The Null Space Property for Sparse Recovery from Multiple Measurement Vectors

Ming-Jun Lai *and Yang Liu†
Department of Mathematics
The University of Georgia
Athens, GA 30602

May 6, 2010

Abstract

We prove a null space property for the uniqueness of the sparse solution vectors recovered from a minimization in ℓ_p quasi-norm subject to multiple systems of linear equations, where $p \in (0, 1]$. Furthermore, we show that the null space property is equivalent to the null space property for the standard ℓ_p minimization subject to a single linear system. This answers the questions raised in [Foucart and Gribonval'09, [18]]. Finally we explain that when the restricted isometry cnstant $\delta_{2s+1} < 1$, then the ℓ_p minimization will find the s-sparse solution if $p > 0$ is small enough.

1 Introduction

Recently, one of the central problems in the compressed sensing for the sparse solution recovery of under-determined linear systems has been extended to the sparse solution vectors for multiple measurement vectors (MMV). That is, letting A be a sensing matrix of size $m \times N$ with $m \ll N$ and given multiple measurement vectors $\mathbf{b}^{(k)}, k = 1, \cdots, r$, we are looking for solution vectors $\mathbf{x}^{(k)}, 1, \cdots, r$ such that

$$A\mathbf{x}^{(k)} = \mathbf{b}^{(k)}, \quad k = 1, \cdots, r \tag{1}$$

and the vectors $\mathbf{x}^{(k)}, k = 1, \cdots, r$ are jointly sparse, i.e. have nonzero entries at the same locations and have as few nonzero entries as possible. Such problems arise in source localization (cf. [23]), neuromagnetic imaging (cf. [13]), and equalization of sparse communication channels (cf. [14, 16]).

*mjlai@math.uga.edu. This author is partly supported by the National Science Foundation under grant DMS-0713807

†yliu@math.uga.edu.

A popular approach to find the sparse solution for multiple measurement vectors (MMV) is to solve the following optimization:

$$\text{minimize}_{\substack{\mathbf{x}^{(k)} \in \mathbb{R}^N \\ k=1,\cdots,r}} \left\{ \left(\sum_{j=1}^{N} \|(x_{1,j}, \cdots, x_{r,j})\|_q^p \right)^{1/p} : \text{subject to} A\mathbf{x}^{(k)} = \mathbf{b}^{(k)}, k = 1, \cdots, r \right\}, \tag{2}$$

where $\mathbf{x}^{(k)} = (x_{k,1}, \cdots, x_{k,N})^T$ for all $k = 1, \cdots, r$ and $\|(x_1, \cdots, x_r)\|_q = \left(\sum_{j=1}^{r} |x_j|^q \right)^{1/q}$ is the standard ℓ_q norm for $q \geq 1$ and $p \geq 1$. Clearly, it is a generalization of the standard ℓ_1 minimization approach for the sparse solution. That is, when $r = 1$, one finds the sparse solution $\mathbf{x}$ solving the following minimization problem:

$$\text{minimize}_{\mathbf{x} \in \mathbb{R}^N} \{\|\mathbf{x}\|_1 : \text{subject to} A\mathbf{x} = \mathbf{b}\}, \tag{3}$$

where $\|\mathbf{x}\|_p = \left(\sum_{j=1}^{N} |x_j|^p \right)^{1/p}$ is the standard ℓ_p norm for $p \geq 1$. Such a minimization problem (3) has been studied for many years. See, e.g. [8] and [19] and references therein. In the literature, there are also several studies for various combinations of $p \geq 1$ and $q \geq 1$ in (2). See, e.g., references [14], [11], [23], [26]–[27].

In particular, the well-known null space property (cf. [15] and [20]) for the standard ℓ_1 minimization has been extended to this setting (2) for multiple measurement vectors. In [3], the following result is proved.

Theorem 1.1 *Let A be a real matrix of $m \times N$ and $S \subset \{1, 2, \cdots, N\}$ be a fixed index set. Denote by S^c the complement set of S in $\{1, 2, \cdots, N\}$. Let $\|\cdot\|$ be any norm. Then all $\mathbf{x}^{(k)}$ with support $\mathbf{x}^{(k)}$ in S for $k = 1, \cdots, r$ can be uniquely recovered using the following*

$$\textit{minimize}_{\substack{\mathbf{x}^{(k)} \in \mathbb{R}^N \\ k=1,\cdots,r}} \left\{ \sum_{j=1}^{N} \|(x_{1,j}, \cdots, x_{r,j})\| : \textit{subject to} A\mathbf{x}^{(k)} = \mathbf{b}^{(k)}, k = 1, \cdots, r \right\} \tag{4}$$

if and only if all vectors $(\mathbf{u}^{(1)}, \cdots, \mathbf{u}^{(r)}) \in (N(A))^r \backslash \{(0, 0, \cdots, 0)\}$ satisfy the following

$$\sum_{j \in S} \|(u_{1,j}, \cdots, u_{r,j})\| < \sum_{j \in S^c} \|(u_{1,j}, \cdots, u_{r,j})\|, \tag{5}$$

where $N(A)$ stands for the null space of A.

In [18], Foucart and Gribonval studied the MMV setting when $r = 2$, $q = 2$ and $p = 1$. They gave another nice explanation of the problem of MMV. When $r = 2$, one can view that the sparse solution $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ are two components of a complex solution $\mathbf{y} = \mathbf{x}^{(1)} + i\mathbf{x}^{(2)}$ of $A\mathbf{y} = \mathbf{c}$ with $\mathbf{c} = \mathbf{b}^{(1)} + i\mathbf{b}^{(2)}$. Then they recognize that the null space property for $A\mathbf{y} = \mathbf{c}$ for the solution as complex vector is the same as the null space property for $A\mathbf{x} = \mathbf{b}$ for solution as real vector. That is, they proved the following

Theorem 1.2 *Let A be a matrix of size $m \times N$ and $S \subset \{1, \cdots, N\}$ be the support of the sparse vector $\mathbf{y}$. The complex null space property: for any $\mathbf{u} \in N(A), \mathbf{w} \in N(A)$ with $(\mathbf{u}, \mathbf{w}) \neq 0$,*

$$\sum_{j \in S} \sqrt{u_j^2 + w_j^2} < \sum_{j \in S^c} \sqrt{u_j^2 + w_j^2}, \tag{6}$$

where $\mathbf{u} = (u_1, u_2, \cdots, u_N)^T$ and $\mathbf{w} = (w_1, w_2, \cdots, w_N)^T$ is equivalent to the following standard null space property: for any $\mathbf{u}$ in the null space $N(A)$ with $\mathbf{u} \neq 0$,

$$\sum_{j \in S} |u_j| < \sum_{j \in S^c} |u_j|. \tag{7}$$

Furthermore, the researchers in [18] raised two questions. One is to extend their result from $r = 2$ to any $r \geq 3$ and the other one is what happen when $q = 2$ and $p < 1$. These motivate us to study the joint sparse solution recovery.

The study of the ℓ_1 minimization in (3) was generalized to the following ℓ_p setting:

$$\min_{\mathbf{x} \in \mathbb{R}^N} \{\|\mathbf{x}\|_p^p, \quad A\mathbf{x} = \mathbf{b}\} \tag{8}$$

for a fixed number $p \in (0, 1]$ (see for instance, [20], [9] and [19]), in which $\|\mathbf{x}\|_p = \left(\sum_{j=1}^N |x_j|^p\right)^{1/p}$ is the standard ℓ_p quasi-norm when $p \in (0, 1)$. Therefore, we may consider a joint recovery from multiple measurement vectors via

$$\text{minimize} \sum_{j=1}^N \left(\sqrt{x_{1,j}^2 + \cdots + x_{r,j}^2}\right)^p : \text{ subject to } A\mathbf{x}^{(1)} = \mathbf{b}^{(1)}, \cdots, A\mathbf{x}^{(r)} = \mathbf{b}^{(r)} \tag{9}$$

for a given $0 < p \leq 1$, where $\mathbf{x}^{(k)} = (x_{k,1}, \cdots, x_{k,N})^T \in \mathbb{R}^N$ for all $k = 1, \cdots, r$, and this is actually (2) for when $q = 2$. Note that when $p \to 0_+$, we have $\left(\sqrt{x_{1,j}^2 + \cdots + x_{r,j}^2}\right)^p \to 1$ if any of $x_{1,j}, \cdots, x_{r,j}$ is nonzero and hence,

$$\sum_{i=1}^N \left(\sqrt{x_{1,j}^2 + \cdots + x_{r,j}^2}\right)^p \to s$$

which is the joint sparsity of the solution vectors $\mathbf{x}^{(k)}, k = 1, \cdots, r$. Thus, the minimization in (9) makes sense. In fact, the minimization (9) has one advantage over the minimization in (2) when $p = q = 1$. That is, a few measurements are needed for exact recovery by using the ℓ_p minimization with $p < 1$ than the standard ℓ_1 convex minmization. Indeed, in [9], Chartrand demonstrated this fact by numerical examples with Gaussian random matrices and in [10], Chartrand and Staneva showed in theory that the ℓ_p minimization (8) can recover the exact sparse solution by using a less number of measurements when $p \to 0_+$ than the standard convex ℓ_1 minimization. In Section 3 we will give a weaker condition on the sensing matrix for the exact recovery by using the minimization (9) when $r \geq 2$. However, when $p < 1$, the ℓ_p minimization is a nonconvex minimization. Its computation is not as well understood as that of the ℓ_1 minimization. Indeed, the ℓ_1 minimization is convex and is equivalent to the standard linear programming problem which has two matured computational approaches: the interior point method and the simplex method.

In this paper we mainly prove the following

Theorem 1.3 *Let A be a real matrix of size $m \times N$ and $S \subset \{1, 2, \cdots, N\}$ be a fixed index set. Fix $p \in (0, 1]$ and $r \geq 1$. Then the following conditions are equivalent:*
(a) All $\mathbf{x}^{(k)}$ with support in S for $k = 1, \cdots, r$ can be uniquely recovered using (9) ;
(b) For all vectors $\left(\mathbf{u}^{(1)}, \cdots, \mathbf{u}^{(r)}\right) \in (N(A))^r \backslash \{(0, 0, \cdots, 0)\}$

$$\sum_{j \in S} \left(\sqrt{u_{1,j}^2 + \cdots + u_{r,j}^2}\right)^p < \sum_{j \in S^c} \left(\sqrt{u_{1,j}^2 + \cdots + u_{r,j}^2}\right)^p; \tag{10}$$

(c) For all vector $\mathbf{z} \in N(A)$ with $\mathbf{z} \neq 0$,

$$\sum_{j \in S} |z_j|^p < \sum_{j \in S^c} |z_j|^p, \tag{11}$$

where $\mathbf{z} = (z_1, \cdots, z_N)^T \in \mathbb{R}^N$.

That is, it is enough to check (11) for all $\mathbf{z} \in N(A)$ in order to see the uniqueness of the joint sparse solution vectors. This significantly reduces the complexity of verification of (10). Also, our results extend the results in Theorem 1.1 from the norm in it to the quasi-norm. These results completely answer the questions raised in [18].

The paper is organized as follows. In addition to the Introduction above, we shall prove the main results in the next section. Finally, we end the paper with some remarks in §3, where we explain that the proof in [18] can not be extended to prove the second part of the main results in Theorem 1.3.

2 The Proof of Theorem 1.3

We divide the proof of Theorem 1.3 into two parts. The first part is to show that (10) is an if and only if condition for the uniqueness of the joint sparse solution vectors, i.e. (a) and (b) are equivalent. The proof is a straightforward generalization of the arguments in [21]. We spell out the detail as follows.

Let $\mathbf{x}^{(k)}, k = 1, \cdots r$ be the joint sparse solution vectors of the minimization (9) with the assumption that the support of each $\mathbf{x}^{(k)}$ are contained in S. For any vectors $\mathbf{u}^{(1)}, \cdots, \mathbf{u}^{(r)}$ in $N(A)$ with an assumption that they are not simultaneously zero, we easily have, for $0 < p \leq 1$,

$$\sum_{j \in S} \|(x_{1,j}, \cdots, x_{r,j})\|_2^p \leq \sum_{j \in S} \|(u_{1,j}, \cdots, u_{r,j})\|_2^p + \sum_{j \in S} \|(x_{1,j} + u_{1,j}, \cdots, x_{r,j} + u_{r,j})\|_2^p \tag{12}$$

since $0 < p \leq 1$. By the property (10), we have

$$\sum_{j \in S} \|(x_{1,j}, \cdots, x_{r,j})\|_2^p < \sum_{j \in S^c} \|(u_{1,j}, \cdots, u_{r,j})\|_2^p + \sum_{j \in S} \|(x_{1,j} + u_{1,j}, \cdots, x_{r,j} + u_{r,j})\|_2^p. \quad (13)$$

But the support of the vectors $\mathbf{x}^{(k)}, k = 1, \cdots, r$ is contained in S. Hence,

$$\begin{aligned} \sum_{j=1}^{N} \|(x_{1,j}, \cdots, x_{r,j})\|_2^p &= \sum_{j \in S} \|(x_{1,j}, \cdots, x_{r,j})\|_2^p \\ < \ & \sum_{j \in S^c} \|(u_{1,j}, \cdots, u_{r,j})\|_2^p + \sum_{j \in S} \|(x_{1,j} + u_{1,j}, \cdots, x_{r,j} + u_{r,j})\|_2^p \\ = \ & \sum_{j \in S^c} \|(x_{1,j} + u_{1,j}, \cdots, x_{r,j} + u_{r,j})\|_2^p + \sum_{j \in S} \|(x_{1,j} + u_{1,j}, \cdots, x_{r,j} + u_{r,j})\|_2^p. \end{aligned}$$

So $\mathbf{x}^{(k)}, k = 1, \cdots r$ are the unique solution to the minimization problem (9).

For the converse, assume that there are vectors $\mathbf{u}^{(1)}, \cdots, \mathbf{u}^{(r)}$ in $N(A)$ which do not satisfy (10). Let us say $\mathbf{u}^{(1)}, \cdots, \mathbf{u}^{(r)}$ are in $N(A)$ which are not all zero vectors satisfying

$$\sum_{j \in S} \left(\sqrt{u_{1,j}^2 + \cdots + u_{r,j}^2}\right)^p \geq \sum_{j \in S^c} \left(\sqrt{u_{1,j}^2 + \cdots + u_{r,j}^2}\right)^p. \quad (14)$$

We can choose $\mathbf{x}^{(k)} \in \mathbb{R}^N$ such that the entries of $\mathbf{x}^{(k)}$ restricted on S are equal to those of $\mathbf{u}^{(k)}$, and the remaining entries are zeros. Then for multiple measurement vectors $\mathbf{b}^{(k)} := A\mathbf{x}^{(k)}, k = 1, \cdots, r$,

$$A\mathbf{x}^{(k)} = A\mathbf{x}^{(k)} - A\mathbf{u}^{(k)} = A\left(\mathbf{x}^{(k)} - \mathbf{u}^{(k)}\right)$$

and

$$\begin{aligned} \sum_{j=1}^{N} \|(x_{1,j}, x_{2,j}, \cdots, x_{r,j})\|_2^p &= \sum_{j \in S} \|(x_{1,j}, x_{2,j}, \cdots, x_{r,j})\|_2^p \\ &= \sum_{j \in S} \|(u_{1,j}, u_{2,j}, \cdots, u_{r,j})\|_2^p \\ &\geq \sum_{j \in S^c} \|(u_{1,j}, u_{2,j}, \cdots, u_{r,j})\|_2^p \\ &= \sum_{j \in S} \|(x_{1,j} - u_{1,j}, x_{2,j} - u_{2,j}, \cdots, x_{r,j} - u_{r,j})\|_2^p \end{aligned}$$

which contradicts with the uniqueness of the recovery of the new measurement vectors $A\mathbf{x}^{(k)}, k = 1, \cdots, r$. This finishes the proof for the equivalence between (a) and (b).

To prove the second part of the main theorem, let us first show the following

Lemma 2.1 *Let $S \subset \{1, 2, \cdots, N\}$ be an index set with $|S| = s$. Given $0 < p \le 1$ and a matrix $B \in \mathbb{R}^{2\times N}$ with columns $\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_N \in \mathbb{R}^2$, if*

$$\|(x,y)B_S\|_p < \|(x,y)B_{S^c}\|_p \tag{15}$$

for all $(x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}$, then

$$\sum_{k\in S} \|\mathbf{c}_k\|_2^p < \sum_{k\in S^c} \|\mathbf{c}_k\|_2^p . \tag{16}$$

Proof. For convenience, let $B =: (b_{i,j})_{2\times N}$. Without loss of generality we can assume $S := \{1, 2, \cdots, s\}$. We show (16) holds for $s = 1$ first.

If $s = 1$, by the assumption of Lemma 2.1,

$$|b_{1,1}x + b_{2,1}y|^p = \|(x,y)B_S\|_p < \|(x,y)B_{S^c}\|_p = \sum_{j=2}^{N} |b_{1,j}x + b_{2,j}y|^p \tag{17}$$

for all $(x, y) \in \mathbb{R}^2\setminus\{(0, 0)\}$. Choosing $x = \frac{b_{1,1}}{\sqrt{b_{1,1}^2+b_{2,j}^2}}$ and $y = \frac{b_{2,1}}{\sqrt{b_{1,j}^2+b_{2,1}^2}}$ in (17) and applying the Cauchy-Schwartz inequality, we have

$$\begin{aligned}\left(\sqrt{b_{1,1}^2 + b_{2,1}^2}\right)^p &< \sum_{j=2}^{N} \left| \frac{1}{\sqrt{b_{1,1}^2 + b_{2,1}^2}} (b_{1,j}b_{1,1} + b_{2,j}b_{2,1}) \right|^p \\ &\le \sum_{j=2}^{N} \left(\sqrt{b_{1,j}^2 + b_{2,j}^2}\right)^p .\end{aligned}$$

Thus the claim (16) for $s = 1$ follows.

For the case when $s \ge 2$, we have

$$\sum_{j=1}^{s} |b_{1,j}x + b_{2,j}y|^p < \sum_{j=s+1}^{N} |b_{1,j}x + b_{2,j}y|^p \tag{18}$$

for all $(x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}$ by the assumption.

Let $v_j := \left(\frac{b_{1,j}}{\sqrt{b_{1,j}^2+b_{2,j}^2}}, \frac{b_{2,j}}{\sqrt{b_{1,j}^2+b_{2,j}^2}}\right) \in \mathbb{S}^1$ for $j = 1, 2, \cdots, N$, where $\mathbb{S}^1$ stands for the unit circle. Then (18) becomes

$$\sum_{j=1}^{s} \left(\sqrt{b_{1,j}^2 + b_{2,j}^2}\right)^p |\langle v_j, \xi\rangle|^p < \sum_{j=s+1}^{N} \left(\sqrt{b_{1,j}^2 + b_{2,j}^2}\right)^p |\langle v_j, \xi\rangle|^p \tag{19}$$

for all unit vector $\xi \in \mathbb{S}^1$. Taking the integral of (19) over $\mathbb{S}^1$, we have

$$\sum_{j=1}^{s} \left(\sqrt{b_{1,j}^2 + b_{2,j}^2}\right)^p \int_{\mathbb{S}^1} |\langle v_j, \xi \rangle|^p \, d\xi < \sum_{j=s+1}^{N} \left(\sqrt{b_{1,j}^2 + b_{2,j}^2}\right)^p \int_{\mathbb{S}^1} |\langle v_j, \xi \rangle|^p \, d\xi. \tag{20}$$

Note that $\int_{\mathbb{S}^1} |\langle \cdot, \xi \rangle|^p \, d\xi$ is a rotation invariant function from the perspective of integral geometry (cf. [1], [2], and [22]). That is, $\int_{\mathbb{S}^1} |\langle v_j, \xi \rangle|^p \, d\xi$ is constant independent of j. In fact we have

Lemma 2.2 *Fix $r \geq 2$. For any $p > 0$,*

$$\int_{\mathbb{S}^{r-1}} |\langle v, \xi \rangle|^p \, d\xi = C$$

for all $v \in \mathbb{S}^{r-1}$, where $C > 0$ is a constant dependent only on p.

Proof. Let U be an orthogonal transformation of $\mathbb{R}^r$. Then for any $v \in \mathbb{S}^{r-1}$, the sphere of the unit ball in $\mathbb{R}^r$, we have

$$\langle U(v), \xi \rangle = \left\langle v, U^{-1}(\xi) \right\rangle \tag{21}$$

for all $\xi \in \mathbb{S}^{r-1}$. It follows from Riesz representation theorem that

$$\mathbb{S}^{r-1} = \{U(v) : U \in O(r)\}, \tag{22}$$

where $O(r)$ denotes the set of all $r \times r$ orthogonal matrices. By change of variables and using the fact that $|\det(U^{-1})| = 1$, we get

$$\begin{aligned} \int_{\mathbb{S}^{r-1}} |\langle U(v), \xi \rangle|^p \, d\xi &= \int_{\mathbb{S}^{r-1}} |\langle v, U^{-1}(\xi) \rangle|^p \, d\xi \\ &= \int_{\mathbb{S}^{r-1}} |\langle v, U^{-1}(\xi) \rangle|^p \, dU^{-1}(\xi) \\ &= \int_{\mathbb{S}^{r-1}} |\langle v, \xi \rangle|^p \, d\xi \end{aligned} \tag{23}$$

for all $U \in O(r)$. Thus we see that $\int_{\mathbb{S}^{r-1}} |\langle v, \xi \rangle|^p \, d\xi \equiv C$ for some $C > 0$ and for all $v \in \mathbb{S}^{r-1}$. ■

Therefore we apply Lemma 2.2 to (20) to get

$$\sum_{j=1}^{s} \left(\sqrt{b_{1,j}^2 + b_{2,j}^2}\right)^p < \sum_{j=s+1}^{N} \left(\sqrt{b_{1,j}^2 + b_{2,j}^2}\right)^p. \tag{24}$$

With $\mathbf{c}_j = [b_{1,j}, b_{2,j}]^T$ for $j = 1, \cdots, N$, we complete the proof of Lemma 2.1. ■

We are now ready to prove the second part of Theorem 1.3 in case $r = 2$ and $p \in (0, 1)$. That is, we need to show that the real null space property: for any $\mathbf{z} \in N(A)$ with $\mathbf{z} \neq 0$,

$$\|z_S\|_p < \|z_{S^c}\|_p \tag{25}$$

is equivalent to the complex null space property: for any $(\mathbf{v},\mathbf{w})$ in $(N(A))^2\backslash\{(\mathbf{0},\mathbf{0})\}$,

$$\sum_{j\in S}\left(\sqrt{v_j^2+w_j^2}\right)^p < \sum_{j\in S^c}\left(\sqrt{v_j^2+w_j^2}\right)^p, \tag{26}$$

for all $0<p<1$. Note that when $p=1$, this was proved in [18].

Assume that we have (25). For any pair $(\mathbf{v},\mathbf{w})$ of vectors in $(N(A))^2\backslash(0,0)$, we let $B=[\mathbf{v},\mathbf{w}]^T$ be a matrix in $\mathbb{R}^{2\times N}$. Without loss of generality, we may assume that $\mathbf{v}$ and $\mathbf{w}$ are linearly independent. For any real numbers x,y, $\mathbf{z}=x\mathbf{v}+y\mathbf{w}$ is in $N(A)$, the null space property (25) implies (15) for all x,y with $(x,y)\neq(0,0)$. The conclusion of Lemma 2.1 implies the null space property (26) for $r=2$.

It is obvious one can get (25) from (26) by choosing zero vector for the complex part of a nonzero vector in $N(A)$. See [3] for the case when $p=1$. These complete the proof for $r=2$. We now extend our arguments to the setting $r>2$.

We first generalize the proof of Lemma 2.1 to have a general comparison theorem for any matrix $B\in\mathbb{R}^{r\times N}$ for $r\geq 3$. Specifically, we have

Theorem 2.1 (Comparison Theorem) *Let $S\subset\{1,2,\cdots,N\}$ be an index set with $|S|=s$. Given $0<p\leq 1$ and a matrix $B=[b_{ij}]_{1\leq i\leq r,1\leq j\leq N}\in\mathbb{R}^{r\times N}$, if*

$$\|(x_1,\, x_2,\, \cdots,\, x_r)\, B_S\|_p < \|(x_1,\, x_2,\, \cdots,\, x_r)\, B_{S^c}\|_p \tag{27}$$

for all $(x_1,\cdots,x_r)\in\mathbb{R}^r\backslash\{(0,\cdots,0)\}$, then

$$\sum_{j\in S}\left(\sqrt{b_{1,j}^2+\cdots+b_{r,j}^2}\right)^p < \sum_{j\in S^c}\left(\sqrt{b_{1,j}^2+\cdots+b_{r,j}^2}\right)^p. \tag{28}$$

Proof. Let us rewrite (27) as follows.

$$\sum_{j\in S}|b_{1,j}x_1+\cdots+b_{r,j}x_r|^p < \sum_{j\in S^c}|b_{1,j}x_1+\cdots+b_{r,j}x_r|^p \tag{29}$$

for all $(x_1,\, x_2,\, \cdots,\, x_r)\in\mathbb{R}^r\setminus\{(0,\,0,\,\cdots,\,0)\}$. Normalizing $(b_{1,j},\,\cdots,\,b_{r,j})$, we let $v_j:=\frac{1}{\sqrt{b_{1,j}^2+\cdots+b_{r,j}^2}}(b_{1,j},\,\cdots,\,b_{r,j})$. Then we have

$$\sum_{j\in S}\left(\sqrt{b_{1,j}^2+\cdots+b_{r,j}^2}\right)^p|\langle v_j,\xi\rangle|^p < \sum_{j\in S^c}\left(\sqrt{b_{1,j}^2+\cdots+b_{r,j}^2}\right)^p|\langle v_j,\xi\rangle|^p \tag{30}$$

for all vector $\xi=(x_1,x_2,\cdots,x_r)\in\mathbb{R}^r\setminus\{(0,\,0,\,\cdots,\,0)\}$.

Taking the integral of (30) over the unit $(r-1)$-sphere $\mathbb{S}^{r-1}$, we have

$$\begin{aligned} &\textstyle\sum_{j\in S}\left(\sqrt{b_{1,j}^2+\cdots+b_{r,j}^2}\right)^p\int_{\mathbb{S}^{r-1}}|\langle v_j,\xi\rangle|^p\,d\xi \\ <\ &\textstyle\sum_{j\in S^c}\left(\sqrt{b_{1,j}^2+\cdots+b_{r,j}^2}\right)^p\int_{\mathbb{S}^{r-1}}|\langle v_j,\xi\rangle|^p\,d\xi. \end{aligned} \tag{31}$$

By using Lemma 2.2, $\int_{\mathbb{S}^{r-1}} |\langle \cdot, \xi\rangle|^p \, d\xi$ is a positive constant. Therefore, (28) follows. Thus we have proved the general comparison theorem for any matrix $B \in \mathbb{R}^{r\times N}$ for $r \geq 3$. ■

The remaining arguments are similar to the case $r = 2$. Assume that we have (25) $r \geq 3$. For any $\left(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \cdots, \mathbf{u}^{(r)}\right)$ of vectors in $(N(A))^r \backslash \{(\mathbf{0}, \mathbf{0}, \cdots, \mathbf{0})\}$, we let $B = [\mathbf{u}^{(1)}, \cdots, \mathbf{u}^{(r)}]^T$ be a matrix in $\mathbb{R}^{r\times N}$. For any $(x_1, x_2, \cdots, x_r) \in \mathbb{R}^r \setminus \{(0, 0, \cdots, 0)\}$, $\mathbf{z} = (x_1, x_2, \cdots, x_r)\, B$ is in $N(A) \setminus \{\mathbf{0}\}$, the null space property (25) implies (27) for all $(x_1, x_2, \cdots, x_r) \in \mathbb{R}^r \setminus \{(0, 0, \cdots, 0)\}$. The conclusion (28) of Theorem 2.1 implies the null space property (26) for $r \geq 3$. The above discussions show that (c) implies (b) in Theorem 1.3 for $r \geq 3$.

On the other hand, we know that it is trivial to get (c) from (b), because we can just choose $\left(\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \cdots, \mathbf{u}^{(r)}\right) = \left(\mathbf{u}^{(1)}, \mathbf{0} \cdots, \mathbf{0}\right) \in (N(A))^r \backslash \{(0, 0, \cdots, 0)\}$. Hence (b) and (c) are equivalent, and thus we have completed the proof of the main theorem.

3 A Condition for the Exact Recovery using (9)

We now present a sufficient condition to recover the exact sparse solution by using the ℓ_q minimization (9). Let us start with $r = 1$, the setting of single measurement vectors by reviewing some related literature. Let α_s and β_s be the best constants such that

$$\alpha_s \|\mathbf{x}\|_2 \leq \|A\mathbf{x}\|_2 \leq \beta_s \|\mathbf{x}\|_2, \quad \forall \mathbf{x} \in \mathbb{R}^N \text{ with } \|\mathbf{x}\|_0 \leq s.$$

In terms of the well-known restricted isometry constant δ_s (cf. [7] and [6]), we have $\alpha_s^2 = 1 - \delta_s$ and $\beta_s^2 = 1 + \delta_s$. Thus, $\alpha_s > 0$ is equivalent to $\delta_s < 1$. It is easy to see that $\alpha_s > 0$ is a necessary condition to ensure that the sparse solution of linear system $A\mathbf{x} = \mathbf{b}$ can be found by solving a $s \times s$ sub-linear system. Note that α_s is monotonically decreasing as s increase. That is, if $\alpha_{2s+2} > 0$, we have $\alpha_s > 0$. In this case, we have $\gamma_{2s+2} = \dfrac{\beta_{2s+2}^2}{\alpha_{2s+2}^2} < \infty$ since $\beta_{2s+2} \leq \|A\|_2$. By Corollary 2.2 in [19], every s sparse solution is exactly recovered by using the ℓ_p minimization for $p > 0$ small enough. Furthermore, if $\alpha_{2s+1} > 0$, Corollary 2 of [9] showed that the solution of the ℓ_p minimization is the exact solution for a $p \in (0, 1)$. As $\alpha_{2s+1}^2 = 1 - \delta_{2s+1}$, this implies that as long as $\delta_{2s+1} < 1$, the ℓ_p minimization method can find the s-sparse solution.

We now consider the case $r \geq 2$. Let $X = [\mathbf{x}^1, \cdots, \mathbf{x}^r]$ be a matrix of columns $\mathbf{x}^{(i)} = (x_{i,1}, x_{i,2}, \cdots, x_{i,N})^T$ for $i = 1, \cdots, r$. We use the Frobenius norm, i.e.,

$$\|X\|_F = \left(\sum_{j=1}^N (x_{1,j})^2 + \cdots + (x_{r,j})^2\right)^{1/2}.$$

Also, let

$$\|X\|_{2,p} = \left(\sum_{j=1}^N \left((x_{1,j})^2 + \cdots + (x_{r,j})^2\right)^{p/2}\right)^{1/p}$$

be the $(2,p)$-norm for matrix X when $p \leq 1$. Note that $\|X\|_{2,p}$ is a quasi-norm for $0 < p < 1$. We have

Theorem 3.1 *For any measurement matrix A, if there is an integer s such that $\alpha_{2s+1} > 0$ or $\delta_{2s+1} < 1$, then the minimizer of (9) is the joint s sparse solution satisfying (1) if $p > 0$ small enough.*

Proof. Let $X^* = [\mathbf{x}^{*,1}, \cdots, \mathbf{x}^{*,r}]$ be the minimizer of (9) and $X = [\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(r)}]$ be the joint s- sparse solution satisfying (1). Then each column of $V = X - X^*$ is in the null space of A. Writing $V = [(v_{j,1}, \cdots, v_{j,r}), j = 1, \cdots, N]$, we let

$$u_j = (v_{j,1})^2 + \cdots + (v_{j,r})^2, \quad j = 1, \cdots, N$$

and $\mathbf{u} = (u_1, u_2, \cdots, u_N)^T$. We divide the indices set $\{1, \cdots, N\}$ into $S_0 \cup S_1 \cup S_2 \cup \cdots$ with S_0 being the index set of nonzero entries of the s sparse solution X and for $j \geq 1$, S_j being the $s+1$ largest entries of the vector $\mathbf{u}$ over the index set $(S_0 \cup \cdots \cup S_{j-1})^c$, where S_0^c stands for the complement of S_0 in $\{1, 2, \cdots, N\}$ and similar for $(S_0 \cup \cdots \cup S_{j-1})^c$. Note that $AV = 0$ and $AV_{S_0 \cup S_1} = -AV_{(S_0 \cup S_1)^c}$. We have

$$\begin{aligned}
\alpha_{2s+1}\|V_{S_0}\|_F &\leq \alpha_{2s+1}\|V_{S_0} + V_{S_1}\|_F \leq \|AV_{S_0 \cup S_1}\|_F \\
&= \left(\sum_{i=1}^r \langle AV^{(i)}_{(S_0 \cup S_{1,i})^c}, AV^{(i)}_{(S_0 \cup S_{1,i})^c} \rangle \right)^{1/2} \\
&\leq \left(\sum_{i=1}^r \sum_{j,k \geq 2} \|AV^{(i)}_{S_j}\|_2 \|AV_{S_k}\|_2 \right)^{1/2} \\
&\leq \beta_{s+1} \left(\sum_{i=1}^r \sum_{j,k \geq 2} \|V^{(i)}_{S_j}\|_2 \|V^{(i)}_{S_k}\|_2 \right)^{1/2} \\
&\leq \beta_{s+1} \left(\sum_{j,k \geq 2} \|\mathbf{u}_{S_j}\|_2 \|\mathbf{u}_{S_k}\|_2 \right)^{1/2} = \beta_{s+1} \sum_{j \geq 2} \|\mathbf{u}_{S_j}\|_2.
\end{aligned}$$

It is easy to see that $\|\mathbf{u}_{S_j}\|_2 \leq \sqrt{s+1}\|\mathbf{u}_{S_j}\|_\infty \leq (s+1)^{1/2-1/p}\|\mathbf{u}_{S_{j-1}}\|_p$ for $j \geq 2$. Thus we have

$$\sum_{j \geq 2} \|\mathbf{u}_{S_j}\|_2 \leq (s+1)^{1/2-1/p}\|\mathbf{u}_{S_0^c}\|_p.$$

Note that $\|\mathbf{u}_{S_0^c}\|_p = \|V_{S_0^c}\|_{2,p}$. By a standard derivation (cf. [19]), we have $\|V_{S_0^c}\|_{2,p} \leq \|V_{S_0}\|_{2,p}$ and by Cauchy-Schwartz inequality, $\|V_{S_0}\|_{2,p} \leq s^{1/p-1/2}\|V_{S_0}\|_F$. Combining these inequalities above yields

$$\|V_{S_0}\|_F \leq \frac{\beta_{s+1}}{\alpha_{2s+1}} \left(\frac{s}{s+1} \right)^{1/p-1/2} \|V_{S_0}\|_F.$$

When p is small enough the right-hand side of the above inequality can be strictly smaller than $\|V_{S_0}\|_F$ which forces $V_{S_0} = 0$. Then by $\|V_{S_0^c}\|_{2,p} \leq \|V_{S_0}\|_{2,p}$, we have $V_{S_0^c} = 0$ and thus, $X^* = X$. ∎

That is, the minimizer of (9) is the joint sparse solution satisfying (1) under the assumption $\delta_{2s+1} < 1$ on the sensing matrix A as long as p is small enough. That is, if all submatrices with $2s+1$ columns of A is of full rank, we have $\alpha_{2s+1} > 0$ and hence, $\delta_{2s+1} < 1$. However, by the ℓ_1 minimization approach, one has to make δ_{2s} very small, e.g., $\delta_{2s} < \sqrt{2} - 1 \approx 0.414$(cf. [5]) and $\delta_{2s} < 2/(3+\sqrt{3}) \approx 0.4531$ (cf. [19]) in order to find the s-sparse solution. More slightly better sufficient conditions on δ_{2s} can be found in [17], [4], and [23].

Acknowledgement 1 *We would like to thank the anonymous reviewers for their valuable comments which helped greatly improve the presentation of paper.*

References

[1] Alesker, S., Continuous rotation invariant valuations on convex sets, Annals of Mathematics, 149 (1999), 977–1005.

[2] Alesker, S. and J. Bernstein, Range characterization of the cosine transform on higher Grassmannians, Advances in Mathematics, 184 (2004), 367–379.

[3] van den Berg, E. and M. P. Friedlander, Joint-sparse recovery from multiple measurements, arXiv:0904.2051v1, 2009.

[4] Cai, T., L. Wang, G. Xu, Shifting inequality and recovery of sparse signals, IEEE Trans. Signal Process., in press.

[5] Candès, E. J., The restricted isometry property and its implications for compressed sensing, Comptes Rendus de l'Académie des Sciences, Série I, 346 (2008), 589–592.

[6] Candès, E. J., J. K. Romberg, and T. Tao, Stable signal recovery from incomplete and inaccurate measurements. Comm. Pure Appl. Math. 59 (2006), 1207–1223.

[7] Candès, E. J. and T. Tao, Decoding by linear programming, IEEE Trans. Inform. Theory, 51 (2005), no. 12, 4203–4215.

[8] Candès, E. J. and T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies, IEEE Trans. Inform. Theory 52 (2006), no. 12, 5406–5425.

[9] Chartrand, R., Exact reconstruction of sparse signals via nonconvex minimization, IEEE Signal Process. Lett., 14 (2007), 707–710.

[10] Chartrand, R. and V. Staneva, Restricted isometry properties and nonconvex compressive sensing, Inverse Problems, 24(2008) no. 3. doi: 10.1088/0266-5611/24/3/035020.

[11] Chen, J. and X. Huo, Theoretical results on sparse representations of multiple-measurement vectors. IEEE Transactions on Signal Processing, 54:4634–4643, December 2006.

[12] Chen, S. S., D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. SIAM Journal on Scientific Computing, 20(1):33–61, 1998.

[13] Cotter, S. F. and B. D. Rao. Sparse channel estimation via matching pursuit with application to equalization. IEEE Transactions on Communications, 50(3), March 2002.

[14] Cotter, S. F., B. D. Rao, K. Engang, and K. Kreutz-Delgado, Sparse solutions to linear inverse problems with multiple measurement vectors, IEEE Transactions on Signal Processing, 53:2477–2488, July 2005.

[15] Donoho, D. L. and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization, PNAS, 100(5):2197–2202, March 2003.

[16] Fevrier, I. J., S. B. Gelfand, and M. P. Fitz. Reduced complexity decision feedback equalization for multi-path channels with large delay spreads. IEEE Transactions on Communications, 47(6):927–937, June 1999.

[17] Foucart, S., A note on guaranteed sparse recovery via ℓ_1-minimization, to appear in Appl. Comput. Harmon. Anal. (2010).

[18] Foucart, S. and R. Gribonval, Real vs. complex null space properties for sparse vector recovery, Comptes Rendus de l'Academie des Sciences, 2009.

[19] Foucart, S. and M. J. Lai, Sparsest Solutions of Underdetermined Linear Systems via ℓ_q minimization for $0 < q \leq 1$, Harmonic Analysis, 26(2009) 395–407.

[20] Gribonval, R. and M. Nielsen, Sparse representations in unions of bases, IEEE Transactions on Information Theory, 49(12):3320–3325, December 2003.

[21] Gribonval, R. and M. Nielsen, Highly sparse representations from dictionaries are unique and independents of the sparseness measure, Applied and Computational Harmonic Analysis, 22(2007), 335–355.

[22] Lonke, Y., Derivatives of the L_p-cosine transform, Advances in Mathematics, 176 (2003) 175–186.

[23] Malioutov, D., M. Cetin, and A. S. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays. IEEE Transactions on Signal Processing, 53(2005), 3010–3022.

[23] Rudelson, M. and R. Vershyniny, Non-asymptotic theory of random matrices: extreme singular values, Proceedings of the International Congress of Mathematicians Hyderabad, India, 2010.

[24] Spitzer, F. and H. Widom, The circumference of a convex polygon, Proceedings of the American Mathematical Society, 12 (1961), 506–509.

[25] Tropp, J. A., Recovery of short, complex linear combinations via ℓ_1 minimization, IEEE Transactions on Information Theory, 51(4):1568–1570, April 2005.

[26] Tropp, J. A., Algorithms for simultaneous sparse approximation: Part II: Convex relaxation. Signal Processing, 86:589–602, 2006.

[27] Tropp, J. A., A. C. Gilbert, and M. J. Strauss, Algorithms for simultaneous sparse approximation: Part I: Greedy pursuit. Signal Processing, 86:572–588, 2006.