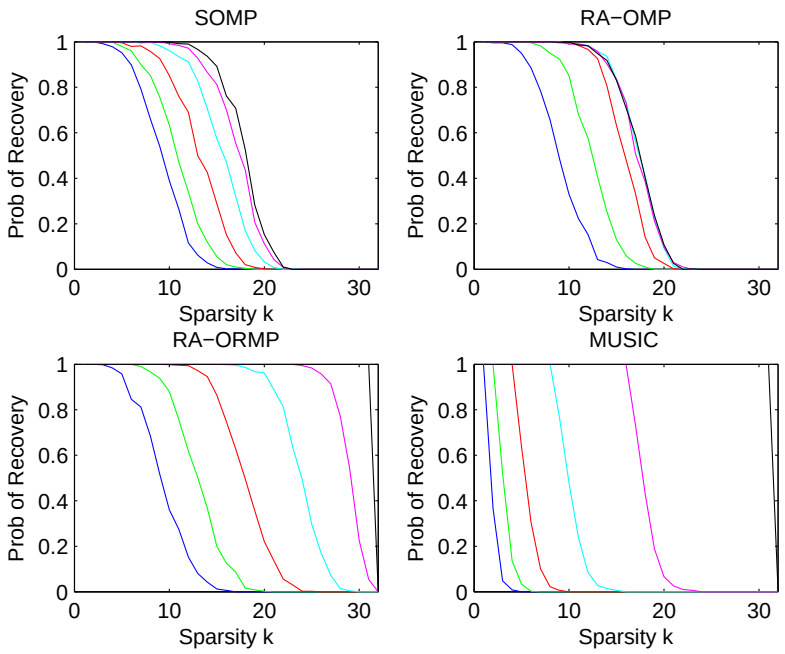

Fig. 1. The empirical probability of recovery for SOMP, RA-OMP, RA-ORMP and MUSIC as a function of sparsity level k. The curves in each plot relate to $l = 1, 2, 4, 8, 16$ and 32 (from left to right).

rank case. For example when $l = 16$ correct recovery is maintained up to $k = 23$.

In Figure 2 we show explicitly the improvement in recovery performance as a function of l for the same setup as above with a fixed sparsity level of $k = m/2 = 16$. Note that for $l \leq 16$ the rank of $\mathbf{Y}$ is equal to l, while for $l > 16$ the rank remains constant at 16. For this level of sparsity none of the algorithms achieve significant recovery rates in the SMV case. However, as the number of measurements and the rank of $\mathbf{Y}$ is increased all algorithms improve. The figure highlights that the rank aware methods are clearly able to exploit rank information and the recovery rate grows to 100% when the data matrix $\mathbf{Y}$ achieves maximal rank. What is particularly striking though is how quickly rank information improves the recovery rate in the RA-ORMP algorithm even when the rank of $\mathbf{Y}$ is significantly below the maximal rank.

In contrast, the recovery rates for SOMP and RA-OMP (recall this is not fully rank aware) do not even reach 100%. Interestingly these plots suggest that the rank information is still playing some role in the recovery performance of SOMP as the dominant increase in its performance occurs during the range of increasing rank. When $l > 16$ and the rank remains fixed the performance appears to plateau at around 80% recovery. Additional simulations (not shown) indicate that further increasing the number of measurement vectors does not enable SOMP to pass this bound. Curiously, although the performance of RA-OMP increases much more rapidly than that of SOMP at first, it appears to stall at the same recovery rate as that for SOMP, suggesting that the rank degeneration identified in Section VI introduces a bottleneck to the recovery performance for RA-OMP.

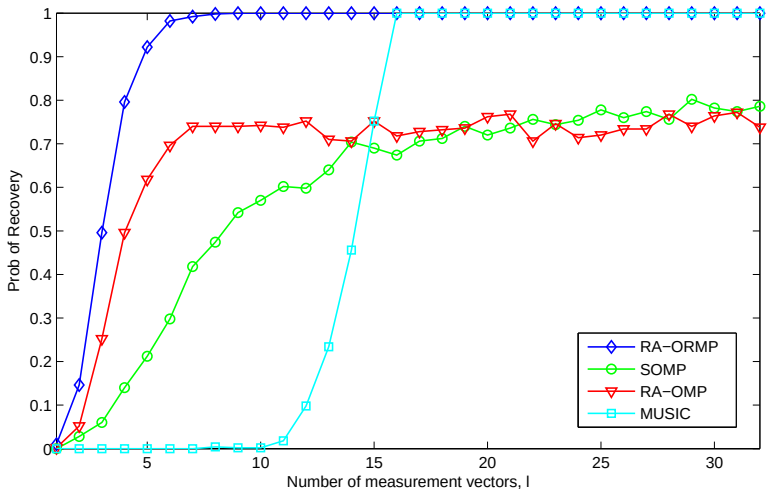

Fig. 2. The empirical probability of recovery for SOMP, RA-OMP, RA-ORMP and MUSIC as a function of number of measurement vectors for a fixed sparsity level, $k = m/2 = 16$.

VIII. Discussion

In this paper we considered the role of the rank of the coefficient matrix in sparse recovery with multiple measurement vectors (MMV). We began by reviewing known sufficient conditions for identifiability of the MMV problem and showed that these conditions are also necessary. We then demonstrated that algorithmically rank information can be used to reduce the complexity of the general combinatorial search procedure.

Next we turned our attention to practical algorithms. The MMV problem has already been studied extensively. However we have shown that the most popular classes of algorithms, q-SOMP and mixed ℓ^1/ℓ^q minimization, are both rank blind. Similar arguments apply to related algorithms (e.g. mixing ℓ^p/ℓ^q minimization) but we have omitted the details. Indeed, to our knowledge, none of the existing popular techniques being used for joint sparse recovery have been shown to be fully rank aware. We therefore developed some rank aware greedy algorithms and derived certain worst case recovery conditions for these methods. One might (rightly) criticise such analysis as overly pessimistic, however we stress that, in the full rank case, the worst case performance of MUSIC and RA-ORMP still significantly outperform the average case performance for the most popular (rank blind) algorithms [14]. Such are the benefits of exploiting the rank information.

In this paper we have focused on greedy rank aware algorithms. An interesting open question is whether a convex rank aware recovery algorithm exist that can interpolate between ℓ_1 minimization when $l = 1$ and guaranteed recovery when $\mathrm{rank}(\mathbf{Y}) = k$. Similarly we could ask whether other popular sparse recovery algorithms such as CoSaMP [8] or Iterative Hard Thresholding [48], [49] can be adapted to create new rank aware recovery algorithms for the MMV problem.

An important direction for future work is to analyze the recovery properties of the proposed algorithms in more

detail for the region $1 < \mathrm{rank}(\mathbf{X}) < k$. It is also important to quantify the typical performance of the algorithms rather than the worst case scenario and to account for the effects of noise. We expect that an analysis similar to that in [25], [14] would be appropriate here.

REFERENCES

[1] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[2] D. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, pp. 227–234, Apr 1995.

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

[5] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal of Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[6] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, pp. 600–616, March 1997.

[7] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[8] D. Needell and J. Tropp, "COSAMP: Iterative signal recovery from incomplete and inaccurate samples.," *Applied Computational Harmonic Analysis*, vol. 26, pp. 301–321, 2008.

[9] T. Blumensath and M. Davies, "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis and Applications, special issue on sparsity*, vol. 14, no. 5, pp. 629–654, 2008.

[10] S. F. Cotter, B. D. Rao, K. Engan, and K. K-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.

[11] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed compressive sensing," tech. rep., Rice University, 2005. Preprint.

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

[13] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Signal Process.*, vol. 56, pp. 4692–4702, Oct. 2008.

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

[15] Y. Lu and M. Do, "A theory for sampling signals from a union of subspaces," *IEEE transactions on signal processing*, vol. 56(6), pp. 2334–2345, 2008.

[16] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 4, pp. 1872–1882, 2009.

[17] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Transactions Information Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

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

[19] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[20] S. Gleichman and Y. Eldar, "Blind Compressed Sensing," To appear in *IEEE Trans. Inf. Theory*; [Online] arXiv 1002.2586.

[21] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Process.*, vol. 57, pp. 993–1009, Mar. 2009.

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

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

[24] M. Fornasier and H. Rauhut, "Recovery algorithms for vector valued data with joint sparsity constraints," *SIAM J. Numer. Anal.*, vol. 46, no. 2, pp. 577–613, 2008.

[25] 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," *Journal of Fourier Analysis and Applications, special issue on Sparsity*, vol. 14, pp. 655–687, 2008.

[26] R. Gribonval, B. Mailhe, H. Rauhut, K. Schnass, and P. Vandergheynst, "Average case analysis of multichannel thresholding," in *Proc. IEEE Intl. Conf. Acoust. Speech Signal Process.*, vol. II, pp. 853–856, 2007.

[27] P. Feng, *Universal Minimum-Rate Sampling and Spectrum-Blind Reconstruction for Multiband Signals*. PhD thesis, University of Illinois, 1998.

[28] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *Proceedings of RADC Spectral Estimation Workshop*, pp. 243–258, 1979.

[29] M. E. Davies and Y. C. Eldar, "Rank awareness and in joint sparse recovery," *ArXiv*, vol. 1004.4529, 2010.

[30] K. Lee and Y. Bresler, "iMUSIC: Iterative MUSIC algorithm for joint sparse recovery with any rank," 2010. Arxiv preprint: arXiv:1004.3071v1.

[31] J. M. Kim, O. K. Lee, and J. C. Ye, "Compressive MUSIC: A missing link between compressive sensing and array signal processing," Arxiv preprint, arXiv:1004.4398v1.

[32] A. Fannjiang, "The MUSIC algorithm for sparse objects: A compressed sensing analysis," *CoRR*, vol. abs/1006.1678, 2010.

[33] J. C. Ye, "Personal communication," September 2010.

[34] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, pp. 471–501, 2010.

[35] D. L. Donoho and M. Elad, "Optimally-sparse representation in general (non-orthogonal) dictionaries via l1 minimization," *Proc. Nat. Aca. Sci.*, vol. 100, pp. 2197–2202, 2003.

[36] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in *Proceedings of IEEE International Conference on the Acoustics, Speech, and Signal Processing*, pp. 1688–1691, 1996.

[37] G. W. Stewart, "Error and perturbation bounds for subspaces associated with certain eigenvalue problems," *SIAM Review*, vol. 15, no. 4, pp. 727–764.

[38] H. L. V. Trees, *Optimum Array Processing*. John Wiley & Sons, Inc., 2002.

[39] A. Lutoborski and V. N. Temlyakov, "Vector greedy algorithms," *J. Complex.*, vol. 19, no. 4, pp. 458–473, 2003.

[40] R. Gribonval and M. Nielsen, "Highly sparse representations are unique and independent of the sparseness measure," Tech. Rep. R2003-16, Dept. of Mathematical Sciences, University of Aalborg, Denmark, 2003.

[41] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," *Journal of the AMS*, vol. 22, pp. 211–231, 2009.

[42] E. v. Berg and M. P. Friedlander, "Theoretical and empirical results for recovery from multiple measurements," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2516–2527, 2010.

[43] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification.," *International Journal of Control*, vol. 50, no. 5, pp. 1873–1896, 1989.

[44] T. Blumensath and M. Davies, "On the difference between orthogonal matching pursuit and orthogonal least squares." unpublished manuscript, available at: http://www.personal.soton.ac.uk/tb1m08/papers/BDOMPvsOLS07.pdf, 2007.

[45] S. K. Oh and C. K. Un, "A sequential estimation approach for performance improvement of eignenstructure-based methods in array processing," *IEEE Transaction on Signal Processing*, vol. 41, pp. 457–463, 1993.

[46] P. Stoica, P. Handel, and A. Nehorai, "Improved sequential MUSIC," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 31, pp. 1230–1238, 1995.

[47] J. C. Mosher and R. M. Leahy, "Source localization using recursively applied and projected (RAP) MUSIC," *IEEE Transaction on Signal Processing*, vol. 47, pp. 332–340, 1999.

[48] T. Blumensath and M. E. Davies, "Normalized iterated hard thresholding; guaranteed stability and performance," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 298 – 309, 2010.

[49] T. Blumensath and M. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.

Mike E. Davies (M'00) received the B.A. (Hons.) degree in engineering from Cambridge University, Cambridge, U.K., in 1989 and the Ph.D. degree in nonlinear dynamics and signal processing from University College London (UCL), London, U.K., in 1993. Mike Davies was awarded a Royal Society University Research Fellowship in 1993 which he held first at UCL and then the University of Cambridge. He acted as an Associate Editor for IEEE Transactions in Speech, Language and Audio Processing, 2003-2007.

Since 2006 Dr. Davies has been with the University of Edinburgh where he holds the Jeffrey Collins SFC funded chair in Signal and Image Processing. He is the Director of the Joint Research Institute in Signal and Image Processing, a the joint venture between the University of Edinburgh and Heriot-Watt university as part of the Edinburgh Research Partnership. His current research focus is on sparse approximation, computational harmonic analysis, compressed sensing and their applications within signal processing. His other research interests include: non-Gaussian signal processing, high-dimensional statistics and information theory.

Yonina C. Eldar (S'98-M'02-SM'07) received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering both from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 1995 and 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2002.

From January 2002 to July 2002, she was a Postdoctoral Fellow at the Digital Signal Processing Group at MIT. She is currently a Professor in the Department of Electrical Engineering at the Technion-Israel Institute of Technology, Haifa. She is also a Research Affiliate with the Research Laboratory of Electronics at MIT and a Visiting Professor at Stanford University, Stanford, CA. Her research interests are in the broad areas of statistical signal processing, sampling theory and compressed sensing, optimization methods, and their applications to biology and optics.

Dr. Eldar was in the program for outstanding students at TAU from 1992 to 1996. In 1998, she held the Rosenblith Fellowship for study in electrical engineering at MIT, and in 2000, she held an IBM Research Fellowship. From 2002 to 2005, she was a Horev Fellow of the Leaders in Science and Technology program at the Technion and an Alon Fellow. In 2004, she was awarded the Wolf Foundation Krill Prize for Excellence in Scientific Research, in 2005 the Andre and Bella Meyer Lectureship, in 2007 the Henry Taub Prize for Excellence in Research, in 2008 the Hershel Rich Innovation Award, the Award for Women with Distinguished Contributions, the Muriel & David Jacknow Award for Excellence in Teaching, and the Technion Outstanding Lecture Award, in 2009 the Technions Award for Excellence in Teaching, in 2010 the Michael Bruno Memorial Award from the Rothschild Foundation, and in 2011 the Weizmann Prize for Exact Sciences. She is a member of the IEEE Signal Processing Theory and Methods technical committee and the Bio Imaging Signal Processing technical committee, an Associate Editor for the SIAM Journal on Imaging Sciences, and on the Editorial Board of Foundations and Trends in Signal Processing. In the past, she served as an associate editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the EURASIP Journal of Signal Processing, and the SIAM Journal on Matrix Analysis and Applications.