Robust Recovery of Signals From a Structured Union of Subspaces

Yonina C. Eldar, *Senior Member, IEEE* and Moshe Mishali, *Student Member, IEEE*

***Abstract*—Traditional sampling theories consider the problem of reconstructing an unknown signal x from a series of samples. A prevalent assumption which often guarantees recovery from 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 lies in a union of subspaces. In this paper we develop a general framework for robust and efficient recovery of such signals from a given set of samples. More specifically, we treat the case in which x lies in a sum of k subspaces, chosen from a larger set of m possibilities. The samples are modelled as inner products with an arbitrary set of sampling functions. To derive an efficient and robust recovery algorithm, we show that our problem can be formulated as that of recovering a block-sparse vector whose non-zero elements appear in fixed blocks. We then propose a mixed ℓ_2/ℓ_1 program for block sparse recovery. Our main result is an equivalence condition under which the proposed convex algorithm is guaranteed to recover the original signal. This result relies on the notion of block restricted isometry property (RIP), which is a generalization of the standard RIP used extensively in the context of compressed sensing. Based on RIP we also prove stability of our approach in the presence of noise and modeling errors. A special case of our framework is that of recovering multiple measurement vectors (MMV) that share a joint sparsity pattern. Adapting our results to this context leads to new MMV recovery methods as well as equivalence conditions under which the entire set can be determined efficiently.**

***Index Terms*—Block sparsity, block restricted isometry property, compressed sensing, mixed-norm recovery, multiple measurement vectors (MMV), union of linear subspaces.**

I. Introduction

Sampling theory has a rich history dating back to Cauchy. Undoubtedly, the sampling theorem that had the most impact on signal processing and communications is that associated with Whittaker, Kotelńikov, and Shannon [1], [2]. Their famous result is that a bandlimited function $x(t)$ can be recovered from its uniform samples as long as the sampling rate exceeds the Nyquist rate, corresponding to twice the highest frequency of the signal [3]. More recently, this basic theorem has been extended to include more general classes of signal spaces. In particular, it can be shown that under mild technical conditions, a signal x lying in a given subspace can be recovered exactly from its linear generalized samples using a series of filtering operations [4]–[7].

Recently, there has been growing interest in nonlinear signal models in which the unknown x does not necessarily lie in a subspace. In order to ensure recovery from the samples, some underlying structure is needed. A general model that captures many interesting cases is that in which x lies in a union of subspaces. In this setting, x resides in one of a set of given subspaces $\mathcal{V}_i$, however, a priori it is not known in which one. A special case of this framework is the problem underlying the field of compressed sensing (CS), in which the goal is to recover a length N vector $\mathbf{x}$ from $n < N$ linear measurements, where $\mathbf{x}$ has no more than k non-zero elements in some basis [8], [9]. Many algorithms have been proposed in the literature in order to recover $\mathbf{x}$ in a stable and efficient manner [9]–[13]. A variety of conditions have been developed to ensure that these methods recover $\mathbf{x}$ exactly. One of the main tools in this context is the restricted isometry property (RIP) [9], [14], [15]. In particular, it can be shown that if the measurement matrix satisfies the RIP with an appropriate constant, then $\mathbf{x}$ can be recovered by solving an ℓ_1 minimization algorithm.

Another special case of a union of subspaces is the setting in which the unknown signal $x = x(t)$ has a multiband structure, so that its Fourier transform consists of a limited number of bands at unknown locations [16], [17]. By formulating this problem within the framework of CS, explicit sub-Nyquist sampling and reconstruction schemes were developed in [16], [17] that ensure perfect-recovery at the minimal possible rate. This setup was recently generalized in [18], [19] to deal with sampling and reconstruction of signals that lie in a finite union of shift-invariant subspaces. By combining ideas from standard sampling theory with CS results [20], explicit low-rate sampling and recovery methods were developed for such signal sets. In [21], an extension was considered to a special case of an infinite union of shit-invariant subspaces. The infinite union is a result of the fact that each generator of the space has an unknown time delay. Another example of a union of subspaces is the set of finite rate of innovation signals [21]–[23], that are modelled as a weighted sum of shifts of a given generating function, where the shifts are unknown.

In this paper, our goal is to develop a unified framework for efficient recovery of signals that lie in a structured union of subspaces. Our emphasis is on computationally efficient methods that are stable in the presence of noise and modeling errors. In contrast to our previous work [16]–[19], [21], here we consider unions of finite-dimensional subspaces. Specifi-

Manuscript received July 25, 2008; revised April 3, 2009. First published MMMMMM DD, YYYY; current version published MMMMMM DD, YYYY. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Justin Romberg.

The authors are with the Technion—Israel Institute of Technology, Haifa 32000, Israel. Phone: +972-4-8293256, fax: +972-4-8295757, E-mail: yonina@ee.technion.ac.il, moshiko@tx.technion.ac.il. 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).

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

Digital Object Identifier XX.XXXX/TSP.XXXX.XXXXXX

cally, we restrict our attention to 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 treated in [24], [25]. Conditions under which unique and stable sampling are possible were developed in [24], [25]. 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 propose a convex optimization algorithm that will often recover the true underlying x, and develop explicit conditions under which perfect recovery is guaranteed. Furthermore, we prove that our method is stable and robust in the sense that the reconstruction error is bounded in the presence of noise and mismodeling, namely when x does not lie exactly in the union. Our results rely on a generalization of the RIP which fits the union setting we treat here.

Our first contribution is showing that the problem of recovering x in a structured union of subspaces can be cast as a sparse recovery problem, in which it is desired to recover a sparse vector $\mathbf{c}$ that has a particular sparsity pattern: the nonzero values appear in fixed blocks. 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 of the sparsity pattern, recovery may be possible under more general conditions.

Next, we develop a concrete algorithm to determine a block-sparse vector from given measurements, which is based on minimizing a mixed ℓ_2/ℓ_1 norm. This problem can be cast as a convex second order cone program (SOCP), and solved efficiently using standard software packages. A mixed norm approach for block-sparse recovery was also considered in [26], [27]. By analyzing the measurement operator's null space, it was shown that asymptotically, as the signal length grows to infinity, and under ideal conditions (no noise or modeling errors), perfect recovery is possible with high probability. However, no robust equivalence results were established between the output of the algorithm and the true block-sparse vector for a given finite-length measurement vector, or in the presence of noise and mismodeling.

Generalizing the concept of RIP to our setting, we introduce the block RIP, which is a less stringent requirement. We then prove that if the measurement matrix satisfies the block RIP, then our proposed convex algorithm will recover the underlying block sparse signal. Furthermore, under block RIP, our algorithm is stable in the presence of noise and mismodeling errors. Using ideas similar to [13], [28] 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 the standard RIP. These results establish that a signal x that lies in a finite structured union can be recovered efficiently and stably with overwhelming probability if the measurement matrix is constructed from a random ensemble.

An interesting special case of the block-sparse model is the multiple measurement vector (MMV) problem, in which there is a set of unknown vectors that share a joint sparsity pattern. MMV recovery algorithms were studied in [20], [29]–[32]. Equivalence results based on mutual coherence for a mixed ℓ_p/ℓ_1 program were derived in [30]. These results 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 ℓ_2/ℓ_1 method translates into minimizing the sum of the ℓ_2 row-norms of the unknown matrix representing the MMV set. Our general results lead to RIP-based equivalence conditions for this algorithm. Furthermore, our framework suggests a different type of 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 when compared with other popular MMV algorithms.

The remainder of the paper is organized as follows. In Section II we describe the general problem of sampling from a union of subspaces. The relationship between our problem and that of block-sparse recovery is developed in Section III. In Section IV we explore stability and uniqueness issues which leads to the definition of block RIP. We also present a non-convex optimization algorithm with combinatorial complexity whose solution is the true unknown x. A convex relaxation of this algorithm is proposed in Section V. We then derive equivalence conditions based on block RIP. The concept of block RIP is further used to establish robustness and stability of our algorithm in the presence of noise and modeling errors. This approach is specialized to MMV sampling in Section VI. Finally, in Section VII we prove that random ensembles tend to satisfy the block RIP with high probability.

Throughout the paper, we denote vectors in an arbitrary Hilbert space $\mathcal{H}$ by lower case letters *e.g.,* x, and sets of vectors in $\mathcal{H}$ by calligraphic letters, *e.g.,* $\mathcal{S}$. Vectors in $\mathbb{R}^N$ are written as boldface lowercase letters *e.g.,* $\mathbf{x}$, and matrices as boldface uppercase letters *e.g.,* $\mathbf{A}$. The identity matrix of appropriate dimension is written as $\mathbf{I}$ or $\mathbf{I}_d$ when the dimension is not clear from the context, and $\mathbf{A}^T$ is the transpose of the matrix $\mathbf{A}$. The ith element of a vector $\mathbf{x}$ is denoted by $\mathbf{x}(i)$. Linear transformations from $\mathbb{R}^n$ to $\mathcal{H}$ are written as upper case letters $A : \mathbb{R}^n \to \mathcal{H}$. The adjoint of A is written as A^*. The standard Euclidean norm is denoted $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T\mathbf{x}}$ and $\|\mathbf{x}\|_1 = \sum_i |\mathbf{x}(i)|$ is the ℓ_1 norm of $\mathbf{x}$. The Kronecker product between matrices $\mathbf{A}$ and $\mathbf{B}$ is denoted $\mathbf{A} \otimes \mathbf{B}$, and $\mathrm{vec}(\mathbf{Y})$ is the vector obtained by concatenating the columns of $\mathbf{Y}$. The following variables are used in the sequel: n is the number of samples, N is the length of the input signal $\mathbf{x}$ when it is a vector, k is the sparsity or block sparsity (to be defined later on) of a vector $\mathbf{c}$, and m is the number of subspaces. For ease of notation we assume throughout that all scalars are defined over the field of real numbers; however, the results are also valid over the complex domain with appropriate modifications.

II. Union of Subspaces

A. Subspace Sampling

Traditional sampling theory deals with the problem of recovering an unknown signal $x \in \mathcal{H}$ from a set of n samples

$y_i = f_i(x)$ where $f_i(x)$ is some function of x. The signal x can be a function of time $x = x(t)$, or can represent a finite-length vector $x = \mathbf{x}$. The most common type of sampling is linear sampling in which

$$y_i = \langle s_i, x \rangle, \quad 1 \le i \le n, \tag{1}$$

for a set of functions $s_i \in \mathcal{H}$ [4], [33]–[39]. Here $\langle x, y \rangle$ denotes the standard inner product on $\mathcal{H}$. For example, if $\mathcal{H} = L_2$ is the space of real finite-energy signals then

$$\langle x, y \rangle = \int_{-\infty}^{\infty} x(t)y(t)dt. \tag{2}$$

When $\mathcal{H} = \mathbb{R}^N$ for some N,

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{N} \mathbf{x}(i)\mathbf{y}(i). \tag{3}$$

Nonlinear sampling is treated in [40]. 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 $s_i = \mathbf{s}_i$ 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 $\mathbf{s}_i$. 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 $S : \mathbb{R}^n \to \mathcal{H}$ corresponding to sampling vectors $\{s_i \in \mathcal{H}, 1 \le i \le n\}$ is defined by

$$S\mathbf{c} = \sum_{i=1}^{n} \mathbf{c}(i)s_i \tag{4}$$

for all $\mathbf{c} \in \mathbb{R}^n$. From the definition of the adjoint, if $\mathbf{c} = S^*x$, then $\mathbf{c}(i) = \langle s_i, x \rangle$. 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

$$\mathbf{y} = S^*x, \tag{5}$$

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_i do not span the entire space $\mathcal{H}$, then there are many possible signals x consistent with $\mathbf{y}$. More specifically, if we define by $\mathcal{S}$ the sampling space spanned by the vectors s_i, 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 (5) will result in the same samples $\mathbf{y}$. However, by exploiting prior knowledge on x, in many cases uniqueness can be guaranteed. A prior very often assumed is that x lies in a given subspace $\mathcal{A}$ of $\mathcal{H}$ [4]–[7]. 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 $\mathbf{y}$ [6], [7], [41].

B. Union of Subspaces

When subspace information is available, perfect reconstruction can often be guaranteed. Furthermore, recovery can be implemented by a simple linear transformation of the given samples (5). However, there are many practical scenarios in which we are given prior information about x that is not necessarily in the form of a subspace. One such case studied in detail in [41] is that in which x is known to be smooth. Here we focus our attention on the setting where x lies in a union of subspaces

$$\mathcal{U} = \bigcup_i \mathcal{V}_i \tag{6}$$

where each $\mathcal{V}_i$ is a subspace. Thus, x belongs to one of the $\mathcal{V}_i$, but we do not know a priori to which one [24], [25]. Note that the set $\mathcal{U}$ is no longer a subspace. Indeed, if $\mathcal{V}_i$ is, for example, a one-dimensional space spanned by the vector $\mathbf{v}_i$, then $\mathcal{U}$ contains vectors of the form $\alpha\mathbf{v}_i$ for some i but does not include their linear combinations. 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 [24], [25]. 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 (6) 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

$$\mathcal{V}_i = \bigoplus_{|j|=k} \mathcal{A}_j, \tag{7}$$

where $\{\mathcal{A}_j, 1 \le j \le m\}$ are a given set of disjoint subspaces, and $|j| = k$ denotes a sum over k indices. Thus, each subspace $\mathcal{V}_i$ corresponds to a different choice of k subspaces $\mathcal{A}_j$ that comprise the sum. We assume throughout the paper that m and the dimensions $d_i = \dim(\mathcal{A}_i)$ of the subspaces $\mathcal{A}_i$ are finite. Given n samples

$$\mathbf{y} = S^*x \tag{8}$$

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.

An alternative interpretation of our model is as follows. Given an observation vector $\mathbf{y}$, we seek a signal x for which $\mathbf{y} = S^*x$ and in addition x can be written as

$$x = \sum_{i=1}^{k} x_i, \tag{9}$$

where each x_i lies in $\mathcal{A}_j$ for some index j.

A special case is the standard CS problem in which $x = \mathbf{x}$ is a vector of length N, that has a sparse representation in a given basis defined by an invertible matrix $\mathbf{W}$. Thus, $\mathbf{x} = \mathbf{W}\mathbf{c}$ where $\mathbf{c}$ is a sparse vector that has at most k nonzero elements. This fits our framework by choosing $\mathcal{A}_i$ as the space spanned by the ith column of $\mathbf{W}$. In this setting $m = N$, and there are $\binom{N}{k}$ subspaces comprising the union.

Another example is the block sparsity model [26], [42] in which $\mathbf{x}$ is divided into equal-length blocks of size d, and at

most k blocks can be non zero. Such a vector can be described in our setting with $\mathcal{H} = \mathbb{R}^N$ by choosing $\mathcal{A}_i$ to be the space spanned by the corresponding d columns of the identity matrix. Here $m = N/d$ and there are $\binom{N/d}{k}$ subspaces in the union.

A final example is the MMV problem [20], [29]–[32] in which our goal is to recover a matrix $\mathbf{X}$ from measurements $\mathbf{Y} = \mathbf{M}\mathbf{X}$, for a given sampling matrix $\mathbf{M}$. The matrix $\mathbf{X}$ is assumed to have at most k non-zero rows. Thus, not only is each column $\mathbf{x}_i$ k-sparse, but in addition the non-zero elements of $\mathbf{x}_i$ share a joint sparsity pattern. This problem can be transformed into that of recovering a k-block sparse signal by stacking the rows of $\mathbf{X}$ and $\mathbf{Y}$, leading to the relationship

$$\mathrm{vec}(\mathbf{Y}^T) = (\mathbf{M} \otimes \mathbf{I})\,\mathrm{vec}(\mathbf{X}^T). \tag{10}$$

The structure of $\mathbf{X}$ leads to a vector $\mathrm{vec}(\mathbf{X}^T)$ that is k-block sparse.

C. Problem Formulation and Main Results

Given k and the subspaces $\mathcal{A}_i$, we would like to address the following questions:

1) What are the conditions on the sampling vectors $s_i, 1 \leq i \leq n$ in order to guarantee that the sampling is invertible and stable?
2) How can we recover the unique x (regardless of computational complexity)?
3) How can we recover the unique x in an efficient and stable manner?

The first question was addressed in [24], [25] in the more general context of unions of spaces (without requiring a particular structure such as (7)). However, no concrete methods were proposed in order to recover x. Here we provide efficient convex algorithms that recover x in a stable way for arbitrary k under appropriate conditions on the sampling functions s_i and the spaces $\mathcal{A}_i$.

Our results are based on an equivalence between the union of subspaces problem assuming (7) and that of recovering block-sparse vectors. This allows us to determine x from the given samples by first treating the problem of recovering a block k-sparse vector $\mathbf{c}$ from a given set of measurements. This relationship is established in the next section. In the reminder of the paper we therefore focus on the block k-sparse model and develop our results in that context. In particular, we introduce a block RIP condition that ensures uniqueness and stability of our sampling problem. We then suggest an efficient convex optimization problem which approximates an unknown block-sparse vector $\mathbf{c}$. Based on block RIP we prove that $\mathbf{c}$ can be recovered exactly in a stable way using the proposed optimization program. Furthermore, in the presence of noise and modeling errors, our algorithm can approximate the best block-k sparse solution.

III. Connection with Block Sparsity

Consider the model of a signal x in the union of k out of m subspaces $\mathcal{A}_i$, with $d_i = \dim(\mathcal{A}_i)$ as in (6) and (7). To write x explicitly, we choose a basis for each $\mathcal{A}_i$. Denoting by $A_i : \mathbb{R}^{d_i} \to \mathcal{H}$ the set transformation corresponding to a basis for $\mathcal{A}_i$, any such x can be written as

$$\mathbf{c}^T = $$

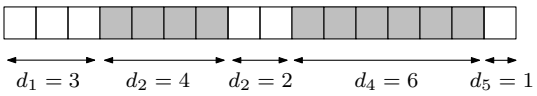

Fig. 1. A block-sparse vector $\mathbf{c}$ over $\mathcal{I} = \{d_1, \ldots, d_5\}$. The gray areas represent 10 non-zero entries which occupy two blocks.

$$x = \sum_{|i|=k} A_i \mathbf{c}_i, \tag{11}$$

where $\mathbf{c}_i \in \mathbb{R}^{d_i}$ are the representation coefficients in $\mathcal{A}_i$, and $|i| = k$ denotes a sum over a set of k indices. The choice of indices depend on the signal x and are unknown in advance.

To develop the equivalence with block sparsity, it is useful to introduce some further notation. First, we define $A : \mathbb{R}^N \to \mathcal{H}$ as the set transformation that is a result of concatenating the different A_i, with

$$N = \sum_{i=1}^{m} d_i. \tag{12}$$

Next, we define the ith sub-block $\mathbf{c}[i]$ of a length-N vector $\mathbf{c}$ over $\mathcal{I} = \{d_1, \ldots, d_m\}$. The ith sub-block is of length d_i, and the blocks are formed sequentially so that

$$\mathbf{c}^T = [\underbrace{c_1 \ \ldots \ c_{d_1}}_{\mathbf{c}[1]} \ \ldots \ \underbrace{c_{N-d_m+1} \ \ldots \ c_N}_{\mathbf{c}[m]}]^T. \tag{13}$$

We can then define A by

$$A\mathbf{c} = \sum_{i=1}^{m} A_i \mathbf{c}[i]. \tag{14}$$

When $\mathcal{H} = \mathbb{R}^N$ for some N, $A_i = \mathbf{A}_i$ is a matrix and $A = \mathbf{A}$ is the matrix obtained by column-wise concatenating $\mathbf{A}_i$. If for a given x the jth subspace $\mathcal{A}_j$ does not appear in the sum (7), or equivalently in (11), then $\mathbf{c}[j] = \mathbf{0}$.

Any x in the union (6), (7) can be represented in terms of k of the bases A_i. Therefore, we can write $x = A\mathbf{c}$ where there are at most k non-zero blocks $\mathbf{c}[i]$. Consequently, our union model is equivalent to the model in which x is represented by a sparse vector $\mathbf{c}$ in an appropriate basis. However, the sparsity pattern here has a unique form which we will exploit in our conditions and algorithms: the non-zero elements appear in blocks.

Definition 1: A vector $\mathbf{c} \in \mathbb{R}^N$ is called block k-sparse over $\mathcal{I} = \{d_1, \ldots, d_m\}$ if $\mathbf{c}[i]$ is nonzero for at most k indices i where $N = \sum_i d_i$.

An example of a block-sparse vector with $k = 2$ is depicted in Fig. 1. When $d_i = 1$ for each i, block sparsity reduces to the conventional definition of a sparse vector. Denoting

$$\|\mathbf{c}\|_{0,\mathcal{I}} = \sum_{i=1}^{m} I(\|\mathbf{c}[i]\|_2 > 0), \tag{15}$$

where $I(\|\mathbf{c}[i]\|_2 > 0)$ is an indicator function that obtains the value 1 if $\|\mathbf{c}[i]\|_2 > 0$ and 0 otherwise, a block k-sparse vector $\mathbf{c}$ can be defined by $\|\mathbf{c}\|_{0,\mathcal{I}} \leq k$.

Evidently, there is a one-to-one correspondence between a vector x in the union, and a block-sparse vector $\mathbf{c}$. The measurements (5) can also be represented explicitly in terms of $\mathbf{c}$ as

$$\mathbf{y} = S^* x = S^* A \mathbf{c} = \mathbf{D}\mathbf{c}, \tag{16}$$

where $\mathbf{D}$ is the $n \times N$ matrix defined by

$$\mathbf{D} = S^* A. \tag{17}$$

We can therefore phrase our problem in terms of $\mathbf{D}$ and $\mathbf{c}$ as that of recovering a block-k sparse vector $\mathbf{c}$ over $\mathcal{I}$ from the measurements (16).

Note that the choice of basis A_i for each subspace does not affect our model. Indeed, choosing alternative bases will lead to $x = A\mathbf{W}\mathbf{c}$ where $\mathbf{W}$ is a block diagonal matrix with blocks of size d_i. Defining $\tilde{\mathbf{c}} = \mathbf{W}\mathbf{c}$, the block sparsity pattern of $\tilde{\mathbf{c}}$ is equal to that of $\mathbf{c}$.

Since our problem is equivalent to that of recovering a block sparse vector over $\mathcal{I}$ from linear measurements $\mathbf{y} = \mathbf{D}\mathbf{c}$, in the reminder of the paper we focus our attention on this problem.

IV. Uniqueness and Stability

In this section we study the uniqueness and stability of our sampling method. These properties are intimately related to the RIP, which we generalize here to the block-sparse setting.

The first question we address is that of uniqueness, namely conditions under which a block-sparse vector $\mathbf{c}$ is uniquely determined by the measurement vector $\mathbf{y} = \mathbf{D}\mathbf{c}$.

Proposition 1: There is a unique block-k sparse vector $\mathbf{c}$ consistent with the measurements $\mathbf{y} = \mathbf{D}\mathbf{c}$ if and only if $\mathbf{D}\mathbf{c} \neq \mathbf{0}$ for every $\mathbf{c} \neq \mathbf{0}$ that is block $2k$-sparse.

Proof: The proof follows from [24, Proposition 4]. ∎

We next address the issue of stability. A sampling operator is stable for a set $\mathcal{T}$ if and only if there exists constants $\alpha > 0$, $\beta < \infty$ such that

$$\alpha \|x_1 - x_2\|_{\mathcal{H}}^2 \leq \|S^* x_1 - S^* x_2\|_2^2 \leq \beta \|x_1 - x_2\|_{\mathcal{H}}^2, \tag{18}$$

for every x_1, x_2 in $\mathcal{T}$. The ratio $\kappa = \beta/\alpha$ provides a measure for stability of the sampling operator. The operator is maximally stable when $\kappa = 1$. In our setting, S^* is replaced by $\mathbf{D}$, and the set $\mathcal{T}$ contains block-k sparse vectors. The following proposition follows immediately from (18) by noting that given two block-k sparse vectors $\mathbf{c}_1, \mathbf{c}_2$ their difference $\mathbf{c}_1 - \mathbf{c}_2$ is block-$2k$ sparse.

Proposition 2: The measurement matrix $\mathbf{D}$ is stable for every block k-sparse vector $\mathbf{c}$ if and only if there exists $C_1 > 0$ and $C_2 < \infty$ such that

$$C_1 \|\mathbf{v}\|_2^2 \leq \|\mathbf{D}\mathbf{v}\|_2^2 \leq C_2 \|\mathbf{v}\|_2^2, \tag{19}$$

for every $\mathbf{v}$ that is block $2k$-sparse.

It is easy to see that if $\mathbf{D}$ satisfies (19) then $\mathbf{D}\mathbf{c} \neq \mathbf{0}$ for all block $2k$-sparse vectors $\mathbf{c}$. Therefore, this condition implies both invertibility and stability.

A. Block RIP

Property (19) is related to the RIP used in several previous works in CS [9], [14], [15]. A matrix $\mathbf{D}$ of size $n \times N$ is said to have the RIP if there exists a constant $\delta_k \in [0, 1)$ such that for every k-sparse $\mathbf{c} \in \mathbb{R}^N$,

$$(1 - \delta_k)\|\mathbf{c}\|_2^2 \leq \|\mathbf{D}\mathbf{c}\|_2^2 \leq (1 + \delta_k)\|\mathbf{c}\|_2^2. \tag{20}$$

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

Definition 2: Let $\mathbf{D} : \mathbb{R}^N \to \mathbb{R}^n$ be a given matrix. Then $\mathbf{D}$ has the block RIP over $\mathcal{I} = \{d_1, \ldots, d_m\}$ with parameter $\delta_{k|\mathcal{I}}$ if for every $\mathbf{c} \in \mathbb{R}^N$ that is block k-sparse over $\mathcal{I}$ we have that

$$(1 - \delta_{k|\mathcal{I}})\|\mathbf{c}\|_2^2 \leq \|\mathbf{D}\mathbf{c}\|_2^2 \leq (1 + \delta_{k|\mathcal{I}})\|\mathbf{c}\|_2^2. \tag{21}$$

By abuse of notation, we use δ_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 [25]. From Proposition 1 it follows that if $\mathbf{D}$ satisfies the RIP (21) with $\delta_{2k} < 1$, then there is a unique block-sparse vector $\mathbf{c}$ consistent with (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 (21) must hold for all $2M$-sparse vectors $\mathbf{c}$. Since we only require the RIP for block sparse signals, (21) only 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 δ_M (where M depends on k; for blocks with equal size d, $M = kd$).

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

$$\mathbf{D} = \left(\begin{array}{cc|cc|cc} -1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 2 & -1 & 0 & 0 & 3 \\ 0 & 3 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 & 1 \end{array}\right) \cdot \mathbf{B}, \tag{22}$$

where $\mathbf{B}$ is a diagonal matrix that results in unit-norm columns of $\mathbf{D}$, *i.e.,* $\mathbf{B} = \operatorname{diag}(1, 15, 1, 1, 1, 12)^{-1/2}$. In this example $m = 3$ and $\mathcal{I} = \{d_1 = 2, d_2 = 2, d_3 = 2\}$. Suppose that $\mathbf{c}$ is block-1 sparse, which corresponds to at most two non-zero values. Brute-force calculations show that the smallest value of δ_2 satisfying the standard RIP (20) is $\delta_2 = 0.866$. On the other hand, the block-RIP (21) corresponding to the case in which the two non-zero elements are restricted to occur in one block is satisfied with $\delta_{1|\mathcal{I}} = 0.289$. Increasing the number of non-zero elements to $k = 4$, we can verify that the standard RIP (20) does not hold for any $\delta_4 \in [0, 1)$. Indeed, in this example there exist two 4-sparse vectors that result in the same measurements. In contrast, $\delta_{2|\mathcal{I}} = 0.966$ satisfies the lower bound in (21) when restricting the 4 non-zero values to two blocks. Consequently, the measurements $\mathbf{y} = \mathbf{D}\mathbf{c}$ uniquely specify a single block-sparse $\mathbf{c}$.

In the next section, we will see that the ability to recover $\mathbf{c}$ in a computationally efficient way depends on the constant $\delta_{2k|\mathcal{I}}$ in the block RIP (21). The smaller the value of $\delta_{2k|\mathcal{I}}$,

the fewer samples are needed in order to guarantee stable recovery. Both standard and block RIP constants $\delta_k, \delta_{k|\mathcal{I}}$ are by definition increasing with k. Therefore, it was suggested in [13] to normalize each of the columns of $\mathbf{D}$ to 1, so as to start with $\delta_1 = 0$. In the same spirit, we recommend choosing the bases for $\mathcal{A}_i$ such that $\mathbf{D} = S^* A$ has unit-norm columns, corresponding to $\delta_{1|I} = 0$.

B. Recovery Method

We have seen that if $\mathbf{D}$ satisfies the RIP (21) with $\delta_{2k} < 1$, then there is a unique block-sparse vector $\mathbf{c}$ consistent with (16). The question is how to find $\mathbf{c}$ in practice. Below we present an algorithm that will in principle find the unique $\mathbf{c}$ from the samples $\mathbf{y}$. Unfortunately, though, it has exponential complexity. In the next section we show that under a stronger condition on δ_{2k} we can recover $\mathbf{c}$ in a stable and efficient manner.

Our first claim is that $\mathbf{c}$ can be uniquely recovered by solving the optimization problem

$$\begin{aligned} \min_{\mathbf{c}} \quad & \|\mathbf{c}\|_{0,\mathcal{I}} \\ \text{s.t.} \quad & \mathbf{y} = \mathbf{D}\mathbf{c}. \end{aligned} \tag{23}$$

To show that (23) will indeed recover the true value of $\mathbf{c}$, suppose that there exists a $\mathbf{c}'$ such that $\mathbf{D}\mathbf{c}' = \mathbf{y}$ and $\|\mathbf{c}'\|_{0,\mathcal{I}} \leq \|\mathbf{c}\|_{0,\mathcal{I}} \leq k$. Since both $\mathbf{c}, \mathbf{c}'$ are consistent with the measurements,

$$\mathbf{0} = \mathbf{D}(\mathbf{c} - \mathbf{c}') = \mathbf{D}\mathbf{d}, \tag{24}$$

where $\|\mathbf{d}\|_{0,\mathcal{I}} \leq 2k$ so that $\mathbf{d}$ is a block $2k$-sparse vector. Since $\mathbf{D}$ satisfies (21) with $\delta_{2k} < 1$, we must have that $\mathbf{d} = \mathbf{0}$ or $\mathbf{c} = \mathbf{c}'$.

In principle (23) can be solved by searching over all possible sets of k blocks whether there exists a $\mathbf{c}$ that is consistent with the measurements. The invertibility condition (21) ensures that there is only one such $\mathbf{c}$. However, clearly this approach is not efficient.

V. Convex Recovery Algorithm

A. Noise-Free Recovery

We now develop an efficient convex optimization problem instead of (23) to approximate $\mathbf{c}$. As we show, if $\mathbf{D}$ satisfies (21) with a small enough value of δ_{2k}, then the method we propose will recover $\mathbf{c}$ exactly.

Our approach is to minimize the sum of the energy of the blocks $\mathbf{c}[i]$. To write down the problem explicitly, we define the mixed ℓ_2/ℓ_1 norm over the index set $\mathcal{I} = \{d_1, \ldots, d_m\}$ as

$$\|\mathbf{c}\|_{2,\mathcal{I}} = \sum_{i=1}^{m} \|\mathbf{c}[i]\|_2. \tag{25}$$

The algorithm we suggest is then

$$\begin{aligned} \min_{\mathbf{c}} \quad & \|\mathbf{c}\|_{2,\mathcal{I}} \\ \text{s.t.} \quad & \mathbf{y} = \mathbf{D}\mathbf{c}. \end{aligned} \tag{26}$$

Problem (26) can be written as an SOCP by defining $t_i = \|\mathbf{c}[i]\|_2$. Then (26) is equivalent to

$$\begin{aligned} \min_{\mathbf{c}, t_i} \quad & \sum_{i=1}^{m} t_i \\ \text{s.t.} \quad & \mathbf{y} = \mathbf{D}\mathbf{c} \\ & t_i \geq \|\mathbf{c}[i]\|_2, \quad 1 \leq i \leq m \\ & t_i \geq 0, \quad 1 \leq i \leq m, \end{aligned} \tag{27}$$

which can be solved using standard software packages.

The next theorem establishes that the solution to (26) is the true $\mathbf{c}$ as long as δ_{2k} is small enough.

Theorem 1: Let $\mathbf{y} = \mathbf{D}\mathbf{c}_0$ be measurements of a block k-sparse vector $\mathbf{c}_0$. If $\mathbf{D}$ satisfies the block RIP (21) with $\delta_{2k} < \sqrt{2} - 1$ then

1) there is a unique block-k sparse vector $\mathbf{c}$ consistent with $\mathbf{y}$;
2) the SOCP (27) has a unique solution;
3) the solution to the SOCP is equal $\mathbf{c}_0$.

Before proving the theorem we note that it provides a gain over standard CS results. Specifically, it is shown in [15] that if $\mathbf{c}$ is k-sparse and the measurement matrix $\mathbf{D}$ satisfies the standard RIP with $\delta_{2k} < \sqrt{2} - 1$, then $\mathbf{c}$ can be recovered exactly from the measurements $\mathbf{y} = \mathbf{D}\mathbf{c}$ via the linear program:

$$\begin{aligned} \min_{\mathbf{c}} \quad & \|\mathbf{c}\|_1 \\ \text{s.t.} \quad & \mathbf{y} = \mathbf{D}\mathbf{c}. \end{aligned} \tag{28}$$

Since any block k-sparse vector is also M-sparse with M equal to the sum of the k largest values of d_i, we can find $\mathbf{c}_0$ of Theorem 1 by solving (28) if δ_{2M} is small enough. However, this standard CS approach does not exploit the fact that the non-zero values appear in blocks, and not in arbitrary locations within the vector $\mathbf{c}_0$. On the other hand, the SOCP (27) explicitly takes the block structure of $\mathbf{c}_0$ into account. Therefore, the condition of Theorem 1 is not as stringent as that obtained by using equivalence results with respect to (28). Indeed, the block RIP (21) bounds the norm of $\|\mathbf{D}\mathbf{c}\|$ over block sparse vectors $\mathbf{c}$, while the standard RIP considers all possible choices of $\mathbf{c}$, also those that are not $2k$-block sparse. Therefore, the value of δ_{2k} in (21) can be lower than that obtained from (20) with $k = 2M$, as we illustrated by an example in Section III. This advantage will also be seen in the context of a concrete example at the end of the section.

Our proof below is rooted in that of [15]. However, some essential modifications are necessary in order to adapt the results to the block-sparse case. These differences are a result of the fact that our algorithm relies on the mixed ℓ_2/ℓ_1 norm rather than the ℓ_1 norm alone. This adds another layer of complication to the proof, and therefore we expand the derivations in more detail than in [15].

Proof: We first note that $\delta_{2k} < 1$ guarantees uniqueness of $\mathbf{c}_0$ from Proposition 1. To prove parts 2) and 3) we show that any solution to (26) has to be equal to $\mathbf{c}_0$. To this end let $\mathbf{c}' = \mathbf{c}_0 + \mathbf{h}$ be a solution of (26). The true value $\mathbf{c}_0$ is non-zero over at most k blocks. We denote by $\mathcal{I}_0$ the block indices for which $\mathbf{c}_0$ is nonzero, and by $\mathbf{h}_{\mathcal{I}_0}$ the restriction of

$\mathbf{h}$ to these blocks. Next we decompose $\mathbf{h}$ as

$$\mathbf{h} = \sum_{i=0}^{\ell-1} \mathbf{h}_{\mathcal{I}_i}, \tag{29}$$

where $\mathbf{h}_{\mathcal{I}_i}$ is the restriction of $\mathbf{h}$ to the set $\mathcal{I}_i$ which consists of k blocks, chosen such that the norm of $\mathbf{h}_{\mathcal{I}_0^c}$ over $\mathcal{I}_1$ is largest, the norm over $\mathcal{I}_2$ is second largest and so on. Our goal is to show that $\mathbf{h} = \mathbf{0}$. We prove this by noting that

$$\|\mathbf{h}\|_2 = \|\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1} + \mathbf{h}_{(\mathcal{I}_0\cup\mathcal{I}_1)^c}\|_2 \leq \|\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2 + \|\mathbf{h}_{(\mathcal{I}_0\cup\mathcal{I}_1)^c}\|_2. \tag{30}$$

In the first part of the proof we show that $\|\mathbf{h}_{(\mathcal{I}_0\cup\mathcal{I}_1)^c}\|_2 \leq \|\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2$. In the second part we establish that $\|\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2 = 0$, which completes the proof.

Part I: $\|\mathbf{h}_{(\mathcal{I}_0\cup\mathcal{I}_1)^c}\|_2 \leq \|\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2$

We begin by noting that

$$\|\mathbf{h}_{(\mathcal{I}_0\cup\mathcal{I}_1)^c}\|_2 = \left\|\sum_{i=2}^{\ell-1} \mathbf{h}_{\mathcal{I}_i}\right\|_2 \leq \sum_{i=2}^{\ell-1} \|\mathbf{h}_{\mathcal{I}_i}\|_2. \tag{31}$$

Therefore, it is sufficient to bound $\|\mathbf{h}_{\mathcal{I}_i}\|_2$ for $i \geq 2$. Now,

$$\|\mathbf{h}_{\mathcal{I}_i}\|_2 \leq k^{1/2}\|\mathbf{h}_{\mathcal{I}_i}\|_{\infty,\mathcal{I}} \leq k^{-1/2}\|\mathbf{h}_{\mathcal{I}_{i-1}}\|_{2,\mathcal{I}}, \tag{32}$$

where we defined $\|\mathbf{a}\|_{\infty,\mathcal{I}} = \max_i \|\mathbf{a}[i]\|_2$. The first inequality follows from the fact that for any block k-sparse $\mathbf{c}$,

$$\|\mathbf{c}\|_2^2 = \sum_{|i|=k} \|\mathbf{c}[i]\|_2^2 \leq k\|\mathbf{c}\|_{\infty,\mathcal{I}}^2. \tag{33}$$

The second inequality in (32) is a result of the fact that the norm of each block in $\mathbf{h}_{\mathcal{I}_i}$ is by definition smaller or equal to the norm of each block in $\mathbf{h}_{\mathcal{I}_{i-1}}$. Since there are at most k nonzero blocks, $k\|\mathbf{h}_{\mathcal{I}_i}\|_{\infty,\mathcal{I}} \leq \|\mathbf{h}_{\mathcal{I}_{i-1}}\|_{2,\mathcal{I}}$. Substituting (32) into (31),

$$\begin{aligned}\|\mathbf{h}_{(\mathcal{I}_0\cup\mathcal{I}_1)^c}\|_2 &\leq k^{-1/2}\sum_{i=1}^{\ell-2} \|\mathbf{h}_{\mathcal{I}_i}\|_{2,\mathcal{I}} \\ &\leq k^{-1/2}\sum_{i=1}^{\ell-1} \|\mathbf{h}_{\mathcal{I}_i}\|_{2,\mathcal{I}} = k^{-1/2}\|\mathbf{h}_{\mathcal{I}_0^c}\|_{2,\mathcal{I}},\end{aligned} \tag{34}$$

where the equality is a result of the fact that $\|\mathbf{c}_1 + \mathbf{c}_2\|_{2,\mathcal{I}} = \|\mathbf{c}_1\|_{2,\mathcal{I}} + \|\mathbf{c}_2\|_{2,\mathcal{I}}$ if $\mathbf{c}_1$ and $\mathbf{c}_2$ are non-zero on disjoint blocks.

To develop a bound on $\|\mathbf{h}_{\mathcal{I}_0^c}\|_{2,\mathcal{I}}$ note that since $\mathbf{c}'$ is a solution to (26), $\|\mathbf{c}_0\|_{2,\mathcal{I}} \geq \|\mathbf{c}'\|_{2,\mathcal{I}}$. Using the fact that $\mathbf{c}' = \mathbf{c}_0 + \mathbf{h}_{\mathcal{I}_0} + \mathbf{h}_{\mathcal{I}_0^c}$ and $\mathbf{c}_0$ is supported on $\mathcal{I}_0$ we have

$$\begin{aligned}\|\mathbf{c}_0\|_{2,\mathcal{I}} &\geq \|\mathbf{c}_0 + \mathbf{h}_{\mathcal{I}_0}\|_{2,\mathcal{I}} + \|\mathbf{h}_{\mathcal{I}_0^c}\|_{2,\mathcal{I}} \\ &\geq \|\mathbf{c}_0\|_{2,\mathcal{I}} - \|\mathbf{h}_{\mathcal{I}_0}\|_{2,\mathcal{I}} + \|\mathbf{h}_{\mathcal{I}_0^c}\|_{2,\mathcal{I}},\end{aligned} \tag{35}$$

from which we conclude that

$$\|\mathbf{h}_{\mathcal{I}_0^c}\|_{2,\mathcal{I}} \leq \|\mathbf{h}_{\mathcal{I}_0}\|_{2,\mathcal{I}} \leq k^{1/2}\|\mathbf{h}_{\mathcal{I}_0}\|_2. \tag{36}$$

The last inequality follows from applying Cauchy-Schwarz to any block k-sparse vector $\mathbf{c}$:

$$\|\mathbf{c}\|_{2,\mathcal{I}} = \sum_{|i|=k} \|\mathbf{c}[i]\|_2 \cdot 1 \leq k^{1/2}\|\mathbf{c}\|_2. \tag{37}$$

Substituting (36) into (34):

$$\|\mathbf{h}_{(\mathcal{I}_0\cup\mathcal{I}_1)^c}\|_2 \leq \|\mathbf{h}_{\mathcal{I}_0}\|_2 \leq \|\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2, \tag{38}$$

which completes the first part of the proof.

Part II: $\|\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2 = 0$

We next show that $\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}$ must be equal to $\mathbf{0}$. In this part we invoke the RIP.

Since $\mathbf{D}\mathbf{c}_0 = \mathbf{D}\mathbf{c}' = \mathbf{y}$, we have $\mathbf{D}\mathbf{h} = \mathbf{0}$. Using the fact that $\mathbf{h} = \mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1} + \sum_{i\geq 2} \mathbf{h}_{\mathcal{I}_i}$,

$$\|\mathbf{D}\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2^2 = -\sum_{i=2}^{\ell-1} \langle \mathbf{D}(\mathbf{h}_{\mathcal{I}_0} + \mathbf{h}_{\mathcal{I}_1}), \mathbf{D}\mathbf{h}_{\mathcal{I}_i}\rangle. \tag{39}$$

From the parallelogram identity and the block-RIP it can be shown that

$$|\langle \mathbf{D}\mathbf{c}_1, \mathbf{D}\mathbf{c}_2\rangle| \leq \delta_{2k}\|\mathbf{c}_1\|_2\|\mathbf{c}_2\|_2, \tag{40}$$

for any two block k-sparse vectors with disjoint support. The proof is similar to [15, Lemma 2.1] for the standard RIP. Therefore,

$$|\langle \mathbf{D}\mathbf{h}_{\mathcal{I}_0}, \mathbf{D}\mathbf{h}_{\mathcal{I}_i}\rangle| \leq \delta_{2k}\|\mathbf{h}_{\mathcal{I}_0}\|_2\|\mathbf{h}_{\mathcal{I}_i}\|_2, \tag{41}$$

and similarly for $\langle \mathbf{D}\mathbf{h}_{\mathcal{I}_1}, \mathbf{D}\mathbf{h}_{\mathcal{I}_i}\rangle$. Substituting into (39),

$$\begin{aligned}\|\mathbf{D}\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2^2 &= \left|\sum_{i=2}^{\ell-1} \langle \mathbf{D}(\mathbf{h}_{\mathcal{I}_0} + \mathbf{h}_{\mathcal{I}_1}), \mathbf{D}\mathbf{h}_{\mathcal{I}_i}\rangle\right| \\ &\leq \sum_{i=2}^{\ell-1} (|\langle \mathbf{D}\mathbf{h}_{\mathcal{I}_0}, \mathbf{D}\mathbf{h}_{\mathcal{I}_i}\rangle| + |\langle \mathbf{D}\mathbf{h}_{\mathcal{I}_1}, \mathbf{D}\mathbf{h}_{\mathcal{I}_i}\rangle|) \\ &\leq \delta_{2k}(\|\mathbf{h}_{\mathcal{I}_0}\|_2 + \|\mathbf{h}_{\mathcal{I}_1}\|_2)\sum_{i=2}^{\ell-1} \|\mathbf{h}_{\mathcal{I}_i}\|_2.\end{aligned} \tag{42}$$

From the Cauchy-Schwarz inequality, any length-2 vector $\mathbf{a}$ satisfies $\mathbf{a}(1) + \mathbf{a}(2) \leq \sqrt{2}\|\mathbf{a}\|_2$. Therefore,

$$\|\mathbf{h}_{\mathcal{I}_0}\|_2 + \|\mathbf{h}_{\mathcal{I}_1}\|_2 \leq \sqrt{2}\sqrt{\|\mathbf{h}_{\mathcal{I}_0}\|_2^2 + \|\mathbf{h}_{\mathcal{I}_1}\|_2^2} = \sqrt{2}\|\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2, \tag{43}$$

where the last equality is a result of the fact that $\mathbf{h}_{\mathcal{I}_0}$ and $\mathbf{h}_{\mathcal{I}_1}$ have disjoint support. Substituting into (42) and using (32), (34) and (36),

$$\begin{aligned}\|\mathbf{D}\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2^2 &\overset{(32),(34)}{\leq} \sqrt{2}k^{-1/2}\delta_{2k}\|\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2\|\mathbf{h}_{\mathcal{I}_0^c}\|_{2,\mathcal{I}} \\ &\overset{(36)}{\leq} \sqrt{2}\delta_{2k}\|\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2\|\mathbf{h}_{\mathcal{I}_0}\|_2 \\ &\leq \sqrt{2}\delta_{2k}\|\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2^2,\end{aligned} \tag{44}$$

where the last inequality follows from $\|\mathbf{h}_{\mathcal{I}_0}\|_2 \leq \|\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2$. Combining (44) with the RIP (21) we have

$$(1-\delta_{2k})\|\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2^2 \leq \|\mathbf{D}\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2^2 \leq \sqrt{2}\delta_{2k}\|\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2^2. \tag{45}$$

Since $\delta_{2k} < \sqrt{2} - 1$, (45) can hold only if $\|\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2 = 0$, which completes the proof. ■

We conclude this subsection by pointing out more explicitly the differences between the proof of Theorem 1 and that of [15]. The main difference begins in (32); in our formulation each of the subvectors $\mathbf{h}_{\mathcal{I}_i}$ may have a different number of non-zero elements, while the equivalent equation in [15] (Eq. (10)) relies on the fact that the maximal number of non-zero elements in each of the subvectors is the same. This requires the use of several mixed-norms in our setting. The rest of the proof follows the spirit of [15] where in some of the

inequalities conventional norms are used, while in others the adaptation to our setting necessitates mixed norms.

B. Robust Recovery

We now treat the situation in which the observations are noisy, and the vector $\mathbf{c}_0$ is not exactly block-k sparse.

Specifically, suppose that the measurements (16) are corrupted by bounded noise so that

$$\mathbf{y} = \mathbf{D}\mathbf{c} + \mathbf{z}, \tag{46}$$

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

$$\begin{aligned} \min_{\mathbf{c}} \quad & \|\mathbf{c}\|_{2,\mathcal{I}} \\ \text{s.t.} \quad & \|\mathbf{y} - \mathbf{D}\mathbf{c}\|_2 \leq \epsilon. \end{aligned} \tag{47}$$

In addition, given a $\mathbf{c} \in \mathbb{R}^N$, we denote by $\mathbf{c}^k$ the best approximation of $\mathbf{c}$ by a vector with k non-zero blocks, so that $\mathbf{c}^k$ minimizes $\|\mathbf{c} - \mathbf{d}\|_{2,\mathcal{I}}$ over all block k-sparse vectors $\mathbf{d}$. Theorem 2 shows that even when $\mathbf{c}$ is not block k-sparse and the measurements are noisy, the best block-k approximation can be well approximated using (47).

Theorem 2: Let $\mathbf{y} = \mathbf{D}\mathbf{c}_0 + \mathbf{z}$ be noisy measurements of a vector $\mathbf{c}_0$. Let $\mathbf{c}^k$ denote the best block k-sparse approximation of $\mathbf{c}_0$, such that $\mathbf{c}^k$ is block k-sparse and minimizes $\|\mathbf{c}_0 - \mathbf{d}\|_{2,\mathcal{I}}$ over all block k-sparse vectors $\mathbf{d}$, and let $\mathbf{c}'$ be a solution to (47). If $\mathbf{D}$ satisfies the block RIP (21) with $\delta_{2k} < \sqrt{2} - 1$ then

$$\begin{aligned} \|\mathbf{c}_0 - \mathbf{c}'\|_2 \leq \frac{2(1-\delta_{2k})}{1-(1+\sqrt{2})\delta_{2k}} k^{-1/2} \|\mathbf{c}_0 - \mathbf{c}^k\|_{2,\mathcal{I}} \\ + \frac{4\sqrt{1+\delta_{2k}}}{1-(1+\sqrt{2})\delta_{2k}} \epsilon. \end{aligned} \tag{48}$$

Before proving the theorem, note that the first term in (48) is a result of the fact that $\mathbf{c}_0$ is not exactly k-block sparse. The second expression quantifies the recovery error due to the noise.

Proof: The proof is very similar to that of Theorem 1 with a few differences which we indicate. These changes follow the proof of [15, Theorem 1.3], with appropriate modifications to address the mixed norm.

Denote by $\mathbf{c}' = \mathbf{c}_0 + \mathbf{h}$ the solution to (47). Due to the noise and the fact that $\mathbf{c}_0$ is not block k-sparse, we will no longer obtain $\mathbf{h} = \mathbf{0}$. However, we will show that $\|\mathbf{h}\|_2$ is bounded. To this end, we begin as in the proof of Theorem 1 by using (30). In the first part of the proof we show that $\|\mathbf{h}_{(\mathcal{I}_0 \cup \mathcal{I}_1)^c}\|_2 \leq \|\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\|_2 + 2e_0$ where $e_0 = k^{-1/2}\|\mathbf{c}_0 - \mathbf{c}_{\mathcal{I}_0}\|_{2,\mathcal{I}}$ and $\mathbf{c}_{\mathcal{I}_0}$ is the restriction of $\mathbf{c}_0$ onto the k blocks corresponding to the largest ℓ_2 norm. Note that $\mathbf{c}_{\mathcal{I}_0} = \mathbf{c}^k$. In the second part, we develop a bound on $\|\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\|_2$.

Part I: Bound on $\|\mathbf{h}_{(\mathcal{I}_0 \cup \mathcal{I}_1)^c}\|_2$

We begin by decomposing $\mathbf{h}$ as in the proof of Theorem 1. The inequalities until (35) hold here as well. Instead of (35) we have

$$\begin{aligned} \|\mathbf{c}_0\|_{2,\mathcal{I}} &\geq \|\mathbf{c}_{\mathcal{I}_0} + \mathbf{h}_{\mathcal{I}_0}\|_{2,\mathcal{I}} + \|\mathbf{c}_{\mathcal{I}_0^c} + \mathbf{h}_{\mathcal{I}_0^c}\|_{2,\mathcal{I}} \\ &\geq \|\mathbf{c}_{\mathcal{I}_0}\|_{2,\mathcal{I}} - \|\mathbf{h}_{\mathcal{I}_0}\|_{2,\mathcal{I}} + \|\mathbf{h}_{\mathcal{I}_0^c}\|_{2,\mathcal{I}} - \|\mathbf{c}_{\mathcal{I}_0^c}\|_{2,\mathcal{I}}. \end{aligned} \tag{49}$$

Therefore,

$$\|\mathbf{h}_{\mathcal{I}_0^c}\|_{2,\mathcal{I}} \leq 2\|\mathbf{c}_{\mathcal{I}_0^c}\|_{2,\mathcal{I}} + \|\mathbf{h}_{\mathcal{I}_0}\|_{2,\mathcal{I}}, \tag{50}$$

where we used the fact that $\|\mathbf{c}_0\|_{2,\mathcal{I}} - \|\mathbf{c}_{\mathcal{I}_0}\|_{2,\mathcal{I}} = \|\mathbf{c}_{\mathcal{I}_0^c}\|_{2,\mathcal{I}}$. Combining (34), (37) and (50) we have

$$\|\mathbf{h}_{(\mathcal{I}_0 \cup \mathcal{I}_1)^c}\|_2 \leq \|\mathbf{h}_{\mathcal{I}_0}\|_2 + 2e_0 \leq \|\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\|_2 + 2e_0, \tag{51}$$

where $e_0 = k^{-1/2}\|\mathbf{c}_0 - \mathbf{c}_{\mathcal{I}_0}\|_{2,\mathcal{I}}$.

Part II: Bound on $\|\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\|_2$

Using the fact that $\mathbf{h} = \mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1} + \sum_{i \geq 2} \mathbf{h}_{\mathcal{I}_i}$ we have

$$\|\mathbf{D}\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\|_2^2 = \langle \mathbf{D}\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}, \mathbf{D}\mathbf{h} \rangle - \sum_{i=2}^{\ell-1} \langle \mathbf{D}(\mathbf{h}_{\mathcal{I}_0} + \mathbf{h}_{\mathcal{I}_1}), \mathbf{D}\mathbf{h}_{\mathcal{I}_i} \rangle. \tag{52}$$

From (21),

$$\begin{aligned} |\langle \mathbf{D}\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}, \mathbf{D}\mathbf{h} \rangle| &\leq \|\mathbf{D}\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\|_2 \|\mathbf{D}\mathbf{h}\|_2 \\ &\leq \sqrt{1+\delta_{2k}} \|\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\|_2 \|\mathbf{D}\mathbf{h}\|_2. \end{aligned} \tag{53}$$

Since both $\mathbf{c}'$ and $\mathbf{c}_0$ are feasible

$$\|\mathbf{D}\mathbf{h}\|_2 = \|\mathbf{D}(\mathbf{c}_0 - \mathbf{c}')\|_2 \leq \|\mathbf{D}\mathbf{c}_0 - \mathbf{y}\|_2 + \|\mathbf{D}\mathbf{c}' - \mathbf{y}\|_2 \leq 2\epsilon, \tag{54}$$

and (53) becomes

$$|\langle \mathbf{D}\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}, \mathbf{D}\mathbf{h} \rangle| \leq 2\epsilon\sqrt{1+\delta_{2k}} \|\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\|_2. \tag{55}$$

Substituting into (52),

$$\begin{aligned} \|\mathbf{D}\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\|_2^2 &\leq |\langle \mathbf{D}\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}, \mathbf{D}\mathbf{h} \rangle| \\ &\quad + \sum_{i=2}^{\ell-1} |\langle \mathbf{D}(\mathbf{h}_{\mathcal{I}_0} + \mathbf{h}_{\mathcal{I}_1}), \mathbf{D}\mathbf{h}_{\mathcal{I}_i} \rangle| \\ &\leq 2\epsilon\sqrt{1+\delta_{2k}} \|\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\|_2 \\ &\quad + \sum_{i=2}^{\ell-1} |\langle \mathbf{D}(\mathbf{h}_{\mathcal{I}_0} + \mathbf{h}_{\mathcal{I}_1}), \mathbf{D}\mathbf{h}_{\mathcal{I}_i} \rangle|. \end{aligned} \tag{56}$$

Combining with (42) and (44),

$$\begin{aligned} \|\mathbf{D}\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\|_2^2 \leq \\ \left(2\epsilon\sqrt{1+\delta_{2k}} + \sqrt{2}\delta_{2k} k^{-1/2} \|\mathbf{h}_{\mathcal{I}_0^c}\|_{2,\mathcal{I}}\right) \|\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\|_2. \end{aligned} \tag{57}$$

Using (37) and (50) we have the upper bound

$$\begin{aligned} \|\mathbf{D}\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\|_2^2 \leq \\ \left(2\epsilon\sqrt{1+\delta_{2k}} + \sqrt{2}\delta_{2k}(\|\mathbf{h}_{\mathcal{I}_0}\| + 2e_0)\right) \|\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\|_2. \end{aligned} \tag{58}$$

On the other hand, the RIP results in the lower bound

$$\|\mathbf{D}\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\|_2^2 \geq (1-\delta_{2k}) \|\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\|_2^2. \tag{59}$$

From (58) and (59),

$$(1-\delta_{2k}) \|\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\|_2 \leq 2\epsilon\sqrt{1+\delta_{2k}} + \sqrt{2}\delta_{2k}(\|\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\| + 2e_0), \tag{60}$$

or

$$\|\mathbf{h}_{\mathcal{I}_0 \cup \mathcal{I}_1}\|_2 \leq \frac{2\sqrt{1+\delta_{2k}}}{1-(1+\sqrt{2})\delta_{2k}} \epsilon + \frac{2\sqrt{2}\delta_{2k}}{1-(1+\sqrt{2})\delta_{2k}} e_0. \tag{61}$$

The condition $\delta_{2k} < \sqrt{2}-1$ ensures that the denominator in (61) is positive. Substituting (61) results in

$$\|\mathbf{h}\|_2 \leq \|\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2 + \|\mathbf{h}_{(\mathcal{I}_0\cup\mathcal{I}_1)^c}\|_2 \leq 2\|\mathbf{h}_{\mathcal{I}_0\cup\mathcal{I}_1}\|_2 + 2e_0, \tag{62}$$

which completes the proof of the theorem. ■

To summarize this section we have seen that as long as $\mathbf{D}$ satisfies the block-RIP (21) with a suitable constant, any block-k sparse vector can be perfectly recovered from its samples $\mathbf{y} = \mathbf{D}\mathbf{c}$ using the convex SOCP (26). This algorithm is stable in the sense that by slightly modifying it as in (47) it can tolerate noise in a way that ensures that the norm of the recovery error is bounded by the noise level. Furthermore, if $\mathbf{c}$ is not block k-sparse, then its best block-sparse approximation can be approached by solving the SOCP. These results are summarized in Table I. In the table, δ_{2k} refers to the block RIP constant.

TABLE I
COMPARISON OF ALGORITHMS FOR SIGNAL RECOVERY FROM $\mathbf{y} = \mathbf{D}\mathbf{c}_0 + \mathbf{z}$

	Algorithm (26)	Algorithm (47)
$\mathbf{c}_0$	block k-sparse	arbitrary
Noise $\mathbf{z}$	none ($\mathbf{z}=\mathbf{0}$)	bounded $\|\mathbf{z}\|_2 \leq \epsilon$
Condition on $\mathbf{D}$	$\delta_{2k} \leq \sqrt{2}-1$	$\delta_{2k} \leq \sqrt{2}-1$
Recovery $\mathbf{c}'$	$\mathbf{c}' = \mathbf{c}_0$	$\|\mathbf{c}_0 - \mathbf{c}'\|_2$ small; see (48)

C. Advantage of Block Sparsity

The standard sparsity model considered in CS assumes that $\mathbf{x}$ has at most k non-zero elements, however it does not impose any further structure. In particular, the non-zero components can appear anywhere in the vector. There are many practical scenarios in which the non-zero values are aligned to blocks, meaning they appear in regions, and are not arbitrarily spread throughout the vector. One example in the structured union of subspaces model we treat in this paper. Other examples are considered in [27].

Prior work on recovery of block-sparse vectors [26] assumed consecutive blocks of the same size. It was sown that in this case, when n, N go to infinity, the algorithm (26) will recover the true block-sparse vector with overwhelming probability. Their analysis is based on characterization of the null space of $\mathbf{D}$. In contrast, our approach relies on RIP which allows the derivation of uniqueness and equivalence conditions for finite dimensions and not only in the asymptotic regime. In addition, Theorem 2 considers the case of mismodeling and noisy observations while in [26] only the ideal noise-free setting is treated.

To demonstrate the advantage of our algorithm over standard basis pursuit (28), consider the matrix $\mathbf{D}$ of (22). In Section V, the standard and block RIP constants of $\mathbf{D}$ were calculated and it was shown that block RIP constants are smaller. This suggests that there are input vectors $\mathbf{x}$ for which the mixed ℓ_2/ℓ_1 method of (26) will be able to recover them exactly from measurements $\mathbf{y} = \mathbf{D}\mathbf{c}$ while standard ℓ_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 (23), finds the correct $\mathbf{x}$ while standard ℓ_1 minimization fails in this case; its output is $\hat{\mathbf{x}} = [-0.0289, 0, 0.9134, -1.0289, -1.0289, 0]$.

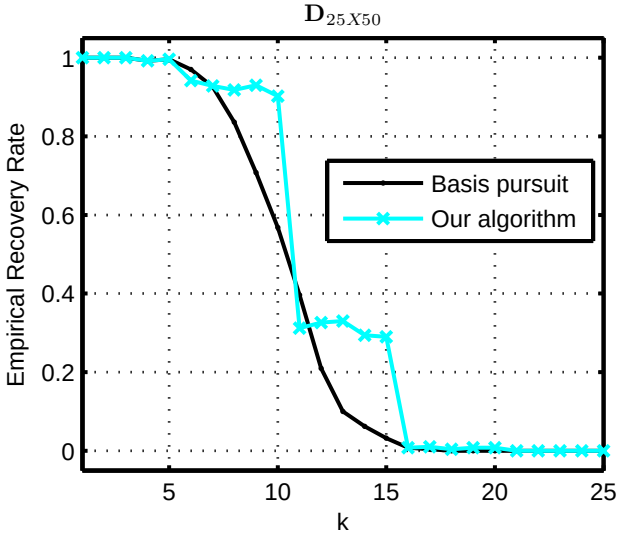

Fig. 2. Recovery rate of block-sparse signals using standard ℓ_1 minimization (basis pursuit) and the mixed ℓ_2/ℓ_1 algorithm.

We further compare the recovery performance of ℓ_1 minimization (28) and our algorithm (26) for an extensive set of random signals. In the experiment, we draw a matrix $\mathbf{D}$ of size 25×50 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.

VI. 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 if we choose the worst possible $\mathbf{X}$, improved performance over the single measurement setting can be guaranteed.

A. Equivalence Results

As we have seen in Section II, a special case of block sparsity is the MMV model, in which we are given a matrix of measurements $\mathbf{Y} = \mathbf{M}\mathbf{X}$ where $\mathbf{X}$ is an unknown $L \times d$ matrix that has at most k non-zero rows. Denoting by $\mathbf{c} = \mathrm{vec}(\mathbf{X}^T), \mathbf{y} = \mathrm{vec}(\mathbf{Y}^T)$, $\mathbf{D} = \mathbf{M}^T \otimes \mathbf{I}_d$ we can express the vector of measurements $\mathbf{y}$ as $\mathbf{y} = \mathbf{D}\mathbf{c}$ where $\mathbf{c}$ is a block sparse vector with consecutive blocks of length d. Therefore, the results of Theorems 1 and 2 can be specified to this problem.

Recovery algorithms for MMV using convex optimization programs were studied in [30], [32] and several greedy algorithms were proposed in [29], [31]. Specifically, in [29]–[32] the authors study a class of optimization programs, which we refer to as M-BP:

$$\text{M-BP}(\ell_q): \qquad \min \sum_{i=1}^{L} \|\mathbf{X}^i\|_q^p \quad \text{s.t. } \mathbf{Y} = \mathbf{M}\mathbf{X}, \tag{63}$$

where $\mathbf{X}^i$ is the ith row of $\mathbf{X}$. The choice $p = 1, q = \infty$ was considered in [32], while [30] treated the case of $p = 1$ and arbitrary q. Using $p \leq 1$ and $q = 2$ was suggested in [29], [43], leading to the iterative algorithm M-FOCUSS. For $p = 1, q = 2$, the program (63) has a global minimum which M-FOCUSS is proven to find. A nice comparison between these methods can be found in [32]. Equivalence for MMV algorithms based on RIP analysis does not appear in previous papers. The most detailed theoretical analysis can be found in [30] which establishes equivalence results based on mutual coherence. The results imply equivalence for (63) 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.

In our recent work [20], we suggested an alternative approach to solving MMV problems. Our strategy is based on merging the d measurement columns with random coefficients and in such a way transforming the multiple measurement problem into a single measurement counterpart. As proved in [20], this technique preserves the non-zero location set with probability one thus reducing computational complexity. Moreover, we showed that this method can be used to boost the empirical recovery rate by repeating the random merging several times.

Using the block-sparsity approach we can alternatively cast any MMV model as a single measurement vector problem by deterministically transforming the multiple measurement vectors into the single vector model $\mathrm{vec}(\mathbf{Y}^T) = (\mathbf{M} \otimes \mathbf{I}_d)\,\mathrm{vec}(\mathbf{X}^T)$. Here $\mathbf{c} = \mathrm{vec}(\mathbf{X}^T)$ is block-k sparse with consecutive blocks of length d. In contrast to [20] this does not reduce the number of unknowns so that the computational complexity of the resulting algorithm is on the same order as previous approaches, and also does not offer the opportunity for boosting. However, as we see in the next subsection, with an appropriate choice of measurement matrix this approach results in improved recovery capabilities.

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 (26) to the effective measurement vector $\mathbf{y}$ is the same as solving (63) with $p = 1, q = 2$. Thus the equivalence conditions we develop below relate to this program. Next, if $\mathbf{z} = \mathbf{D}\mathbf{c}$ where $\mathbf{c}$ 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 ith row is equal to $\mathbf{c}[i]$, and similarly for $\mathbf{Z}$. The block sparsity of $\mathbf{c}$ implies that $\mathbf{X}$ has at most $2k$ non-zero rows. The squared ℓ_2 norm $\|\mathbf{z}\|_2^2$ is equal to the squared ℓ_2 norm of the rows of $\mathbf{Z}$ which can be written as

$$\|\mathbf{z}\|_2^2 = \|\mathbf{Z}\|_F^2 = \mathrm{Tr}(\mathbf{Z}^T\mathbf{Z}), \tag{64}$$

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

$$(1-\delta_{2k})\,\mathrm{Tr}(\mathbf{X}^T\mathbf{X}) \leq \mathrm{Tr}(\mathbf{X}^T\mathbf{M}^T\mathbf{M}\mathbf{X}) \leq (1+\delta_{2k})\,\mathrm{Tr}(\mathbf{X}^T\mathbf{X}), \tag{65}$$

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

We now show that (65) is equivalent to the standard RIP condition

$$(1 - \delta_{2k})\|\mathbf{x}\|_2^2 \leq \|\mathbf{M}\mathbf{x}\|_2^2 \leq (1 + \delta_{2k})\|\mathbf{x}\|_2^2, \tag{66}$$

for any length L vector $\mathbf{x}$ that is $2k$-sparse. To see this, suppose first that (65) is satisfied for every matrix $\mathbf{X}$ with at most $2k$ non-zero rows and let $\mathbf{x}$ be an arbitrary $2k$-sparse vector. If we define $\mathbf{X}$ to be the matrix whose columns are all equal to $\mathbf{x}$, then $\mathbf{X}$ will have at most $2k$ non-zero rows and therefore satisfies (65). Since the columns of $\mathbf{X}$ are all equal, $\mathrm{Tr}(\mathbf{X}^T\mathbf{X}) = d\|\mathbf{x}\|_2^2$ and $\mathrm{Tr}(\mathbf{X}^T\mathbf{M}^T\mathbf{M}\mathbf{X}) = d\|\mathbf{M}\mathbf{x}\|_2^2$ so that (66) holds. Conversely, suppose that (66) is satisfied for all $2k$-sparse vectors $\mathbf{x}$ and let $\mathbf{X}$ be an arbitrary matrix with at most $2k$ non-zero rows. Denoting by $\mathbf{x}_j$ the columns of $\mathbf{X}$, each $\mathbf{x}_j$ is $2k$-sparse and therefore satisfies (66). Summing over all values j results in (65).

To summarize, if $\mathbf{M}$ satisfies the conventional RIP condition (66), then the algorithm (63) 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 ℓ_1 approach (28). 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 (63) 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.

B. Improved MMV Recovery

We have seen that the pessimistic equivalence results for MMV algorithms is a consequence of the fact that in the

worst-case scenario in which $\mathbf{x}_i = \mathbf{x}$, using a separable measurement strategy will render all observation vectors equal. In this subsection we 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.

One way to improve the analytical results is to consider an average case analysis instead of a worst-case approach. In [44] we show that if the unknown vectors $\mathbf{x}_i$ are generated randomly, then the performance improves with increasing number of measurement vectors. The advantage stems from the fact that the situation of equal vectors has zero probability and therefore does not affect the average performance. Here we take a different route which does not involve randomness in the unknown vectors, and leads to improved results even in the worst-case (namely without requiring an average analysis).

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 $\mathbf{y} = \mathbf{D}\mathbf{c}$ where $\mathbf{c} = \mathrm{vec}(\mathbf{X}^T)$ 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 (26) with a dense $\mathbf{D}$ to that of (63) with a block diagonal $\mathbf{D}$, we compute the empirical recovery rate of the methods in the same way performed in [20]. 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 ℓ 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 (63) for different choices of q and p, ReMBO [20] and our algorithm (26) 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 the performance advantage is due to the joint measurement process rather than a new recovery algorithm.

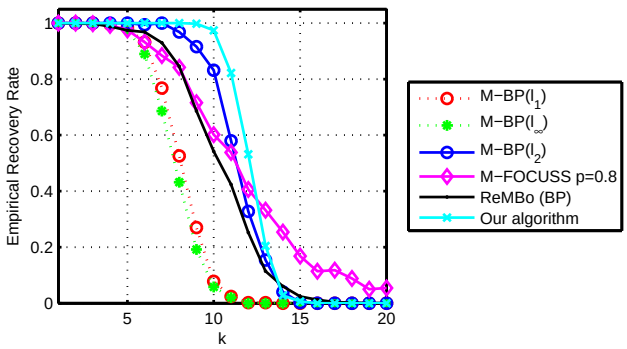

Fig. 3. Recovery rate for different number k of non-zero rows in $\mathbf{X}$. Each point on the graph represents an average recovery rate over 500 simulations.

VII. 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 $\mathbf{c}$. 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 $\mathbf{c}$. Our approach relies on results of [13], [28] developed for standard RIP, however, exploiting the block structure of $\mathbf{c}$ leads to a much faster decay rate.

Proposition 3: 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 (21) over $\mathcal{I} = \{d_1 = d, \ldots, d_m = d\}$, 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)

$$\mathrm{Prob}\left(\sqrt{1+\delta_{k|\mathcal{I}}} > 1 + (1+\epsilon) f(r)\right) \leq 2e^{-NH(r)\epsilon} \cdot e^{-m(d-1)H(r)}. \tag{67}$$

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 $0 < q < 1$.

The assumption that $d_i = d$ simplifies the calculations in the proof. Following the proof, we shortly address the more difficult case in which the blocks have varying lengths. We note that Proposition 3 reduces to the result of [13] 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 (67) is responsible for this behavior.

Proof: Let $\lambda = (1+\epsilon) f(r)$ and define

$$\bar{\sigma} = \max_{|T|=k,d} \sigma_{\max}(\mathbf{D}_T), \quad \underline{\sigma} = \min_{|T|=k,d} \sigma_{\min}(\mathbf{D}_T), \tag{68}$$

where $\sigma_{\max}(\mathbf{D}_T), \sigma_{\min}(\mathbf{D}_T)$, are the largest and the smallest singular values of $\mathbf{D}_T$, respectively. We use $|T| = k, d$ to denote a column subset of $\mathbf{D}$ consisting of k blocks of length d. For brevity we omit subscripts and denote $\delta = \delta_{k|\mathcal{I}}$. The inequalities in the definition of block-RIP (21) imply that

$$1 + \delta \geq \bar{\sigma}^2 \tag{69}$$

$$1 - \delta \leq \underline{\sigma}^2. \tag{70}$$

Since δ is the smallest number satisfying these inequalities we have that $1+\delta = \max(\bar{\sigma}^2, 2-\underline{\sigma}^2)$. Therefore,

$$\begin{aligned}
&\text{Prob}\left(\sqrt{1+\delta} > 1+\lambda\right) \\
&\quad = \text{Prob}\left(\sqrt{\max(\bar{\sigma}^2, 2-\underline{\sigma}^2)} > 1+\lambda\right) && (71)\\
&\quad \le \text{Prob}(\bar{\sigma} > 1+\lambda) + \text{Prob}(\sqrt{2-\underline{\sigma}^2} > 1+\lambda). && (72)
\end{aligned}$$

Noting that $\underline{\sigma} \ge 1-\lambda$ implies $\sqrt{2-\underline{\sigma}^2} \le 1+\lambda$ we conclude that

$$\text{Prob}\left(\sqrt{1+\delta} > 1+\lambda\right) \le \text{Prob}(\bar{\sigma} > 1+\lambda) + \text{Prob}(\underline{\sigma} < 1-\lambda). \quad (73)$$

We now bound each term in the right-hand-side of (73) using a result of Davidson and Szarek [45] regarding the concentration of the extreme singular values of a Gaussian matrix. It was proved in [45] that an $m \times n$ matrix $\mathbf{X}$ with $n \ge m$ satisfies

$$\text{Prob}(\sigma_{\max}(\mathbf{X}) > 1 + \sqrt{m/n} + t) \le e^{-nt^2/2} \quad (74)$$

$$\text{Prob}(\sigma_{\min}(\mathbf{X}) < 1 - \sqrt{m/n} - t) \le e^{-nt^2/2}. \quad (75)$$

Applying a union bound leads to

$$\begin{aligned}
&\text{Prob}\left(\bar{\sigma} > 1 + \sqrt{\frac{kd}{n}} + t\right) \\
&\quad \le \sum_{|T|=k,d} \text{Prob}\left(\sigma_{\max}(\mathbf{D}_T) > 1 + \sqrt{\frac{kd}{n}} + t\right) && (76)\\
&\quad \le \sum_{|T|=k,d} e^{-nt^2/2} && (77)\\
&\quad = \binom{m}{k} e^{-nt^2/2}. && (78)
\end{aligned}$$

Using the well-known bound on the binomial coefficient (for sufficiently large m)

$$\binom{m}{k} \le e^{mH(k/m)}, \quad (79)$$

we conclude that

$$\text{Prob}\left(\bar{\sigma} > 1 + \sqrt{\frac{kd}{n}} + t\right) \le e^{mH(k/m)} e^{-nt^2/2}. \quad (80)$$

To utilize this result in (73) we rearrange

$$\begin{aligned}
1+\lambda &= 1 + (1+\epsilon) f(r) && (81)\\
&= 1 + (1+\epsilon)\left(\sqrt{\frac{kd}{n}} + \sqrt{\frac{2N}{n}H(r)}\right) && (82)\\
&\ge 1 + \sqrt{\frac{kd}{n}} + \sqrt{(1+\epsilon)\frac{2N}{n}H(r)} && (83)
\end{aligned}$$

and obtain that

$$\begin{aligned}
&\text{Prob}\left(\bar{\sigma} > 1+\lambda\right) && (84)\\
&\quad \le \text{Prob}\left(\bar{\sigma} > 1 + \sqrt{\frac{kd}{n}} + \sqrt{(1+\epsilon)\frac{2N}{n}H(r)}\right).
\end{aligned}$$

Using (80) leads to

$$\begin{aligned}
\text{Prob}\left(\bar{\sigma} > 1+\lambda\right) &\le e^{mH(k/m)} e^{-\frac{n(1+\epsilon)2NH(r)}{2n}} && (85)\\
&= e^{NH(r) - m(d-1)H(r) - (1+\epsilon)NH(r)} && (86)\\
&\le e^{-NH(r)\epsilon} e^{-m(d-1)H(r)}. && (87)
\end{aligned}$$

Similar arguments are used to bound the second term in (73), completing the proof. ∎

The proof of Proposition 3 can be adapted to the case in which d_i are not equal. In this case, the notation $|T| = k, d$ is replaced by $|T| = k|\mathcal{I}$ and has the following meaning: T indicates a column subset of $\mathbf{D}$ consisting of k blocks from $\mathcal{I}$. Since $\mathcal{I}$ contains variable-length blocks, $|T|$ is not constant and depends on the particular column subset. Consequently, in order to apply the union bounds in (76) we need to consider the worst-case scenario corresponding to the maximal block length in $\mathcal{I}$. Proposition 3 thus holds for $d = \max(d_i)$. However, it is clear that the resulting probability bound will not be as stringent as in the case of equal $d_i = d$, especially when the ratio $\max(d_i)/\min(d_i)$ is large.

Proposition 3 holds as is for matrices $\mathbf{D}$ from the Bernoulli ensemble, namely $[\mathbf{D}]_{ik} = \pm\frac{1}{\sqrt{n}}$ with equal probability. In fact, the proposition is true for any ensemble for which the concentration of extreme singular values holds.

The following corollary emphasizes the asymptotic behavior of block-RIP constants per given number of samples.

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

$$\text{Prob}\left(\sqrt{1+\delta_{k|\mathcal{I}}} > 1 + (1+\epsilon)g(r)\right) \le 2e^{-mH(r)\epsilon}. \quad (88)$$

Proof: Let $\lambda = (1+\epsilon)g(r)$. The result then follows by replacing (81)-(83) with

$$1 + \lambda \ge 1 + \sqrt{\frac{kd}{n}} + \sqrt{(1+\epsilon)\frac{2N}{nd}H(r)}, \quad (89)$$

which leads to $\text{Prob}(\bar{\sigma} > 1+\lambda) \le e^{-mH(r)\epsilon}$. ∎

To evaluate the asymptotic behavior of block-RIP we note that for every $\epsilon > 0$ the right-hand side of (88) goes to zero when $N = md \to \infty$. Consequently, for fixed d

$$\delta_{k|\mathcal{I}} < \rho(r) \stackrel{\triangle}{=} -1 + [1 + g(r)]^2, \quad (90)$$

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.

Although Fig. 4 shows advantage for block-RIP, the absolute sparsity ratios predicted by the theory are pessimistic as also noted in [13], [28] in the case of $d = 1$. 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 plots the results and qualitatively affirms 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

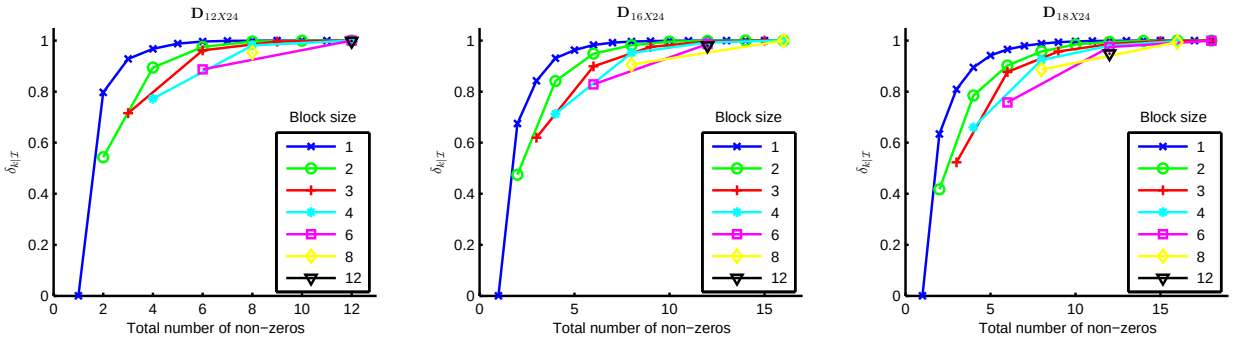

Fig. 5. The standard and block-RIP constants $\delta_{k|\mathcal{I}}$ for three different dimensions n, N. Each graph represent an average over 10 instances of random matrix $\mathbf{D}$. Each instance of $\mathbf{D}$ is scaled by a factor such that (18) is satisfied with $\alpha + \beta = 2$.

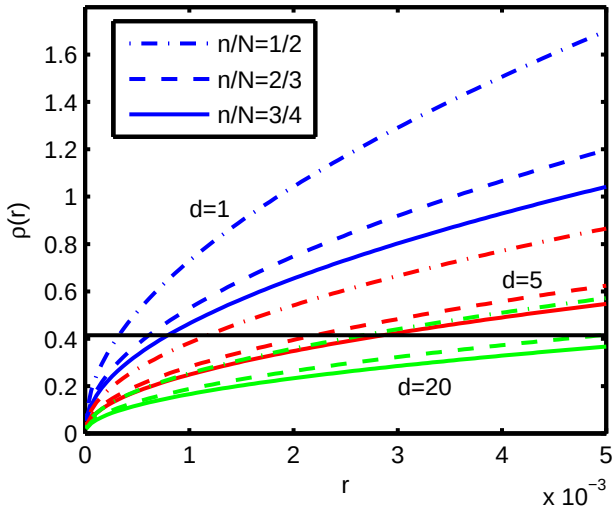

Fig. 4. The upper bound on $\delta_{k|\mathcal{I}}$ as a function of the sparsity ratio r, for three sampling rates n/N, and three block structures $d = 1, 5, 20$. The horizontal threshold is fixed on $\rho^* = \sqrt{2} - 1$ representing the threshold for equivalence derived in Theorem 1.

addressed in the following proposition, which quotes a result from [46] based on the proofs of [47]; we rephrase the result to match our notation.

Proposition 4 ([46, Theorem 3.3]): Consider the setting of Proposition 3, namely a random Gaussian matrix $\mathbf{D}$ of size $n \times N$ and block sparse signals over $\mathcal{I} = \{d_1 = d, \ldots, d_m = d\}$, where $N = md$ for some integer m. Let $t > 0$ and $0 < \delta < 1$ be constant numbers. If

$$n \geq \frac{36}{7\delta}\left(\ln(2L) + kd\ln\left(\frac{12}{\delta}\right) + t\right), \tag{91}$$

where $L = \binom{m}{k}$, then $\mathbf{D}$ satisfies the block-RIP (21) with restricted isometry constant $\delta_{k|\mathcal{I}} = \delta$, with probability at least $1 - e^{-t}$.

As observed in [46], the first term in (91) has the dominant impact on the required number of measurements in an asymptotic sense. This term quantifies the amount of measurements that are needed to code the exact subspace in which the sparse signal resides. Specifically, for block sparse signals

$$(m/k)^k \leq L = \binom{m}{k} \leq (e\,m/k)^k. \tag{92}$$

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. Block-sparsity reduces the total number of subspaces and therefore requires d times less measurements to code the signal subspace. The second term in (91) has a smaller contribution to the number of measurements. This term is proportional to kd, which is the number of nonzero values. Since the number of nonzeros is the same regardless of the sparsity structure, this term is not reduced in the block setting.

Note that Corollary 4 puts the emphasis on the required problem dimensions to satisfy a given RIP level. In contrast, Proposition 3 provides a tail bound on the expected isometry constant for given problem dimensions.

VIII. Conclusion

In this paper, we studied the problem of recovering an unknown signal x in an arbitrary Hilbert space $\mathcal{H}$, from a given set of n linear samples. The signal x is known to lie in a union of subspaces, so that $x \in \mathcal{V}_i$ where each of the subspaces $\mathcal{V}_i$ is a sum of k subspaces $\mathcal{A}_i$ chosen from an ensemble of m possibilities. While previous treatments of this model considered invertibility conditions, here we provide concrete recovery algorithms for a signal over a structured union.

We began by showing that recovering x can be reduced to a sparsity problem in which the goal is to recover a block-sparse vector $\mathbf{c}$ from measurements $\mathbf{y} = \mathbf{D}\mathbf{c}$ where the non-zero values in $\mathbf{c}$ are grouped into blocks. The measurement matrix $\mathbf{D}$ is equal to S^*A where S^* is the sampling operator and A is a set transformation corresponding to a basis for the sum of all $\mathcal{A}_i$. To determine $\mathbf{c}$ we suggested a mixed ℓ_2/ℓ_1 convex optimization program that takes on the form of an SOCP. Relying on the notion of block-RIP, we developed sufficient conditions under which $\mathbf{c}$ can be perfectly recovered using the proposed algorithm. We also proved that under the same conditions, the unknown $\mathbf{c}$ can be stably approximated in the presence of noise. Furthermore, if $\mathbf{c}$ is not exactly

block-sparse, then its best block-sparse approximation can be approached using the proposed method. We then showed that when $\mathbf{D}$ is chosen at random, the recovery conditions are satisfied with high probability.

Specializing the results to MMV systems, we proposed a new method for sampling in MMV problems. In this approach each measurement vector depends on all the unknown vectors. As we showed, this can lead to better recovery rate. Furthermore, we established equivalence results for a class of MMV algorithms based on RIP.

Throughout the paper, 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. Recovery methods for sparse signals in infinite dimensions have been addressed in some of our previous work [16]–[20]. In particular, we have shown that a signal lying in a union of shift-invariant subspaces can be recovered efficiently from certain sets of sampling functions. A first step in the direction of treating infinite unions of infinite subspaces is the example studied in [21] in which we treat an infinite union resulting from unknown time delays. 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 [48] that was posted online after the submission of this paper proposes a new framework called model-based compressive sensing (MCS). The MCS approach assumes a vector 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 propose generalizations of several known greedy algorithms. The union model developed in this paper is broader than the block-sparse setting treated in [48] 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 (6), (7). In addition, we aim at optimization-based recovery algorithms (26),(47) which require selecting the objective in order to promote the model properties.

REFERENCES

[1] C. E. Shannon, "Communications in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, Jan 1949.

[2] A. J. Jerri, "The Shannon sampling theorem–Its various extensions and applications: A tutorial review," *Proc. of the IEEE*, vol. 65, no. 11, pp. 1565–1596, Nov. 1977.

[3] H. Nyquist, "Certain topics in telegraph transmission theory," *EE Trans.*, vol. 47, pp. 617–644, Jan. 1928.

[4] M. Unser, "Sampling—50 years after Shannon," *IEEE Proc.*, vol. 88, pp. 569–587, Apr. 2000.

[5] 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.

[6] Y. C. Eldar and T. Michaeli, "Beyond bandlimited sampling: Nonlinearities, smoothness and sparsity," to appear in *IEEE Signal Proc. Magazine*.

[7] 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.

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

[9] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[10] J. Mosher, P. Lewis, and R. Leahy, "Coherence and MUSIC in biomagnetic source localization," in *Biomagnetism: Fundamental Research and Clinical Applications: Proceedings of the 9th International Conference on Biomagnetism*. IOS Press, 1995, p. 330.

[11] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM J. Scientific Computing*, vol. 20, pp. 33–61, 1999.

[12] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[13] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

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

[15] 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.

[16] 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.

[17] ——, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *arXiv 0902.4291;* submitted to *IEEE Selcted Topics on Signal Process.*, 2009.

[18] Y. C. Eldar, "Compressed sensing of analog signals in shit-invariant spaces," to appear in *IEEE Trans. on Signal Proc.*

[19] ——, "Uncertainty relations for analog signals," submitted to *IEEE Trans. on Information Theory*, 2008.

[20] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Signal Processing*, vol. 56, no. 10, pp. 4692–4702, Oct. 2008.

[21] K. Gedalyahu and Y. C. Eldar, "Low rate sampling schemes for time delay estimation," submitted to *IEEE Trans. on Signal Processing*.

[22] M. V. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Processing*, vol. 50, pp. 1417–1428, June 2002.

[23] P. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix," *IEEE Trans. Sig. Proc.*, vol. 55, no. 5, pp. 1741–1757, May 2007.

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

[25] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Trans. Inform. Theory*, to appear.

[26] 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.

[27] 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.

[28] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[29] 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 Processing*, vol. 53, no. 7, pp. 2477–2488, July 2005.

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

[31] 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.

[32] ——, "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.

[33] S. G. Mallat, "A theory of multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, pp. 674–693, 1989.
[34] I. Djokovic and P. P. Vaidyanathan, "Generalized sampling theorems in multiresolution subspaces," *IEEE Trans. Signal Processing*, vol. 45, pp. 583–599, Mar. 1997.
[35] 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.
[36] O. Christansen and Y. C. Eldar, "Oblique dual frames and shift-invariant spaces," *Applied and Computational Harmonic Analysis*, vol. 17, no. 1, 2004.
[37] Y. C. Eldar, "Sampling and reconstruction in arbitrary spaces and oblique dual frame vectors," *J. Fourier Analys. Appl.*, vol. 1, no. 9, pp. 77–96, Jan. 2003.
[38] ——, "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.
[39] Y. C. Eldar and O. Christansen, "Characterization of oblique dual frame pairs," *J. Applied Signal Processing*, pp. 1–11, 2006, article ID 92674.
[40] T. G. Dvorkind, Y. C. Eldar, and E. Matusiak, "Nonlinear and non-ideal sampling: Theory and methods," *IEEE Trans. Signal Processing*, to appear.
[41] T. Michaeli and Y. C. Eldar, *Convex Optimization in Signal Processing and Communications*. Cambridge Univ. Press, 2009, ch. Optimization Techniques in Modern Sampling Theory.
[42] Y. C. Eldar and H. Bölcskei, "Compressed sensing for block-sparse signals: Uncertainty relations, coherence, and efficient recovery," 2009, in preparation.
[43] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.
[44] Y. C. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," submitted to *IEEE Trans. on Inform. Theory*.
[45] S. J. Szarek, "Condition numbers of random matrices," *J. Complexity*, vol. 7, no. 2, pp. 131–149, 1991.
[46] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, 2009.
[47] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.
[48] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *Preprint*, 2008.

Moshe Mishali (S'07) received the B.Sc. degree in electrical engineering (summa cum laude) from the Technion – Israel Institute of Technology, Haifa, in 2000, where he is currently pursuing the Ph.D. degree in electrical engineering.

From 1996 to 2000, he was a member of the Technion Program for Exceptionally Gifted Students. Since 2006, he has been a Research Assistant and Project Supervisor with the Signal and Image Processing Lab, Electrical Engineering Department, Technion. His research interests include theoretical aspects of signal processing, compressed sensing, sampling theory, and information theory.

Mr. Mishali received the Hershel Rich Innovation Award in 2008.

Yonina C. Eldar (S'98–M'02–SM'07) Yonina C. Eldar received the B.Sc. degree in Physics in 1995 and the B.Sc. degree in Electrical Engineering in 1996 both from Tel-Aviv University (TAU), Tel-Aviv, Israel, and the Ph.D. degree in Electrical Engineering and Computer Science in 2001 from the Massachusetts Institute of Technology (MIT), Cambridge.

From January 2002 to July 2002 she was a Postdoctoral Fellow at the Digital Signal Processing Group at MIT. She is currently an Associate Professor in the Department of Electrical Engineering at the Technion - Israel Institute of Technology, Haifa, Israel. She is also a Research Affiliate with the Research Laboratory of Electronics at MIT. Her research interests are in the general areas of signal processing, statistical signal processing, and computational biology.

Dr. Eldar was in the program for outstanding students at TAU from 1992 to 1996. In 1998, she held the Rosenblith Fellowship for study in Electrical Engineering at MIT, and in 2000, she held an IBM Research Fellowship. From 2002-2005 she was a Horev Fellow of the Leaders in Science and Technology program at the Technion and an Alon Fellow. In 2004, she was awarded the Wolf Foundation Krill Prize for Excellence in Scientific Research, in 2005 the Andre and Bella Meyer Lectureship, in 2007 the Henry Taub Prize for Excellence in Research, and in 2008 the Hershel Rich Innovation Award, the Award for Women with Distinguished Contributions, and the Muriel & David Jacknow Award for Excellence in Teaching. She is a member of the IEEE Signal Processing Theory and Methods technical committee and the Bio Imaging Signal Processing technical committee, an Associate Editor for the IEEE Transactions on Signal Processing, the EURASIP Journal of Signal Processing, and the SIAM Journal on Matrix Analysis and Applications, and on the Editorial Board of Foundations and Trends in Signal Processing.