Theoretical Results on Sparse Representations of Multiple-Measurement Vectors

Jie Chen, *Student Member, IEEE*, and Xiaoming Huo, *Senior Member, IEEE*

***Abstract*—The sparse representation of a multiple-measurement vector (MMV) is a relatively new problem in sparse representation. Efficient methods have been proposed. Although many theoretical results that are available in a simple case—single-measurement vector (SMV)—the theoretical analysis regarding MMV is lacking. In this paper, some known results of SMV are generalized to MMV. Some of these new results take advantages of additional information in the formulation of MMV. We consider the uniqueness under both an ℓ_0-norm-like criterion and an ℓ_1-norm-like criterion. The consequent equivalence between the ℓ_0-norm approach and the ℓ_1-norm approach indicates a computationally efficient way of finding the sparsest representation in a redundant dictionary. For greedy algorithms, it is proven that under certain conditions, orthogonal matching pursuit (OMP) can find the sparsest representation of an MMV with computational efficiency, just like in SMV. Simulations show that the predictions made by the proved theorems tend to be very conservative; this is consistent with some recent advances in probabilistic analysis based on random matrix theory. The connections will be discussed.**

***Index Terms*—Basis pursuit, multiple-measurement vector (MMV), orthogonal matching pursuit (OMP), sparse representation.**

I. Introduction

THE problem of finding sparse representations of multiple-measurement vectors (MMV) in a redundant dictionary was motivated by a neuromagnetic inverse problem that arises in magnetoencephalography (MEG)—a modality of imaging the possible activation regions in the brain. We refer to Cotter *et al.* [1], [2] and a historic paper [3] for more details and other potential applications. The problem of MMV can also be considered as how to achieve sparse representations for single-measurement vectors (SMVs) *simultaneously* [4]–[6]. In this paper, we focus on the theoretical development of the MMV problem, instead of its applications.

Given a multiple-measurement vector B and a dictionary A, an MMV problem solves the system of equations

$$AX = B$$

where $A \in \mathbb{R}^{m\times n}$, $X \in \mathbb{R}^{n\times L}$, and $B \in \mathbb{R}^{m\times L}$. Each column of the matrix A is associated with an *atom*. A set of all atoms is called a *dictionary* (see Mallat's book [7]), which is denoted by Ω. A *sparse representation* means that matrix X (or a vector, if one has an SMV: $L = 1$) has a small number of rows that contain nonzero entries. A mathematical definition of the *sparsity* of a matrix X will be provided later.

A redundant dictionary simply means that $m < n$. Usually, we have $m \ll n$ and $L < m$. As we mentioned earlier, when $L = 1$, we have the case of an SMV. Matrices X and B can be rewritten as $X = [x^{(1)}, x^{(2)}, \ldots, x^{(L)}]$, $B = [b^{(1)}, b^{(2)}, \ldots, b^{(L)}]$, where $x^{(l)}$'s and $b^{(l)}$'s, $1 \le l \le L$, are column vectors. Evidently, the system of equations $AX = B$ can be rewritten as $Ax^{(l)} = b^{(l)}$, where $l = 1, \ldots, L$. For simplicity, we assume that the columns of A have been normalized; hence all the diagonal entries of the Gram matrix $G = A^T A$ are equal to ones and all the off-diagonal entries are in the interval $[-1, 1]$.

In the case of SMV, there are abundant results on the sparsest representations in a redundant dictionary. We refer to [8]–[14]. The introduction of Donoho, Elad, and Temlyakov [15] gives a comprehensive depiction on many important applications. In MMV, we replace x and b by the upper-case letters, X and B, emphasizing that they are *matrices* instead of *column vectors*.

In SMV, the sparsity of a representation is defined as the ℓ_0 quasi-norm of the vector x, which is denoted by $\|x\|_0$. The quantity $\|x\|_0$ is simply the number of nonzero elements in the vector x. Without loss of accuracy, for simplicity, throughout this paper, we will call the quantity $\|x\|_0$ an ℓ_0 norm, instead of an ℓ_0-quasi-norm; similarly, we will say an ℓ_0-norm-like criterion, instead of an ℓ_0-quasi-norm-like criterion. The sparsest representation in SMV is the solution to the following optimization problem:

$$(\mathbf{Q0}): \quad \min \|x\|_0, \quad \text{subject to } Ax = b.$$

The above problem can be convexified as a minimizing-the-ℓ_1-norm problem

$$(\mathbf{Q1}): \quad \min \|x\|_1, \quad \text{subject to } Ax = b$$

where $\|x\|_1$ is the sum of the absolute values of the entries of vector x, i.e., for $x = [x_1, x_2, \ldots, x_n]^T$, we have $\|x\|_1 = \sum_{i=1}^{n} |x_i|$. Readers may compare the objective functions in **(Q0)** and **(Q1)**. Note that **(Q1)** can be solved via *linear programming*.

The problem **(Q0)** is essentially a combinatorial optimization problem, which in general is difficult to solve. We hope that the solution to problem **(Q1)** is, in some situations, close enough to the solution to **(Q0)**. The equivalence of the solutions between

Manuscript received October 11, 2004; revised November 30, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Helmut Boelcskei. This work has been partially supported by National Science Foundation Grants DMS 0140587 and DMS 03-46307. Part of this work was done when Dr. Huo was a visiting Fellow at the Institute of Pure and Applied Mathematics, Los Angeles, CA, in fall 2004.

The authors are with the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205 USA (e-mail: jchen@isye.gatech.edu; xiaoming@isye.gatech.edu).

Digital Object Identifier 10.1109/TSP.2006.881263

(Q0) and **(Q1)** has been proved under various conditions, and the more recent work was done by many researchers, including Donoho and Elad [12], Tropp [16], and Fuchs [14]. Evidently, the equivalence between the two solutions is very important in computing the sparsest representation in SMV. In this paper, we extend the corresponding theorems from SMV to MMV.

Another way to obtain a sparse representation is through a greedy algorithm, e.g., orthogonal matching pursuit (OMP). It has been proved by Donoho, Elad, and Temlyakov [15] and Tropp [16] independently that under certain conditions, the OMP can find the sparsest representation of the signal. In this paper, we extend this theory to MMV as well.

In the present paper, we consider a *noiseless* case: an SMV b or an MMV B is a linear combination of atoms without noise, i.e., $b = Ax$ or $B = AX$. It will be a different mathematical problem when additive noise is considered in the formulation. For *noisy* cases, we refer to [15]–[17] for results in SMV and [5], [6] for results in MMV.

In our generalization from SMV to MMV, it is shown that the generalization can be very broad: the inner vector norm can be any vector norm in a Euclidean space. Moreover, in the case of minimizing the ℓ_0-norm, less stringent requirements to guarantee uniqueness can be derived, compared with those for SMV.

The rest of the paper is organized as follows. Section II describes the uniqueness of the solutions to the minimizing-the-ℓ_0 -norm problems. Section III describes the situations in which the solutions to the minimizing-the-ℓ_1-norm problems are identical with the solutions to the minimizing-the-ℓ_0-norm problems. Section IV describes the property of the sparsest representations that are computed from a greedy algorithm—the OMP. Conditions under which OMP gives the sparsest representations are given. Section V describes some simulations, which indicate that the theoretical bounds are conservative. Section VI gives the discussion on related works, possible extensions, and future research topics. Section VII makes some concluding remarks.

II. Minimizing the ℓ_0 Norm

A. Formulation

We describe our formulation of MMV. The following quantity is the number of rows (in a matrix X) that contain nonzero entry(ies):

$$\mathcal{R}(X) = \|(m(x_i))_{n\times 1}\|_0$$

where $x_i \in \mathbb{R}^L$ is the transpose of the ith row of the matrix X, i.e., $X = [x_1, x_2, \ldots, x_n]^T$, $m(\cdot)$ is any vector norm in $\mathbb{R}^L$, and vector $(m(x_i))_{n\times 1}$ has the ith entry equal to $m(x_i)$, $1 \le i \le n$. Symbol $\mathcal{R}$ stands for a sparsity *rank*. A noiseless sparse representation problem in MMV can be written as

$$(\mathbf{P0}): \quad \min \mathcal{R}(X), \quad \text{subject to } AX = B.$$

Readers can compare this with **(Q0)**. In fact, if $L = 1$, the above optimization problem becomes **(Q0)**.

In general, solving **(P0)** requires enumerating all the subsets of set $\{1, 2, \ldots, n\}$. The complexity of such a subset-search algorithm grows *exponentially* with the dictionary size n.

B. Uniqueness in ℓ_0-Norm Minimization

We restrict our attention to the case when the solution to **(P0)** is unique. It is provable that having sufficient sparsity is a sufficient condition for the solution (i.e., representation) to be the unique sparsest one. We give some conditions under which the solution to the problem **(P0)** is unique. This is a necessary preparation for a subsequent result, i.e., equivalence of solutions between the ℓ_0-norm minimization problem and the ℓ_1-norm minimization problem.

The following generalizes the result of Donoho and Elad [12] to MMV. We start with the concept of Spark [18].

Definition 2.1 (Spark): Given a matrix A, the quantity Spark, which is denoted by $\mathrm{Spark}(A)$ (or σ), is the smallest possible integer such that there exist σ columns of matrix A that are linearly dependent.

In [12], $\mathrm{Spark}(A)/2$ is a threshold of the sparsity: if the signal is made by less than $\mathrm{Spark}(A)/2$ atoms, or in other words, if the signal is a linear combination of less than $\mathrm{Spark}(A)/2$ columns of matrix A, then the solution to **(P0)** is exactly the atoms that are included in this linear combination. For MMV, with the above mentioned $\mathcal{R}(\cdot)$, we can draw the following conclusion. It is interesting that the result holds for any vector norm $m(\cdot)$.

Theorem 2.2: Matrix X will be the unique solution of the problem **(P0)**, if $B = AX$ and

$$\mathcal{R}(X) < \frac{\mathrm{Spark}(A)}{2}.$$

Compared with the SMV cases, the above theorem has the same upper bound.

Remark 2.3: From the known results in SMV, Theorem 2.2 can be proved as a direct extension. The argument is as follows. If $\mathcal{R}(X) < \mathrm{Spark}(A)/2$, obviously $\|X^{(j)}\|_0 < \mathrm{Spark}(A)/2$, $1 \le j \le L$, where $X^{(j)}$ is the jth column of matrix X. Hence, the solution to the following optimization problem:

$$\min \|Y^{(j)}\|_0, \text{subject to } AY^{(j)} = B^{(j)},\ j = 1, 2, \ldots, L$$

where $Y \in \mathbb{R}^{n\times L}$ and $Y^{(j)}$ is the jth column of Y, should give exactly the jth column of matrix X; Recall $B^{(j)}$ is the jth column of matrix B. This renders the fact that X is the unique solution to **(P0)**.

Readers can easily derive a rigorous proof by following the above idea. Next, we present a different proof, which we think is more straightforward.

Proof: Suppose matrices X_1 and $X_2 \in \mathbb{R}^{n\times L}$ are the solutions to **(P0)** with property $\max\{\mathcal{R}(X_1), \mathcal{R}(X_2)\} < \mathrm{Spark}(A)/2$. We have

$$\mathcal{R}(X_1 - X_2) \le \mathcal{R}(X_1) + \mathcal{R}(X_2) < \mathrm{Spark}(A). \tag{2.1}$$

On the other hand, because $0 = A(X_1 - X_2)$, if we consider $(X_1 - X_2)^{(1)}$, which is the first column of matrix $X_1 - X_2$, we have $0 = A(X_1 - X_2)^{(1)}$. It leads to $\|(X_1 - X_2)^{(1)}\|_0 \ge \mathrm{Spark}(A)$. Therefore, we have $\mathcal{R}(X_1 - X_2) \ge \mathrm{Spark}(A)$, which contradicts (2.1). This contradiction proves the theorem. ■

If we are willing to consider the additional feature of matrix B—to take advantage of the MMV formulation—a more general condition can be derived. A precedent is Lemma 1 in Cotter *et al.* [1]. The following result is more general.

Theorem 2.4: Let $\mathrm{Rank}(B)$ denote the rank of matrix B. Apparently $\mathrm{Rank}(B) \leq L$. Matrix X will be the unique solution to the problem **(P0)**, if $B = AX$ and

$$\mathcal{R}(X) < \frac{[\mathrm{Spark}(A) - 1 + \mathrm{Rank}(B)]}{2}.$$

Proof: Recall $B \in \mathbb{R}^{m \times L}$. Suppose we have $B = AX_1 = AX_2$, where $X_1, X_2 \in \mathbb{R}^{n \times L}$, and $X_1 \neq X_2$. Let $d(\mathrm{Null}(B))$ denote the dimension of the (right) null space of matrix B: $\{x : Bx = 0\}$. Similarly, $d(\mathrm{Null}(X_1))$ and $d(\mathrm{Null}(X_2))$ denote the dimensions of the (right) null spaces of matrices X_1 and X_2. We have

$$d(\mathrm{Null}(X_1)) \leq d(\mathrm{Null}(B))$$

and

$$d(\mathrm{Null}(X_2)) \leq d(\mathrm{Null}(B)).$$

Recall $\mathrm{Rank}(B)$ denotes the rank of matrix B. Similarly let $\mathrm{Rank}(X_1)$ and $\mathrm{Rank}(X_2)$ denote the ranks of matrices X_1 and X_2. We have

$$\mathrm{Rank}(X_1) \geq \mathrm{Rank}(B) \tag{2.2}$$

and

$$\mathrm{Rank}(X_2) \geq \mathrm{Rank}(B). \tag{2.3}$$

Consider a matrix $[A_1, A_{12}, A_2]$, where submatrix $[A_1, A_{12}]$ is made by the columns of matrix A that correspond to the nonzero rows of matrix X_1 and submatrix $[A_{12}, A_2]$ is made by the columns of matrix A that correspond to the nonzero rows of matrix X_2. Let r_1 and r_2 denote the numbers of nonzero rows in matrices X_1 and X_2, respectively. Note that matrix A_{12} corresponds to columns where matrices X_1 and X_2 have nonzero rows simultaneously. Let r_{12} denote the number of columns of matrix A_{12}. Matrix X_{11} (respectively, Matrix X_{22}) consists of the nonzero rows of matrix X_1 (respectively, X_2) corresponding to A_1 (respectively, A_2). Matrix X_{12} (respectively, Matrix X_{21}) consists of the nonzero rows of X_1 (respectively, X_2) corresponding to the columns in A_{12}. We have

$$B = [A_1, A_{12}] \cdot \begin{bmatrix} X_{11} \\ X_{12} \end{bmatrix} = [A_{12}, A_2] \cdot \begin{bmatrix} X_{21} \\ X_{22} \end{bmatrix}$$

and

$$0 = A(X_1 - X_2) = [A_1, A_{12}, A_2] \cdot \begin{bmatrix} X_{11} \\ X_{12} - X_{21} \\ -X_{22} \end{bmatrix}. \tag{2.4}$$

From (2.4), we have

$$d(\mathrm{Null}([A_1, A_{12}, A_2])) \geq \mathrm{Rank}\left(\begin{bmatrix} X_{11} \\ X_{12} - X_{21} \\ -X_{22} \end{bmatrix} \right). \tag{2.5}$$

It is easy to see

$$\mathrm{Rank}\left(\begin{bmatrix} X_{11} \\ X_{12} - X_{21} \\ -X_{22} \end{bmatrix} \right) \geq \max\{\mathrm{Rank}(X_{11}), \mathrm{Rank}(X_{22})\}. \tag{2.6}$$

Without loss of generality, we consider X_{11} only. It is easy to see that

$$d(\mathrm{Null}(X_{11})) \leq d(\mathrm{Null}(X_1)) + r_{12}. \tag{2.7}$$

The above is true because if we consider two systems of linear equations: for variable y, $X_1 \cdot y = 0$ or $X_{11} \cdot y = 0$; the former has r_{12} more constrains, so its solution space(the null space of matrix X_1) is at most reduced by r_{12} dimensions, which is (2.7). Inequality (2.7) immediately leads to

$$\mathrm{Rank}(X_1) - r_{12} \leq \mathrm{Rank}(X_{11}). \tag{2.8}$$

Combining (2.5), (2.6), (2.8), and one of (2.2) and (2.3), we have

$$d(\mathrm{Null}([A_1, A_{12}, A_2])) \geq \mathrm{Rank}(B) - r_{12}. \tag{2.9}$$

By the definition of Spark, we have

$$\begin{aligned} \mathrm{Rank}([A_1, A_{12}, A_2]) &\geq \mathrm{Spark}([A_1, A_{12}, A_2]) - 1 \\ &\geq \mathrm{Spark}(A) - 1. \end{aligned} \tag{2.10}$$

Combining all the above, we have

$$\begin{aligned} r_1 + r_2 - r_{12} &= \#Cols([A_1, A_{12}, A_2]) \\ &= \mathrm{Rank}([A_1, A_{12}, A_2]) \\ &\quad + d(\mathrm{Null}([A_1, A_{12}, A_2])) \\ &\geq \mathrm{Spark}(A) - 1 + \mathrm{Rank}(B) - r_{12}. \end{aligned}$$

Hence, $r_1 + r_2 \geq \mathrm{Spark}(A) - 1 + \mathrm{Rank}(B)$. The last inequality is based on (2.9) and (2.10). It is easy to see that the above proves the theorem. ■

It turns out that to study the theoretical property of the ℓ_0-norm approach, we only need the fact that $m(x) = 0$ if and only if $x = \vec{0}$, where $\vec{0}$ is an all zero vector in $\mathbb{R}^L$. It is evident that Theorem 2.2 is a special case of Theorem 2.4, as $\mathrm{Rank}(B) \geq 1$.

For the upper bound of Theorem 2.4, it is conceptually interesting to ask whether the rank of matrix B can be replaced with the rank of matrix X. Because $B = AX$, it is evident that $\mathrm{Null}(X) \subset \mathrm{Null}(B)$. Hence, $d(\mathrm{Null}(X)) \leq d(\mathrm{Null}(B))$. Therefore, $\mathrm{Rank}(X) \geq \mathrm{Rank}(B)$. Given this, it is not clear how to utilize the existing approach to generate an upper bound that is based on $\mathrm{Rank}(X)$. Further exploitation in this direction will be a future research topic.

C. Mutual Incoherence and $\mu_{1/2}(G)$

A difficulty associated with an upper bound with $\mathrm{Spark}(A)$ is that the quantity Spark is hard to calculate, as pointed out by Donoho and Elad [12]. Up to now, there is no good algorithm to compute $\mathrm{Spark}(A)$ besides enumerating all the possible subsets. For practical use, we introduce other quantities: *mutual incoherence* and $\mu_{1/2}(G)$. These quantities have appeared in

previous papers, e.g., [9], [10], and [12]. They provide upper bounds that are lower than the one that is built on Spark. However, these quantities are easy to compute.

Definition 2.5: Mutual incoherence (denoted by M) is the maximum absolute inner product between two column vectors of matrix A, i.e.,

$$M = M(A) = \max_{1\le i,j\le n, i\neq j} |G(i,j)|$$

where $G(i,j)$ is the (i,j)th entry of the Gram matrix G: $G = A^T A$.

Note that quantities M and Spark have the following relation, which has been proved in Donoho and Elad [12, Theorem 7]:

$$\mathrm{Spark}(A) \ge \left(1 + \frac{1}{M}\right).$$

Therefore, an upper bound with $\mathrm{Spark}(A)$ is better. In fact, the above inequality together with Theorem 2.4 gives a one-line proof of the following corollary. We omit the proof.

Corollary 2.6: If $B = AX$ and

$$\mathcal{R}(X) < \frac{(M^{-1} + \mathrm{Rank}(B))}{2}$$

then matrix X is the unique solution to the problem **(P0)**.

We consider another quantity.

Definition 2.7: For a Gram matrix G, which is symmetric, let $\mu_{1/2}(G)$ denote the smallest number m, such that the sum of a collection of m off-diagonal magnitudes in a single row or column of the Gram matrix G is at least 1/2.

In [12, Theorem 6 and sec. 4.2], we can find the following relation: $\mathrm{Spark}(A) \ge 2\mu_{1/2}(G) + 1$. Combining with Theorem 2.4, we immediately have the following.

Corollary 2.8: If $B = AX$ and

$$\mathcal{R}(X) < \mu_{1/2}(G) + \frac{\mathrm{Rank}(B)}{2}$$

then matrix X is the unique solution to the problem **(P0)**.

We conclude our analysis of the uniqueness in the minimizing-the-ℓ_0-norm approach.

III. Minimizing the ℓ_1 Norm

A. Formulation

Recall that we have defined a sparsity rank of matrix $X \in \mathbb{R}^{n\times L}$

$$\mathcal{R}(X) = \|(m(x_i))_{n\times 1}\|_0$$

where $m(x_i)$ is a vector norm in $\mathbb{R}^L$. In this section, we consider a relaxation to the above quantity.

We consider the following function as a relaxation of the quantity $\mathcal{R}(X)$:

$$\mathrm{Relax}(X) = \|(m(x_i))_{n\times 1}\|_1.$$

Note that the only difference between $\mathcal{R}(X)$ and $\mathrm{Relax}(X)$ is that the outside ℓ_0 norm is replaced by an ℓ_1 norm. The corresponding optimization problem becomes

$$(\mathbf{P1}): \quad \min \quad \mathrm{Relax}(X), \quad \text{subject to } B = AX.$$

The above formulation includes many known works. For example, in Tropp [6], $m(\cdot)$ is the ℓ_∞ norm; in Malioutov *et al.* [19], $m(\cdot)$ is the ℓ_2 norm.

Besides the ℓ_1 norm, other functions of X have been proposed as objective functions. In the pioneer works on MMV [1], [20], [21], the following diversity measure on sparsity was proposed:

$$J^{(p,q)}(x) = \sum_{i=1}^{n} (\|x^{(i)}\|_q)^p, \quad 0 \le p \le 1, q \ge 1$$

where p and q are parameters, and vector $x^{(i)}$ is the ith row of matrix X. The norm of a row is given by $\|x^{(i)}\|_q = (\sum_{j=1}^{L} |x_{ij}|^q)^{1/q}$. An algorithm, which was named M-FOCUSS, is proposed to minimize the above objective [1]. The M-FOCUSS, for $q = 2$, $p \le 1$ is an iterative algorithm that uses the idea of Lagrange multipliers. A disadvantage of the above objective function is that it could have more than one local minima, e.g., when $p < 1$. An iterative algorithm could be trapped by a local minimum. With $p = 1$ in the above objective, we obtain the ℓ_1-norm minimization problem **(P1)** with ℓ_q norm inside.

B. Uniqueness Under the ℓ_1 Norm

We consider an optimal solution to the problem **(P1)**. Let $B = AX^*$, where X^* is the optimal solution to the problem **(P0)**. Let S be an index set that contains the rows of X^* where $m(x_i^*) > 0$. Here x_i^* denotes the ith row of matrix X^*. Let A_S denote a matrix that is made by the columns of A with indexes from S. We can write $B = A_S X_S^*$, where matrix X_S^* is made by the nonzero rows of X^*. Without loss of generality, we can assume that A_S is of full column rank; otherwise, the number of nonzero rows of X^* can be reduced, which contradicts the optimality. We define the generalized inverse of A_S to be $A_S^+ = (A_S^T A_S)^{-1} A_S^T$. Based on the fact that A_S is of full column rank, the generalized inverse is well defined. We present a sufficient condition of the sparsity of X^* in the following.

Theorem 3.1: A sufficient condition for X^* to be the unique solution to **(P1)** is that

$$\|A_S^+ A_j\|_1 < 1, \forall j \notin S. \tag{3.11}$$

Note that the above is the exact recovery condition in Tropp's [16]. (See also [14].) It turns out that it is also a sufficient condition for the uniqueness under the ℓ_1 norm for MMV, *with an arbitrary inner vector norm* $m(\cdot)$. Readers may want to revisit the formulation of **(P1)**.

As a preparation for the proof of theorem 3.1, the following is a well-known result for norms in the Euclidean space. We present it without a proof.

Proposition 3.2: For a linear combination $\sum_{i=1}^{k} c_i x_i$, where $c_i \in \mathbb{R}$, $x_i \in \mathbb{R}^L$, and k is an integer, for any norm $m(\cdot)$ in $\mathbb{R}^L$, we have

$$m\left(\sum_{i=1}^{k} c_i x_i\right) \le \sum_{i=1}^{k} |c_i| \cdot m(x_i).$$

Proof of Theorem 3.1: Suppose there are two representations: $B = A_S X_S^* = A_{S'} Y_{S'}$, where $S \neq S'$ and set S' includes

the indexes of the nonzero rows of the matrix $Y \in \mathbb{R}^{n\times L}$. We only need to show that

$$\text{Relax}(X^*) < \text{Relax}(Y). \tag{3.12}$$

Recall

$$\text{Relax}(X^*) = \|(m(x_i^*))_{n\times 1}\|_1 = \sum_{i=1}^{n} m(x_i^*) = \sum_{i\in S} m(x_i^*).$$

Because $X_S^* = (A_S^+ A_{S'})Y_{S'}$, we have

$$x_i^* = \sum_k (A_S^+ A_{S'})_{ik}(Y_{S'})_k$$

where $(A_S^+ A_{S'})_{ik}$ is the (i,k)th entry of the matrix $A_S^+ A_{S'}$, and $(Y_{S'})_k$ is the kth row of $Y_{S'}$. Note that the above is a linear combination. From Proposition 3.2, we have

$$m(x_i^*) \le \sum_k |(A_S^+ A_{S'})_{ik}| m((Y_{S'})_k).$$

Taking $\sum_i$ on both sides, we have

$$\sum_i m(x_i^*) \le \sum_i \sum_k |(A_S^+ A_{S'})_{ik}| m((Y_{S'})_k) < \sum_k m(Y_k).$$

The last inequality is based on two facts (see Acknowledgment). The first fact is that S' contains at least one column that does not appear in S; otherwise, S' would be a strict subset of S, which contradicts the minimality of S. Therefore, there must exist some k, such that $\sum_i |(A_S^+ A_{S'})_{ik}| < 1$, based on (3.11). The other fact is that $\|A_S^+ A_j\|_1 \le 1$ for every column j in matrix A. Hence, we prove (3.12). ■

C. Equivalence

In [16, Theorem B and Corollary 3.6], we know whenever one of the following conditions is satisfied:

$$\mathcal{R}(X^*) < \frac{\left(1+\frac{1}{M}\right)}{2} \tag{3.13}$$

or

$$\mathcal{R}(X^*) < \mu_{1/2}(G) \tag{3.14}$$

$\max_{j\notin S} \|A_S^+ A_j\|_1 < 1$ holds for any signal with $\mathcal{R}(X^*)$ atoms in its optimal representation. Therefore, according to Theorem 3.1, when (3.13) or (3.14) holds, X^* is the unique solution to **(P1)**.

On the other hand, according to [12], we have the following relation: $\text{Spark}(A)/2 > \mu_{1/2}(G) \ge 1/2M$. Thus, according to Theorem 2.2, if $B = AX$ and $\mathcal{R}(X) < (1+1/M)/2$ or $\mathcal{R}(X) < \mu_{1/2}(G)$, X is the unique sparest solution to **(P0)**, i.e., $X = X^*$.

From all the above, we have the following theorem.

Theorem 3.3 (Equivalence): For a dictionary A with Gram matrix $G = A^T A$. If $AX = B$ and

$$\mathcal{R}(X) < \frac{\left(1+\frac{1}{M}\right)}{2}$$

or

$$\mathcal{R}(X) < \mu_{1/2}(G)$$

then matrix X is the unique solution to **(P1)**, and this solution is identical with the solution to **(P0)**.

Note that our condition of equivalence in the above theorem is identical with the one in SMV. Recall that by taking into account of the property of matrix B, a stronger uniqueness condition is achieved in the ℓ_0-like norm. The difficulty in getting a stronger equivalence condition for MMV is that the uniqueness of the ℓ_1-norm approach does not seem to depend on the matrices B or X.

It is interesting to realize that the proof of SMV still works for any norm $m(\cdot)$ in $\mathbb{R}^L$.

D. Comparison Between SMV and MMV

In the minimizing-the-ℓ_0-norm problem, by taking advantage of the formulation of MMV, we can raise the upper bound in the uniqueness condition from $\text{Spark}(A)/2$ to $[\text{Spark}(A) - 1 + \text{Rank}(B)]/2$.

There is no evidence that between condition $\mathcal{R}(x) < \text{Spark}(A)/2$ and condition $\max_{j\notin S} \|A_S^+ A_j\|_1 < 1$, one is able to dominate the other. In principle, if $\max_{j\notin S} \|A_S^+ A_j\|_1 < 1$ and $\text{Spark}(A)/2 < \mathcal{R}(x) < [\text{Spark}(A) - 1 + \text{Rank}(B)]/2$, we can claim the equivalence between ℓ_0 norm and ℓ_1 norm for MMV, and this is not achievable by simply concatenating SMV problems.

Here is another difference between an MMV problem and an SMV problem. Note that if we find the sparsest representation for B under the condition in Theorem 3.1, we do *not* have enough evidence that each column of X^* can be obtained by solving an SMV problem for each column of B. The reason is that from $\max_{j\notin S} \|A_S^+ A_j\|_1 < 1$, where S consists of the atoms in the optimal representation of matrix B, it is not necessary to have $\max_{j\notin S_i} \|A_{S_i}^+ A_j\|_1 < 1$, where S_i consists of the atoms in the optimal representation of vector $b^{(i)}$. This is because the number of the atoms in the optimal representation of vector $b^{(i)}$ may be less than the number of the atoms in the optimal representation of matrix B. In summary, the uniqueness conditions under the ℓ_1 norm differ between in the formulation of an MMV and in the formulation of a combination of several SMVs. (The description here is speculative and consequently raises an open question: What can we say about the relationship between the set S and the group of sets $\{S_i\}$? We leave this question to be studied in the future.)

IV. Orthogonal Matching Pursuit

Matching pursuit (MP) [22] is proposed as an efficient numerical method to decompose a signal. As an improvement of MP, OMP [23], [24] has been introduced. OMP overcomes some drawbacks of MP. Unfortunately, counterexamples show that both methods could be trapped by the initial selection of a "bad" atom (see Chen *et al.* [25]). For the MMV problem, many variants of OMP have been proposed. A subset of them are [1], [4], [5], [26]–[29].

For this section, we propose our OMP with an $\ell_q (q \ge 1)$ norm of the inner product. Note that in MMV, the inner product becomes a vector. A condition that guarantees the exact recovery of OMP is derived. This condition is identical to the corresponding exact recovery condition in SMV (see [16]). Again,

it is interesting to see that an existing condition holds for a large class of vector norms.

A. OMP Algorithm for MMV

An OMP in MMV, which is denoted by OMPMMV, works as follows.

Orthogonal Matching Pursuit for MMV (OMPMMV)

1) Initialization: residual $R_0 = B$ and subset $S_0 = \emptyset$.
2) At the tth iteration:
   a) choose the atom a_{k_t}, which satisfies $a_{k_t} = \arg\max_{a_k} ||z_k||_q$, where $z_k = R_{t-1}^T a_k$ and $q \geq 1$;
   b) let $S_t = [S_{t-1}, a_{k_t}]$, and $X^* = \arg\min_X ||S_t X - B||_F^2$, $y_t = S_t X^*$;
   c) Set $R_t = B - y_t$.

Readers can find that, except taking the ℓ_q norm of the vector z_k in Step 2a), the remaining components in the above algorithm are standard in an OMP.

In [4] and [5], Tropp *et al.* proposed ℓ_1 norm in Step 2a). In [26]–[29], ℓ_2 and ℓ_∞ are proposed for weak matching pursuit and weak matching pursuit for the MMV problem. In [1], ℓ_2 norm is applied. We will prove that, when the coefficient matrix of B is very sparse, no matter what the ℓ_q norm is, an OMP with the ℓ_q norm in Step 2a) can recover the sparsest representation.

B. Matrix Norm Preparation

Before providing the proof, we introduce some necessary notations and results that will be used in this section.

Definition 4.1: The (p, q) matrix (or operator) norm of A is defined as

$$||A||_{p,q} = \max_{x \neq 0} \frac{||Ax||_q}{||x||_p} = \max_{||x||_p = 1} ||Ax||_q.$$

Several of the (p, q) matrix norms can be computed easily (see also [5] and [30]).

Lemma 4.2: Consider matrix A.

1) The $(1, q)$ matrix norm is the maximum ℓ_q norm of the columns of A.
2) The (2, 2) matrix norm is the maximum singular value of A.
3) The (p, ∞) norm is the maximum $\ell_{p'}$ norm of the rows of A, where $1/p + 1/p' = 1$.

The following property regarding (p, q) matrix norm can be easily derived from the definitions or results mentioned above.

Lemma 4.3: For matrix A, we have

1) $||Ax||_q \leq ||A||_{p,q} \cdot ||x||_p$, and
2) $||A^T||_{\infty,\infty} = ||A||_{1,1}$.

In particular, the (p, ∞) matrix norm has the following property.

Lemma 4.4: For matrices A and B, and $p > 0$, we have

$$||AB||_{p,\infty} \leq ||A||_{\infty,\infty} ||B||_{p,\infty}.$$

Proof: The following are direct applications of some previous results:

$$\begin{aligned} ||AB||_{p,\infty} &= \max_{||x||_p=1} ||A(Bx)||_\infty \\ &\leq \max_{||x||_p=1} ||A||_{\infty,\infty} ||(Bx)||_\infty \\ &= ||A||_{\infty,\infty} \max_{||x||_p=1} ||(Bx)||_\infty \\ &= ||A||_{\infty,\infty} ||B||_{p,\infty}. \end{aligned}$$

We prove the lemma. ■

C. Main Result

Note that OMP never chooses the same atom twice because the residual is to the atoms that have already been selected. If at each step, OMP selects the atoms in the optimal representation, after $\mathcal{R}(X^*)$ steps, the residual must become zero, and the algorithm stops. Note that since we only consider the *noiseless* formulation, we are allowed to use such an idealistic argument.

According to our notation, in Step 2a) in OMPMMV, we have $\max_{a_k} ||z_k||_q = \max_{a_k} ||a_k^T R_{t-1}||_q = ||A^T R_{t-1}||_{p,\infty}$, where $1/p + 1/q = 1$. Thus, at $(t+1)$th step, we can select the atom in the optimal representation if and only if $||A_S^T R_t||_{p,\infty} > ||A_{\bar{S}}^T R_t||_{p,\infty}$, where $\bar{S}$ is the complement of S in the dictionary Ω. Following this idea, we have the following theorem with the same notations used in the previous sections.

Theorem 4.5: A sufficient condition for OMPMMV to recover a representation of matrix B associated with atom indexes S is

$$\max_{j \notin S} ||A_S^+ A_j||_1 < 1.$$

Readers can see that the above is again the Exact Recovery Condition in [16]. In fact, readers can see that the following proof is modified from the corresponding proof in [16].

Proof: At each iteration t

$$\begin{aligned} \rho_t &= \frac{||A_S^T R_t||_{p,\infty}}{||A_{\bar{S}}^T R_t||_{p,\infty}} \\ &= \frac{||A_S^T R_t||_{p,\infty}}{||A_{\bar{S}}^T (A_S^+)^T A_S^T R_t||_{p,\infty}} \\ &\geq \frac{1}{||A_{\bar{S}}^T (A_S^+)^T||_{\infty,\infty}}. \end{aligned}$$

In the above, we use the equality

$$R_t = (A_S^+)^T A_S^T R_t. \tag{4.15}$$

Recall $(A_S^+)^T A_S^T = A_S (A_S^T A_S)^{-1} A_S^T$, which is a projection matrix to the subspace spanned by the columns of matrix A_S. Because A_S is the optimal set, the columns of R_t is in the subspace spanned by the columns of A_S. Hence, we have (4.15).

To pick the atom in the optimal representation, we want $\rho_t > 1$, that is $||A_{\bar{S}}^T (A_S^+)^T||_{\infty,\infty} < 1$. Moreover, we have

$$||A_{\bar{S}}^T (A_S^+)^T||_{\infty,\infty} = ||(A_S^+) A_{\bar{S}}||_{1,1} = \max_{j \in \bar{S}} ||A_S^+ A_j||_1 < 1.$$

This completes the proof. □

Applying the same argument that has been used in the ℓ_1 norm, we have the following corollary.

Corollary 4.6: If $AX = B$ and

$$\mathcal{R}(X) < \frac{\left(1 + \frac{1}{M}\right)}{2}$$

or

$$\mathcal{R}(X) < \mu_{1/2}(G)$$

matrix X is the unique sparsest solution to **(P0)**, and OMPMMV can recover this representation exactly.

Compared with Theorem 4.5, the conditions in the above corollary are much easier to check. We can calculate matrix X through OMPMMV first, and then check if such a matrix X satisfies the conditions.

V. Simulation

A. Exact Recovery of OMPMMV and (P1)

Simulations are conducted to bring insights on when OMPMMV and **(P1)** can exactly find the *original* signal. Two experiments are conducted. In the first experiment, matrix $A \in \mathbb{R}^{m\times n}$ has dimensions $m = 20$ and $n = 30$. We set $L = 5$ and $m(\cdot) = \ell_1$. The entries of matrix A are independently sampled from the standard normal $N(0, 1)$. We compute the multiple-measurement vector, B, using $B = AX_0$, where the N nonzero rows of matrix $X_0 \in \mathbb{R}^{30\times 5}$ are randomly chosen, and the values of the nonzero entries of matrix X_0 are assigned again by independently sampling from the standard normal distribution. The value of N is ranged from 1 to $\lceil (1 + 1/M)/2 \rceil + 15$. For each generated pairs of matrices B and A, matrix X is solved via both OMPMMV and **(P1)**. The solution X is compared with the original matrix X_0. If $X \equiv X_0$, an 'exact recovery' is obtained. The above simulation is executed for 1000 times for the same matrix A. The proportion of exact recoveries among 1000 times of simulation via OMPMMV (respectively, **(P1)**) for the N nonzero rows of matrix X_0 is reported as "the empirical probability of exact recovery" for the value N via OMPMMV (respectively, **(P1)**) in Fig. 1(a). We observed that the OMPMMV performs slightly better. Symbol $*$ indicates where the theoretical upper bound for uniqueness is (i.e., $\lfloor (1 + 1/M)/2 \rfloor$). In the figures, it is not very evident where the above proportions are equal to 1—the proportions of exact recoveries are very close to 1, but not 1. We introduce two symbols to indicate the positions when the proportion are exactly equal to 1: symbol $\oplus$ indicates the largest value of N while OMPMMV finds the original X_0 among all simulations; symbol $\ominus$ indicates the largest value of N while the solutions of **(P1)** are identical with matrix X_0 for all simulations.

In the second experiment, matrix A is generated by concatenating two orthonormal bases: $[I, H]$, where matrix I is an identity matrix and matrix H is a Hadamard matrix. We choose $m = 16$ and $n = 32$. Matrix X_0 has $L = 3$ columns. All the other settings are the same as in the first experiment. Again, we observe that the OMPMMV performs slightly better.

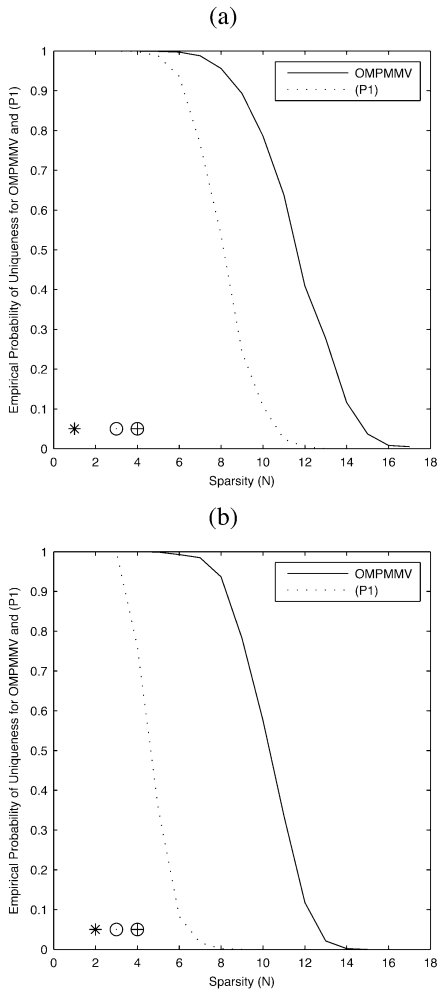


Fig. 1. (a) First experiment of exact recovery, in which $A \in \mathbb{R}^{m\times n}$, $X_0 \in \mathbb{R}^{n\times L}$, $m = 20, n = 30, L = 5$, where entries of matrices A and X_0 are independently sampled from $N(0, 1)$. Symbol $*$ is marked at 1. For the OMPMMV, the $\oplus$ is marked at $N = 4$; while for (P1), $\ominus$ is marked at $N = 3$. (b) We now have matrix $A = [I, H]$ where matrix $A \in \mathbb{R}^{16\times 32}$ and submatrix H is a 16×16 Hadamard matrix. Matrix I is a 16×16 identity matrix. Matrix X_0 is chosen in the same way, with L equal to 3 and N being the number of nonzero rows. In this case, symbol $*$ is marked at $N = 2$. Symbols $\oplus$ and $\ominus$ are at $N = 4$ and $N = 3$, respectively.

In both cases, we observe that the exact recovery can occur when the value of N is above the theoretical threshold ($\lfloor (1 + 1/M)/2 \rfloor$) that is given in this paper. Based on this, we say that the theoretical upper bound is pessimistic.

B. Comparison of Different Vector Norms in (P1)

The settings in the subsection are the same as those in the last subsection. In this subsection, we do the simulation for the ℓ_1 norm method with different $m(\cdot)$ norms. First, for matrix A randomly generated from $\mathrm{Normal}(0, 1)$, where $m = 30$, $n = 20$, $L = 5$, we do the simulation for $m(\cdot) = \ell_1$ and $m(\cdot) = \ell_\infty$, respectively, and draw their empirical probabilities of exact recoveries on one plot. [See the results in Fig. 2(a).] Second, for $A = [I, H]$, where submatrix H is a 16×16 Hadamard matrix, matrix I is a 16×16 identity matrix, and $L = 3$, we do the same simulation as above. [See the results in Fig. 2(b).]

From the simulations, we see that the curves in each plot are similar. This might imply that for the method of ℓ_1 norm , there is no significant difference among different inside vector norms.

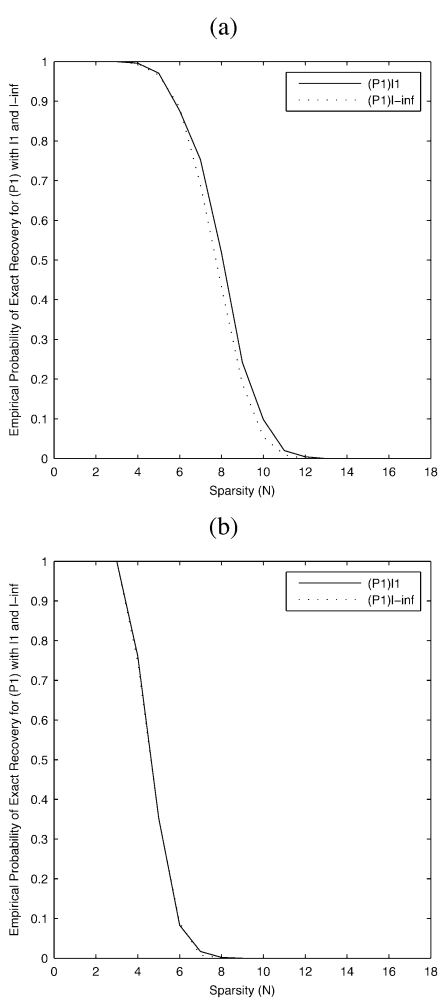


Fig. 2. (a) Consider the case $A \in \mathbb{R}^{m\times n}$, $X_0 \in \mathbb{R}^{n\times L}$, $m = 20, n = 30$, $L = 5$, where entries of matrices A and X_0 are independently sampled from $N(0,1)$. The theoretical upper bound for the equivalence is 1. Let $N_i, i = 1, \infty$ denote the largest value of N when the solutions of (P1) with $m(\cdot)$ being the ℓ_i norm are identical with matrix X_0 among all of the 1000 simulations. We have $N_1 = N_\infty = 3$. (b) We now consider matrix $A = [I, H]$, where submatrix H is a 16×16 Hadamard matrix and submatrix I is a 16×16 identity matrix. We have $L = 3$. The theoretical upper bound for equivalence is 2. We obtain $N_1 = N_\infty = 3$.

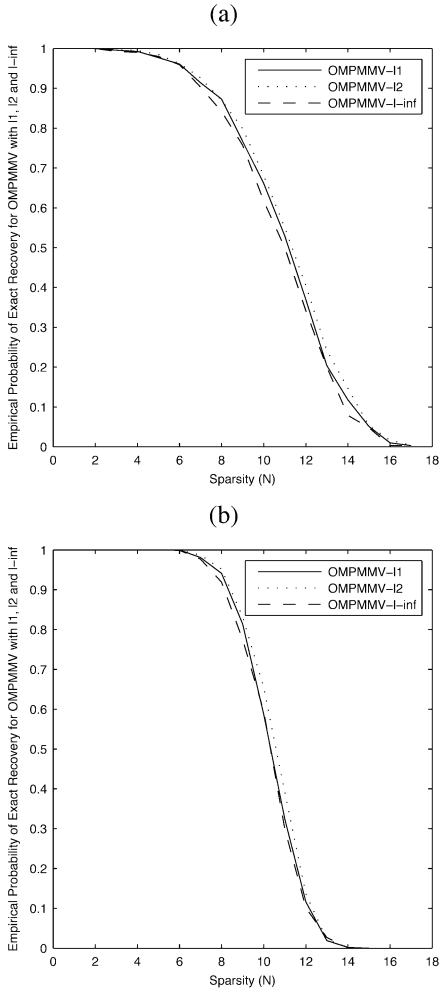


Fig. 3. (a) We consider $A \in \mathbb{R}^{m\times n}$, $X_0 \in \mathbb{R}^{n\times L}$, $m = 20, n = 30, L = 5$, where entries of matrices A and X_0 are independently sampled from $N(0,1)$. The theoretical upper bound for equivalence is 1. Notation N_i, $i = 1, 2, \infty$, denotes the largest value of N while OMPMMV with ℓ_i norm in Step 2a) finds the original X_0 among all the 1000 trials. We have $N_1 = N_2 = N_\infty = 2$. (b) We have matrix $A = [I, H]$, where submatrix H is a 16×16 Hadamard matrix and submatrix I is a 16×16 identity matrix. We have $L = 3$. The theoretical upper bound for equivalence is 2. We obtain $N_1 = N_2 = 6$ and $N_\infty = 5$.

C. Comparison of Different Vector Norms in OMPMMV

The settings in this subsection are the same as those in the previous subsections. In this subsection, we do the simulation for the OMPMMV method with different ℓ_q norms. First, for matrix A whose entries are randomly generated from $\mathrm{Normal}(0,1)$, where $m = 30$, $n = 20$, $L = 5$, we do the simulation for ℓ_1, ℓ_2 and ℓ_∞, respectively, and draw their empirical probabilities of exact recoveries on one plot. [See the results in Fig. 3(a).] Second, for $A = [I, H]$, where submatrix H is a 16×16 Hadamard matrix, I is a 16×16 identity matrix, and $L = 3$, we do the same simulation as above. [Results are shown in Fig. 3(b).]

From the simulations, we see that the curves in each plot are similar. This might demonstrate that for the same method (ℓ_1 norm or OMP), among different vector norms, there is no significant difference.

VI. Discussion

A. Better Vector Norms in MMV?

In both sparsity rank $\mathcal{R}(X)$ and its relaxation $\mathrm{Relax}(X)$ of matrix $X \in \mathbb{R}^{n\times L}$, we choose an arbitrary norm in $\mathbb{R}^L$. One logic question is whether or not one norm can consistently outperform another norm. To be more specific, we introduce the following dominance concept.

Definition 6.1 (Dominance): We say that a norm $m_1(x)$ is dominated by a norm $m_2(x)$ in $\mathbb{R}^L$, if and only if for any $x, y \in \mathbb{R}^L$, $m_1(x) < m_1(y)$ leads to $m_2(x) < m_2(y)$.

If norm m_2 dominates norm m_1, then m_2 should always be used. The reason is as following. Denote two relaxations $\mathrm{Relax}_1(X) = \sum_{i=1}^n m_1(x_i)$ and $\mathrm{Relax}_2(X) = \sum_{i=1}^n m_2(x_i)$. Whenever **(P1)** with relaxation $\mathrm{Relax}_1(X)$ finds the unique sparest solution, i.e., $\mathrm{Relax}_1(X^*) < \mathrm{Relax}_1(Y)$ for any other Y, **(P1)** with relaxation $\mathrm{Relax}_2(X)$ finds the unique sparsest solution too, i.e., $\mathrm{Relax}_2(X^*) < \mathrm{Relax}_2(Y)$ for any other Y.

For norms in a Euclidean space, the following result demonstrates that no norm can dominate another. The only special case is that they are equivalent.

Lemma 6.2: If norm m_2 dominates norm m_1, then there exists a constant $C > 0$, such that $m_1(x) = C \cdot m_2(x)$, $\forall x \in \mathbb{R}^L$.

Proof: First of all, for any pair $x, y \in \mathbb{R}^L$ satisfying $m_1(x) = m_1(y)$, we prove that $m_2(x) = m_2(y)$. This can be seen from the following. It is easy to see that for $\tau > 0$

$$m_1((1-\tau)y < m_1(x) < m_1((1+\tau)y).$$

From the dominance, we have

$$m_2((1-\tau)y < m_2(x) < m_2((1+\tau)y).$$

Let $\tau \to 0$, we have $m_2(x) = m_2(y)$.

In the second step, we choose a special $x_0 \in \mathbb{R}^L$, such that $m_1(x_0) = 1$. Because

$$m_1\left(\frac{1}{m_1(y)}y\right) = 1 = m_1(x_0)$$

we have

$$m_2\left(\frac{1}{m_1(y)}y\right) = m_2(x_0).$$

Hence

$$m_2(y) = m_1(y) \cdot m_2(x_0).$$

Note that $m_2(x_0)$ is a constant, we have proved the lemma. □

Note our notion of "equivalent" differs from some uses in the literature, e.g., [31, page 269]. More specifically, the result such as Theorem 5.4.4 in [31] is typically associated with the equivalence. The foregoing lemma is similar to some description in the [31, sec. 5.5]; however, we failed to find a direct reference.

The above demonstrates that there is no optimal relaxation when SMV is generalized to MMV. Note that we consider the optimality in the worst case. If we know some properties about X or B, some norms may work better than other norms in function Relax(X), e.g., on statistical average. We leave it as an open question.

B. Simulation

In the simulation, we verify the criterion of "exact recovery," instead of the sparsest representation as formulated in **(P0)**. *exact recovery* in many applications is a more interesting problem. On the other hand, this approach seems to be adopted by most publications in the field—perhaps due to the numerical difficulty to verify the most sparsity.

C. Other Numerical Approaches

The work of Couvreur and Bresler [32] on *backward elimination* and related analysis has strong similarity with some of the results that we developed here for MMV.

Short papers [33], [34] proposed various heuristics to achieve sparse representations . They give a flavor on algorithms that have been adopted in signal processing.

D. Probability, Random Matrices

Recently, in the case of SMV, some very inspiring new results are obtained. Recall in SMV, we have $b = Ax_0$, where $A \in \mathbb{R}^{m\times n}$, $x_0 \in \mathbb{R}^n$, and $b \in \mathbb{R}^m$, $m < n$. Donoho in [35] shows that when $\|x_0\|_0 = O(n)$ and the matrix A is random, with the probability nearly equal to 1, the minimizing ℓ_1 norm approach (i.e., **(Q1)**) gives the solution being equal to x_0.

In general, the upper bounds that are given in this paper are lower than $O(n)$. The cases that are considered here are the *worst cases*. It is shown that these worst-case results are extremely *conservative.*

A similar result regarding noisy data was reported in [36]. At the same time, Candès gave several talks with similar results, based on his joint work with Tao and Romberg [41].

There are interesting developments in the random matrix. Recall that the mutual incoherence M has been used in several upper bounds of underlying sparsity, for both uniqueness and equivalence. Roughly, the upper bounds are $\sim M^{-1}/2$. Historically, it is of particular interest to study the case when matrix A is a concatenation of two square matrices $A = [O_1, O_2]$, where matrices O_1, O_2 are orthogonal. Apparently, the mutual incoherence is the maximum magnitude of the entries of matrix $O_1^T O_2$. Jiang in [37] derived the asymptotic distribution of this quantity. Basically, if $O_1 \in \mathbb{R}^{n\times n}$, he proved that $M^{-1}/2$ is *almost surely* between $(1/2\sqrt{6})\sqrt{n/\log n}$ and $(1/4)\sqrt{n/\log n}$.

In another work of Jiang [38], the limit distribution of the maximal off-diagonal entry in a correlation matrix was derived. It can have similar applications as the above result in analyzing the behavior of $M^{-1}/2$ in other scenarios.

We would like to point out that the worst-case analysis (which eventually produces $M^{-1}/2$) is not powerful enough to produce the probabilistic results that are stated at the beginning of this subsection.

E. Related Publications

Some preliminary results in this paper were reported in a conference paper [39] and a manuscript [40]. The latter was downloadable online. This paper is an extensive revision of [40].

VII. Conclusion

We showed that most of the results on sparse representations of simple measurement vectors can be generalized to the case of multiple-measurement vectors. Our generalization is broad: the inside norm $m(\cdot)$ in **(P0)** and **(P1)** can be *any* vector norm.

When additional information is available in multiple-measurement vectors, better upper bounds for uniqueness in **(P0)** (and hopefully for equivalence, referring to our discussion) become possible. An incarnation of this is Theorem 2.4.

We showed that a greedy algorithm—OMP—under certain conditions, can achieve the sparsest representation, just like the result in SMV. We realize that the generalization can be achieved in a broad sense; more specifically, the inner vector norm in the Step 2a) of OMPMMV can be ℓ_q norm for any $q \geq 1$.

These results provide useful insights in designing numerical solutions to find the sparse representations of multiple-measurement vectors.

Acknowledgment

The authors would like to thank one of the reviewers for giving the two facts noted in the paper and making the inequality more clear.

REFERENCES

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

[2] B. Rao, K. Engan, and S. Cotter, "Diversity measure minimization based method for computing sparse solutions to linear inverse problems with multiple measurement vectors," presented at the Int. Conf. Acoustics, Speech, Signal Process. (ICASSP), Montreal, QC, Canada, May 2004.

[3] I. F. Gorodnitsky, J. S. George, and B. D. Rao, "Neuromagnetic source imaging with FOCUSS: A recursive weighted minimum norm algorithm," *Electroencephalogr. Clin. Neurophysiol.*, vol. 95, no. 4, pp. 231–251, Oct. 1995.

[4] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Simultaneous sparse approximation via Greedy pursuit," presented at the Int. Conf. Acoustics, Speech, Signal Process. (ICASSP), Philadelphia, PA, Mar. 2005.

[5] ——, "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.

[6] J. A. Tropp, "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.

[7] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic, 1998.

[8] D. L. Donoho and P. Stark, "Uncertainty principles and signal recovery," *SIAM J. Appl. Math.*, vol. 49, no. 3, pp. 906–931, 1989.

[9] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2845–2862, Nov. 2001.

[10] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 48, pp. 2558–2567, 2002.

[11] A. Feuer and A. Nemirovski, "On sparse representation in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1579–1581, 2003.

[12] D. L. Donoho and M. Elad, "Optimally sparse representation in general (non-) dictionaries via ℓ^1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, pp. 2197–2202, 2002.

[13] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, pp. 3320–3325, 2003.

[14] J. J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inf. Theory*, vol. 50, pp. 1341–1344, Jun. 2004.

[15] D. L. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[16] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2231–2242, Oct. 2004.

[17] ——, "Just relax: Convex programming methods for identifying sparse signals," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[18] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Its Appl.*, vol. 18, no. 2, pp. 95–138, 1977.

[19] D. M. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[20] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 187–200, Jan. 1999.

[21] K. Kreutz-Delgado and B. D. Rao, "Measures and algorithms for best basis selection," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Seattle, WA, May 1998, vol. 3, pp. 1881–1884.

[22] S. Mallat and Z. Zhang, "Matching pursuit in a time-frequency dictionary," *IEEE Trans. Signal Process.*, vol. 41, pp. 3397–3415, 1993.

[23] G. Davis, S. Mallat, and Z. Zhang, "Adaptive time-frequency decompositions," *Opt. Eng.*, vol. 33, pp. 2183–2191, 1994.

[24] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, A. Singh, Ed., "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Systems Computers*, Los Alamitos, CA, 1993.

[25] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.

[26] D. Leviatan and V. N. Temlyakov, "Simultaneous Approximation by Greedy Algorithms," Univ. of South Carolina at Columbia, IMI Tech. Rep. 2003:02, 2003.

[27] ——, "Simultaneous Greedy Approximation in Banach Spaces," Univ. of South Carolina at Columbia, IMI Tech. Rep. 2003:26, 2003.

[28] A. Lutoborski and V. N. Temlyakov, "Vector Greedy algorithms," *J. Complexity*, vol. 19, pp. 458–473, 2004.

[29] V. N. Temlyakov, "A remark on simultaneous Greedy approximation," *East J. Approx.*, vol. 10, 2004.

[30] G. H. Golub and C. F. Van Loan, *Matrix Computation*. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.

[31] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[32] C. Couvreur and Y. Bresler, "On the optimality of the backward Greedy algorithm for the subset selection problem," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 3, pp. 797–808, 2000.

[33] G. Harikumar, C. Couvreur, and Y. Bresler, "Fast optimal and suboptimal algorithms for sparse solutions to linear inverse problems," presented at the Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), Seattle, WA, May 1998.

[34] G. Harikumar and Y. Bresler, "A new algorithm for computing sparse solutions to linear inverse problems," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Atlanta, GA, May 1996, vol. 3, pp. 1331–1134.

[35] D. L. Donoho, "For most large underdetermined systems of linear equations, the minimal ℓ_1-norm solution is also the sparsest solution," 2004 [Online]. Available: http://www-stat.stanford.edu/~donoho/Reports/2004/l1l0EquivCorrected.pdf

[36] ——, "For most large underdetermined systems of linear equations, the minimal ℓ_1-norm near-solution approximates the sparsest near-solution," 2004 [Online]. Available: http://www-stat.stanford.edu/~donoho/Reports/2004/l1l0approx.pdf

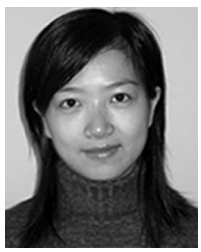
[37] T. Jiang, "Maxima of entries of haar distributed matrices," *Probab. Theory Related Fields*, vol. 131, no. 1, pp. 121–144, 2005.

[38] ——, "The asymptotic distributions of the largest entries of sample correlation matrices," *Ann. Appl. Probab.*, vol. 14, no. 2, pp. 865–880, 2004.

[39] J. Chen and X. Huo, "Sparse representations for multiple measurement vectors (MMV) in an over-complete dictionary," presented at the Int. Conf. Acoustics, Speech, Signal Process. (ICASSP), Philadelphia, PA, Mar. 2005.

[40] ——, "Theoretical results about finding the sparsest representations of multiple measurement vectors (MMV) in an over-complete dictionary, using ℓ_1-norm minimization and greedy algorithms," 2004 [Online]. Available: http://www.isye.gatech.edu/~xiaoming/publication/pdfs/mmv101204.pdf

[41] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 489–509, Feb. 2006.



Jie Chen (S'05) received the B.S. degree in computational math from Nanjing University, China, in 2003. She is currently working towards the Ph.D. degree at the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta.

Her research interests are applied statistics and data mining.

Xiaoming Huo (S'96–M'99–SM'04) received the B.S. degree in mathematics from the University of Science and Technology, China, in 1993 and the M.S. degree in electrical engineering and the Ph.D. degree in statistics from Stanford University, Stanford, CA, in 1997 and 1999, respectively.

He is an Associate Professor with the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta. His research interests include statistics and multiscale image analysis.

Dr. Huo received first prize in the Thirtieth International Mathematical Olympiad (IMO), which was held in Braunschweig, Germany.