Theoretical Results on Sparse Representations of Multiple Measurement Vectors

Jie Chen and Xiaoming Huo

***Abstract*— Multiple measurement vector (MMV) is a relatively new problem in sparse representations. Efficient methods have been proposed. Considering many theoretical results that are available in a simple case – single measure 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 an over-complete 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 theoretical advances in probability. The connections will be discussed.

Keywords: Multiple measurement vector, basis pursuit, orthogonal matching pursuit, sparse representation.

EDICS Category: 2-ALGO.

I. Introduction

The problem of finding sparse representations of multiple measurement vectors (MMV) in an over-complete dictionary was motivated by a neuro-magnetic 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 SMVs *simultaneously* [4], [5], [6]. In this paper, we focus on the theoretical development of MMV problems, instead of their 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 X is called an *atom*. A set of all atoms is called a *dictionary*, 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.

Following a convention in applied mathematics, we consider that the columns of the matrix A form a dictionary (see Mallat's book [7]). An over-complete dictionary simply means that $m < n$. Usually, we have $m \ll n$ and $L < m$. As mentioned earlier, when $L = 1$, we have the case of single measurement vector (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, $1 \leq l \leq L$, 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 an over-complete dictionary. We refer to [8], [9], [10], [11], [12], [13], [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 letters, X and B, emphasizing that they are *matrices* instead of *column vectors*.

In an 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 non-zero 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:

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

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

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

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 **(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 **(Q0)** and **(Q1)** has been proved under various conditions, and the most recent work was done by

Jie Chen and Xiaoming Huo are with the School of ISyE, Georgia Institute of Technology, Atlanta, GA 30332-0205. Emails: jchen, xiaoming@isye.gatech.edu.

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.

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 representations in SMV. In this paper, we extend corresponding theorems for SMV to MMV.

Another way to get 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 too.

In the present paper, we consider a *noiseless* case; i.e., an SMV, b, or an MMV, B, is an linear combination of atoms without noise, i.e., $b = Ax$ or $B = AX$. It is a harder mathematical problem when additive noise is integrated in the formulation. For *noisy* cases, we refer to [15], [16], [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 general: the inner vector norm can be any vector norm in a Euclidean space. Moreover in the case of minimizing-the-ℓ_0-norm, an upper bound, which is higher than the corresponding upper bound in SMV, can be derived.

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 – OMP. Conditions under which the OMP gives the sparsest representation are given. Section V describes some simulations, which indicate that the theoretical bounds are conservative. Section VI gives 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 matrix X, i.e., $X = [x_1, x_2, \cdots, x_n]^T$, $m(\cdot)$ is any vector norm in $\mathbb{R}^L$, and vector $(m(x_i))_{n\times 1}$ has the ith entry being equal to $m(x_i), 1 \le i \le n$. Note that $m(x)$ can be any vector norm in $\mathbb{R}^L$. It turns out that to study the theoretical property of the ℓ_0-norm and the ℓ_1-norm approaches, we only need the fact that m(x)=0 if and only if $x = \vec{0}$, where $\vec{0}$ is a all zero vector in $\mathbb{R}^L$. Symbol $\mathcal{R}$ stands for a sparsity *rank*. A noiseless sparse representation problem in MMV can be written as

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

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

In general, the solution to **(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 a highly sparse representation is sufficiently the sparsest possible representation. 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 between the ℓ_0-norm minimization and the ℓ_1-norm minimization.

The following generalizes the result of Donoho and Elad [12] to MMV. We start with the concept of *Spark*. (The following quoted from [15] describes an origin of $Spark$: "After [12] appeared, we learned that Kruskal [18] worked with a related notion in the context of fitting trilinear arrays by simple models; this notion has later been called the Kruskal rank..." Note the reference numbers in the quotation are changed to be consistent with the references of this paper.)

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

In [12], $Spark(A)/2$ is a threshold of the sparsity: if the signal is made by less than $Spark(A)/2$ atoms, or in other words, if the signal is a linear combination of less than $Spark(A)/2$ 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

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

Comparing with the SMV cases, we can see that the above theorem has the same upper bound.

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

$$\min \|X^{(j)}\|_0, \text{ subject to: } AX^{(j)} = B^{(j)}, \; j = 1, 2, \ldots, L,$$

should give exactly the jth column of the 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 $\max\{\mathcal{R}(X_1), \mathcal{R}(X_2)\} < Spark(A)/2$. We have

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

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 $\|(X_1 - X_2)^{(1)}\|_0 \geq Spark(A)$. Therefore, we have $\mathcal{R}(X_1 - X_2) \geq Spark(A)$, which contradicts (2.1). This contradiction proves the theorem. □

If we are willing to consider 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 $Rank(Cols(B))$ denote the column rank of matrix B. Apparently $Rank(Cols(B)) \leq L$. Matrix X will be the unique solution to the problem **(P0)**, if $B = AX$ and

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

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

$$d(Null(X_1)) \leq d(Null(B)),$$

and

$$d(Null(X_2)) \leq d(Null(B)).$$

Recall $Rank(Cols(B))$ denotes the column rank of matrix B. Similarly let $Rank(Cols(X_1))$ and $Rank(Cols(X_2))$ denote the column ranks of matrices X_1 and X_2. We have

$$Rank(Cols(X_1)) \geq Rank(Cols(B)), \quad (2.2)$$

and

$$Rank(Cols(X_2)) \geq Rank(Cols(B)). \quad (2.3)$$

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 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}. Decompose X_1 and X_2 as

$$X_1 = \begin{bmatrix} X_{11} \\ X_{12} \end{bmatrix}, X_2 = \begin{bmatrix} X_{21} \\ X_{22} \end{bmatrix},$$

where X_{11} (resp. X_{22}) are the rows of matrix X_1 (resp. X_2) corresponds to A_1 (resp. A_2), matrices X_{12} and X_{21} are the rows of X_1 and X_2 corresponding to the columns in A_{12}. We have the following:

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

From (2.4), we have

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

It is easy to see

$$\begin{aligned} & Rank(Cols(\begin{bmatrix} X_{11} \\ X_{12} - X_{21} \\ -X_{22} \end{bmatrix})) \\ \geq \; & \max\{Rank(Cols(X_{11})), Rank(Cols(X_{22}))\} \quad (2.6) \end{aligned}$$

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

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

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 (null) space is at most reduced by r_{12} dimensions, which is (2.7). Inequality (2.7) immediately leads to

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

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

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

By the definition of $Spark$, we have

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

Combining all the above, we have

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

The last inequality is based on (2.9) and (2.10). It is easy to see that the above proves the theorem. □

It is evident that Theorem 2.2 is a special case of Theorem 2.4, by having $Rank(Cols(B)) = 1$.

One referee asks that in the upper bound of Theorem 2.4, whether it is possible to replace the rank of matrix B with a rank of matrix X. Because $B = AX$, it is evident that $Null(X) \subset Null(B)$. Hence $d(Null(X)) < d(Null(B))$. Therefore, $Rank(Cols(X)) \geq Rank(Cols(B))$. Given this, it is not clear how to utilize the existing approach to generate an upper bound that is based on $Rank(Cols(X))$. Other approaches may lead to an upper bound with the column rank of matrix X; we leave it as a future research topic.

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

A difficulty associated with an upper bound with $Spark(A)$ 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 $Spark(A)$ 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], [12]. They provide upper bounds that are slightly 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.,

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

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]:

$$Spark(A) \geq (1 + 1/M).$$

Therefore, an upper bound with $Spark(A)$ 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

$$\mathcal{R}(X) < (M^{-1} + Rank(Cols(B)))/2,$$

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 Section 4.2], we can find the following relation: $Spark(A) \geq 2\mu_{1/2}(G) + 1$. Combining with Theorem 2.4, we immediately have the following.

Corollary 2.8: If $B = AX$ and

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

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 $X \in \mathbb{R}^{m \times L}$,

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

where $m(x)$ is a vector norm in $\mathbb{R}^L$. In this section, we consider a relaxation to the above quantity. As a preparation, the following is a well-known result for norms in the Euclidean space. We present it without a proof.

Proposition 3.1: 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

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

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

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

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

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

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 has been proposed as objective functions. In pioneer works on MMV [1], [20], [21], the following diversity measure on sparsity was proposed:

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

where p and q are parameters, 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 $q = 2, p \leq 1$, 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 columns of A with indices from S. We can write $B = A_S X_S^*$, where matrix X_S^* is made by 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 3.2: 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 the 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)**.

Proof. Suppose there are two representations: $B = A_S X_S^* = A_{S'} Y_{S'}$, where $S \neq S'$ and set S' includes the indices of the nonzero rows of the matrix $Y \in \mathbb{R}^{n \times L}$. We only need to show that

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

Recall

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

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 3.1, 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 (3.11). 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:

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

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 3.2, 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: $Spark(A)/2 > \mu_{1/2}(G) \ge \frac{1}{2M}$. 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

$$\mathcal{R}(X) < (1 + 1/M)/2$$

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 $Spark(A)/2$ to $[Spark(A) - 1 + Rank(Cols(B))]/2$.

There is no evidence that between condition $\mathcal{R}(x) < Spark(A)/2$ 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 $Spark(A)/2 < \mathcal{R}(x) < [Spark(A) - 1 + Rank(Cols(B))]/2$, 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 3.2, 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 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 atoms in the optimal representation of matrix $B^{(i)}$. This is because the number of the atoms in the optimal representation of matrix $B^{(i)}$ 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 under the formulation of an MMV and under the formulation of a combination of several SMVs.

IV. Orthogonal Matching Pursuit

Matching pursuit (MP) [22] is proposed as an efficient numerical method to decompose a signal. As an improvement of MP, orthogonal matching pursuit (OMP) [23], [24] has been introduced. OMP overcomes some drawbacks of MP. Unfortunately, counterexamples show that both methods could be trapped by initial selection of a 'bad' atom, see Chen et al. [25]. For MMV problem, many variants of OMP have been proposed. A subset of them are [1], [4], [5], [26], [27], [28], [29].

For this section, we propose our OMP with an $\ell_q (q \ge 1)$ norm of the inner product. Note under 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 $a_{k_t} = \text{argmax}_{a_k} \|z_k\|_q$, where $z_k = R_{t-1}^T a_k$ and $q \ge 1$;

b) Let $S_t = [S_{t-1}, a_{k_t}]$, and $X^* = \text{argmin}_X \|S_t X - B\|_F^2, y_t = S_t X^*$;
c) Set $R_t = B - y_t$.

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

In [4], [5], Tropp et al. proposed ℓ_1 norm in step 2)-a). In [26], [27], [28], [29], ℓ_2 and ℓ_∞ are proposed for week matching pursuit and week orthogonal matching pursuit for 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 2)-a) 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 [30], [5].

Lemma 4.2: Consider matrix A.

1) The $(1, q)$ matrix norm is the maximum ℓ_q norm of 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 rows of A, where $1/p + 1/p' = 1$.

The following property regarding (p, q) matrix norm can be easily derived from 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 orthogonal 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 since we only consider the *noiseless* formulation, we are allowed to use such an idealistic argument.

According to our notation, in step 2)-a) 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 optimal atoms 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 section.

Theorem 4.5: A sufficient condition for OMPMMV to recover a representation of matrix B associated with atom indices 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 equality

$$R_t = (A_S^+)^T A_S^T R_t. \quad (4.15)$$

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, 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

$$\mathcal{R}(X) < (1 + 1/M)/2$$

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, it is much easier to check the conditions in the above corollary. We can calculate X through OMPMMV first, and then check if such an X satisfies the conditions.

V. Simulation

A. *Exact recovery of OMPMMV and* **(P1)**

Simulations are conducted to bring insights on when the OMPMMV and **(P1)** can exactly find the *original* signal. Two experiments are conducted. In the first experiment, matrix $A \in \mathcal{R}^{m\times n}$ has dimensions $m = 20$ and $n = 30$. We set $L = 5$. The entries of matrix A are independently sampled from the standard normal $N(0,1)$. We compute an observed MMV, B, as $B = AX_0$, where N rows of matrix $X_0 \in \mathcal{R}^{30\times 5}$ are randomly chosen. The values of the nonzero entries 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 the OMPMMV and the **(P1)**. The solution X is compared with the original matrix X_0. If $X \equiv X_0$, an 'exact recovery' is obtained. The proportion of exact recoveries among $1,000$ times of simulation is reported as 'empirical probabilities of exact recovery' in Figure 1(a). We observed that the OMPMMV performs slightly better.

In the second experiment, matrix A is generated by concatenating two ortho-normal bases: $[I, H]$, where matrix I is an identity matrix and matrix H is the 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.

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. Firstly, for A randomly generated from 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 probability of exact recovery on one plot. See the results in Figure 2 (a). Secondly, for $A = [I, H]$, where sub-matrix H is a 16 by 16 Hadamard matrix, matrix I is a 16 by 16 identity matrix, and $L = 3$, we do the same simulation as above. See the results in Figure 2 (b).

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. Firstly, for matrix A whose entries are randomly generated from 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 probability of exact recovery on one plot. See the results in Figure 3 (a). Secondly, for $A = [I, H]$, where sub-matrix H is a 16 by 16 Hadamard matrix, I is a 16 by 16 identity matrix, and $L = 3$, we do the same simulation as above. Results are shown in Figure 3 (b).

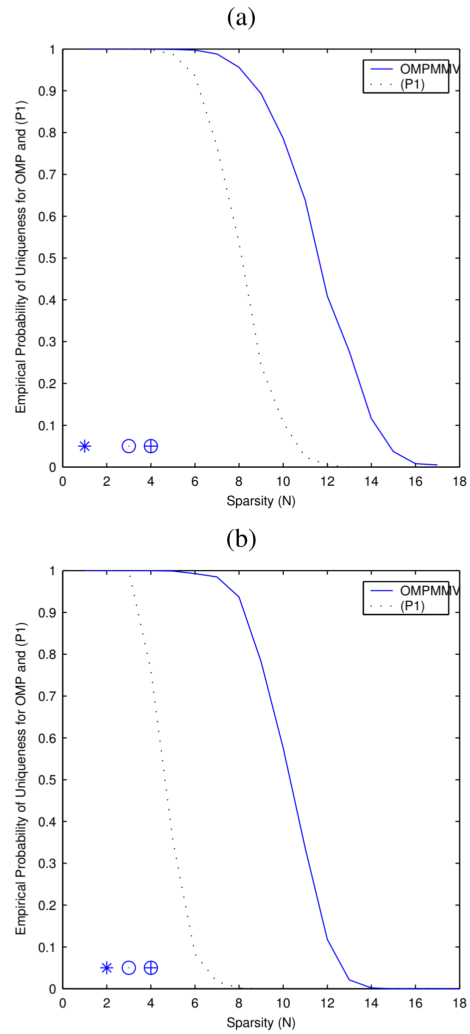

Fig. 1. **(a)** First experiment of exact recovery, in which $A \in \mathcal{R}^{m\times n}, X_0 \in \mathcal{R}^{n\times L}, m = 20, n = 30, L = 5$, where entries of matrices A and X_0 are independently sampled from $N(0,1)$. Symbol $*$ indicates where the theoretical upper bound for uniqueness is (i.e., $\lfloor (1+1/M)/2 \rfloor$). Symbol $\oplus$ indicates the largest value of N while the OMPMMV finds the original X_0 among all simulations; similarly, symbol $\ominus$ indicates the largest value of N while the solutions of **(P1)** are identical with the matrix X_0 for all simulations. 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 $A \in \mathcal{R}^{16\times 32}$ and sub-matrix H is a 16 by 16 Hadamard matrix. Matrix I is a 16 by 16 identity matrix. Matrix X_0 is chosen in the same way, with N being the number of nonzero rows. The symbols $*$, $\oplus$ and $\ominus$ have the same meaning. In this case, symbol $*$ is marked at $N = 2$. Symbols $\oplus$ and $\ominus$ are at $N = 4$ and $N = 3$ respectively.

From the simulations, we see that the curves in each plot are similar. This demonstrates that in 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 $Relax(X)$ 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.

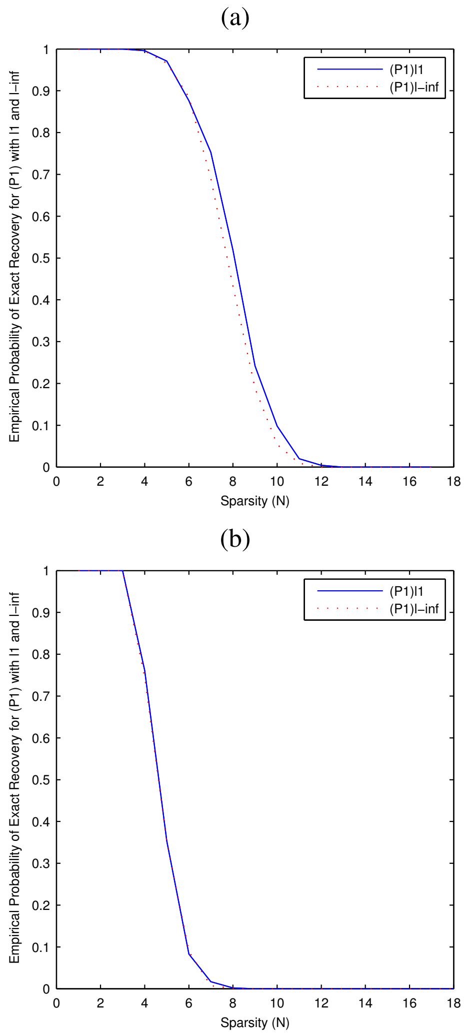

Fig. 2. **(a)** Consider the case $A \in \mathcal{R}^{m\times n}, X_0 \in \mathcal{R}^{n\times L}$, m=20, n=30, L=5, where entries of matrices A and X_0 are independently sampled from $N(0,1)$. The theoretical upper bound for the equivalence is 1. Let $N_i, i = 1, \infty$ denote the largest values of N when the solutions of **(P1)** with $m(\cdot)$ being the ℓ_i norm are identical with the 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 by 16 Hadamard matrix and submatrix I is a 16 by 16 eye matrix. We have $L = 3$. The theoretical upper bound for equivalence is 2. We obtain $N_1 = N_\infty = 3$.

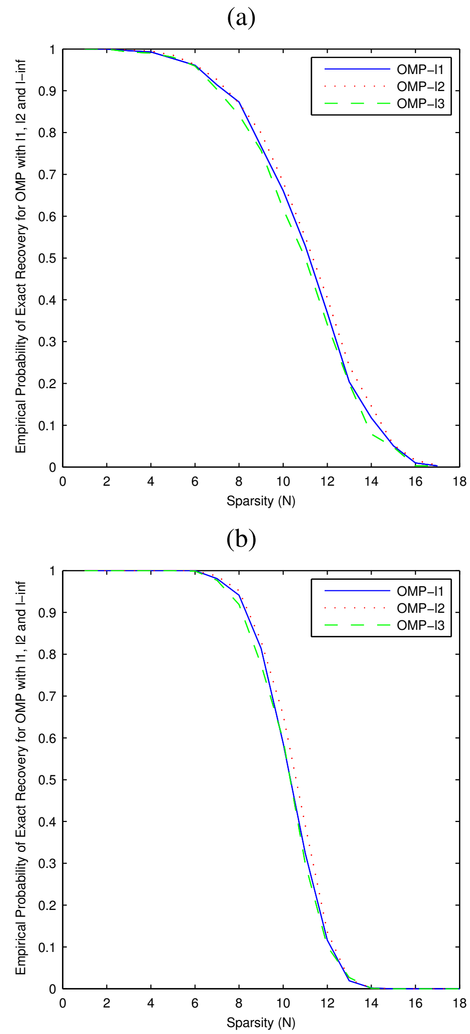

Fig. 3. **(a)** We consider $A \in \mathcal{R}^{m\times n}, X_0 \in \mathcal{R}^{n\times L}, m = 20, n = 30, L = 5$, where entries of matrices A and X_0 are independently sampled from $N(0,1)$. The theoretical upper bound for equivalence is 1. Notation N_i, $i = 1, 2, \infty$, denote the largest values of N while the OMPMMV with ℓ_i norm in step 2)-a) find 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 by 16 Hadamard matrix and submatrix I is a 16 by 16 eye matrix. We have $L = 3$. The theoretical upper bound for equivalence is 2. We obtain $N_1 = N_2 = 6$ and $N_\infty = 5$.

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 $R_1(X) = \sum_{i=1}^n m_1(x_i)$ and $R_2(X) = \sum_{i=1}^n m_2(x_i)$. whenever **(P1)** with relaxation $\mathcal{R}_1(X)$ finds the sparest solution, i.e., $R_1(X^*) < R_1(Y)$ for all other Y, **(P1)** with relaxation $\mathcal{R}_2(X)$ finds the sparsest solution too, i.e., $R_2(X^*) < R_2(Y)$ for all other Y.

Regarding norms in an 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. This is a standard result in vector norms. It is one incarnation of the slogan "All the norms are equivalent in the Euclidean space." We sketch the proof for the completeness.

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. □

The above demonstrates that there is no optimal relaxation while generalizing SMV to MMV. Note that we consider 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)**. On one hand, *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 [31] on *backward elimination* and related analysis has strong similarity with some of the results that we developed here for MMV.

Short papers that proposed various heuristics to achieve sparse representations are [32], [33]. 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 $A \in \mathcal{R}^{m\times n}, x_0 \in \mathcal{R}^n$, and $b \in \mathcal{R}^m$, $m < n$. Donoho in [34] shows that even when $\|x_0\|_0 = O(n)$, with nearly 1 probability, 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 is 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 [35]. At the same time, E. Candés gave several talks with similar results, based on his joint work with T. Tao and J. Romberg. Unfortunately, we do not have access to their paper yet.

There are interesting developments in 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 interests to study the case when the matrix A is a concatenation of two orthogonal square matrices: $A = [O_1, O_2]$, where matrices O_1, O_2 are orthogonal. Apparently the *mutual incoherence* is the maximum magnitude of entries of the matrix $O_1^T O_2$. Jiang in [36] derives the asymptotic distribution of this quantity. Basically, if $O_1 \in \mathcal{R}^{n\times n}$, he proves that $M^{-1}/2$ is *almost surely* between $\frac{1}{2\sqrt{6}}\sqrt{\frac{n}{\log n}}$ and $\frac{1}{4}\sqrt{\frac{n}{\log n}}$.

In another work of Jiang [37], 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 was reported in a conference paper [38] and a manuscript [39]. The latter was downloadable online. This paper is an extensively revised version of [39].

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 quite general: 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 very general sense; more specifically, the inner vector norm in the step 2)-a) of OMPMMV can be ℓ_q norm for any $q \geq 1$.

These results provide useful insights in designing numerical solutions to find sparse representations for *multiple measurement vectors.*

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 Processing*, 2004, Accepted.

[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," in *Proceedings of ICASSP*, Montreal, May 2004.

[3] I. F. Gorodnitsky, J. S. George, and B. D. Rao, "Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm," *Electroencephalography and Clinical Neurophysiology*, vol. 95, no. 4, pp. 231–251, October 1995.

[4] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Simultaneous sparse approximation via greedy pursuit," in *Proceedings of ICASSP 2005*, Philadelphia, March 2005.

[5] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. part i: Greedy pursuit," *EURASIP Journal on Applied Signal Processing*, April 2005, Accepted by special issue: Sparse approximations in signal and image processing.

[6] J. A. Tropp, "Algorithms for simultaneous sparse approximation. part ii: Convex relaxation," *EURASIP Journal on Applied Signal Processing*, April 2005, Accepted by special issue: Sparse approximations in signal and image processing.

[7] S. Mallat, *A wavelet tour of signal processing*, Academic Press, Inc., San Diego, CA, 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 Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, November 2001.

[10] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2558–2567, 2002.

[11] A. Feuer and A. Nemirovski, "On sparse representation in pairs of bases," *IEEE Trans. Inform. Theory*, vol. 49, no. 6, pp. 1579–1581, 2003.

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

[13] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3320–3325, 2003.

[14] J. J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1341–1344, June 2004.

[15] D. L. Donoho, M. Elad, and V. Temlyakov, *Stable recovery of sparse overcomplete representations in the presense of noise*, Stanford University and University of South Carolina, 2004, Submitted manuscript.

[16] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231–2242, October 2004.

[17] J. A. Tropp, "Just relax: Convex programming methods for subset selection and sparse approximation," Tech. Rep., ICES Report 04-04, UT-Austin, 2004.

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

[19] D. M. Malioutov, Mujdat Cetin, and Alan S. Willsky, "Sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, 2005, to appear.

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

[21] K. Kreutz-Delgado and B. D. Rao, "Measures and algorithms for best basis selection," in *Proc. of ICASSP*, Seattle, Washington, May 1998, vol. 3, pp. 1881–1884.

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

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

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

[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, Selected from *SIAM J. Sci. Comput.*, vol 20, no. 1, 33–61, 1998.

[26] D. Leviatan and V. N. Temlyakov, "Simultaneous approximation by greedy algorithms," Tech. Rep., IMI Report 2003:02, Univ. of South Carolina at Columbia, 2003.

[27] D. Leviatan and V. N. Temlyakov, "Simultaneous greedy approximation in Banach spaces," Tech. Rep., IMI Report 2003:26, Univ. of South Carolina at Columbia, 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*, the Johns Hopkins University Press, Baltimore, 1996.

[31] 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.

[32] G. Harikumar, C. Couvreur, and Y. Bresler, "Fast optimal and suboptimal algorithms for sparse solutions to linear inverse problems," in *Proc. of ICASSP*, Seattle, Washington, May 1998.

[33] G. Harikumar and Y. Bresler, "A new algorithm for computing sparse solutions to linear inverse problems," in *Proc. ICASSP*, Atlanta, GA, May 1996, vol. 3, pp. 1331–1134.

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

[35] D. L. Donoho, "For most large underdetermined systems of linear equations, the minimal ℓ_1-norm near-solution approximates the sparsest near-solution," http://www-stat.stanford.edu/∼donoho/Reports/2004/l1l0approx.pdf, August 2004.

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

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

[38] J. Chen and X. Huo, "Sparse representations for multiple measurement vectors (MMV) in an over-complete dictionary," in *Proceedings of ICASSP 2005*, Philadelphia, March 2005.

[39] J. Chen and X. Huo, "Theoretical results about finding the sparsest representations of multiple measurement vectors (MMV) in an over-complete dictionary, using ℓ_1-morm minimization and greedy algorithms," http://www.isye.gatech.edu/∼xiaoming/publication/pdfs/mmv101204.pdf, August 2004.