Recovery Guarantees for Rank Aware Pursuits

Jeffrey D. Blanchard and Mike E. Davies

Abstract

This paper considers sufficient conditions for sparse recovery in the sparse multiple measurement vector (MMV) problem for some recently proposed rank aware greedy algorithms. Specifically we consider the compressed sensing framework with random measurement matrices and show that the rank of the measurement matrix in the sparse MMV problem allows such algorithms to reduce the effect of the $\log n$ term that is present in traditional OMP recovery.

Index Terms

Multiple Measurement Vectors, Greedy Algorithm, Othogonal Matching Pursuit, rank.

I. Introduction

Sparse signal representations provide a general signal model that make it possible to solve many ill-posed problems such as source separation, denoising and most recently compressed sensing [1] by exploiting the additional sparsity constraint. While the general problem of finding the sparsest $\mathbf{x} \in \mathbb{R}^n$ given an observation vector $\mathbf{y} = \Phi\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^m$ is known to be NP-hard [2] a number of suboptimal stategies have been shown to be able to recover $\mathbf{x}$ when $m \sim Ck \log(n/k)$ if Φ is chosen judiciously.

An interesting extension of the sparse recovery problem is the sparse multiple measurement vector (MMV) problem, $\mathbf{Y} = \Phi\mathbf{X}$, $\mathbf{Y} \in \mathbb{R}^{m\times l}$, $\mathbf{X} \in \mathbb{R}^{n\times l}$, which has also received much attention, e.g. [3], [4], [5]. Initially the algorithms proposed for this problem were straightforward extensions of existing single measurement vector (SMV) solutions. However, most of these are unable to exploit the additional information available through the rank of Y. In contrast, some new greedy algorithms for joint sparse

J.D. Blanchard is with the Department of Mathematics and Statistics, Grinnell College, Grinnell, Iowa USA. T:+1 801 585 1644, F:+1 801 581 4148 (jeff@math.grinnell.edu). JDB was supported as an NSF International Research Fellow (NSF OISE 0854991) and by NSF DMS 1112612.

M.E. Davies is with IDCOM, the University of Edinburgh, Edinburgh, UK, and the Joint Research Institute for Signal and Image Processing, part of the Edinburgh Research Partnership in Engineering and Mathematics. (m.davies@ed.ac.uk). This work is was supported in part by EPSRC grant EP/F039697/1 and the European Commission through the SMALL project under FET-Open, grant number 225913.

recovery have been proposed [6], [7], [8] based around the MUSIC (MUltiple SIgnal Classification) algorithm [9] from array signal processing which provides optimal recovery in the maximal rank scenario $r = k$ [10].

The aim of this paper is to analyse the recovery performance of two Rank Aware algorithms when the observation matrix does not have maximal rank, rank$(\mathbf{Y}) < k$. Our approach follows the recovery analysis of [11] where it was shown that Orthogonal Matching Pursuit (OMP) can recover k-sparse vectors from $m \gtrsim Ck \log n$ measurements with high probability. We extend this analysis to the MMV sparse recovery problem and show joint k-sparse matrices, X can be recovered from $m \gtrsim Ck(\frac{1}{r}\log n + 1)$ MMVs using a rank aware algorithm.[1] This implies that the $\log n$ penalty term observed for OMP recovery can be essentially removed with very modest values of rank, $r \gtrsim \log n$.

II. Notation and Problem Formulation

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: supp$(\mathbf{X}) := \bigcup_i \text{supp}(\mathbf{x}_i)$. A matrix $\mathbf{X}$ is called k joint sparse if $|\text{supp}(\mathbf{X})| \leq k$. We make use of the subscript notation $\mathbf{\Phi}_\Omega$ to 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. Thus denoting by $|\Omega|$ the cardinality of Ω, the matrix $\mathbf{X}_{\Omega,:}$ is $|\Omega|$-sparse.

We can now formally define the sparse MMV problem. Consider the observation matrix $\mathbf{Y} = \Phi\mathbf{X}$, $\mathbf{Y} \in \mathbb{R}^{m\times l}$ where $\mathbf{\Phi} \in \mathbb{R}^{m\times n}$ with $m < n$ is the dictionary matrix and $\mathbf{X} \in \mathbb{R}^{n\times l}$ is assumed to be jointly k-sparse. The task is then to recover X from Y given Φ. We will further assume that rank$(Y) = r$ and without loss of generality that $r = l$.

III. Greedy MMV Algorithms

Despite the fact that the rank of the observation matrix $\mathbf{Y}$ can be exploited to improve recovery performance, to date most popular techniques have ignored this fact and have been shown to be "rank-blind" [6]. In contrast, a discrete version of the MUSIC algorithm [10], [13] is able to recover X from Y under mild conditions on Φ whenever $m \geq k + 1$ if we are in the maximal rank case, i.e. rank$(Y) = k$.

While MUSIC provides guaranteed recovery for the MMV problem in the maximal rank case there are no performance guarantees for when rank$(\mathbf{X}) < k$ and empirically MUSIC does not perform well in this scenario. This motivated a number of works [6], [7], [8] to investigate the possibility of an algorithm

[1] These results were previously announced at the "SMALL" workshop, London, January 2011 [12].

that in some way interpolates between a classical greedy algorithm for the SMV problem and MUSIC when $\text{rank}(\mathbf{X}) = k$.

The approach proposed in [7], [8] was to use a greedy selection algorithm to find the first $t = k - r$ coefficients. The remaining components can then be found by applying MUSIC to an augmented data matrix $[\mathbf{Y}, \Phi_{\Omega}^{(t)}]$ which under identifiability assumptions will span the range of Φ_Λ.

In [6] two "rank aware" (RA) algorithms were presented. In RA-OMP the greedy selection step was modified to measure the distance of the columns of Φ from the subspace spanned by the residual matrix at iteration t by measuring the correlation of columns of Φ with an orthonormal basis of the residual matrix: $\mathbf{U}^{(j-1)} = \text{ortho}(\mathbf{R}^{(j-1)})$.[2] However, the recovery performance was shown to deteriorate even in the maximal rank scenario as the algorithm selected more coefficients. To compensate for this, the column normalization used in Order Recursive Matching Pursuit (ORMP) was included. Specifically at the start of the tth iteration, if we have a selected support set $\Omega^{(t)}$, a new column Φ_i is then chosen based upon the following selection rule:

$$i^{(t)} = \arg\max_i \frac{\|\varphi_i^T \mathbf{U}^{(t)}\|_2}{\|\mathbf{P}_{\Omega^{(t)}}^{\perp} \varphi_i\|_2}, \tag{1}$$

where $\mathbf{P}_{\Omega^{(t)}}^{\perp}$ denotes the orthogonal projection onto the null space of $\Phi_{\Omega^{(t)}}$. The righthand side of (1) measures the distance of the *normalized* vector $\mathbf{P}_{\Omega^{(t)}}^{\perp}\varphi_i / \|\mathbf{P}_{\Omega^{(t)}}^{\perp}\varphi_i\|$ from the subspace spanned by $\mathbf{U}^{(t)}$. This ensures that correct selection is maintained at each iteration in the maximal rank scenario.

The full description of the RA-OMP and RA-ORMP are summarized in Algorithm 1. Note in this form they only differ at the index selection step.

In the next section the sparse recovery guarantees for RA-ORMP and RA-OMP+MUSIC (using RA-OMP to select the first $k-r$ indices followed by the modified MUSIC algorithm of [7], [8]) are examined and shown to exploit the rank of $\mathbf{Y}$ very efectively. The RA-OMP+MUSIC algorithm is very similar to the subspace MUSIC proposed by [7] although in [7] only a single orthogonalization of the observation matrix $\mathbf{Y}$ was performed followed by simultaneous OMP (SOMP). [8] considered SOMP+MUSIC but without an initial orthogonalization.[3] Both theoretical and empirical recovery performance of SOMP+MUSIC is very limited due to the "rank-blind" property of SOMP.

[2] In practice, observation noise needs to be considered and an orthonormal basis for the signal subspace should be estimated [9].

[3] While writing up this work we became aware of an updated version of [8] where the authors have switched to considering the RA-OMP+MUSIC proposed here instead of SOMP+MUSIC.

Algorithm 1 RA-OMP / RA-ORMP

initialization: $\mathbf{R}^{(0)} = \mathbf{Y}, \mathbf{X}^{(0)} = \mathbf{0}, \Omega^0 = \emptyset$
for $j = 1; j := j + 1$ **until stopping criterion do**
  Calculate $\mathbf{U}^{(j-1)} = \text{ortho}(\mathbf{R}^{(j-1)})$,
  if (RA-OMP) **then**
    $i^{(j)} = \arg\max_i \|\varphi_i^T \mathbf{U}^{(j-1)}\|_2$
  else if (RA-ORMP) **then**
    $i^{(j)} = \arg\max_{i \notin \Omega^{(j-1)}} \|\varphi_i^T \mathbf{U}^{(j-1)}\|_2 / \|\mathbf{P}^{\perp}_{\Omega^{(j-1)}} \varphi_i\|_2$
  end if
  $\Omega^{(j)} = \Omega^{(j-1)} \cup i^{(j)}$
  $\mathbf{X}^{(j)}_{\Omega^{(j)},:} = \Phi^{\dagger}_{\Omega^{(j)}} \mathbf{Y}$
  $\mathbf{R}^{(j)} = \mathbf{Y} - \Phi \mathbf{X}^{(j)}$
end for

IV. Sparse MMV recovery by RA-OMP

Correct selection by the RA-OMP algorithm at the jth iteration is characterized by the following quantity.

Definition 1 (Greedy Selection Ratio for RA-OMP)**.** In iteration j of RA-OMP, let $\Lambda = \text{supp}(X)$ and define the *greedy selection ratio for RA-OMP* as

$$\rho(j) = \frac{\max_{i' \in \Lambda^c} \|\varphi_{i'}^T \mathbf{U}^{(j)}\|_2}{\max_{i \in \Lambda} \|\varphi_i^T \mathbf{U}^{(j)}\|_2}. \tag{2}$$

The following observation is obvious.

Lemma 1. *In iteration j, RA-OMP will correctly select an atom $\varphi_j, j \in \Lambda = supp(\mathbf{X})$ if and only if $\rho(j) < 1$.*

To bound $\rho(j)$ we introduce the following lemmas.

Lemma 2. *Suppose $\mathbf{U} \in \mathbb{R}^{m \times r}$ with columns $\mathbf{U}_i \in range(\Phi_\Lambda)$ and $\|\mathbf{U}_i\|_2 = 1$ for all i. Let $\alpha = \sigma_{min}(\Phi_\Lambda)$ be the smallest singular value of Φ_Λ with $|\Lambda| = k < m$. Then*

$$\max_{i \in \Lambda} \|\varphi_i^T \mathbf{U}\|_2 \geq \alpha \sqrt{\frac{r}{k}}. \tag{3}$$

Proof: We can write

$$\|\varphi_i^T \mathbf{U}\|_2^2 = \|\mathbf{e}_i^T \Phi_\Lambda^T \mathbf{U}\|_2^2 \tag{4}$$

where $\mathbf{e}_i$ is the standard (dirac) basis in $\mathbb{R}^k$.

Define $\tilde{\mathbf{U}} = \Phi_\Lambda^T \mathbf{U}$. Since $\mathbf{U}_i$ is in the range of Φ_Λ we have the following bound:

$$\|\tilde{\mathbf{U}}_i\|_2^2 \geq \sigma_{min}(\Phi_\Lambda)^2 \|\mathbf{U}_i\|_2^2 = \alpha^2 \tag{5}$$

Hence

$$\begin{aligned}\max_{i\in\Lambda} \|\varphi_i^T \mathbf{U}\|_2^2 = \max_{i\in\Lambda} \|\mathbf{e}_i^T \tilde{\mathbf{U}}\|_2^2 &= \max_{i\in\Lambda} \sum_{l=1}^{r} [\tilde{\mathbf{U}}]_{i,l}^2 \\ &\geq \underset{i}{\text{mean}}\ [\tilde{\mathbf{U}}]_{i,l}^2 = \frac{1}{k}\sum_{i=1}^{k}\sum_{l=1}^{r} [\tilde{\mathbf{U}}]_{i,l}^2 \\ &= \frac{1}{k}\sum_{l=1}^{r} \|\tilde{\mathbf{U}}_l\|_2^2 \geq \frac{r}{k}\alpha^2\end{aligned} \tag{6}$$

where we have bounded the maximum by the mean and then swapped the order of summation. □

Lemma 3. *If $\Phi \in \mathbb{R}^{m\times n}$ with entries draw i.i.d. from $\mathcal{N}(0, m^{-1})$, $\Lambda \subset \{1, \dots, n\}$ is an index set with $|\Lambda| = k$, and $\mathbf{U} \in \mathbb{R}^{m\times r}$ is a matrix with orthonormal columns, rank$(\mathbf{U}) = r$ and with span$(\mathbf{U}) \subset$ span(Φ_Λ), then*

$$\mathbb{P}\left\{\max_{i\in\Lambda^c} \|\varphi_i^T \mathbf{U}\|_2^2 < \mu^2\right\} \geq 1 - (n-k)e^{-(m\mu^2 - 2r)/4}. \tag{7}$$

Proof: Let $z = \phi_i^T \mathbf{U}$ then $z \in \mathbb{R}^r$ and for $i \notin \Lambda$ $z \sim \mathcal{N}(0, m^{-1})$. We can now use the Laplace transform method [14] to bound $\|z\|_2$.

$$\begin{aligned}\mathbb{P}\Big\{\|z\|_2^2 \geq \mu^2\Big\} &\leq e^{-\lambda\mu^2}\mathbb{E}\Big\{e^{\lambda\|z\|_2^2}\Big\} \\ &= e^{-\lambda\mu^2 + r\ln\left(\frac{m}{m-2\lambda}\right)}\end{aligned} \tag{8}$$

for any $\lambda > 0$. Selecting $\lambda = m/4$ gives

$$\mathbb{P}(\|z\|_2^2 \geq \mu^2) \leq e^{-(m\mu^2 - 2r)/4}. \tag{9}$$

Applying the union bound completes the result. □

We will also require that the residual matrix, $\mathbf{R}^{(j)}$, retains generic rank (equivalent to the row degenerancy condition in [7]) which is given by:

Lemma 4. *Let $\Phi \in \mathbb{R}^{m\times n}$ with entries draw i.i.d. from $\mathcal{N}(0, m^{-1})$ and $\mathbf{X}$ be a joint sparse matrix with support $\Lambda \subset \{1, \dots, n\}$, $|\Lambda| = k$, such that $\mathbf{X}_{\Lambda,:}$ is in general postion. If $\Omega^{(j)} \subset \Lambda$ for $j \leq k - r$, then rank$(\mathbf{R}^{(j)}) = r$.*

Proof: Note that:

$$\mathbf{R}^{(j)} = \mathbf{P}^{\perp}_{\Omega^{(j)}} Y = \mathbf{P}^{\perp}_{\Omega^{(j)}} \Phi_{\Lambda-\Omega^{(j)}} \mathbf{X}_{\Lambda-\Omega^{(j)},:} \ . \tag{10}$$

Since $\mathbf{X}_{\Lambda,:}$ is in general position $\text{rank}(\mathbf{X}_{\Lambda-\Omega^{(j)},:}) = \min\{r, k-j\} = r$, and since Φ is i.i.d. Gaussian then $\mathbf{P}^{\perp}_{\Omega^{(j)}} \Phi_{\Lambda-\Omega^{(j)}}$ will have maximal rank with probability 1. Therefore $\text{rank}(\mathbf{R}^{(j)}) = r$. □

We can combine the above lemmata to give:

Lemma 5. *Suppose that after $j < k - r$ iterations of RA-OMP, $\Omega^{(j)} \subset \Lambda = supp(\mathbf{X})$ with $|\Lambda| = k$. Define $\mathbf{R}^{(j)} = \mathbf{Y} - \Phi\mathbf{X}^{(j)}$ and assume $\mathbf{X}$ is in general position and $\Phi \in \mathbb{R}^{m\times n}$ with entries drawn i.i.d. from $\mathcal{N}(0, m^{-1})$. Then, for $\alpha = \sigma_{min}(\Phi_\Lambda)$,*

$$\mathbb{P}\left\{\Omega^{(j+1)} \subset \Lambda\right\} \geq 1 - (n-k)e^{-(m\alpha^2\frac{k}{r} - 2r)/4}. \tag{11}$$

Proof: Let $\mathbf{U}^{(j)} = \text{ortho}(\mathbf{R}^{(j)})$ then from Lem. 4 we have $\text{rank}(\mathbf{U}^{(j)}) = \text{rank}(\mathbf{R}^{(j)}) = \text{rank}(\mathbf{X}) = r$. Now, Lem. 1 and the assumption that $\Omega^{(j)} \subset \Lambda$ allow us to rewrite the probability statement as

$$\mathbb{P}\left\{\Omega^{(j+1)} \subset \Lambda\right\} = \mathbb{P}\left\{\rho(j+1) < 1\right\}.$$

Lemmas 2 and 3 with $\mu^2 = \alpha^2 r/k$ combine to show that

$$\begin{aligned}
\mathbb{P}\left\{\rho(j+1) < 1\right\} &= \mathbb{P}\left\{\max_{j'\in\Lambda^c} \|\varphi_{j'}^T \mathbf{U}\|_2 < \max_{j\in\Lambda} \|\varphi_j^T \mathbf{U}\|_2\right\} \\
&\geq \mathbb{P}\left\{\max_{j'\in\Lambda^c} \|\varphi_{j'}^T \mathbf{U}\|_2^2 < \frac{\alpha^2 r}{k}\right\} \\
&\geq 1 - (n-k)e^{-(m\alpha^2\frac{k}{r} - 2r)/4}.
\end{aligned}$$

□

We can now state our main theorem for RA-OMP.

Theorem 6 (RA-OMP + MUSIC recovery)**.** *Assume $\mathbf{X} \in \mathbb{R}^{n\times r}$, $supp(\mathbf{X}) = \Lambda$, $|\Lambda| = k > r$ with $\mathbf{X}_\Lambda$ in general position and let Φ be a random matrix, independent of $\mathbf{X}$, with i.i.d. entries $\Phi_{i,j} \sim \mathcal{N}(0, m^{-1})$. Then, for some C and with probability greater than $1 - \delta$, RA-OMP + MUSIC will recover $\mathbf{X}$ from $\mathbf{Y} = \Phi\mathbf{X}$ if:*

$$m \geq Ck\big(\frac{1}{r}\log(n/\sqrt{\delta}) + 1\big). \tag{12}$$

Proof: It is sufficient to bound the probability of making $q \leq k - r$ successive correct selections after which MUSIC is guaranteed to recover the remaining coefficients [7], [8]. Suppose $\sigma_{\min}(\Phi_\Lambda) = \alpha$, then

$$\begin{aligned}\mathbb{P}\{\max_{t\leq q}\rho(t)<1\} &\geq \prod_{t\leq q}\mathbb{P}\{\rho(t)<1\} \\ &\geq \left[1-(n-k)e^{-(m\alpha^2 r/k-2r)/4}\right]^q \\ &\geq 1-q(n-k)e^{-(m\alpha^2 r/k-2r)/4}.\end{aligned} \tag{13}$$

Now recall [11] that $\mathbb{P}\{\sigma_{\min}(\Phi_\Lambda)>0.5\}\geq 1-e^{-cm}$ for some $c>0$. Hence:

$$\begin{aligned}\mathbb{P}\{\max_{t\leq q}\rho(t)<1\} &\geq \left(1-n^2e^{-(m\alpha^2 r/k-2r)/4}\right)\left(1-e^{-cm}\right) \\ &\geq 1-n^2e^{-(mr/k-r/2)/4}-e^{-cm} \\ &\geq 1-n^2e^{-C(mr/k-r/2)}\end{aligned} \tag{14}$$

where we have used the fact that $q(n-k)<n^2$. Now choosing $\delta\geq n^2e^{-C(mr/k-r/2)/2}$ and rearranging gives (12). □

V. SPARSE MMV RECOVERY BY RA-ORMP

Given that the only difference between RA-OMP and RA-ORMP is in the selection step we can use similar aguments for RA-ORMP recovery. The key difference to section IV is the need to control the normalization term $\|P^{\perp}_{\Omega^{(j)}}\varphi_i\|_2$ within the selection step. We will use the following.

Lemma 7. *If $\Phi\in\mathbb{R}^{m\times n}$ with entries draw i.i.d. from $\mathcal{N}(0,m^{-1})$, $\Lambda\subset\{1,\ldots,n\}$, $|\Lambda|=k$, and $\Omega^{(j)}\subset\Lambda$, $|\Omega^{(j)}|=j$, then for $i\notin\Omega^{(j)}$ and $m\geq 2j$ we have:*

$$\mathbb{P}\left\{\|P^{\perp}_{\Omega^{(j)}}\varphi_i\|_2^2\geq\frac{1}{4}\right\}\geq 1-e^{-m/32} \tag{15}$$

Proof: Since $P^{\perp}_{\Omega^{(j)}}$ and φ_i are independent, $z:=P^{\perp}_{\Omega^{(j)}}\varphi_i$ is a Gaussian random vector within the $m-j$ dimensional subspace Null$(\Phi^T_{\Omega^{(j)}})$ with variance m^{-1}. Hence we can use the following concentration of measure bound [14]:

$$\mathbb{P}\left\{\|z\|_2^2\geq(1-\epsilon)\frac{m-j}{m}\right\}\geq 1-e^{-\epsilon^2(m-j)/4} \tag{16}$$

Selecting $\epsilon=1/2$ and noting that by assumption $m-j\geq m/2$ gives the required result. □

We could similarly upper bound $\|P^{\perp}_{\Omega^{(j)}}\varphi_i\|_2$, however, we only require the crude bound $\|P^{\perp}_{\Omega^{(j)}}\varphi_i\|_2\leq 1$.

Theorem 8 (RA-ORMP recovery)**.** *Assume $\mathbf{X}\in\mathbb{R}^{n\times r}$, supp$(\mathbf{X})=\Lambda$, $|\Lambda|=k>r$ with $\mathbf{X}_\Lambda$ in general position and let Φ be a random matrix, independent of $\mathbf{X}$, with i.i.d. entries $\Phi_{i,j}\sim\mathcal{N}(0,m^{-1})$. Then, for* *some C and with probability greater than $1-\delta$, RA-ORMP will recover $\mathbf{X}$ from $\mathbf{Y}=\Phi\mathbf{X}$ if m satisfies* (12).

Proof: We first note that if $\Omega^{(j)} \subset \Lambda$, $j < k-r$ and $\alpha = \sigma_{min}(\Phi_\Lambda)$, then

$$\begin{aligned} &\mathbb{P}\left\{\Omega^{(j+1)} \subset \Lambda\right\} \\ &= \mathbb{P}\left\{\arg\max_{i'\in\Lambda^c} \frac{\|\varphi_{i'}^T\mathbf{U}^{(j)}\|_2^2}{\|\mathbf{P}_{\Omega^{(j)}}^\perp \varphi_{i'}\|_2^2} \le \arg\max_{i\in\Lambda} \frac{\|\varphi_{i}^T\mathbf{U}^{(j)}\|_2^2}{\|\mathbf{P}_{\Omega^{(j)}}^\perp \varphi_{i}\|_2^2}\right\} \\ &\geq \mathbb{P}\left\{\arg\max_{i'\in\Lambda^c} \frac{\|\varphi_{i'}^T\mathbf{U}^{(j)}\|_2^2}{\|\mathbf{P}_{\Omega^{(j)}}^\perp \varphi_{i'}\|_2^2} \le \frac{\alpha^2 r}{k}\right\} \\ &\geq \mathbb{P}\left\{\arg\max_{i'\in\Lambda^c} \|\varphi_{i'}^T\mathbf{U}^{(j)}\|_2^2 \le \frac{\alpha^2 r}{4k}\right\} \\ &\quad \times \mathbb{P}\left\{\arg\max_{i'\in\Lambda^c} \|P_{\Omega^{(j)}}^\perp \varphi_{i}\|_2^2 \geq \frac{1}{4}\right\}. \end{aligned} \tag{17}$$

Now using the bounds from Lemma 3 with $\mu^2 = \alpha^2 r/4k$ and Lemma 7 gives:

$$\begin{aligned} &\mathbb{P}\left\{\Omega^{(j+1)} \subset \Lambda\right\} \\ &\geq \left(1-(n-k)e^{-(m\alpha^2\frac{k}{r}-2r)/4}\right)\left(1-(n-k)e^{-m/32}\right) \\ &\geq 1-(n-k)\left(e^{-(m\alpha^2\frac{k}{r}-2r)/4}+e^{-m/32}\right) \\ &\geq 1-(n-k)e^{-C(m\alpha^2\frac{k}{r}-2r)}. \end{aligned} \tag{18}$$

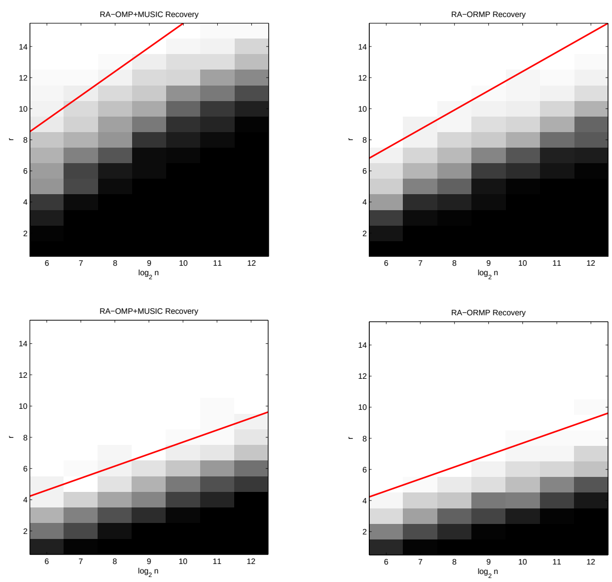
To complete the proof for the correct selection of $\Omega^{(j)}$ for all $j < k-r$ we can again apply the union bound and remove the dependence on α as in (13) and (14) above. We leave the details to the reader.

For the selection of the remaining coefficients we note that for $j \geq k-r$ the original RA-ORMP task is equivalent to solving $\mathbf{R}^{(j)} = [\mathbf{P}_{\Omega^{(j)}}^\perp \Phi_{\Lambda-\Omega^{(j)}}]X_{\Lambda-\Omega^{(j)},:}$ using RA-ORMP. However by our assumptions on $\mathbf{X}$, rank$(\mathbf{X}_{\Lambda-\Omega^{(j)},:}) = r = |\Lambda - \Omega^{(j)}|$ therefore we are in the maximal rank scenario and from [6] we have guaranteed recovery by RA-ORMP. □

VI. Numerical Results

Here we demonstrate empirically that the $(\log n)/r$ term in our recovery result appears to accurately capture the effect of rank on the recovery performance. To this end we performed a number of experiments using Gaussian random matrices for both Φ and $\mathbf{X}_\Lambda$. The parameters m and k were held fixed, first at $m = 3k/2 = 30$ and then with $m = 2k = 40$. We then varied the number of channels of $\mathbf{X}$ from $r = 1, \ldots, 15$ and n in powers of 2 from 64 to 4096. For each set of $\{k, r, m, n\}$ we performed 100 trials and calculated the empirical probability of recovery.

Figure 1 shows the recovery plots for the recovery algorithms RA-OMP + MUSIC and RA-ORMP. In each case the "phase transition" appears to show a clear linear dependency between r and $\log n$ as highlighted by the red line.



(a) (b)

Fig. 1. Sparse MMV recovery plots showing the "phase transitions" for RA-OMP+MUSIC (a) and RA-ORMP (b) with $m = 3k/2 = 30$ (top) and $m = 2k = 40$ (bottom) while varying the size of the dictionary $n = 64, 128, \ldots, 4096$ and number of channels, $r = 1, 2, \ldots, 15$. The red line indicates a linear relation between r and $\log n$.

VII. Conclusion

Our theoretical results predict that the rank of the coefficient matrix in the sparse MMV recovery problem can be successfully exploited in RA-OMP+MUSIC and RA-ORMP to enable joint sparse recovery when $m \gtrsim Ck((\log n)/r + 1)$ and thus to remove the $\log n$ penalty term that is observed in OMP even when there is only a modest number of mutliple measurement vectors $r \gtrsim \log n$. Numerical experiments suggest that this form accurately characterizes the recovery behaviour in practice. Empirically the RA-ORMP algorithm appears to perform slightly better than RA-OMP+MUSIC however this comes with an additional computational expense.

References

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

[2] B. K. Natarajan, “Sparse approximate solutions to linear systems,” *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, Apr 1995.

[3] 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.

[4] 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.

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

[6] M. E. Davies and Y. C. Eldar, “Rank awareness and in joint sparse recovery,” *IEEE Trans. Information Theory*, vol. to appear, 2011.

[7] K. Lee and Y. Bresler, “imusic: Iterative music algorithm for joint sparse recovery with any rank,” 2010, arxiv preprint: arXiv:1004.3071v1.

[8] J. Kim, O. Lee, and J. C. Ye, “Compressive music: A missing link between compressive sensing and array signal processing,” arxiv preprint, arXiv:1004.4398v1.

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

[10] P. Feng, “Universal minimum-rate sampling and spectrum-blind reconstruction for multiband signals,” Ph.D. dissertation, University of Illinois, 1998.

[11] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Transactions on Information Theory*, vol. 53(12), no. 12, pp. 4655–4666, 2007.

[12] M. E. Davies. (2011, Jan.) Rank aware algorithms for joint sparse recovery. SMALL Workshop on sparse dictionary learning. [Online]. Available: http://small-project.eu/small-fs/videos/

[13] 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*, 1996, pp. 1688–1691.

[14] A. Barvinok. (2005) Measure concentration lecture notes. [Online]. Available: www.math.lsa.umich.edu/ barvinok/total710.pdf