SPARSE REPRESENTATIONS FOR MULTIPLE MEASUREMENT VECTORS (MMV) IN AN OVER-COMPLETE DICTIONARY

Jie Chen and Xiaoming Huo

Georgia Institute of Technology, School of Industrial and Systems Engineering
765 Ferst Dr., Atlanta, GA 30332-0205

ABSTRACT

Multiple Measurement Vector (MMV) is a newly emerged problem in sparse representation in an over-complete dictionary; it poses new challenges. Efficient methods have been designed to search for sparse representations [1, 2]; however, we have not seen substantial development in the theoretical analysis, considering what has been done in a simpler case—Single Measurement Vector (SMV)—in which many theoretical results are known, e.g., [9, 3, 4, 5, 6]. This paper extends the known results of SMV to MMV.

Our theoretical results show the fundamental limitation on when a sparse representation is unique. Moreover, the relation between the solutions of ℓ_0-norm minimization and the solutions of ℓ_1-norm minimization indicates a computationally efficient approach to find a sparse representation. Interestingly, simulations show that the predictions made by these theorems tend to be conservative.

1. INTRODUCTION

The problem of sparse representations for Multiple Measurement Vectors (MMV) in an over-complete dictionary is motivated by a neuro-magnetic inverse problem that arises in Magnetoencephalography (MEG)—a modality for imaging the possible activation regions in the brain [1].

An MMV problem can be described as the following linear algebra problem: given multiple measurement vectors B and a dictionary A, one wants to solve the following systems of equations:

$$AX = B,$$

where $A \in \mathcal{R}^{m\times n}, X \in \mathcal{R}^{n\times L}$, and $B \in \mathcal{R}^{m\times L}$. Following tradition, we consider the columns of the matrix A forming a *dictionary*. An over-complete dictionary simply means that $n > m$. Usually, we have $m \ll n$ and $L < m$. When $L = 1$, we have the case of Simple Measurement Vector (SMV). We assume that the columns of A have been normalized. Matrices X and B can be represented in the form: $X = [x^{(1)}, x^{(2)}, \ldots, x^{(L)}]$, $B = [b^{(1)}, b^{(2)}, \ldots, b^{(L)}]$, where $x^{(l)}$'s and $b^{(l)}$'s, $1 \le l \le L$, are column vectors. Obviously, the systems of equations can be rewritten as: $Ax^{(l)} = B^{(l)}, \quad l = 1, \ldots, L$.

The desired property of a solution matrix X (or a vector, if one has SMV) is that the number of rows containing nonzero entries is small ([1]). A mathematical definition will appear later.

There is abundant literature on the searching of sparse representations in over-complete dictionaries in the SMV case. The Introduction section of [7] gives a comprehensive description of many important applications. In the SMV case, we replace X and B by their lower letters—x and b, emphasizing that they are vectors instead of matrices.

In SMV, the sparsity of a representation is defined as the ℓ_0 norm of the vector x, which is denoted by $\|x\|_0$. The $\|x\|_0$ is equal to the number of non-zero elements in the vector x. A sparse representation can be found by solving the following non-convex optimization problem:

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

The above problem can be convexified as a ℓ_1-norm minimization problem, and solved via linear programming. The ℓ_1-norm minimization problem is

$$(\mathbf{Q1}): \quad \min \|x\|_1, \quad \text{s.t. } Ax = b.$$

Note that the problem **(Q0)** is a combinatorial optimization problem, which can be extremely difficult to solve. We hope that the solution to the problem **(Q1)** is, in some sense, close enough to the solution of **(Q0)**. The equivalence of the solutions between **(Q0)** and **(Q1)** has been proved under certain conditions, and the most recent work is [6]. Such an equivalence is very important in searching for sparse representations in SMVs. In this paper, we generalize the corresponding results to the case of MMV.

Another way to get a sparse representation is through an algorithm named Orthogonal Matching Pursuit (OMP). It has been proved that OMP can find the exact sparsest representation of a signal in certain cases [7]. In this paper, we generalize this result to the cases of MMV as well. Moreover, a modified version is introduced, to improve its theoretical property.

The rest of the paper is organized as follows. Section 2 describes the uniqueness of the solutions when the most straightforward approach—minimizing the ℓ_0 norm — is adopted. Section 3 describes when the solutions of the ℓ_1-norm minimization approach are identical to the solutions of the ℓ_0-norm minimization approach. Section 4 describes the property of the sparse representations that are provided by a greedy algorithm—Orthogonal Matching Pursuit (OMP) and a modified version.

Due to space, most of the proofs are omitted or simplified as a narrative description.

2. ℓ_0 NORM SOLUTIONS

2.1. Formulation

A noiseless sparse representation problem in MMV can be written as

$$(\mathbf{P0}): \quad \min \quad \left\| \left(\sum_{j=1}^{L} |x_{ij}| \right)_{i\in\Omega} \right\|_0, \quad \Omega = \{1, \cdots, m\}.$$
$$\text{s.t} \quad AX = B,$$

The sum of the absolute values of entries at each row is used to measure the common sparsity of all the columns; i.e., $\|(\sum_{j=1}^{L} |x_{ij}|)_{i\in\Omega}\|_0$ counts only the number of rows that contain nonzeros. Readers can compare this with **(Q0)**.

In general, to find the solution of **(P0)** requires enumerating all the subsets of the set $\{1, 2, \ldots, n\}$. The complexity of such a subset-search algorithm grows exponentially ¿¡DEFANGED.76284 with the dictionary size n.

2.2. Uniqueness of ℓ_0 Minimization

We restrict our attention to the case when the solution to **(P0)** is unique. We prove that a highly sparse representation must be the sparsest possible representation, and give some conditions under which the solution to **(P0)** is unique. This is a necessary preparation for the subsequent results (e.g., equivalence between two methods).

The following is highly parallel to the results in paper [6]. We start with the concept of *Spark*.

Definition 2.1 *Spark: Given a matrix A, its Spark $S(A)$ $(= \sigma)$ is the smallest possible integer such that there exist σ columns of matrix A that are linearly dependent.*

In [6], $S(A)/2$ is a threshold of the sparsity: if the signal is represented by no more than $S(A)/2$ atoms—i.e. if the signal is a linear combination of no more than $S(A)/2$ columns of matrix A—then the solution of **(Q0)** corresponds exactly to this linear combination.

For MMV, an identical conclusion can be made, although the proof is slightly harder. Due to space, we only state the result, and leave all the proofs to our article [8].

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

$$\left\| (\sum_{j=1}^{L} |x_{ij}|)_{i\in\Omega} \right\|_0 < S(A)/2.$$

The problem with the above upper bound, $S(A)/2$, is that the quantity Spark itself is hard to calculate. Up to now, there was no efficient algorithm for computing it except enumerating all the possible subsets. For practical use, we introduce another quantity: *mutual coherence*. This quantity has appeared in previous papers, e.g., [3, 4, 6]. It gives an upper bound that is lower than Spark's.

Definition 2.3 *Mutual coherence (denoted by M) is the maximum absolute inner product between two atoms; i.e.,*

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

where Gram matrix $G = A^T A$.

Similar to Theorem 2.2, we have the following result for mutual coherence.

Theorem 2.4 *Matrix X is the unique solution of* **(P0)***, if $B = AX$ and $\|(\sum_{j=1}^{L} |x_{ij}|)_{i\in\Omega}\|_0 < (1 + M^{-1})/2$.*

Note that M and Spark have the following relationship [6]:

$$S(A) \ge (1 + 1/M).$$

Therefore, the bound of $S(A)/2$ is better. We can use Theorem 2.2 and this relationship to prove Theorem 2.4. However, in our paper [8], Theorem 2.4 is proved without using Spark.

3. ℓ_1 NORM SOLUTIONS

3.1. Model

Due to the computational complexity of the ℓ_0-norm minimization problem, minimizing the ℓ_1-norm was proposed as an alternative. The ℓ_1-norm minimization problem can be formulated as follows:

$$(\mathbf{P1}): \quad \min \quad \sum_{i=1}^{n} \sum_{j=1}^{L} |x_{ij}| = \left\| \left(\sum_{j=1}^{L} |x_{ij}| \right)_{i\in\Omega} \right\|_1.$$
$$\text{s.t} \quad AX = B,$$

This is a convex optimization problem, and can be cast as a linear programming problem, which is solvable via, e.g., simplex or interior point methods. **(P1)** can be viewed as a convexification of **(P0)**. This idea is well documented in the literature, e.g., [9, 3, 4, 5, 6, 7, 10].

Besides the ℓ_1 norm, other functions of X have been proposed as objective functions. In works that are related to MMV, e.g., [1] and [11], 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, 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, with $q = 2, p >< DEFANGED.76285 \le 1$. It uses the idea of Lagrange multipliers, as well as iterations. A disadvantage of M-FOCUSS is that this iterative algorithm could be trapped by a local minimum, since its objective function is generally non-convex. With $p = 1, q = 1$ in the above objective, we obtain the ℓ_1 norm minimization problem **(P1)**. An immediate advantage is that **(P1)** can be solved via linear programming, which has been extensively developed and has handy softwares. Furthermore, under certain conditions, we will prove that the solutions of **(P1)** and **(P0)** are equivalent.

3.2. Uniqueness and Equivalence of ℓ_1 Norm Minimization

We give the conditions under which the ℓ_0 solution and ℓ_1 solutions are identical. Three different thresholds for this purpose are introduced. These three thresholds also apply to the SMV problem, [6, 10]. However, for MMV, we need to consider the sum of absolute values of each row in addition. Therefore the proofs are different. Again, for proofs please refer to [8].

Using a previously defined quantity—mutual coherence M—the uniqueness of ℓ_1 norm minimization as well as the equivalence between the ℓ_1 solution and ℓ_0 solution can be established as follows.

Theorem 3.1 *For a dictionary A and its Gram matrix G, let M be its mutual coherence that was defined earlier. If $AX = B$ and $\|(\sum_{j=1}^{L} |x_{ij}|)_{i\in\Omega}\|_0 < (1+1/M)/2$, then X is the unique solution of* **(P1)***, and this solution is identical to the solution of* **(P0)***.*

Note that the upper bound, $(1+1/M)/2$, is the same as the one in Theorem 2.4.

In the following, two more conditions are provided. They emphasize different aspects of the problem.

Definition 3.2 *For G a symmetric matrix, let $\mu_{1/2}(G)$ denote the smallest number m such that some collection of m off-diagonal magnitudes in a single row or column of the matrix G sums at least to $1/2$.*

Recall that in a Gram matrix, all diagonal entries are equal to 1. The following establishes a new condition that is based on quantity $\mu_{1/2}(G)$.

Theorem 3.3 *For a dictionary A and its Gram matrix G, if $AX = B$ and $\|(\sum_{j=1}^{L} |x_{ij}|)_{i\in\Omega}\|_0 < \mu_{1/2}(G)$, then X is the unique solution of* **(P1)***, and this solution is identical to the solution of* **(P0)***.*

It is straightforward to derive the following relation between M and $\mu_{1/2}(G)$: $\frac{1}{2M} \le \mu_{1/2}$. Note that none of the thresholds, $\mu_{1/2}(G)$ and $(1 + 1/M)/2$, can dominate the other one.

The following condition assumes that an optimal solution of **(P0)** is known. We consider X^* is the unique solution to **(P1)**. Let S be the index set of rows of X^* that contain nonzero entries. Let A^S denote a matrix that is made by the columns of the matrix A with index set S. Let matrix $(A^S)^+$ be the generalized inverse of matrix A^S: $(A^S)^+ A^S = I_\sigma$, where $\sigma = \mathcal{R}(X^*)$, which also is the size of set S. Apparently, we can assume that matrix A^S is of the full column rank, then we have the following lemma.

Theorem 3.4 *(1) Consider a column index j of the matrix A, satisfying $j \notin S$. Assume the columns of matrix A^S are linearly independent. If*

$$\|(A^S)^+ A_j\|_1 < 1, \quad \forall j \notin S, \tag{1}$$

then X^* is the unique solution of **(P1)***.*

(2) If for any matrix $A^{S'}$ ($A^{S'} \ne A^S$) that is made by σ columns of matrix A, there is a nonzero vector $c \in \mathcal{R}^\sigma$, such that $\sum_{i=1}^{L} |c_i| > 0$ (i.e., there is a nonzero element among the first L entries of vector c) and

$$c^T P \cdot (A^S)^+ \cdot A^{S'} = \vec{0}_{1\times\sigma},$$

where $\vec{0}_{1\times\sigma}$ is a 1 by σ vector that is made by zeros and P is a matrix that was specified earlier, then matrix X^* is the unique solution of problem **(P0)***.*

(3)If conditions that were described in (1) and (2) are satisfied, then matrix X^* is a unique solution to both **(P0)** *and* **(P1)***.*

4. ORTHOGONAL MATCHING PURSUIT

To solve an MMV problem, in [1], Orthogonal Matching Pursuit (OMP) was introduced. Due to space, we omit all the details, and refer to the original paper. In [8], it is proved that the original OMP is equivalent to the following procedure. The notation in the following is the same as in the original description [1].

An Equivalent Form of the Original OMP

1. Initialization: residual $R_0 = B$.

2. At the pth iteration:

   (a) Choose the atom a_{k_p}, which satisfies
   $a_{k_p} = \text{argmax}_{a_k} \|z_k\|_2$, where $z_k = R_{p-1}^H a_k$;

   (b) Let $S_p = [S_{p-1}, a_{k_p}]$, $S_0 = \emptyset$, and
   $X^* = \text{argmin}_X \|S_p X - B\|_F^2, \quad y_p = S_p X^*$.

   (c) Set $R_p = B - y_p$.

We prove that the above is equivalent to the original OMP [8]. This equivalent form of OMP is then used to prove the following theorem, which shows that OMP can recover the sparsest representation in MMV exactly.

Definition 4.1 *Support(A): The support of a matrix A is the set of the indices of rows containing nonzero entries.*

Theorem 4.2 *By OMP, when*

$$\left\| \left(\sum_{j=1}^{L} |x_{ij}^*|\right)_{i\in\Omega} \right\|_0 \le \frac{1+M}{(1+\sqrt{L})M},$$

we have Support(X_{OMP})=Support(X_{OPT}), where X_{OPT} is the sparest representation, and X_{OMP} is the result of OMP.

The above can be called a theorem of "exact recovery". We propose a modified version of OMP. With this modified algorithm, the threshold of the sparsity for the 'exact recovery' can be improved.

Our modified OMP makes a small change in the first step of the equivalent form of OMP. At every iteration, we choose the atom a_{k_p} that satisfies $\max_{a_k} \|R_{p-1}^H a_k\|_1$ instead of $\max_{a_k} \|R_{p-1}^H a_k\|_2$ as in the original OMP.

Similar to the theorem for the original OMP, the theorem about the condition of the 'exact recovery' of the sparest representation for the modified OMP can be established.

Theorem 4.3 *By the modified OMP, when*

$$\left\| \left(\sum_{j=1}^{L} |x_{ij}^*|\right)_{i\in\Omega} \right\|_0 \le \frac{1+M}{2M},$$

we have Support(X_{MOMP})=Support(X_{OPT}), where X_{MOMP} is the result of the modified OMP.

Compared with the upper bound, $\frac{1+M}{(1+\sqrt{L})M}$, in the original OMP, the new upper bound of the modified OMP, $\frac{1+M}{2M}$, is better.

5. SIMULATION AND DISCUSSION

In our simulations, we applied ℓ_1-norm minimization and the modified OMP to the simulated data, and it was found that the above mentioned upper bounds using M are usually too relaxed. In many finite situations, even though the sparsity of the optimal solution is well above these theoretical upper bounds, the uniqueness and equivalence can still hold. Due to space, we have to omit a detailed description.

What we have considered is the *noiseless* version of the sparse representation problem. Recently, advances have been made in the *noisy* version of these problems, [7, 12]. The latter is more meaningful, because in practice, numerical solutions always contain round off errors. Developing parallel theorems in the noisy case is an interesting research topic.

Due to space, all proofs are postponed to our paper [8].

6. REFERENCES

[1] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, *Sparse solutions to linear inverse problems with multiple measurement vectors*, UC San Diego, 2004, Submitted manuscript.

[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," Montreal, May 2004, ICASSP.

[3] 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.

[4] 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.

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

[6] 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.

[7] 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.

[8] J. Chen and X. Huo, *Sparse Representations of Multiple Measurement Vectors (MMV) in an Over-complete Dictionary*, Georgia Institute of Technology, Atlanta, GA, 2004, submitted manuscript.

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

[10] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," Tech. Rep., CAM Tech Report, Univ. Texas, 2003.

[11] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. on Signal Processing*, pp. 187–200, Jan. 1999.

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