Recovery of Low Rank and Jointly Sparse Matrices with Two Sampling Matrices

Sampurna Biswas, Hema K. Achanta, Mathews Jacob, Soura Dasgupta, and Raghuraman Mudumbai

***Abstract*—We provide a two-step approach to recover a jointly k-sparse matrix $\mathbf{X}$, (at most k rows of $\mathbf{X}$ are nonzero), with rank $r << k$ from its under sampled measurements. Unlike the classical recovery algorithms that use the same measurement matrix for every column of $\mathbf{X}$, the proposed algorithm comprises two stages, in each of which the measurement is taken by a different measurement matrix. The first stage uses a standard algorithm, [4] to recover any r columns (e.g. the first r) of $\mathbf{X}$. The second uses a new set of measurements and the subspace estimate provided by these columns to recover the rest. We derive conditions on the second measurement matrix to guarantee perfect subspace aware recovery for two cases: First a worst-case setting that applies to all matrices. The second a generic case that works for almost all matrices. We demonstrate both theoretically and through simulations that when $r << k$ our approach needs far fewer measurements. It compares favorably with recent results using dense linear combinations, that do not use column-wise measurements.**

***Index Terms*—Dynamic imaging, joint sparsity, low rank, rank aware ORMP.**

I. Introduction

THE multiple measurement vector (MMV) problem considers the recovery of a matrix $\mathbf{X} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n] \in \mathbb{R}^{m\times n}$ that is *jointly k-sparse* (i.e, only k-rows of $\mathbf{X}$ are nonzero) [1]–[6]. Current schemes sample all the columns of the matrix using the *same measurement matrix* $\mathbf{A} \in \mathbb{R}^{s\times m}$, $s < m$, [1], [2], [4]–[7]:

$$\mathbf{Y} = \mathbf{A}\mathbf{X}. \tag{1}$$

The theoretical results show that $\mathbf{X}$ can be recovered from $\mathbf{Y}$ using combinatorial searches if

$$k < \frac{\text{rank}(\mathbf{X}) + \text{spark}(\mathbf{A}) - 1}{2}, \tag{2}$$

where $\text{spark}(\mathbf{A})$ is the smallest number of linearly dependent columns of $\mathbf{A}$ [1], [5].

Manuscript received March 19, 2015; revised June 09, 2015; accepted June 09, 2015. Date of publication June 19, 2015; date of current version June 25, 2015. This work was supported in part by the U.S. National Science Foundation under Grants EPS-1101284, ECCS-1150801, CNS-1329657, CCF-1302456, CCF-1116067, NIH 1R21HL109710-01A1, ACS RSG-11-267-01-CCE, and ONR N00014-13-1-0202. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Michael Wakin.

The authors are with the Department of Electrical and Computer Engineering, University of Iowa, IA 52242 USA (e-mail: sampurna-biswas@uiowa.edu; hemakumari-achanta@uiowa.edu; mathews-jacob@uiowa.edu; soura-dasgupta@uiowa.edu; raghuraman-mudumbai@uiowa.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2015.2447455

When the signal matrix $\mathbf{Y}$ has full row rank, this approach thus offers a two fold reduction in the number of measurements over the independent recovery of the columns using classical compressed sensing [1].

In several applications, including the recovery of dynamic MRI data motivating this paper [8], [9], $\mathbf{x}_i$ live in a subspace of dimension $r << k$. Consequently both $\mathbf{X}$ and $\mathbf{Y}$ have rank much smaller than k. In such cases the gain offered by the classical MMV schemes over independently recovering the vectors using compressed sensing is small [5], [4], [2]. In this setting, there is another class of previous work that involves recovery of low-rank and jointly sparse matrices from their undersampled measurements [10], [11]. Unlike (1), [10], [11] do not assume column-wise measurements. Such measurement schemes are unrealizable in applications like diffuse optical or fluorescent tomography, and dynamic imaging, where each column of $\mathbf{X}$ corresponds to a frame in the image time series, and each measurement is the linear combination of entries of just one specific column.

Motivated by dynamic imaging applications, we assume the measurement setting of (1) with $r = \text{rank}(\mathbf{X}) << k$. We consider the partition of $\mathbf{X}$, specified by

$$\mathbf{X} = [\,\mathbf{X}_1 \quad \mathbf{X}_2\,], \tag{3}$$

where $\mathbf{X}_1 \in \mathbb{R}^{m\times r}$ and $\mathbf{X}_2 \in \mathbb{R}^{m\times(n-r)}$. We assume the observations to be specified by

$$\mathbf{Y}_1 = \mathbf{A}_1\mathbf{X}_1 \tag{4}$$

$$\mathbf{Y}_2 = \mathbf{A}_2\mathbf{X}_2, \tag{5}$$

where $\mathbf{A}_i \in \mathbb{R}^{s_i\times m}$. We introduce a two stage algorithm to recover $\mathbf{X}$. In the first stage, we measure $\mathbf{X}_1$ as in (4) and recover it from $\mathbf{Y}_1$ using the classical MMV scheme. Since we rely on classical MMV scheme and results from [1], [5] to solve for $\mathbf{X}_1$, our theoretical results for the recovery of $\mathbf{X}_1$ also assume combinatorial search. The recovered columns of $\mathbf{X}_1$ provide an estimate of the r-dimensional subspace spanned by the $\mathbf{x}_i$. Once the subspace of $\mathbf{X}$ is estimated from $\mathbf{X}_1$, the subspace aware recovery of $\mathbf{X}_2$ follows from a matrix inversion. To guarantee that any $\mathbf{X}_1 \in \mathbb{R}^{m\times r}$ captures the entire subspace of $\mathbf{X}$, we require $\text{spark}(\mathbf{X}) = r+1$. This condition is not too restrictive, since $\mathbf{x}_i$ drawn from an r-dimensional subspace will generically satisfy this requirement. In this two step recovery method, the results of the second stage potentially depends on the accuracy of the support and subspace recovered in the first step. Our simulations show that the error in the first step is small with more measurements. Since only few columns are recovered in the first step,

increasing the number of measurements in the first step does not considerably increase the total number of measurements.

The main contributions of this paper are formulating this two stage recovery approach and providing sufficient conditions for subspace aware recovery of $\mathbf{X}_2$. In particular, with respect to the latter, we show the following.

(a) We show that $\text{spark}(\mathbf{A}_2) \geq k$ is a sufficient condition to guarantee the subspace aware recovery of any jointly k-sparse $\mathbf{X}_2$ (assuming that the first step has successfully recovered $\mathbf{X}_1$), with spark $r+1$, from (4) and (5). We term this condition as the *worst-case* sufficient condition.

(b) We show that for any $\mathbf{X}_2$ of rank r, a matrix $\mathbf{A}_2$ with r rows that does not uniquely recover $\mathbf{X}_2$, lies on a set of zero measure. This condition is termed as the *generic case* sufficient condition.

The worst case yields a two-fold gain over traditional schemes. The generic case result shows that the worst case conditions are pessimistic: in practice, perfect recovery is possible with far fewer measurements with almost all r row measurement matrices. This is so as for a given $\mathbf{X}_2$ an $\mathbf{A}_2$ with r rows that does not provide unique recovery lies on a set of measure zero. Thus such a randomly selected $\mathbf{A}_2$ should effect exact recovery with high probability. Indeed as is well known in the vast literature on random matrices, an $\mathbf{A}_2$ with elements from i.i.d. Gaussian distributions, [13] should effect recovery with high probability. The number of measurements that are sufficient to recover almost all matrices compares favorably with that quantified in the dense sampling setting of [10], [11].

The intuition behind the above results on subspace recovery of $\mathbf{X}_2$ can be easily illustrated using a rank-1 matrix $\mathbf{X}$. If the first column $\mathbf{x}_1 = \mathbf{X}_1$ is available, the remaining columns $\mathbf{X}_2$ can be recovered using only one *non-zero* measurement per column, by choosing $\mathbf{A}_2$ to be a row vector, provided $\mathbf{A}_2\mathbf{x}_1 \neq 0$. The set of all row vectors $\mathbf{A}_2$, which satisfy $\mathbf{A}_2\mathbf{X}_2 = 0$, has zero measure for any specific $\mathbf{X}_2$; this is the *generic case sufficient condition*. Thus, one measurement per column (e.g. using a row vector, whose entries are e.g. Gaussian random variables) will suffice to recover $\mathbf{X}_2$. The condition $\mathbf{A}_2\mathbf{x}_1 \neq 0$ can be ensured for all k-sparse $\mathbf{x}_1$ if the spark of $\mathbf{A}_2$ is at least $k+1$, which implies that $\mathbf{A}_2$ has at least k rows; this is the *worst-case sufficient condition* on $\mathbf{A}_2$, which will guarantee the recovery of any $\mathbf{X}_2$, including the worst possible $\mathbf{X}_2$.

Section II formulates the MMV problem and justifies our approach. A practical algorithm is presented in Section III. Section IV gives simulations and Section V concludes.

II. The Multiple Measurement Vector Problem

Since MMV scheme applies the same measurement matrix to all columns of $\mathbf{X}$, the total number of measurements in (1) is at least $\mathcal{N}_{MMV} = (2k-r+2)n$; r is the rank of $\mathbf{Y}$. Thus in the best case scenario ($r = \text{spark}(\mathbf{A})-1$), MMV can provide a factor of two reduction in the number of measurements over the independent recovery of the vectors $\mathcal{N}_{SSV} = 2kn$. However, the gain is minimal when $\text{rank}(\mathbf{X}) << k$ and hence $\text{rank}(\mathbf{Y}) << k$. Yet intuitively, when $\text{rank}(\mathbf{X})$ is small, there is a great deal of redundancy in the matrix $\mathbf{X}$, which should substantially reduce the number of measurements. We provide a way to achieve this reduction.

A. Proposed approach

We make the following assumption.

Assumption 1: In (3) $\mathbf{X}_1 \in \mathbb{R}^{m\times r}$, $\mathbf{X}_2 \in \mathbb{R}^{m\times(n-r)}$, $\mathbf{X}$ has at most k nonzero rows, $\text{rank}(\mathbf{X}) = r$ and $\text{spark}(\mathbf{X}) = r+1$.

Thus any r columns of $\mathbf{X}$ are linearly independent and span the column space of $\mathbf{X}$. We show that $\mathbf{A}_2$ can have far fewer rows than $\mathbf{A}_1$.

B. Recovery of $\mathbf{X}_1$

As $\text{rank}(\mathbf{X}_1) = r$, the method of [1] can be used to recover $\mathbf{X}_1$ if $\text{spark}(\mathbf{A}_1) > 2k - r + 1$; The original result in [1] relied on $\text{rank}(\mathbf{Y}_1) = r$, which was later shown to be equivalent to $\text{rank}(\mathbf{X}_1) = r$ [5].

As $s_1 \geq \text{spark}(\mathbf{A}_1)$ and the same matrix is used for r columns in $\mathbf{X}_1$, the total number of measurements required to estimate $\mathbf{X}_1$ is at least $s_1 r$, and equals that required by the scheme of [1]. The gains in our approach is in the recovery of $\mathbf{X}_2$ especially when $r << k$. Thus $s_{1,\min} = 2k - r + 2$ is the fewest measurements required to get the subspace.

C. Recovery of $\mathbf{X}_2$

As $\text{spark}(\mathbf{X}) = r+1$ and $\text{rank}(\mathbf{X}) = r$, every column of $\mathbf{X}$ is a linear combination of the r columns of $\mathbf{X}_1$. Thus there exist $\mathbf{Q} \in \mathbb{R}^{r\times(n-r)}$ such that $\mathbf{X}_2 = \mathbf{X}_1\mathbf{Q}$ i.e.

$$\mathbf{Y}_2 = \mathbf{A}_2\mathbf{X}_1\mathbf{Q}. \tag{6}$$

As $\mathbf{X}_1$ has been determined in step one, the recovery of $\mathbf{X}_2$ entails estimating $\mathbf{Q}$. The unique recoverability in particular is equivalent to $\mathbf{A}_2\mathbf{X}_1$ having full column rank. Recovery entails

$$\mathbf{Q} = \left(\mathbf{X}_1^\top\mathbf{A}_2^\top\mathbf{A}_2\mathbf{X}_1\right)^{-1}\mathbf{X}_1^\top\mathbf{A}_2^\top\mathbf{Y}_2. \tag{7}$$

We first provide a worst-case condition on $\mathbf{A}_2$ that ensures that $\text{rank}(\mathbf{A}_2\mathbf{X}_1) = r$ whenever $\mathbf{X}_1$ obeys Assumption 1.

Theorem 1: Under Assumption 1 consider $\mathbf{A}_2 \in \mathbb{R}^{s_2\times m}$. If $\text{spark}(\mathbf{A}_2) \geq k+1$ then $\text{rank}(\mathbf{A}_2\mathbf{X}_1) = r$. If $\text{spark}(\mathbf{A}_2) \leq k$ then there is an $\mathbf{X}_1$ obeying Assumption 1 for which $\text{rank}(\mathbf{A}_2\mathbf{X}_1) < r$.

Proof: Suppose $\text{spark}(\mathbf{A}_2) \geq k+1$ but $\text{rank}(\mathbf{A}_2\mathbf{X}_1) < r$. Then there exists a nonzero $\boldsymbol{\delta} \in \mathbb{R}^r$ such that

$$\mathbf{A}_2\mathbf{X}_1\boldsymbol{\delta} = 0. \tag{8}$$

As $\mathbf{X}_1$ has at most k nonzero rows so does $\mathbf{X}_1\boldsymbol{\delta}$. As $\mathbf{X}_1$ has full column rank, $\mathbf{X}_1\boldsymbol{\delta} \neq 0$. Thus (8) implies that $\mathbf{A}_2$ has k linearly dependent columns. Thus $\text{spark}(\mathbf{A}_2) \leq k$. The contradiction proves the first part of the theorem.

Suppose $\text{spark}(\mathbf{A}_2) \leq k$ i.e. $\mathbf{A}_2$ has at most k linearly dependent columns. Without loss of generality suppose these are the first k. Thus for some $\mathbf{A}_{21} \in \mathbb{R}^{s_2\times k}$ one has $\mathbf{A}_2 = [\mathbf{A}_{21}, \mathbf{A}_{22}]$ where for some $0 \neq \boldsymbol{\eta} \in \mathbb{R}^k$, $\mathbf{A}_{21}\boldsymbol{\eta} = 0$. As $k \geq r$ one can find $\mathbf{X}_{11} \in \mathbb{R}^{k\times r}$ with rank r that has η in its range space. In particular, $\mathbf{X}_{11}$ is such that there exists a $\boldsymbol{\delta} \neq 0$ such that $\mathbf{X}_{11}\boldsymbol{\delta} = \boldsymbol{\eta}$. Choose

$$\mathbf{X}_1 = \begin{bmatrix} \mathbf{X}_{11} \\ \mathbf{0}_{(m-k)\times r} \end{bmatrix}.$$

Clearly $\mathbf{X}_1$ obeys Assumption 1. As

$$\mathbf{A}_2\mathbf{X}_1\boldsymbol{\delta} = [\mathbf{A}_{21}, \mathbf{A}_{22}] \begin{bmatrix} \mathbf{X}_{11}\boldsymbol{\delta} \\ \mathbf{0}_{(m-k)\times r} \end{bmatrix} = \mathbf{A}_{21}\boldsymbol{\eta} = 0,$$

$\operatorname{rank}(\mathbf{A}_2\mathbf{X}_1) < r$. This proves the second part of the theorem. □

Under the above conditions, the recovery of $\mathbf{Q}$ through (7) is well defined. Combining the number of measurement needed at each step of the two-step algorithm, the total number of measurements that are sufficient for the recovery of all matrices is $\mathcal{N}_{\text{worst}} = r(2k - r + 2) + k(n - r)$. When the number of snapshots are much greater than the rank ($n >> r$), this setting provides approximately a factor of two savings over $\mathcal{N}_{SMV} = 2kn$, with $r << k$. Contrast this with the classical MMV scheme which provides a factor of two savings only when $r = k$. When the support is specified, $\mathbf{X}$ can be represented using $\mathcal{N}_{\text{freedom}} = r(k + n - r)$ parameters; comparing the worst-case result with the degrees of freedom, the gap between the two is given by $\mathcal{N}_{\text{worst}} - \mathcal{N}_{\text{freedom}} = n(k - r) + 2r$.

This worst case bound is pessimistic. The theorem below provides sufficient conditions for the recovery with almost all matrices.

Theorem 2: Under Assumption 1, the matrix $\mathbf{A}_2\mathbf{X}_1$ is nonsingular for almost all matrices $\mathbf{A}_2 \in \mathbb{R}^{r\times m}$.

Proof: Observe that $\det(\mathbf{A}_2\mathbf{X}_1)$ is a polynomial in the entries of $\mathbf{A}_2$ that is either the zero polynomial or takes nonzero values every where, except on a manifold of zero volume [12]. Thus it suffices to show that there is at least one $\mathbf{A}_2$, possibly complex, for which $\det(\mathbf{A}_2\mathbf{X}_1) \neq 0$.

Under Assumption 1, $\mathbf{X}_1 = \mathbf{W}_1\boldsymbol{\Lambda}\mathbf{W}_2^H$. Here $\mathbf{W}_1 \in \mathbb{C}^{m\times r}$ and $\mathbf{W}_2 \in \mathbb{C}^{r\times r}$ obey $\mathbf{W}_1^H\mathbf{W}_1 = \mathbf{I}$, $\mathbf{W}_2^H\mathbf{W}_2 = \mathbf{I}$ and $\boldsymbol{\Lambda} \in \mathbb{R}^{r\times r}$ is a nonsingular diagonal matrix. With $\mathbf{A}_2 = \mathbf{W}_1^H$, $\mathbf{A}_2\mathbf{X}_1 = \boldsymbol{\Lambda}\mathbf{W}_2^H$ is invertible. This proves the result. □

The above result shows that $\mathbf{A}_2\mathbf{X}_1$ has full column rank for almost all $\mathbf{A}_2 \in \mathbb{R}^{s_2\times m}$, for $s_2 \geq r$. Combining the required measurements at each step, we obtain $\mathcal{N}_{generic} = (2k - r + 2)r + r(n - r)$. For $r << k$ of course $\mathcal{N}_{\text{generic}} << \mathcal{N}_{worst}$. Thus, the gain offered by the proposed framework over classical settings is quite significant, especially since $r << k$ and the second step dominates. Unsurprisingly, the generic number of measurements per snapshot that are required to recover $\mathbf{X}$ approaches r as the number of snapshots $n \to \infty$. The difference between the degrees of freedom and the sufficient number of measurements for the recovery of almost all matrices is given by $\mathcal{N}_{generic} - \mathcal{N}_{\text{freedom}} = r(k - r) + 2r$. Since $r << k$, the gap is considerably smaller than the worst-case considered above.

III. Recovery Algorithm

We now describe the two-step sequential recovery algorithm to estimate $\mathbf{X}_1$ and $\mathbf{X}_2$. The analysis above for the perfect recovery of $\mathbf{X}_1$ and $\mathbf{X}_2$ assumes that measurements are not corrupted by noise. Moreover, since ℓ_0 recovery is computationally infeasible, surrogates such as greedy optimization or ℓ_1 minimization algorithms will be used. Hence, we expect the number of measurements required for practical and stable recovery algorithms to be significantly larger than the theoretically predicted values above. However, the same remarks also apply to classical MMV recovery, and our two-step approach still delivers commensurate improvements over recovery algorithms based on the classical MMV model.

One cannot use the algorithm specified by (2) to estimate $\mathbf{X}_1$ from $\mathbf{Y}_1$ as it involves a computationally infeasible combinatorial search. If $\mathbf{Y}_1$ were full rank, the solutions can be obtained using the MUSIC algorithm. As we assume $r << k$, this approach is infeasible. We use a greedy algorithm to determine $\mathbf{X}_1$. As classical greedy algorithms for joint sparse recovery fail to exploit the subspace structure of the problem, we use the rank aware order recursive matching pursuit (RA-ORMP) algorithm to improve the recovery [5].

We estimate $\mathbf{X}_2$ using a pseudo-inverse as $\mathbf{X}_2 = \mathbf{X}_1\mathbf{Q}^*$ where

$$\mathbf{Q}^* = \arg\min_{\mathbf{Q}} \|\mathbf{Y}_2 - \mathbf{A}_2\mathbf{X}_1\mathbf{Q}\|^2. \tag{9}$$

IV. Numerical simulations

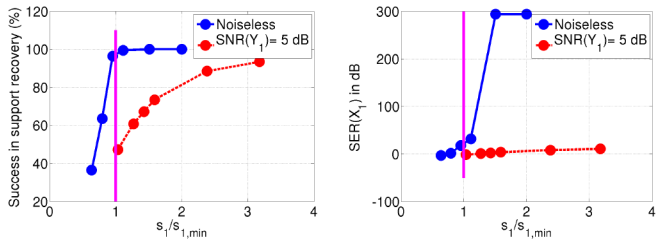
The main focus of this section is to demonstrate the ability of the two step recovery algorithm to considerably reduce the number of measurements using numerical simulations. We use the greedy RA-ORMP scheme [5] for the joint sparse recovery of $\mathbf{X}_1$. Hence, our simulations are not in full agreement with MMV guarantees that assume combinatorial optimization [1]–[6]. As with classical MMV schemes, for robustness we use more than the postulated minimum number of observations. Our goal is to show that our approach achieves the same accuracy as the algorithms using single measurement matrices, with far fewer measurements.

We consider several random realizations $\mathbf{X} \in \mathbb{R}^{m\times n}$, each of which are of rank r and are jointly k-sparse. We generate these random signal realizations as $\mathbf{X} = \mathbf{U}\mathbf{V}$ by setting k randomly selected rows of the matrix $\mathbf{U} \in \mathbb{R}^{m\times r}$, where entries of $\mathbf{U}$ and $\mathbf{V} \in \mathbb{R}^{r\times n}$ are zero mean Gaussian random variables with unit variance. All matrices satisfy the condition $\operatorname{spark}(\mathbf{X}) = r + 1$. We assume $m = 240$, $n = 130$, $r = 4$, $k = 33$. The columns of the above matrices are grouped as $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$, where $\mathbf{X}_1 \in \mathbb{R}^{m\times r}$ and $\mathbf{X}_2 \in \mathbb{R}^{m\times n-r}$. We consider the recovery of $\mathbf{X}$ from $\mathbf{Y}_1 = \mathbf{A}_1\mathbf{X}_1$ and $\mathbf{Y}_2 = \mathbf{A}_2\mathbf{X}_2$, where $\mathbf{A}_1 \in \mathbb{R}^{s_1\times m}$ and $\mathbf{A}_2 \in \mathbb{R}^{s_2\times m}$ are measurement matrices whose entries are Gaussian distributed zero-mean random variables with variance of one.

The performance of the joint sparse recovery of $\mathbf{X}_1$ using the RA-ORMP algorithm is studied in Fig. 1. The solid curve corresponds to recovery from noiseless measurements. The dashed curve is the noisy setting, where $\text{SNR}(\mathbf{Y}_1)$ is 5 dB. The percentage of the support indices that are correctly recovered and the signal to error ratio (SER) of the estimated $\mathbf{X}_1$ averaged over 1000 runs, are respectively in the plots on the left and right. SER & the signal to noise ratio (SNR) of the noisy measurements $\mathbf{Y}_1$ are defined as

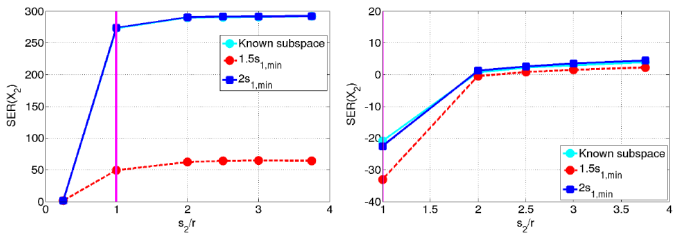
$$\text{SER}(\mathbf{X}_1) = -20\log\frac{\|\mathbf{X}_1 - \mathbf{X}_{1,\text{estimated}}\|_{\text{F}}}{\|\mathbf{X}_1\|_{\text{F}}} \tag{10}$$

$$\text{SNR}(\mathbf{Y}_1) = -20\log\frac{\|\mathbf{Y}_1 - \mathbf{Y}_{1,\text{noisy}}\|_{\text{F}}}{\|\mathbf{Y}_1\|_{\text{F}}}. \tag{11}$$



(a) (b)

Fig. 1. Performance of RA-ORMP algorithm in recovering $\mathbf{X}_1$ (1st step of the algorithm). The left subfigure shows the accuracy of support recovery as a function of normalized measurements $s_1/s_{1,\min}$. The SER($\mathbf{X}_1$) is shown on the right subfigure. The solid curve shows the noiseless case & dashed curve is for SNR($\mathbf{Y}_1$) $= 5$ db. (see (10,11)) (a) Support recovery (b) SER ($\mathbf{X}_1$).



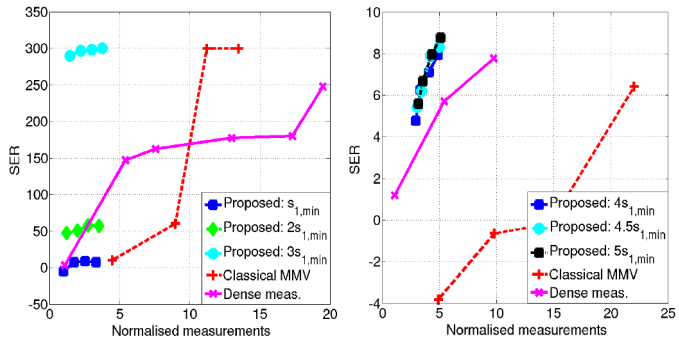
(a) (b)

Fig. 2. Support-aware recovery of $\mathbf{X}_2$ (2nd step of the algorithm) as a function of normalized measurements s_2/r. The SER of the estimate (in dB) as a function of the normalized measurements are shown for the noiseless (left) and noisy setting (right), respectively. (See main text) (a) Noiseless (b) SNR ($\mathbf{Y}_1$) & SNR ($\mathbf{Y}_2$) $= 5$ dB.

The vertical lines correspond to $s_{1,\min} = 2k - r + 2$ measurements, which is the minimum required for successful recovery in the absence of noise (see (2)). With $s_{1,\min}$ measurements the accuracy of the support recovery is close to 100% in the noiseless setting. More measurements are needed for support recovery with lower SER of 5 dB.

We study the performance of the subspace aware recovery of $\mathbf{X}_2$ in Fig. 2. The plot on the left is the noiseless setting, while the one on the right corresponds to the noisy case. The three curves in each of the plots correspond to recovery with different support estimates. The solid line marked with circles corresponds to the case when the support is accurately known, while the dashed curve corresponds to the support estimated using ORMP with $1.5s_{1,\min}$ measurements. Similarly, the square marked curve corresponds to the support estimate where $2s_{1,\min}$ measurements are used by ORMP. The vertical lines correspond to $s_2/r = 1$, which is the minimum predicted by Theorem 2 for subspace aware recovery. We observe that perfect recovery is obtained in the noiseless setting if the support is known perfectly (cyan curve) and when $s_1 = 2s_{1,\min}$. The lower SER of the estimated $\mathbf{X}_2$ when $s_1 = 1.5s_{1,\min}$ resulted from incorrect support estimates during $\mathbf{X}_1$ recovery. Specifically, a small fraction ($\approx 1\%$ of runs) resulted in one of the sparse locations being wrongly estimated. This error propagated to the subspace aware recovery of $\mathbf{X}_2$.

We compare the signal to error ratio of the reconstructions obtained by the proposed scheme against the classical MMV and the convex optimization scheme using dense measurements described in [10] setting in Fig. 3. The dense measurements were obtained by computing the inner-products with Gaussian



(a) (b)

Fig. 3. SER of the entire matrix ($\mathbf{X}$) vs. the number of normalized measurements (normalized to $\mathcal{N}_{generic}$). The SER of the estimates (in dB) are shown for the noiseless (left figure) and noisy setting (right figure), respectively. See text for more details (a) Noiseless (b) SNR ($\mathbf{Y}_1$) & SNR ($\mathbf{Y}_2$) $= 5$ dB.

random matrices. The matrix recovery is posed as the convex optimization scheme with σ^2 as noise variance:

$$\mathbf{X} = \arg\min_{\mathbf{x}} \|\mathbf{X}\|_{\ell_1-\ell_2} + \lambda\|\mathbf{X}\|_* \text{ s.t } \|\mathcal{A}_{\text{dense}}(\mathbf{X}) - \mathbf{y}\|_{\mathbf{F}}^2 \leq \sigma^2, \tag{12}$$

which was implemented using the CVX package. The λ parameter was optimized for each under-sampling factor to yield the best possible results. We normalize the total number of measurements by dividing it by $\mathcal{N}_{\text{generic}}$. We assumed $m = 45, n = 40, r = 2$, and $k = 6$ in these experiments, to keep the computational complexity of the convex optimization scheme tractable. We consider the noiseless setting as well as the noisy case, where SNR($\mathbf{Y}_1$) and SNR($\mathbf{Y}_2$) $= 5$ dB. The classical MMV recovery implemented using ORMP is indicated by the dashed curve marked by plus sign. The curve marked by crosses, corresponds to the recovery using dense measurements using convex optimization. The other curves correspond to the proposed scheme with different number of measurements for recovering $\mathbf{X}_1$, expressed as a factor of $s_{1,\min}$. We observe that the proposed scheme provides good recovery when the number of measurements in the first step equals $2.s_{1,min}$. By contrast, the classical MMV scheme requires more measurements to achieve the same signal to error ratio. We observe that the performance of the proposed scheme is slightly better than the dense measurement scheme, possibly because of the convex optimization prescribed by [10] for the latter. For the dense measurement scheme, a non-convex optimization method may have provided better recovery, as indicated by the theoretical results in [11] but is computationally infeasible.

V. Conclusions and Discussion

A two step scheme is given to recover a jointly k-sparse, low rank matrix $\mathbf{X}$ of rank $r << k$ and spark $r + 1$ from its under sampled measurements, using a separate measurement matrix for each step. The first step recovers r columns using classical MMV schemes providing a basis of $\mathbf{X}$. The basis is used to recover the remaining columns in the second step. Conditions on the second measurement matrix are given for both worst case and generic settings. The use of two, as opposed to one, measurement matrices considerably reduces the number of samples required to recover a jointly sparse, low-rank matrix. Our scheme compares favorably with schemes requiring dense sampling, even though they are inapplicable to a number of settings such as MRI.

REFERENCES

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

[2] J. Kim, O. K. Lee, and J. C. Ye, "Compressive music: A missing link between compressive sensing and array signal processing," *CoRR*, vol. abs/1004.4398, 2010.

[3] J. M. Kim, O. K. Lee, and J. C. Ye, "Improving noise robustness in subspace-based joint sparse recovery," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5799–5809, 2012.

[4] K. Lee, Y. Bresler, and M. Junge, "Subspace methods for joint sparse recovery," *IEEE Trans. Information Theory*, vol. 58, no. 6, pp. 3613–3641, 2012.

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

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

[7] P. K. Y. C. Eldar and H. Bolcskei, "Compressed sensing of block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, 2010.

[8] S. Lingala, Y. Hu, E. Di Bella, and M. Jacob, "Accelerated dynamic mri exploiting sparsity and low-rank structure: Kt slr," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1042–1054, May 2011.

[9] C. Brinegar, Y.-J. L. Wu, L. M. Foley, T. K. Hitchens, Q. Ye, C. Ho, and Z.-P. Liang, "Real-time cardiac mri without triggering, gating, or breath holding," in *EMBC*, 2008.

[10] M. Golbabaee and P. Vandergheynst, "Compressed sensing of simultaneous low-rank and joint-sparse matrices," *arXiv preprint arXiv:1211.5058*, 2012.

[11] S. Oymak, A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi, "Simultaneously structured models with application to sparse and low-rank matrices," *arXiv preprint, arXiv:1212.3753*, 2012.

[12] L. Losaz, "Singular spaces of matrices and their application in combinatorics," *Bull. Brazilian Math. Soc.*, vol. 20, no. 1, pp. 87–89, 1989.

[13] A. M. Tulino and S. Verdu, *Random Matrix Theory and Wireless Communications*. Boston, MA, USA: Now, 2004.

This paper previously published in IEEE Signal Processing Letters