

# BLOCK SPARSITY AND SAMPLING OVER A UNION OF SUBSPACES

Yonina C. Eldar and Moshe Mishali

Technion–Israel Institute of Technology, Haifa, Israel  
Email: {yonina@ee,moshiko@techunix}.technion.ac.il

## ABSTRACT

Sparse signal representations have gained wide popularity in recent years. In many applications the data can be expressed using only a few nonzero elements in an appropriate expansion. In this paper, we study a block-sparse model, in which the nonzero coefficients are arranged in blocks. To exploit this structure, we redefine the standard (NP-hard) sparse recovery problem, based on which we propose a convex relaxation in the form of a mixed  $\ell_2/\ell_1$  program. Isometry-based analysis is used to prove equivalence of the solution to that of the optimal program, under certain mild conditions. We further establish the robustness of our algorithm to mismodeling and bounded noise. We then present theoretical arguments and numerical experiments demonstrating the improved recovery performance of our method in comparison with sparse reconstruction that does not incorporate a block structure. The results are then applied to two related problems. The first is that of simultaneous sparse approximation. Our results can be used to prove isometry-based equivalence properties for this setting. In addition, we propose an alternative approach to acquire the measurements, that leads to performance improvement over standard methods. Finally, we show how our results can be used to sample signals in a finite structured union of subspaces, leading to robust and efficient recovery algorithms.

**Index Terms**— Block sparsity, compressed sensing, multiple measurement vectors (MMV), restricted isometry property, sparse approximation, union of subspaces.

## 1. INTRODUCTION

Recovery of sparse vectors from few measurements is of interest in many different applications. This problem underlies the recent framework of compressed sensing (CS) [1, 2], which treats recovery of a sparse vector  $\mathbf{x}$  from measurements  $\mathbf{y} = \mathbf{D}\mathbf{x}$ , where  $\mathbf{D}$  is a measurement matrix with fewer rows than columns. Many algorithms have been proposed to determine  $\mathbf{x}$  in a stable and efficient manner [2–5]. One of the main analysis tools in this context is the restricted isometry property (RIP) [2, 6]. In particular, if the measurement matrix  $\mathbf{D}$  satisfies the RIP then  $\mathbf{x}$  can be recovered by solving a convex  $\ell_1$  minimization algorithm.

The standard sparsity model assumes that  $\mathbf{x}$  has at most  $k$  nonzero elements, however it does not impose any further structure. In particular, the non-zero components can appear anywhere in  $\mathbf{x}$ . There are many cases in which the non-zero values are aligned to blocks, meaning they appear in regions. Examples include bursty noise profiles, and measurements of gene expression levels [7]. We

refer to such a model as block sparsity. Clearly any block-sparse vector is also sparse in the standard sense. However, by exploiting the block structure, recovery may be possible under more general conditions. In [7, 8] a mixed  $\ell_2/\ell_1$  algorithm was proposed for block-sparse recovery. By analyzing the null space of  $\mathbf{D}$ , it was shown that asymptotically in the signal length, perfect recovery is possible with high probability. However, this result cannot be applied in the finite, non-asymptotic case that is of interest here. Furthermore, robustness and mismodeling issues were not addressed.

In this paper we analyze the recovery ability of the convex  $\ell_2/\ell_1$  program for finite dimensions. We also study robustness of the method in the presence of model uncertainties and noise. Our results rely on a new notion of block RIP, which is less stringent than standard RIP. We prove that if  $\mathbf{D}$  satisfies block RIP, then  $\mathbf{x}$  can be recovered exactly using the mixed  $\ell_2/\ell_1$  program. The proposed algorithm is also shown to be stable in the presence of noise and mismodeling errors under block RIP. We then prove that random matrices satisfy the block RIP with overwhelming probability. Moreover, the probability to satisfy the block RIP is substantially larger than that of satisfying standard RIP. These results establish that a block-sparse vector  $\mathbf{x}$  can be recovered efficiently and stably with overwhelming probability using a random measurement matrix. Furthermore, the recovery performance is higher when exploiting block sparsity. Complementary results based on mutual coherence appear in [9], where equivalence is also established for an extension of orthogonal matching pursuit to the block-sparse case.

An interesting special case of the block-sparse model is the multiple measurement vector (MMV) problem, in which we measure a set of vectors that share a joint sparsity pattern. MMV recovery algorithms were studied in [3, 4, 10–12]. Equivalence results based on mutual coherence for a mixed  $\ell_p/\ell_1$  program were derived in [11], which turn out to be the same as that obtained from a single measurement problem. This is in contrast to the fact that in practice, MMV methods tend to outperform algorithms that treat each of the vectors separately. In order to develop meaningful equivalence results, we cast the MMV problem as one of block-sparse recovery. Our mixed  $\ell_2/\ell_1$  method translates into minimizing the sum of the  $\ell_2$  row-norms of the matrix representing the MMV set. Our general results lead to RIP-based equivalence conditions for this algorithm. Furthermore, our framework suggests a different sampling method for MMV problems which tends to increase the recovery rate. The equivalence condition we obtain in this case is stronger than the single measurement setting. As we show, this method leads to superior recovery rate in comparison to alternative MMV algorithms.

Finally, we show how the block sparsity model can be used in order to treat the problem of sampling signals over a union of subspaces. Traditional sampling theories consider the problem of reconstructing an unknown signal  $x$  in an arbitrary Hilbert space, from a series of samples. A prevalent assumption which often guaran-

---

This work was supported in part by the Israel Science Foundation under Grant no. 1081/07 and by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMMunications NEWCOM++ (contract no. 216715).

tees a unique signal consistent with the given measurements is that  $x$  lies in a known subspace. Recently, there has been growing interest in nonlinear but structured signal models, in which  $x$  is assumed to lie in a union of subspaces [12–18]. In order to ensure stable and efficient recovery from the samples, some underlying structure is needed. Here we focus on the case in which  $x$  lies in a finite union of finite-dimensional spaces and the samples are modeled as inner products with an arbitrary set of sampling functions. Specifically, we consider the case in which  $x$  resides in a sum of  $k$  subspaces, chosen from a given set of  $m$  subspaces  $\mathcal{A}_j, 1 \leq j \leq m$ . However, which subspaces comprise the sum is unknown. This setting is a special case of the more general union model considered in [15, 16]. Conditions under which unique and stable sampling are possible were developed in [15, 16]. However, no concrete algorithm was provided to recover such a signal from a given set of samples in a stable and efficient manner. Here we show that this problem can be formulated as one of recovering a block-sparse vector. Relying on our previous results leads to an efficient convex algorithm that often recovers the original signal from the given samples. In addition, this method is stable and robust in the sense that the reconstruction error is bounded when  $x$  does not lie in the union, and in the presence of noise.

The remaining of the paper is organized as follows. The block-sparse model is described in Section 2. In this section, we also introduce the block RIP and present an algorithm with combinatorial complexity whose solution is the true unknown  $\mathbf{x}$ . A convex relaxation of this algorithm is proposed in Section 3. We then derive equivalence conditions based on the notion of block RIP. We also prove that our method is robust and stable in the presence of noise and modeling errors. This approach is specialized to MMV sampling in Section 4. In Section 5 we prove that random ensembles tend to satisfy the block RIP with high probability. Finally, we apply the results to sampling over a union of subspaces in Section 6.

## 2. BLOCK SPARSITY AND BLOCK RIP

### 2.1. Block Sparsity

We consider the problem of recovering a vector  $\mathbf{x} \in \mathbb{R}^N$  from measurement  $\mathbf{y} \in \mathbb{R}^L$  given by

$$\mathbf{y} = \mathbf{D}\mathbf{x} \quad (1)$$

where  $\mathbf{D}$  is a measurement matrix of size  $L \times N$  with  $L < N$ . We treat  $\mathbf{x}$  as consisting of blocks with given lengths  $d_\ell, 1 \leq \ell \leq M$ . Denoting by  $\mathbf{x}[\ell]$  the  $\ell$ th sub-block of length  $d_\ell$  we can write  $\mathbf{x}$  as

$$\mathbf{x}^T = \underbrace{[x_1 \dots x_{d_1}]}_{\mathbf{x}[1]} \dots \underbrace{[x_{N-d_M+1} \dots x_N]}_{\mathbf{x}[M]}. \quad (2)$$

A vector  $\mathbf{x} \in \mathbb{R}^N$  is block  $k$ -sparse over  $\mathcal{I} = \{d_1, \dots, d_M\}$  with  $N = \sum_\ell d_\ell$  if  $\mathbf{x}[\ell]$  has nonzero norm for at most  $k$  indices  $\ell$ . When  $d_\ell = 1$ , block-sparsity reduces to conventional sparsity. Denoting

$$\|\mathbf{x}\|_{0,\mathcal{I}} = \sum_{\ell=1}^M I(\|\mathbf{x}[\ell]\|_2 > 0), \quad (3)$$

where  $I(\|\mathbf{x}[\ell]\|_2 > 0)$  is an indicator function that obtains the value 1 if  $\|\mathbf{x}[\ell]\|_2 > 0$  and 0 otherwise, a block  $k$ -sparse vector  $\mathbf{x}$  can be defined by  $\|\mathbf{x}\|_{0,\mathcal{I}} \leq k$ . Here  $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T \mathbf{x}}$  is the standard Euclidean norm. An example of a block-sparse vector with  $k = 2$  is depicted in Fig. 1.

Our goal is to provide conditions on  $\mathbf{D}$  ensuring that the block-sparse vector  $\mathbf{x}$  can be reconstructed from measurements of the form



**Fig. 1.** A block-sparse vector  $\mathbf{x}$  over  $\mathcal{I} = \{d_1 = 3, d_2 = 4, d_3 = 2, d_4 = 6, d_5 = 1\}$ . The gray areas represent 10 non-zero entries which occupy two blocks.

(1) through computationally efficient algorithms. We begin by stating a condition under which there is a unique block-sparse  $\mathbf{x}$  consistent with the measurements  $\mathbf{y}$  [19].

**Proposition 1** *The representation (1) is unique if and only if  $\mathbf{D}\mathbf{c} \neq \mathbf{0}$  for every  $\mathbf{c} \neq \mathbf{0}$  that is block  $2k$ -sparse.*

Similarly to (2), we can represent  $\mathbf{D}$  as a concatenation of column-blocks  $\mathbf{D}[\ell]$  of size  $L \times d_\ell$ :

$$\mathbf{D} = \underbrace{[\mathbf{d}_1 \dots \mathbf{d}_{d_1}]}_{\mathbf{D}[1]} \dots \underbrace{[\mathbf{d}_{N-d_M+1} \dots \mathbf{d}_N]}_{\mathbf{D}[M]}. \quad (4)$$

Using these sub-blocks we can write  $\mathbf{y} = \mathbf{D}\mathbf{x}$  as

$$\mathbf{y} = \sum_{\ell=1}^M \mathbf{D}[\ell]\mathbf{x}[\ell]. \quad (5)$$

From Proposition 1 it follows that the columns of  $\mathbf{D}[\ell]$  are linearly independent for all  $\ell$ .

### 2.2. Block RIP

Invertibility and stability are intimately related to RIP, which we generalize here to the block-sparse setting. A matrix  $\mathbf{D}$  of size  $L \times N$  is said to have the RIP [2, 6] if there exists a constant  $\delta_k \in [0, 1)$  such that for every  $k$ -sparse  $\mathbf{x} \in \mathbb{R}^N$ ,

$$(1 - \delta_k)\|\mathbf{x}\|_2^2 \leq \|\mathbf{D}\mathbf{x}\|_2^2 \leq (1 + \delta_k)\|\mathbf{x}\|_2^2. \quad (6)$$

Extending this property to block-sparse vectors leads to the following definition:

**Definition 1** *Let  $\mathbf{D} : \mathbb{R}^N \rightarrow \mathbb{R}^L$  be a given matrix. Then  $\mathbf{D}$  has the block RIP over  $\mathcal{I} = \{d_1, \dots, d_M\}$  with parameter  $\delta_{k|\mathcal{I}}$  if for every  $\mathbf{x} \in \mathbb{R}^N$  that is block  $k$ -sparse over  $\mathcal{I}$  we have that*

$$(1 - \delta_{k|\mathcal{I}})\|\mathbf{x}\|_2^2 \leq \|\mathbf{D}\mathbf{x}\|_2^2 \leq (1 + \delta_{k|\mathcal{I}})\|\mathbf{x}\|_2^2. \quad (7)$$

By abuse of notation, we use  $\delta_k$  for the block-RIP constant  $\delta_{k|\mathcal{I}}$  when it is clear from the context that we refer to blocks. Block-RIP is a special case of the  $\mathcal{A}$ -restricted isometry defined in [16].

Note that a block  $k$ -sparse vector over  $\mathcal{I}$  is  $m$ -sparse in the conventional sense where  $m$  is the sum of the  $k$  largest values in  $\mathcal{I}$ , since it has at most  $m$  nonzero elements. If we require  $\mathbf{D}$  to satisfy RIP for all  $m$ -sparse vectors, then (7) must hold for all  $2m$ -sparse vectors  $\mathbf{x}$ . Since we only require RIP for block sparse signals, (7) has to be satisfied for a certain subset of  $2m$ -sparse signals, namely those that have block sparsity. As a result, the block-RIP constant  $\delta_{k|\mathcal{I}}$  is typically smaller than  $\delta_m$ .

From our previous discussion it follows that if  $\mathbf{D}$  satisfies the block RIP (7) with  $\delta_{2k|\mathcal{I}} < 1$ , then there is a unique value of  $\mathbf{x}$  consistent with (1). The question is how to find  $\mathbf{x}$  in practice. Below we present an algorithm that will in principle find the unique  $\mathbf{x}$  from the samples  $\mathbf{y}$ . Unfortunately, though, it has exponential complexity.

In the next section we show that under a stronger condition on  $\delta_{2k}$  we can recover  $\mathbf{x}$  in a stable and efficient manner.

Our first claim is that  $\mathbf{x}$  can be uniquely recovered by solving

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{x}\|_{0,\mathcal{I}} \\ \text{s. t.} \quad & \mathbf{y} = \mathbf{D}\mathbf{x}. \end{aligned} \quad (8)$$

To show that (8) will indeed recover the true value of  $\mathbf{x}$ , suppose that there exists a  $\mathbf{x}'$  such that  $\mathbf{D}\mathbf{x}' = \mathbf{y}$  and  $\|\mathbf{x}'\|_{0,\mathcal{I}} \leq \|\mathbf{x}\|_{0,\mathcal{I}} \leq k$ . Since both  $\mathbf{x}, \mathbf{x}'$  are consistent with the measurements,

$$\mathbf{0} = \mathbf{D}(\mathbf{x} - \mathbf{x}') = \mathbf{D}\mathbf{c}, \quad (9)$$

where  $\|\mathbf{c}\|_{0,\mathcal{I}} \leq 2k$  so that  $\mathbf{c}$  is a block  $2k$ -sparse vector. Since  $\mathbf{D}$  satisfies (7) with  $\delta_{2k} < 1$ , we must have that  $\mathbf{d} = \mathbf{0}$  or  $\mathbf{x} = \mathbf{x}'$ .

### 3. CONVEX RECOVERY ALGORITHM

#### 3.1. Noise-Free Recovery

We now develop a convex optimization problem instead of (8) to approximate  $\mathbf{x}$ . Our approach is to minimize the sum of the energy in the blocks  $\mathbf{x}[\ell]$ . To write down the problem explicitly, define the mixed  $\ell_2/\ell_1$  norm over the index set  $\mathcal{I} = \{d_1, \dots, d_M\}$  as

$$\|\mathbf{x}\|_{2,\mathcal{I}} = \sum_{\ell=1}^M \|\mathbf{x}[\ell]\|_2. \quad (10)$$

Our proposed reconstruction algorithm is

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{x}\|_{2,\mathcal{I}} \\ \text{s. t.} \quad & \mathbf{y} = \mathbf{D}\mathbf{x}. \end{aligned} \quad (11)$$

Problem (11) is a second order cone program (SOCP) and can therefore be solved efficiently using standard software packages [19]. The next theorem establishes that the solution to (11) is the true  $\mathbf{x}$  as long as  $\delta_{2k}$  is small enough. The proof can be found in [19].

**Theorem 1** *Let  $\mathbf{y} = \mathbf{D}\mathbf{x}_0$  be measurements of a block  $k$ -sparse vector  $\mathbf{x}_0$ . If  $\mathbf{D}$  satisfies the block RIP (7) with  $\delta_{2k} < \sqrt{2} - 1$  then*

1. *there is a unique block  $k$ -sparse vector  $\mathbf{x}$  consistent with  $\mathbf{y}$ ;*
2. *the SOCP (11) has a unique solution;*
3. *the solution to the SOCP is equal to  $\mathbf{x}_0$ .*

Theorem 1 provides a gain over standard CS. Specifically, it is shown in [6] that if  $\mathbf{x}$  is  $k$ -sparse and the measurement matrix  $\mathbf{D}$  satisfies the standard RIP with  $\delta_{2k} < \sqrt{2} - 1$ , then  $\mathbf{x}$  can be recovered exactly from  $\mathbf{y} = \mathbf{D}\mathbf{x}$  via the linear program:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{x}\|_1 \\ \text{s. t.} \quad & \mathbf{y} = \mathbf{D}\mathbf{x}, \end{aligned} \quad (12)$$

where  $\|\mathbf{x}\|_1 = \sum_{\ell} |\mathbf{x}(\ell)|$  is the  $\ell_1$  norm, and  $\mathbf{x}(\ell)$  denotes the  $\ell$ th element of  $\mathbf{x}$ . Since any block  $k$ -sparse vector is also  $m$ -sparse with  $m$  equal to the sum of the  $k$  largest values of  $d_{\ell}$ , we can find  $\mathbf{x}_0$  of Theorem 1 by solving (12) if  $\delta_{2m}$  is small enough. However, this approach does not exploit the fact that the non-zero values appear in blocks. On the other hand, (11) explicitly takes the block structure of  $\mathbf{x}_0$  into account. Therefore, the condition of Theorem 1 is not as stringent as that obtained by using equivalence results with respect to (12). Indeed, the block RIP (7) bounds the norm of  $\|\mathbf{D}\mathbf{x}\|$  over block-sparse vectors  $\mathbf{x}$ , while standard RIP considers all choices of

$\mathbf{x}$ , also those that are not block  $2k$ -sparse. Therefore, the value of  $\delta_{2k|\mathcal{I}}$  in (7) can be lower than  $\delta_k$ , obtained from (6) with  $k = 2m$ .

To emphasize the advantage of block RIP, consider the following matrix, separated into three blocks of two columns each:

$$\mathbf{D} = \left( \begin{array}{cc|cc|cc} -1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 2 & -1 & 0 & 0 & 3 \\ 0 & 3 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 & 1 \end{array} \right) \cdot \mathbf{B}, \quad (13)$$

where  $\mathbf{B}$  is a diagonal matrix that results in unit-norm columns of  $\mathbf{D}$ , i.e.,  $\mathbf{B} = \text{diag}(1, 15, 1, 1, 1, 12)^{-1/2}$ . In this example  $M = 3$  and  $\mathcal{I} = \{d_1 = 2, d_2 = 2, d_3 = 2\}$ . Suppose that  $\mathbf{x}$  is block 1-sparse, which corresponds to at most two non-zero values. Brute-force calculations show that the smallest value of  $\delta_2$  satisfying the standard RIP (6) is  $\delta_2 = 0.866$ . On the other hand, the block-RIP (7) corresponding to the case in which the two-non zero elements are restricted to one block is satisfied with  $\delta_{1|\mathcal{I}} = 0.289$ . Increasing the number of non-zero elements to  $k = 4$ , the standard RIP (6) does not hold for any  $\delta_4 \in [0, 1)$ . This implies that there can be two 4-sparse vectors that result in the same measurements  $\mathbf{y} = \mathbf{D}\mathbf{x}$ . In contrast,  $\delta_{2|\mathcal{I}} = 0.966$  satisfies the lower bound in (7), which allows 4 non-zero values grouped into two blocks. Consequently,  $\mathbf{y}$  uniquely specifies  $\mathbf{x}$  when the non-zeros are restricted to blocks.

#### 3.2. Robust Recovery

Using ideas similar to [6] we can show that even when  $\mathbf{x}$  is not exactly block  $k$ -sparse, and there is noise in the measurements, the recovery algorithm (11) can approximate the best block  $k$ -sparse vector. Specifically, given  $\mathbf{x} \in \mathbb{R}^N$ , we denote by  $\mathbf{x}^k$  the best block  $k$ -sparse approximation of  $\mathbf{x}$  over  $\mathcal{I}$ . Thus,

$$\mathbf{x}^k = \arg \min_{\|\mathbf{d}\|_{0,\mathcal{I}} \leq k} \|\mathbf{x} - \mathbf{d}\|_{2,\mathcal{I}}, \quad (14)$$

over all block  $k$ -sparse vectors  $\mathbf{d}$ . The measurements (1) are corrupted by bounded noise so that

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{z}, \quad (15)$$

where  $\|\mathbf{z}\|_2 \leq \epsilon$ . In order to recover  $\mathbf{x}$  we use the modified SOCP:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{x}\|_{2,\mathcal{I}} \\ \text{s. t.} \quad & \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \leq \epsilon. \end{aligned} \quad (16)$$

Theorem 2, proven in [19], shows that even when  $\mathbf{x}$  is not block  $k$ -sparse and the measurements are noisy, the best  $k$  approximation can be well approximated using (16).

**Theorem 2** *Let  $\mathbf{y} = \mathbf{D}\mathbf{x}_0 + \mathbf{z}$  be noisy measurements of a vector  $\mathbf{x}_0$ . Let  $\mathbf{x}^k$  denote the best block  $k$ -sparse approximation of  $\mathbf{x}_0$ , and let  $\mathbf{x}'$  be a solution to (16). If  $\mathbf{D}$  satisfies the block RIP (7) with  $\delta_{2k} < \sqrt{2} - 1$  then*

$$\begin{aligned} \|\mathbf{x}_0 - \mathbf{x}'\|_2 \leq & \frac{2(1 - \delta_{2k})}{1 - (1 + \sqrt{2})\delta_{2k}} k^{-1/2} \|\mathbf{x}_0 - \mathbf{x}^k\|_{2,\mathcal{I}} \\ & + \frac{4\sqrt{1 + \delta_{2k}}}{1 - (1 + \sqrt{2})\delta_{2k}} \epsilon. \end{aligned} \quad (17)$$

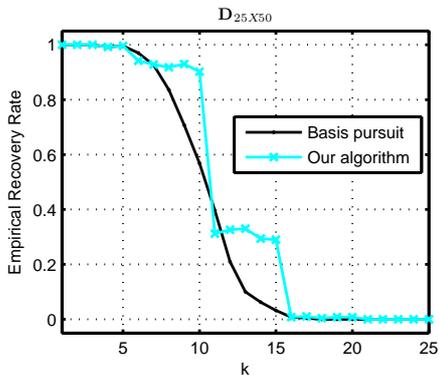
Note that the first term in (17) is a result of the fact that  $\mathbf{x}_0$  is not exactly block  $k$ -sparse. The second expression quantifies the recovery error due to the noise.

In summary, as long as  $\mathbf{D}$  satisfies block-RIP (7) with a suitable constant, any block  $k$ -sparse vector  $\mathbf{x}$  can be perfectly recovered from its samples  $\mathbf{y} = \mathbf{D}\mathbf{x}$  using (11). This algorithm is stable in the sense that by slightly modifying it as in (16) it can tolerate noise in a way that ensures that the norm of the recovery error is bounded by the noise level. Furthermore, if  $\mathbf{x}$  is not exactly block  $k$ -sparse, then its best block  $k$ -sparse approximation can be approached.

### 3.3. Advantage of Block Sparsity

To demonstrate the advantage of our algorithm over (12), consider the matrix  $\mathbf{D}$  of (13). In Section 3 it was shown that the block RIP constants are smaller than the standard block RIP constants for this choice of  $\mathbf{D}$ . This suggests that there are vectors  $\mathbf{x}$  for which (11) will be able to recover them exactly from measurements  $\mathbf{y} = \mathbf{D}\mathbf{x}$  while standard  $\ell_1$  minimization will fail. To illustrate this behavior, let  $\mathbf{x} = [0, 0, 1, -1, -1, 0.1]^T$  be a 4-sparse vector, in which the non-zero elements are known to appear in blocks of length 2. The prior knowledge that  $\mathbf{x}$  is 4-sparse is not sufficient to determine  $\mathbf{x}$  from  $\mathbf{y}$ . In contrast, there is a unique block-sparse vector consistent with  $\mathbf{y}$ . Furthermore, our algorithm which is a relaxed version of (8), finds the correct  $\mathbf{x}$  while standard  $\ell_1$  minimization fails in this case.

We further compare the recovery performance of  $\ell_1$  minimization (12) and our algorithm (11) for an extensive set of random signals. In the experiment, we draw a matrix  $\mathbf{D}$  of size  $25 \times 50$  from the Gaussian ensemble. The input vector  $\mathbf{x}$  is also randomly generated as a block-sparse vector with blocks of length 5. We draw  $1 \leq k \leq 25$  non-zero entries from a zero-mean unit variance normal distribution and divide them into blocks which are chosen uniformly at random within  $\mathbf{x}$ . Each of the algorithms is executed based on the measurements  $\mathbf{y} = \mathbf{D}\mathbf{x}$ . In Fig. 2 we plot the fraction of successful reconstructions for each  $k$  over 500 experiments. The results illustrate the advantage of incorporating the block-sparsity structure into the optimization program. An interesting feature of the graph is that when using the block-sparse recovery approach, the performance is roughly constant over the block-length (5 in this example). This explains the performance advantage over standard sparse recovery.



**Fig. 2.** Recovery rate of block-sparse signals using standard  $\ell_1$  minimization (basis pursuit) and the mixed  $\ell_2/\ell_1$  algorithm.

## 4. APPLICATION TO MMV MODELS

We now specialize our algorithm and equivalence results to the MMV problem. This leads to two contributions which we discuss in this section: The first is an equivalence result based on RIP for

a mixed-norm MMV algorithm. The second is a new measurement strategy in MMV problems that leads to improved performance over conventional MMV methods, both in simulations and as measured by the RIP-based equivalence condition. In contrast to previous equivalence results, for this strategy we show that even in the worst case, improved performance over the single measurement setting can be guaranteed.

### 4.1. Equivalence Results

In an MMV setting we are given a matrix of measurements  $\mathbf{Y}$  that is obtained by measuring a set of  $k$ -sparse vectors  $\mathbf{x}_\ell$  which are the columns of a matrix  $\mathbf{X}$ . The distinguishing feature of the MMV setting is that the non-zero elements of  $\mathbf{x}_\ell$  are assumed to share a joint sparsity pattern. Thus,

$$\mathbf{Y} = \mathbf{M}\mathbf{X}, \quad (18)$$

where  $\mathbf{X}$  has at most  $k$  non-zero rows, and  $\mathbf{M}$  is a given sampling matrix. This problem can be transformed into that of recovering a block  $k$ -sparse signal by noting that if we define  $\mathbf{x} = \text{vec}(\mathbf{X}^T)$  as the vector obtained by stacking the rows of  $\mathbf{X}$ , then  $\mathbf{x}$  is block  $k$ -sparse where each block has length  $d$ . From (18) we have that

$$\text{vec}(\mathbf{Y}^T) = (\mathbf{M} \otimes \mathbf{I}) \text{vec}(\mathbf{X}^T), \quad (19)$$

where  $\mathbf{A} \otimes \mathbf{B}$  is the Kronecker product between matrices  $\mathbf{A}$  and  $\mathbf{B}$ . Denoting by  $\mathbf{y} = \text{vec}(\mathbf{Y}^T)$  and  $\mathbf{D} = \mathbf{M} \otimes \mathbf{I}$ , the measurements can be expressed as  $\mathbf{y} = \mathbf{D}\mathbf{x}$  where  $\mathbf{x}$  is a block-sparse vector with blocks of length  $d$ . Therefore, the uniqueness conditions and the results of Theorems 1 and 2 can be specified to this problem.

Recovery algorithms for MMV using convex optimization programs were studied in [4, 11] and several greedy algorithms were proposed in [3, 10]. Specifically, in [3, 4, 10, 11] the authors study a class of optimization programs, which we refer to as M-BP:

$$\text{M-BP}(\ell_q): \quad \min \sum_{i=1}^L \|\mathbf{X}^i\|_q^p \quad \text{s. t. } \mathbf{Y} = \mathbf{M}\mathbf{X}, \quad (20)$$

where  $\mathbf{X}^i$  is the  $i$ th row of  $\mathbf{X}$ . The choice  $p = 1, q = \infty$  was considered in [4], while [11] treated the case of  $p = 1$  and arbitrary  $q$ . Using  $p \leq 1$  and  $q = 2$  was suggested in [10], leading to the iterative algorithm M-FOCUSS. For  $p = 1, q = 2$ , the program (20) has a global minimum which M-FOCUSS is proven to find. A nice comparison between these methods can be found in [4]. Equivalence for MMV algorithms based on RIP analysis does not appear in previous papers. The most detailed theoretical analysis can be found in [11] which establishes equivalence results based on mutual coherence. The results imply equivalence for (20) with  $p = 1$  under conditions equal to those obtained for the single measurement case. Note that RIP analysis typically leads to tighter equivalence bounds than mutual coherence analysis.

Since we can cast the MMV problem as one of block-sparse recovery, we may apply our equivalence results of Theorem 1 to this setting leading to RIP-based equivalence. To this end we first note that applying the SOCP (11) to the effective measurement vector  $\mathbf{y}$  is the same as solving (20) with  $p = 1, q = 2$ . Thus the equivalence conditions we develop below relate to this program. Next, if  $\mathbf{z} = \mathbf{D}\mathbf{x}$  where  $\mathbf{x}$  is a block  $2k$ -sparse vector and  $\mathbf{D} = \mathbf{M} \otimes \mathbf{I}_d$ , then taking the structure of  $\mathbf{D}$  into account,  $\mathbf{Z} = \mathbf{M}\mathbf{X}$  where  $\mathbf{X}$  is a size  $L \times d$  matrix whose  $i$ th row is equal to  $\mathbf{x}[i]$ , and similarly for  $\mathbf{Z}$ . The block sparsity of  $\mathbf{x}$  implies that  $\mathbf{X}$  has at most  $2k$  non-zero rows.

The squared  $\ell_2$  norm  $\|\mathbf{z}\|_2^2$  is equal to the squared  $\ell_2$  norm of the rows of  $\mathbf{Z}$  which can be written as

$$\|\mathbf{z}\|_2^2 = \|\mathbf{Z}\|_F^2 = \text{Tr}(\mathbf{Z}^T \mathbf{Z}), \quad (21)$$

where  $\|\mathbf{Z}\|_F$  denotes the Frobenius norm. Since  $\|\mathbf{x}\|_2^2 = \|\mathbf{X}\|_F^2$  the RIP condition becomes

$$(1 - \delta_{2k}) \text{Tr}(\mathbf{X}^T \mathbf{X}) \leq \text{Tr}(\mathbf{X}^T \mathbf{M}^T \mathbf{M} \mathbf{X}) \leq (1 + \delta_{2k}) \text{Tr}(\mathbf{X}^T \mathbf{X}), \quad (22)$$

for any  $L \times d$  matrix  $\mathbf{X}$  with at most  $2k$  non-zero rows.

It is straightforward to show (see [19]) that (22) is equivalent to the standard RIP condition

$$(1 - \delta_{2k}) \|\mathbf{x}\|_2^2 \leq \|\mathbf{M}\mathbf{x}\|_2^2 \leq (1 + \delta_{2k}) \|\mathbf{x}\|_2^2, \quad (23)$$

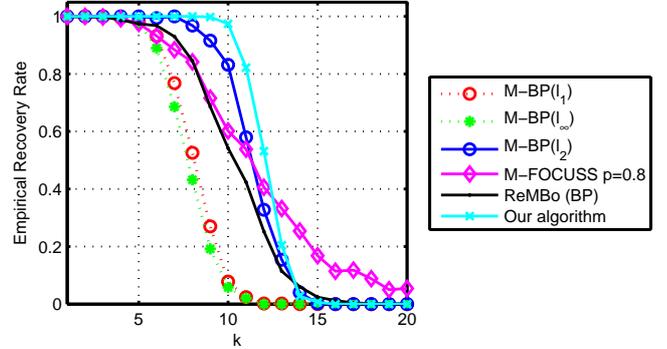
for any length  $L$  vector  $\mathbf{x}$  that is  $2k$ -sparse. Therefore, we conclude that if  $\mathbf{M}$  satisfies the conventional RIP condition (23), then the algorithm (20) with  $p = 1, q = 2$  will recover the true unknown  $\mathbf{X}$ . This requirement reduces to that we would obtain if we tried to recover each column of  $\mathbf{X}$  separately, using the standard  $\ell_1$  approach (12). As we already noted, previous equivalence results for MMV algorithms also share this feature. Although this condition guarantees that processing the vectors jointly does not harm the recovery ability, in practice exploiting the joint sparsity pattern of  $\mathbf{X}$  via (20) leads to improved results. Unfortunately, this behavior is not captured by any of the known equivalence conditions. This is due to the special structure of  $\mathbf{D} = \mathbf{M} \otimes \mathbf{I}$ . Since each measurement vector  $\mathbf{y}_i$  is affected only by the corresponding vector  $\mathbf{x}_i$ , it is clear that in the worst-case we can choose  $\mathbf{x}_i = \mathbf{x}$  for some vector  $\mathbf{x}$ . In this case, all the  $\mathbf{y}_i$ s are equal so that adding measurement vectors will not improve our recovery ability. Consequently, worst-case analysis based on the standard measurement model for MMV problems cannot lead to improved performance over the single measurement case.

## 4.2. Improved MMV Recovery

We now introduce an alternative measurement technique for MMV problems that can lead to improved worst-case behavior, as measured by RIP, over the single channel case. To enhance the performance of MMV recovery, we note that when we allow for an arbitrary (unstructured)  $\mathbf{D}$ , the RIP condition of Theorem 1 is weaker than the standard RIP requirement for recovering  $k$ -sparse vectors. This suggests that we can improve the performance of MMV methods by converting the problem into a general block sparsity problem, and then sampling with an arbitrary unstructured matrix  $\mathbf{D}$  rather than the choice  $\mathbf{D} = \mathbf{M}^T \otimes \mathbf{I}_d$ . The tradeoff introduced is increased computational complexity since each measurement is based on all input vectors. The theoretical conditions will now be looser, since block-RIP is weaker than standard RIP. Furthermore, in practice, this approach often improves the performance over separable MMV measurement techniques as we illustrate in the following example.

In the example, we compare the performance of several MMV algorithms for recovering  $\mathbf{X}$  in the model  $\mathbf{Y} = \mathbf{M}\mathbf{X}$ , with our method based on block sparsity in which the measurements  $\mathbf{y}$  are obtained via  $\mathbf{y} = \mathbf{D}\mathbf{x}$  where  $\mathbf{x} = \text{vec}(\mathbf{X}^T)$  and  $\mathbf{D}$  is a dense matrix. Choosing  $\mathbf{D}$  as a block diagonal matrix with blocks equal to  $\mathbf{M}$  results in the standard MMV measurement model. The effective matrices  $\mathbf{D}$  have the same size in the case in which it is block diagonal and when it is dense. To compare the performance of (11) with a dense  $\mathbf{D}$  to that of (20) with a block diagonal  $\mathbf{D}$ , we compute the empirical recovery rate of the methods in the same way performed in [12]. The matrices  $\mathbf{M}$  and  $\mathbf{D}$  are drawn randomly from a Gaussian ensemble. In our example, we choose  $\ell = 20, L = 30, d = 5$

where  $\ell$  is the number of rows in  $\mathbf{Y}$ . The matrix  $\mathbf{X}$  is generated randomly by first selecting the  $k$  non-zero rows uniformly at random, and then drawing the elements in these rows from a normal distribution. The empirical recovery rates using the methods of (20) for different choices of  $q$  and  $p$ , ReMBO [12] and our algorithm (11) with dense  $\mathbf{D}$  are depicted in Fig. 3. When the index  $p$  is omitted it is equal to 1. Evidently, our algorithm performs better than most popular optimization techniques for MMV systems. We stress that



**Fig. 3.** Recovery rate for different number  $k$  of non-zero rows in  $\mathbf{X}$ . Each point represents an average recovery rate over 500 simulations.

the performance advantage is due to the joint measurement process rather than a new recovery algorithm.

## 5. RANDOM MATRICES

Theorems 1 and 2 establish that a sufficiently small block RIP constant  $\delta_{2k|\mathcal{I}}$  ensures exact recovery of the coefficient vector  $\mathbf{x}$ . We now prove that random matrices are likely to satisfy this requirement. Specifically, we show that the probability that  $\delta_{k|\mathcal{I}}$  exceeds a certain threshold decays exponentially in the length of  $\mathbf{x}$ . Our approach relies on results of [5] developed for standard RIP, however, exploiting the block structure of  $\mathbf{x}$  leads to a much faster decay rate.

**Proposition 2** *Suppose  $\mathbf{D}$  is an  $n \times N$  matrix from the Gaussian ensemble, namely  $[\mathbf{D}]_{ik} \sim \mathcal{N}(0, \frac{1}{n})$ . Let  $\delta_{k|\mathcal{I}}$  be the smallest value satisfying the block RIP (7) over  $\mathcal{I} = \{d_1 = d, \dots, d_M = d\}$ , assuming  $N = md$  for some integer  $m$ . Then, for every  $\epsilon > 0$  the block RIP constant  $\delta_{k|\mathcal{I}}$  obeys (for  $n, N$  large enough, and fixed  $d$ )*

$$\text{Prob} \left( \sqrt{1 + \delta_{k|\mathcal{I}}} > 1 + (1 + \epsilon)f(r) \right) \leq 2e^{-NH(r)\epsilon} \cdot e^{-m(d-1)H(r)}. \quad (24)$$

Here, the ratio  $r = kd/N$  is fixed,  $f(r) = \sqrt{\frac{N}{n}} \left( \sqrt{r} + \sqrt{2H(r)} \right)$ , and  $H(q) = -q \log q - (1 - q) \log(1 - q)$  is the entropy function defined for  $0 < q < 1$ .

Proposition 2 reduces to the result of [5] when  $d = 1$ . However, since  $f(r)$  is independent of  $d$ , it follows that for  $d > 1$  and fixed problem dimensions  $n, N, r$ , block-RIP constants are smaller than the standard RIP constant. The second exponent in the right-hand-side of (24) is responsible for this behavior.

The following corollary allows us to emphasize the asymptotic behavior of block-RIP constants per given number of samples.

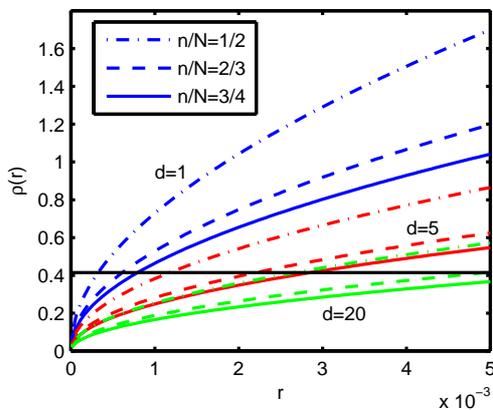
**Corollary 3** Consider the setting of Proposition 2, and define  $g(r) = \sqrt{\frac{N}{n}} \left( \sqrt{r} + \sqrt{2H(r)d^{-1}} \right)$ . Then:

$$\text{Prob} \left( \sqrt{1 + \delta_{k|\mathcal{I}}} > 1 + (1 + \epsilon)g(r) \right) \leq 2e^{-mH(r)\epsilon}. \quad (25)$$

To evaluate the asymptotic behavior of block-RIP we note that for every  $\epsilon > 0$  the right-hand-side of (25) goes to zero when  $N = md \rightarrow \infty$ . Consequently, this means that for fixed  $d$

$$\delta_{k|\mathcal{I}} < \rho(r) \triangleq -1 + [1 + g(r)]^2, \quad (26)$$

with overwhelming probability. In Fig. 4 we compute  $\rho(r)$  for several problem dimensions and compare it with standard-RIP which is obtained when  $d = 1$ . Evidently, as the non-zero entries are forced to block structure, a wider range of sparsity ratios  $r$  satisfy the condition of Theorem 1.



**Fig. 4.** The upper bound on  $\delta_{k|\mathcal{I}}$  as a function of the sparsity ratio  $r$ , for three sampling rates  $n/N$ , and three block structures  $d = 1, 5, 20$ . The horizontal threshold is fixed on  $\rho^* = \sqrt{2} - 1$ .

Although Fig. 4 shows advantage for block-RIP, the absolute sparsity ratios predicted by the theory are pessimistic as also noted in [5] for the case  $d = 1$ . To offer a more optimistic viewpoint, the RIP and block-RIP constants were computed brute-force for several instances of  $\mathbf{D}$  from the Gaussian ensemble. Fig. 5 plot the results and qualitatively affirm that block-RIP constants are more "likely" to be smaller than their standard RIP counterparts, even when the dimensions  $n, N$  are relatively small.

An important question is how many samples are needed roughly in order to guarantee stable recovery. This question is addressed in the following proposition, which quotes a result from [20].

**Proposition 3 ([20, Theorem 3.3])** Consider the setting of Proposition 2, namely a random Gaussian matrix  $\mathbf{D}$  of size  $n \times N$  and block sparse signals over  $\mathcal{I} = \{d_1 = d, \dots, d_m = d\}$ , where  $N = md$  for some integer  $m$ . Let  $t > 0$  and  $0 < \delta < 1$  be constant numbers. If

$$n \geq \frac{36}{7\delta} \left( \ln(2L) + kd \ln \left( \frac{12}{\delta} \right) + t \right), \quad (27)$$

where  $L = \binom{m}{k}$ , then  $\mathbf{D}$  satisfies the block-RIP (7) with restricted isometry constant  $\delta_{k|\mathcal{I}} = \delta$ , with probability at least  $1 - e^{-t}$ .

As observed in [20], the first term in (27) has the dominant impact on the required number of measurements in an asymptotic sense. Specifically, for block sparse signals

$$\left( \frac{m}{k} \right)^k \leq L = \binom{m}{k} \leq (em/k)^k. \quad (28)$$

Thus, for a given fraction of nonzeros  $r = kd/N$ , roughly  $n \approx k \log(m/k) = -k \log(r)$  measurements are needed. For comparison, to satisfy the standard RIP a larger number  $n \approx -kd \log(r)$  is required.

## 6. UNION OF SUBSPACES

We now show how the concept of block sparsity can be used in order to develop algorithms for recovering a signal from a union of subspaces in a stable and efficient manner.

### 6.1. Sampling Over a Union

Traditional sampling theory deals with the problem of recovering an unknown signal  $x \in \mathcal{H}$  from a set of  $n$  samples. The signal  $x$  can be a function of time  $x = x(t)$ , or can represent a finite-length vector  $x = \mathbf{x}$ . Typically it is assumed that  $x$  lies in a given subspace  $\mathcal{A}$  of  $\mathcal{H}$  [21–23]. The most common type of sampling is linear sampling in which

$$y_\ell = \langle s_\ell, x \rangle, \quad 1 \leq \ell \leq n, \quad (29)$$

for a set of functions  $s_\ell \in \mathcal{H}$  [21, 24–28]. Here  $\langle x, y \rangle$  denotes the standard inner product on  $\mathcal{H}$ . Nonlinear sampling is treated in [29]. However, here our focus will be on the linear case.

When  $\mathcal{H} = \mathbb{R}^N$  the unknown  $x = \mathbf{x}$  as well as the sampling functions  $s_\ell = \mathbf{s}_\ell$  are vectors in  $\mathbb{R}^N$ . Therefore, the samples can be written conveniently in matrix form as  $\mathbf{y} = \mathbf{S}^T \mathbf{x}$ , where  $\mathbf{S}$  is the matrix with columns  $\mathbf{s}_\ell$ . In the more general case in which  $\mathcal{H} = L_2$  or any other abstract Hilbert space, we can use the set transformation notation in order to conveniently represent the samples. A set transformation  $S : \mathbb{R}^n \rightarrow \mathcal{H}$  corresponding to sampling vectors  $\{s_\ell \in \mathcal{H}, 1 \leq \ell \leq n\}$  is defined by

$$\mathbf{S}\mathbf{x} = \sum_{\ell=1}^n \mathbf{x}(\ell) s_\ell \quad (30)$$

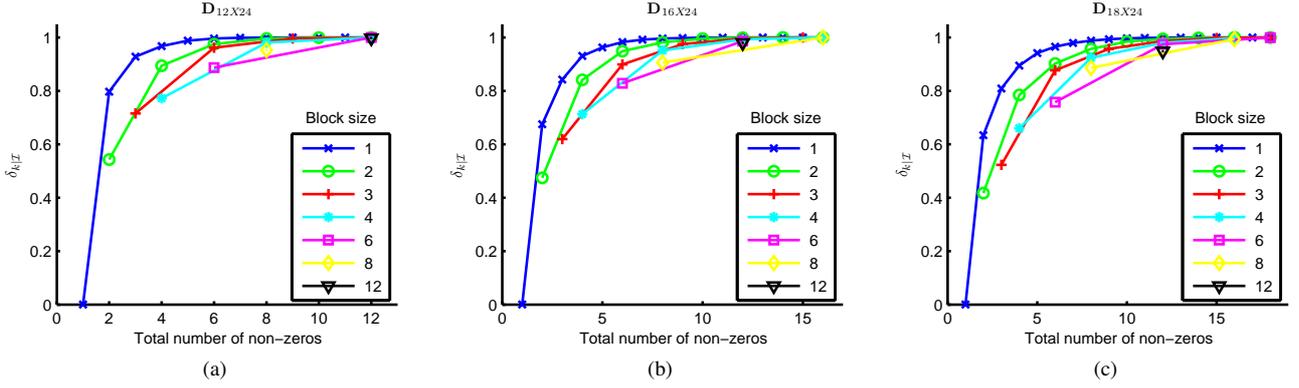
for all  $\mathbf{x} \in \mathbb{R}^n$ . From the definition of the adjoint, if  $\mathbf{x} = S^* x$ , then  $\mathbf{x}(\ell) = \langle s_\ell, x \rangle$ . Note that when  $\mathcal{H} = \mathbb{R}^N$ ,  $S = \mathbf{S}$  and  $S^* = \mathbf{S}^T$ . Using this notation, we can always express the samples as

$$\mathbf{y} = S^* x, \quad (31)$$

where  $S$  is a set transformation for arbitrary  $\mathcal{H}$ , and an appropriate matrix when  $\mathcal{H} = \mathbb{R}^N$ .

Our goal is to recover  $x$  from the samples  $\mathbf{y} \in \mathbb{R}^n$ . If the vectors  $s_\ell$  do not span the entire space  $\mathcal{H}$ , then there are many possible signals  $x$  consistent with the measurements  $\mathbf{y}$ . More specifically, if we define by  $\mathcal{S}$  the sampling space spanned by the vectors  $s_\ell$ , then clearly  $S^* v = 0$  for any  $v \in \mathcal{S}^\perp$ . Therefore, if  $\mathcal{S}^\perp$  is not the trivial space then adding such a vector  $v$  to any solution  $x$  of (31) will result in the same samples  $\mathbf{y}$ . A prior very often assumed is that  $x$  lies in a given subspace  $\mathcal{A}$  of  $\mathcal{H}$  [21–23, 30]. If  $\mathcal{A}$  and  $\mathcal{S}$  have the same finite dimension, and  $\mathcal{S}^\perp$  and  $\mathcal{A}$  intersect only at the 0 vector, then  $x$  can be perfectly recovered from the samples  $\mathbf{y}$  [22, 23].

Evidently, when subspace information is available, perfect reconstruction can often be guaranteed. Furthermore, in all of the cases



**Fig. 5.** The standard and block-RIP constants  $\delta_{k|X}$  for three different dimensions  $n, N$ . Each graph represent an average over 10 instances of random matrix  $\mathbf{D}$ .

above, recovery can be implemented by a simple linear transformation of the given samples (31). However, there are many practical scenarios in which  $x$  does not necessarily lie in a subspace. Nevertheless,  $x$  can often be described as lying in a union of subspaces

$$\mathcal{U} = \bigcup_{\ell} \mathcal{V}_{\ell}, \quad (32)$$

where each  $\mathcal{V}_{\ell}$  is a subspace. Thus,  $x$  belongs to one of the  $\mathcal{V}_{\ell}$ , but we do not know a priori to which one [15, 16]. Our goal is to recover a vector  $x$  lying in a union of subspaces, from a given set of samples. In principle, if we knew which subspace  $x$  belonged to, then reconstruction can be obtained using standard sampling results. However, here the problem is more involved because conceptually we first need to identify the correct subspace and only then can we recover the signal within the space.

Previous work on sampling over a union focused on invertibility and stability results [15, 16]. In contrast, here, our main interest is in developing concrete recovery algorithms that are provably robust. To achieve this goal, we limit our attention to a subclass of (32) for which stable recovery algorithms can be developed and analyzed. Specifically, we treat the case in which each  $\mathcal{V}_i$  has the additional structure

$$\mathcal{V}_i = \bigoplus_{\ell=k} \mathcal{A}_{\ell}, \quad (33)$$

where  $\{\mathcal{A}_{\ell}, 1 \leq \ell \leq M\}$  are a given set of disjoint subspaces, and  $|\ell| = k$  denotes a sum over  $k$  indices. Thus, each subspace  $\mathcal{V}_i$  corresponds to a different choice of  $k$  subspaces  $\mathcal{A}_{\ell}$  that comprise the sum. We assume that  $M$  and the dimensions  $d_{\ell} = \dim(\mathcal{A}_{\ell})$  of the subspaces  $\mathcal{A}_{\ell}$  are finite. Given  $n$  samples (31) and the knowledge that  $x$  lies in exactly one of the subspaces  $\mathcal{V}_i$ , we would like to recover the unknown signal  $x$ . In this setting, there are  $\binom{M}{k}$  possible subspaces comprising the union.

## 6.2. Connection with Block Sparsity

To solve the recovery problem, we show that  $x$  can be represented in terms of a block  $k$ -sparse vector. Our problem can then be reduced to that of recovering a block-sparse vector from a given set of samples.

To this end, we choose a basis for each  $\mathcal{A}_{\ell}$ . Denoting by  $A_{\ell} : \mathbb{R}^{d_{\ell}} \rightarrow \mathcal{H}$  a set transformation corresponding to a basis for  $\mathcal{A}_{\ell}$ , any  $x$  in the union can be written as

$$x = \sum_{|\ell|=k} A_{\ell} \mathbf{c}_{\ell}, \quad (34)$$

where  $\mathbf{c}_{\ell} \in \mathbb{R}^{d_{\ell}}$  are the representation coefficients in  $\mathcal{A}_{\ell}$ , and here  $|\ell| = k$  denotes a sum over a set of  $k$  indices. Our goal is to recover any  $x$  of the form (34) from the samples (31).

Define  $A : \mathbb{R}^N \rightarrow \mathcal{H}$  as the set transformation that is a result of concatenating the different  $A_{\ell}$ , with  $N = \sum_{\ell} d_{\ell}$ . Then

$$\mathbf{Ac} = \sum_{\ell=1}^M A_{\ell} \mathbf{c}[\ell]. \quad (35)$$

When  $\mathcal{H} = \mathbb{R}^N$  for some  $N$ ,  $A_{\ell} = \mathbf{A}_{\ell}$  is a matrix and  $A = \mathbf{A}$  is the matrix obtained by column-wise concatenating  $\mathbf{A}_{\ell}$ . If the  $j$ th subspace  $\mathcal{A}_j$  does not appear in the union, then  $\mathbf{c}[j] = \mathbf{0}$ . Thus,  $x$  can be written as  $x = \mathbf{Ac}$  where there are at most  $k$  non-zero blocks  $\mathbf{c}[\ell]$ . The measurements  $\mathbf{y} = S^*x$  can be expressed in terms of  $\mathbf{c}$  as

$$\mathbf{y} = S^* \mathbf{Ac}. \quad (36)$$

Since  $\mathbf{c}$  is block  $k$ -sparse, we have shown that our problem is equivalent to that of recovering a block-sparse vector with  $\mathbf{D} = S^* \mathbf{A}$ . Therefore, the results we developed can be applied to this setting.

Specifically, we first determine  $\mathbf{c}$  from  $\mathbf{y}$  by solving the SOCP (16) in the presence of noise with  $\mathbf{D} = S^* \mathbf{A}$ . Once  $\mathbf{c}$  is found, we can construct the original signal  $x$  via (34).

## 6.3. Relation to Previous Results

Previous treatments of the union model have considered invertibility and stability conditions. In contrast, here we provide concrete recovery algorithms for a signal over a structured union. We also show that the recovery is robust with respect to noise and mismodeling errors.

In our model, we assumed a finite union of subspaces as well as finite dimension of the underlying spaces. An interesting future direction to explore is the extension of the ideas developed herein to the more challenging problem of recovering  $x$  in a possibly infinite union of subspaces, which are not necessarily finite-dimensional. In some of our previous work, we have shown that a signal lying in a union of shift-invariant subspaces can be recovered efficiently from certain sets of sampling functions [12–14, 17, 18]. In our future work, we intend to combine these results with those in the current paper in order to develop a more general theory for recovery from a union of subspaces.

A recent preprint [31] proposes a new framework called model-based compressive sensing (MCS). The MCS approach assumes a signal model in which only certain predefined sparsity patterns may

appear. In general, obtaining efficient recovery algorithms in such scenarios is difficult, unless further structure is imposed on the sparsity patterns. Therefore, the authors consider two types of sparse vectors: block sparsity as treated here, and a wavelet tree model. For these settings, they generalize two known greedy algorithms: CoSaMP [32] and iterative hard thresholding (IHT) [20]. The union model developed here is broader than the block-sparse setting treated in [31] in the sense that it allows to model linear dependencies between the nonzero values rather than only between their locations, by appropriate choice of subspaces in (32), (33). In addition, we aim at optimization-based recovery algorithms which require selecting the objective in order to promote the model properties.

## 7. REFERENCES

- [1] D. L. Donoho, "Compressed sensing," *IEEE Trans. on Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr 2006.
- [2] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [3] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process. (Special Issue on Sparse Approximations in Signal and Image Processing)*, vol. 86, pp. 572–588, Apr. 2006.
- [4] —, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Process. (Special Issue on Sparse Approximations in Signal and Image Processing)*, vol. 86, pp. 589–602, Apr. 2006.
- [5] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [6] E. Candès, "The restricted isometry property and its implications for compressed sensing," *C. R. Acad. Sci. Paris, Ser. I*, vol. 346, pp. 589–592, 2008.
- [7] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, "Recovering Sparse Signals Using Sparse Measurement Matrices in Compressed DNA Microarrays," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 2, no. 3, pp. 275–285, June 2008.
- [8] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *arxiv 0804.0041*, Mar. 2008.
- [9] Y. C. Eldar and H. Bölcskei, "Compressed sensing for block-sparse signals: Uncertainty relations, coherence, and efficient recovery," 2009, in preparation.
- [10] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, July 2005.
- [11] 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, Dec. 2006.
- [12] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4692–4702, Oct. 2008.
- [13] Y. C. Eldar, "Compressed sensing of analog signals in shift-invariant spaces," to appear in *IEEE Trans. on Signal Proc.*
- [14] —, "Uncertainty relations for analog signals," submitted to *IEEE Trans. on Information Theory*, 2008.
- [15] Y. M. Lu and M. N. Do, "A theory for sampling signals from a union of subspaces," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2334–2345, 2008.
- [16] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Trans. Inf. Theory*, to appear.
- [17] 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.
- [18] —, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *arXiv 0902.4291*; submitted to *IEEE Selected Topics on Signal Process.*, 2009.
- [19] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a union of subspaces," *IEEE Trans. Inf. Theory*, 2008, submitted.
- [20] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *arxiv 0805.0510*, July 2008.
- [21] M. Unser, "Sampling—50 years after Shannon," *IEEE Proc.*, vol. 88, pp. 569–587, Apr. 2000.
- [22] Y. C. Eldar and T. Michaeli, "Beyond bandlimited sampling: Nonlinearities, smoothness and sparsity," to appear in *IEEE Signal Proc. Magazine*.
- [23] Y. C. Eldar and T. G. Dvorkind, "A minimum squared-error framework for generalized sampling," *IEEE Trans. Signal Processing*, vol. 54, no. 6, pp. 2155–2167, Jun. 2006.
- [24] Y. C. Eldar and T. Werther, "General framework for consistent sampling in Hilbert spaces," *International Journal of Wavelets, Multiresolution, and Information Processing*, vol. 3, no. 3, pp. 347–359, Sep. 2005.
- [25] O. Christansen and Y. C. Eldar, "Oblique dual frames and shift-invariant spaces," *Applied and Computational Harmonic Analysis*, vol. 17, no. 1, 2004.
- [26] Y. C. Eldar, "Sampling and reconstruction in arbitrary spaces and oblique dual frame vectors," *J. Fourier Anal. Appl.*, vol. 1, no. 9, pp. 77–96, Jan. 2003.
- [27] —, "Sampling without input constraints: Consistent reconstruction in arbitrary spaces," in *Sampling, Wavelets and Tomography*, A. I. Zayed and J. J. Benedetto, Eds. Boston, MA: Birkhäuser, 2004, pp. 33–60.
- [28] Y. C. Eldar and O. Christansen, "Characterization of oblique dual frame pairs," *J. Applied Signal Processing*, pp. 1–11, 2006, article ID 92674.
- [29] T. G. Dvorkind, Y. C. Eldar, and E. Matusiak, "Nonlinear and non-ideal sampling: Theory and methods," *IEEE Trans. Signal Processing*, to appear.
- [30] P. P. Vaidyanathan, "Generalizations of the sampling theorem: Seven decades after Nyquist," *IEEE Trans. Circuit Syst. I*, vol. 48, no. 9, pp. 1094–1109, Sep. 2001.
- [31] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *Preprint*, 2008.
- [32] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301 – 321, 2009.