**RESEARCH**

**Open Access**

CrossMark

# Exact recovery of sparse multiple measurement vectors by $l_{2,p}$-minimization

Changlong Wang[1] and Jigen Peng[1,2*]

*Correspondence:
jgpengxjtu@126.com
[1]School of Mathematics and
Statistics, Xi'an Jiaotong University,
Xi'an, Shaanxi 710049, China
[2]School of Mathematics and
Information Science, Guangzhou
University, 230 Guangzhou
University City Outer Ring Road,
Guangzhou, 510006, China

**Abstract**

The joint sparse recovery problem is a generalization of the single measurement vector problem widely studied in compressed sensing. It aims to recover a set of jointly sparse vectors, i.e., those that have nonzero entries concentrated at a common location. Meanwhile $l_p$-minimization subject to matrixes is widely used in a large number of algorithms designed for this problem, i.e., $l_{2,p}$-minimization

$$\min_{X \in \mathbb{R}^{n \times r}} \|X\|_{2,p} \quad \text{s.t. } AX = B.$$

Therefore the main contribution in this paper is two theoretical results about this technique. The first one is proving that in every multiple system of linear equations there exists a constant $p^*$ such that the original unique sparse solution also can be recovered from a minimization in $l_p$ quasi-norm subject to matrixes whenever $0 < p < p^*$. The other one is showing an analytic expression of such $p^*$. Finally, we display the results of one example to confirm the validity of our conclusions, and we use some numerical experiments to show that we increase the efficiency of these algorithms designed for $l_{2,p}$-minimization by using our results.

**MSC:** 15A06; 15A15; 15A48

**Keywords:** sparse recovery; multiple measurement vectors; joint sparse recovery; $l_{2,p}$-minimization

## 1 Introduction

In sparse information processing, one of the central problems is to recover a sparse solution of an underdetermined linear system, such as visual coding [1], matrix completion [2], source localization [3], and face recognition [4]. That is, letting $A$ be an underdetermined matrix of size $m \times n$, and $b \in \mathbb{R}^m$ is a vector representing some signal, so the single measurement vector (SMV) problem is popularly modeled into the following $l_0$-minimization:

$$\min_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{s.t. } Ax = b, \tag{1}$$

where $\|x\|_0$ indicates the number of nonzero elements of $x$.

Furthermore, a natural extension of single measurement vector is the joint sparse recovery problem, also known as the multiple measurement vector (MMV) problem, which arises naturally in source localization [3], neuromagnetic imaging [5], and equalization of

sparse-communication channels [6]. Instead of a single measurement $b$, we are given a set of $r$ measurements,

$$Ax^{(k)} = b^{(k)}, \quad k = 1, \ldots, r, \tag{2}$$

in which the vectors $x^{(k)}$ ($k = 1, \ldots, r$) are joint sparse, i.e., the solution vectors share a common support and have nonzero entries concentrated at common locations.

For a given vector $x$, we define a vector function

$$\|x\|_{2,0} = \begin{cases} 1 & \|x\|_2 \neq 0, \\ 0 & \|x\|_2 = 0. \end{cases} \tag{3}$$

Then the MMV problem can be modeled as the following $l_{2,0}$-minimization problem:

$$\min_{X \in \mathbb{R}^{n \times r}} \|X\|_{2,0} = \sum_{i=1}^{n} \|X_{\text{row } i}\|_{2,0} \quad \text{s.t. } AX = B, \tag{4}$$

where $B = [b^{(1)} \ldots b^{(r)}] \in \mathbb{R}^{m \times r}$ and $X_{\text{row } i}$ is defined as the $i$th row of $X$.

In this paper, we define the support of $X$ by $\text{support}(X) = S = \{i : \|X_{\text{row } i}\|_2 \neq 0\}$ and say that the solution $X$ is $k$-sparse when $|S| \leq k$, and we also say that $X$ can be recovered by $l_{2,0}$-minimization if $X$ is the unique solution of an $l_{2,0}$-minimization problem.

It needs to be emphasized that we cannot regard the solution of multiple measurement vector (MMV) as a combination of several solutions of single measurement vectors, i.e., the solution matrix $X$ to $l_{2,0}$-minimization is not always composed of the solution vectors to $l_0$-minimization.

**Example 1** We consider an underdetermined system $AX = B$, where

$$A = \begin{pmatrix} 2 & 0 & 0 & 1 & 0 \\ 0 & 0.5 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & -0.5 \\ 0 & 0 & 0 & -1 & 0.5 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

If we treat the $AX = B = [b_1 \ b_2]$ as a combination of two single measurement vectors $Ax = b_1$ and $Ax = b_2$, it is easy to verify that each sparse solution to these two problems is $x_1 = [0.5\ 2\ 0\ 0]^T$ and $x_2 = [0\ 0\ 1\ 2]^T$. So let $X^* = [x_1\ x_2]$, it is easy to check that $\|X^*\|_{2,0} = 4$. In fact, it is easy to verify that

$$X = \begin{pmatrix} 0.5 & 0.5 \\ 2 & 2 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

is the solution to $l_{2,0}$-minimization since $\|X\|_{2,0} = 3 < \|X^*\|_{2,0} = 4$.

With this simple Example 1, we should be aware that the MMV problem wants a jointly sparse solution, not a solution which is just composed of sparse vectors. Therefore, the

MMV problem is more complex than the SMV one and needs its own theoretical work. Be inspired by $l_p$-minimization, a popular approach to find the sparest solution to the MMV problem, which is to solve the following $l_{2,p}$-minimization optimization problem:

$$\min_{X \in \mathbb{R}^{n \times r}} \|X\|_{2,p} \quad \text{s.t. } AX = B, \tag{5}$$

where the mixed norm $\|X\|_{2,p}^p = \sum_{i=1}^n \|X_{\text{row } i}\|_2^p$.

In [7], $l_0$-minimization has been proved to be NP-hard because of the discrete and discontinuous nature of $\|x\|_0$. Therefore, it is obviously NP-hard to solve $l_{2,0}$-minimization too. Due to the fact that $\|X\|_{2,0} = \lim_{p \to 0} \|X\|_{2,p}^p$, it seems to be more natural to consider $l_{2,p}$-minimization instead of an NP-hard optimization $l_{2,0}$-minimization than others.

## 1.1 Related work

Many researchers have made a lot of contribution related to the existence, uniqueness, and other properties of $l_{2,p}$-minimization (see [8–11]). Eldar [12] gives a sufficient condition for MMV when $p = 1$, and Unser [13] analyses some properties of the solution to $l_{2,p}$-minimization when $p = 1$. Fourcart and Gribonval [9] studied the MMV setting when $r = 2$ and $p = 1$; they give a sufficient and necessary condition to judge whether a $k$-sparse matrix $X$ can be recovered by $l_{2,p}$-minimization. Furthermore, Lai and Liu [10] consider the MMV setting when $r \geq 2$ and $p \in [0, 1]$, they improve the condition in [9] and give a sufficient and necessary condition when $r \geq 2$.

On the other hand, numerous algorithms have been proposed and studied for $l_{2,0}$-minimization (e.g., [14, 15]). Orthogonal matching pursuit (OMP) algorithms are extended to the MMV problem [16], and convex optimization formulations with mixed norm extend to the corresponding SMV solution [17]. Hyder [15] provides us a robust algorithm for joint sparse recovery, which shows a clear improvement in both noiseless and noisy environments. Furthermore, there exists a lot of excellent work (see[18–22]) presenting us algorithms designed for $l_{2,p}$-minimization. However, it is an important theoretical problem whether there exists a general equivalence relationship between $l_{2,p}$-minimization and $l_{2,0}$-minimization.

In the case $r = 1$, Peng [23] has given a definite answer to this theoretical problem. There exists a constant $p(A, b) > 0$ such that every solution to $l_p$-minimization is also the solution to $l_0$-minimization whenever $0 < p < p(A, b)$.

However, Peng only proves the conclusion when $r = 1$, so it is urgent to extend this conclusion to the MMV problem. Furthermore, Peng just proves the existence of such $p$, he does not give us a computable expression of such $p$. Therefore, the main purpose of this paper is not only to prove the equivalence relationship between $l_{2,p}$-minimization and $l_{2,0}$-minimization, but also present an analysis expression of such $p$ in Section 2 and Section 3.

## 1.2 Main contribution

In this paper, we focus on the equivalence relationship between $l_{2,p}$-minimization and $l_{2,0}$-minimization. Furthermore, it is an application problem that an analytic expression of such $p^*$ is needed, especially in designing some algorithms for $l_{2,p}$-minimization.

In brief, this paper gives answers to two problems which urgently need to be solved:

(I). There exists a constant $p^*$ such that every $k$-sparse $X$ can be recovered by $l_{2,0}$-minimization and $l_{2,p}$-minimization whenever $0 < p < p^*$.

(II). We give such an analytic expression of such $p^*$ based on the restricted isometry property (RIP).

Our paper is organized as follows. In Section 2, we present some preliminaries which play a core role in the proof of our main theorem and prove the equivalence relationship between $l_{2,p}$-minimization and $l_{2,0}$-minimization. In Section 3 we focus on proving another main result of this paper. There we present an analytic expression of such $p^*$. Finally, we summarize our findings in the last section.

### 1.3 Notation

For convenience, for $x \in \mathbb{R}^n$, we define its support by $\text{support}(x) = \{i : x_i \neq 0\}$ and the cardinality of set S by $|S|$. Let $\text{Ker}(A) = \{x \in \mathbb{R}^n : Ax = 0\}$ be the null space of matrix A. We also use the subscript notation $x_S$ to denote a vector that is equal to $x$ on the index set $S$ and zero everywhere else and use the subscript notation $X_S$ to denote a matrix whose rows are those of the rows of $X$ that are in the set index S and zero everywhere else. Let $X_{\text{col } i}$ be the $i$th column in $X$, and let $X_{\text{row } i}$ be the $i$th row in $X$, i.e., $X = [X_{\text{col } 1}, X_{\text{col } 2}, \ldots, X_{\text{col } r}] = [X_{\text{row } 1}, X_{\text{row } 2}, \ldots, X_{\text{row } m}]^T$ for $X \in \mathbb{R}^{n \times r}$. We use $\langle A, B \rangle = \text{tr}(A^T B)$ and $\|A\|_F = \sum_{i,j} |a_{ij}|^2$.

## 2 Equivalence relationship between $l_{2,p}$-minimization and $l_{2,0}$-minimization

At the beginning of this section, we introduce a very important property of the measurement matrix $A$.

**Definition 1** ([24]) A matrix $A$ is said to have the restricted isometry property of order $k$ with restricted isometry constant $\delta_k \in (0, 1)$ if $\delta_k$ is the smallest constant such that

$$(1 - \delta_k)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_k)\|x\|_2^2 \tag{6}$$

for all $k$-sparse vector $x$, where a vector $x$ is said to be $k$-sparse if $\|x\|_0 \leq k$.

Next, we will introduce another important concept named M-null space constant, and this concept is the key to proving the equivalence relationship between $l_{2,0}$-minimization and $l_{2,p}$-minimization.

**Definition 2** For $0 \leq p \leq 1$ and a positive integer $k$, the M-null space constant $h(p, A, r, k)$ is the smallest number such that

$$\|X_S\|_{2,p}^p \leq h(p, A, r, k)\|X_{S^C}\|_{2,p}^p$$

for any index set $S \subset \{1, 2, \ldots, n\}$ with $|S| \leq k$ and any $X \in (\text{Ker}(A))^r \setminus \{(\mathbf{0}, \mathbf{0}, \ldots, \mathbf{0})\}$.

M-NSC provides us a sufficient and necessary condition of the solution to $l_{2,0}$-minimization and $l_{2,p}$-minimization, and it is important for proving the equivalence relationship between these two models. Furthermore, we emphasize a few important properties of $h(p, A, r, k)$.

**Proposition 1** For a given matrix $A$, the M-NSC $h(p, A, r, k)$ defined in Definition 2 is nondecreasing in $p \in [0, 1]$.

*Proof* The proof is divided into two steps.

Step 1: To prove $h(p, A, r, k) \leq h(1, A, r, k)$ for any $p \in [0, 1]$.

For any $X \in (N(A))^r \setminus \{(\mathbf{0}, \mathbf{0}, \ldots, \mathbf{0})\}$, without loss of generality, we assume that $\|X_{\text{row }1}\|_2 \geq \|X_{\text{row }2}\|_2 \geq \cdots \geq \|X_{\text{row }n}\|_2$.

We define a function $\theta(p, X, k)$ as

$$\theta(p, X, k) = \frac{\sum_{i=1}^{k} \|X_{\text{row }i}\|_2^p}{\sum_{i=k+1}^{n} \|X_{\text{row }i}\|_2^p}, \tag{7}$$

then it is easy to get that the definition of $h(p, A, r, k)$ is equivalent to

$$h(p, A, r, k) = \max_{|S| \leq k} \sup_{X \in (N(A))^r \setminus \{(\mathbf{0}, \mathbf{0}, \ldots, \mathbf{0})\}} \theta(p, X, k). \tag{8}$$

For any $p \in [0, 1]$, we notice that the function $f(t) = \frac{t^p}{t}$ $(t > 0)$ is a nonincreasing function. For any $j \in \{k + 1, \ldots, n\}$ and $i \in \{1, 2, \ldots, k\}$, we have that

$$\frac{\|X_{\text{row }j}\|_2^p}{\|X_{\text{row }j}\|_2} \geq \frac{\|X_{\text{row }i}\|_2^p}{\|X_{\text{row }i}\|_2}. \tag{9}$$

We can rewrite inequalities (9) into

$$\frac{\|X_{\text{row }i}\|_2^p}{\|X_{\text{row }j}\|_2^p} \leq \frac{\|X_{\text{row }i}\|_2}{\|X_{\text{row }j}\|_2}. \tag{10}$$

Therefore, we can get that

$$\frac{\sum_{i=1}^{k} \|X_{\text{row }i}\|_2^p}{\|X_{\text{row }j}\|_2^p} \leq \frac{\sum_{i=1}^{k} \|X_{\text{row }i}\|_2}{\|X_{\text{row }j}\|_2}. \tag{11}$$

We can conclude that

$$\frac{\sum_{j=k+1}^{n} \|X_{\text{row }j}\|_2^p}{\sum_{i=1}^{k} \|X_{\text{row }i}\|_2^p} \geq \frac{\sum_{j=k+1}^{n} \|X_{\text{row }j}\|_2}{\sum_{i=1}^{k} \|X_{\text{row }i}\|_2} \tag{12}$$

such that $\frac{1}{\theta(p, X, k)} \geq \frac{1}{\theta(1, X, k)}$, i.e., $\theta(p, X, k) \leq \theta(1, X, k)$.

Because $h(p, A, r, k) = \max_{|S| \leq k} \sup_{X \in (N(A))^r \setminus \{(\mathbf{0}, \mathbf{0}, \ldots)\}} \theta(p, X, k)$, we can get that $h(p, A, r, k) \leq h(1, A, r, k)$.

Step 2: To prove $h(pq, A, r, k) \leq h(p, A, r, k)$ for any $p \in [0, 1]$ and $q \in (0, 1)$.

According to the definition of $\theta(p, X, k)$ in Step 1, we have that

$$\theta(pq, X, k) = \frac{\sum_{i=1}^{k} \|X_{\text{row }i}\|_2^{pq}}{\sum_{j=k+1}^{n} \|X_{\text{row }j}\|_2^{pq}} = \frac{\sum_{i=1}^{k} (\|X_{\text{row }i}\|_2^p)^q}{\sum_{j=k+1}^{n} (\|X_{\text{row }j}\|_2^p)^q} \leq \frac{\sum_{i=1}^{n} \|X_{\text{row }i}\|_2^p}{\sum_{j=k+1}^{n} \|X_{\text{row }j}\|_2^p}. \tag{13}$$

It needs to be pointed out that we have proved the fact in Step 1 that

$$\frac{\sum_{i=1}^{k} |u_i|^p}{\sum_{j=k+1}^{n} |u_j|^p} \leq \frac{\sum_{i=1}^{k} |u_i|}{\sum_{j=k+1}^{n} |u_j|} \tag{14}$$

for any $|u_1| \geq |u_2| \ldots |u_n|$.

Therefore, we can get that $\theta(pq, X, k) \leq \theta(p, X, k)$; in other words, $\theta(p_1, X, k) \leq \theta(p_2, X, k)$ as long as $p_1 \leq p_2$.

Because $h(p, A, r, k) = \max_{|S| \leq k} \sup_{X \in (N(A))^r \setminus \{0,0,\ldots,0\}} \theta(p, X, k)$, so we can get that $h(p, A, r, k)$ is nondecreasing in $p \in [0, 1]$

The proof is completed. □

**Proposition 2** For a given matrix $A$, the M-NSC $h(p, A, r, k)$ defined in Definition 2 is a continuous function in $p \in [0, 1]$

*Proof* As has been proved in Proposition 1, $h(p, A, r, k)$ is nondecreasing in $p \in [0, 1]$ such that there is jump discontinuous if $h(p, A, r, k)$ is discontinuous at a point. Therefore, it is enough to prove that it is impossible to have jump discontinuous points of $h(p, A, r, k)$.

For convenience, we still use $\theta(p, X, S)$ which is defined in the proof of Proposition 1, and the following proof is divided into three steps.

Step 1. To prove that there exist $X \in (N(A))^r$ and a set $S \subset \{1, 2, \ldots, n\}$ such that $\theta(p, X, S) = h(p, A, r, k)$.

Let $V = \{X \in ((N(A))^r) : \|X_{\mathrm{row}\, i}\|_2 = 1, i = 1, 2, \ldots, n\}$, and it is easy to get that $h(p, A, r, k) = \max_{|S| \leq k} \sup_{X \in V} \theta(p, X, S)$

It needs to be pointed out that the choice of the set $S \subset 1, 2, \ldots, n$ with $|S| \leq k$ is limited, so there exists a set $S'$ with $|S'| \leq k$ such that $h(p, A, r, k) = \sup_{X \in V} \theta(p, X, S')$.

On the other hand, $\theta(p, X, S')$ is obviously continuous in $X$ on $V$. Because of the compactness of $V$, there exists $X' \in V$ such that $h(p, A, r, k) = \theta(p, X', S')$.

Step 2. To prove that $\lim_{p \to p_0^-} h(p, A, r, k) = h(p_0, A, r, k)$.

We assume that $\lim_{p \to p_0^-} h(p, A, r, k) \neq h(p_0, A, r, k)$. According to Proposition 1, $h(p, A, r, k)$ is nondecreasing in $p \in [0, 1]$, therefore, we can get a sequence of $\{p_n\}$ with $p_n \to p_0^-$ such that

$$\lim_{p_n \to p_0^-} h(p_n, A, r, k) = M < h(p_0, a, r, k). \tag{15}$$

According to the proof in Step 1, there exist $X' \in (N(A))^r$ and $S \subset \{1, 2, \ldots, n\}$ such that $h(p_0, A, r, k) = \theta(p_0, X', S')$. It is easy to get that $\lim_{p \to p_0^-} \theta(p_n, X, S') = \theta(p, X', S') = h(p_0, A, r, k)$.

However, according to the definition of $\theta(p, X, S)$, it is obvious that

$$h(p_n, A, r, k) \geq \theta(p_n, X', S'); \tag{16}$$

however, (15) and (16) contradict each other.

Therefore, we have that $\lim_{p \to p_0^-} h(p, A, r, k) = h(p_0, A, r, k)$.

Step 3. To prove that $\lim_{p \to p_0^+} h(p, A, r, k) = h(p_0, A, r, k)$ for any $p_0 \in [0, 1)$.

We consider a sequence of $\{p_n\}$ with $p_0 \leq p_n < 1$ and $p \to p_0^+$.

According to Step 1, there exist $X_n \in V$ and $|S_n| \leq k$ such that $h(p_n, A, r, k) = \theta(p_n, X_n, S_n)$. Since the choice of $S \subset \{1, 2, \ldots, n\}$ with $|S| \leq k$ is limited, there exist two subsequences $\{p_{n_i}\}$ of $\{p_n\}$, $\{X_{n_i}\}$ of $\{X_n\}$ and a set $S'$ such that $\theta(p_{n_i}, X_{n_i}, S') = h(p_{n_i}, A, r, k)$.

Furthermore, since $X_n \in V$, it is easy to get a subsequence of $X_{n_i}$ which is convergent. Without loss of generality, we assume that $X_{n_i} \to X'$.

Therefore, we can get that $h(p_{n_i}, A, r, k) = \theta(p_{n_i}, X_{n_i}, S') \to \theta(p_0, X', S')$.

**Figure 1 M-NSC in Example 1.**
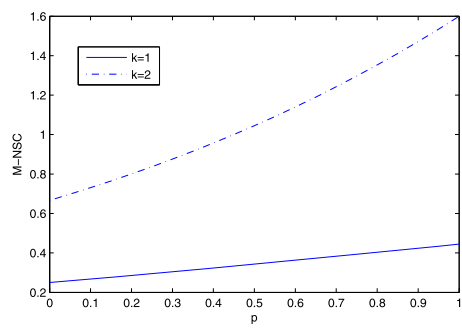
According to the definition of $h(p_0, A, r, k)$, we can get that $\theta(p_0, X', S') \leq h(p_0, A, r, k)$ such that $\lim_{p \to p_0^+} h(p, A, r, k) = h(p_0, A, r, k)$.

Combining Step 2 and Step 3, we show that it is impossible for $h(p, A, r, k)$ to have jump discontinuous.

The proof is completed. □

The concept M-NSC is very important in this paper and it will offer tremendous help in illustrating the performance of $l_{2,0}$-minimization and $l_{2,p}$-minimization; however, M-NSC is difficult to be calculated for a large scale matrix. We show the figure of M-NSC in Example 1 in Figure 1. Combining Propositions 1 and 2, we can get the first main theorem which shows us the equivalence relationship between $l_{2,0}$-minimization and $l_{2,p}$-minimization.

**Theorem 1** *If every $k$-sparse matrix $X$ can be recovered by $l_{2,0}$-minimization, then there exists a constant $p^*(A, B, r)$ such that $X$ also can be recovered by $l_{2,p}$-minimization whenever $0 < p < p^*(A, B, r)$.*

*Proof* First of all, we will prove that $h(0, A, r, k) < 1$ under the assumption. If $h(0, A, r, k) \geq 1$ for some fixed $r$ and $k$, then there exists $X \in (N(A))^r$ such that $\|X_S\|_{2,0} \geq \|X_{S^C}\|_{2,0}$ for a certain set $S \subset \{1, 2, \ldots, n\}$ with $|S| \leq k$. Let $B = AX_S$, then it is obvious that $X_{S^C}$ is a sparser solution than $X_S$, which contradicts the assumption.

By Propositions 1 and 2, since $h(p, A, r, k)$ is continuous and nondecreasing at the point $p = 0$, there exist a constant $p^*(A, B, r)$ and a small enough number $\delta$ that $h(0, A, r, k) < h(p, A, r, k) \leq h(0, A, r, k) + \delta < 1$ for any $p \in (0, p^*(A, B, r))$.

Therefore, for a given $k$-sparse matrix $X^* \in \mathbb{R}^{n \times r}$ and any $0 < p < p^*(A, B, r)$, we have that

$$
\begin{aligned}
\left\| X^* + H \right\|_{2,p}^p &= \left\| \left( X^* + H \right)_{S^*} \right\|_{2,p}^p + \left\| H_{S^{*C}} \right\|_{2,p}^p \\
&\geq \left\| X^* \right\|_{2,p}^p - \left\| H_{S^*} \right\|_{2,p}^p + \left\| H_{S^{*C}} \right\|_{2,p}^p > \left\| X^* \right\|_{2,p}^p.
\end{aligned}
\tag{17}
$$

The last inequality is the result from the inequality $h(p, A, r, k) < 1$, the proof is completed. □

## 3 An analysis expression of such p

In Section 2, we have proved the fact that there exists a constant $p^*(A, B, r)$ such that both $l_{2,p}$-minimization and $l_{2,0}$-minimization have the same solution. However, it is also important to give such an analytic expression of $p^*(A, B, r)$. In Section 3, we focus on giving

an analytic expression of an upper bound of $h(p, A, r, k)$, and we can get the equivalence relationship between $l_{2,p}$-minimization and $l_{2,0}$-minimization as long as $h(p, A, r, k) < 1$ is satisfied. In order to reach our goal, we postpone our main theorems and begin with some lemmas.

**Lemma 1** *For any $X \in \mathbb{R}^{n \times r}$ and $p \in (0, 1]$, we have that*

$$\|X\|_{2,p} \leq \|X\|_{2,0}^{\frac{1}{p}-\frac{1}{2}} \|X\|_F.$$

*Proof* For any $X \in \mathbb{R}^{n \times r}$, without loss of generality, we assume that $\|X_{\text{row } i}\|_2 = 0$ for $i \in \{\|X\|_{2,0} + 1, \dots, n\}$.

According to Hölder's inequality, we can show that

$$\|X\|_{2,p}^p = \sum_{i=1}^{\|X\|_{2,0}} \|X_{\text{row } i}\|_2^p \leq \left( \sum_{i=1}^{\|X\|_{2,0}} \left( \|X_{\text{row } i}\|_2^p \right)^{\frac{2}{p}} \right)^{\frac{p}{2}} \left( \sum_{i=1}^{\|X\|_{2,0}} 1 \right)^{1-\frac{p}{2}} = \|X\|_{2,0}^{1-\frac{p}{2}} \|X\|_F^p, \quad (18)$$

that is, $\|X\|_{2,p} \leq \|X\|_{2,0}^{\frac{1}{p}-\frac{1}{2}} \|X\|_F$. $\square$

**Lemma 2** ([25]) *If $0 < p < q$ and $u_1 \geq \cdots \geq u_k \geq u_{k+1} \geq \cdots \geq u_s \geq u_{s+1} \geq \cdots \geq u_{k+t} \geq 0$, it holds that*

$$\left( \sum_{i=k+1}^{k+t} u_i^q \right)^{\frac{1}{q}} \leq C_{p,q}(k, s, t) \left( \sum_{i=1}^{s} u_i^p \right)^{\frac{1}{p}} \quad (19)$$

*with $C_{p,q}(k, s, t) = \max\{\frac{t^{\frac{p}{q}}}{s}, (\frac{p}{q})^{\frac{p}{q}} (1 - \frac{p}{q})^{1-\frac{p}{q}} k^{\frac{p}{q}-1}\}^{\frac{1}{p}}$.*

**Lemma 3** *For $p \in (0, 1]$, we have that $(\frac{p}{2})^{\frac{1}{2}} (\frac{1}{2-p})^{\frac{1}{2}-\frac{1}{p}} \geq \frac{\sqrt{2}}{2}$.*

*Proof* We denote a function $f(p)$ on the interval $(0, 1]$

$$f(p) = \left( \frac{p}{2} \right)^{\frac{1}{2}} \left( \frac{1}{2-p} \right)^{\frac{1}{2}-\frac{1}{p}}, \quad (20)$$

and we can get that

$$\ln f(p) = \frac{1}{2} \ln \frac{p}{2} - \left( \frac{1}{2} - \frac{1}{p} \right) \ln(2-p). \quad (21)$$

It is easy to get that

$$h(p) = \frac{f'(p)}{f(p)} = \frac{1}{2p} - \left( \frac{1}{p^2} \ln(2-p) - \left( \frac{1}{2} - \frac{1}{p} \right) \frac{1}{2-p} \right)$$

$$= -\frac{1}{p^2} \ln(2-p) \leq 0. \quad (22)$$

Therefore, $f(p)$ is nondecreasing in $p \in (0, 1]$, and we can get $f(p) \geq f(1) = \frac{\sqrt{2}}{2}$.

The proof is completed. $\square$

**Lemma 4** *For any $0 < p \leq 1$, we have that $(1 - \frac{p}{2})^{\frac{1}{p} - \frac{1}{2}} \leq (\frac{\sqrt{2}}{2} - e^{-\frac{1}{2}})p + e^{-\frac{1}{2}}$.*

*Proof* We define a function $\varphi(p)$ on the interval $(0, 1]$ by $\varphi(p) = (1 - \frac{p}{2})^{\frac{1}{p} - \frac{1}{2}}$, and it is easy to get that

$$\ln \varphi(p) = \left(\frac{1}{p} - \frac{1}{2}\right) \ln \left(1 - \frac{p}{2}\right). \tag{23}$$

Now, we take the derivative of both sides of inequality (23) with respect to $p$, it is easy to get that

$$\begin{aligned}
\frac{\varphi'(p)}{\varphi(p)} &= -\frac{1}{p^2} \ln \left(1 - \frac{p}{2}\right) + \left(\frac{1}{p} - \frac{1}{2}\right) \frac{-1}{2 - p} \\
&= \frac{1}{p} \left(\frac{1}{p} \ln \frac{2}{2 - p} - \frac{1}{2}\right) \\
&= \frac{1}{p} \left(\frac{1}{p} \ln \left(1 + \frac{p}{2 - p}\right) - \frac{1}{2}\right) \\
&\geq \frac{1}{p} \left(\frac{1}{p} \left(\frac{p}{2 - p} + \frac{p^2}{2(2 - p)^2}\right) - \frac{1}{2}\right) \\
&= \frac{1}{p} \left(\frac{4 - p}{2(2 - p)^2} - \frac{1}{2}\right) > 0.
\end{aligned} \tag{24}$$

The fourth inequality is due to the fact that $\ln(1 + x) \leq x - \frac{x^2}{2}$ for any $x \in (0, 1]$. The last inequality is the result from a simple inequality $4 - p - (2 - p)^2 = 3p - p^2 > 0$.

Because $\lim_{p \to 0} \varphi(p) = e^{-\frac{1}{2}}$ and $\varphi(1) = \frac{\sqrt{2}}{2}$, we can get that $\varphi(p)$ is an increasing function and $\varphi(p) \leq \frac{\sqrt{2}}{2}$ for $0 < p \leq 1$.

In fact, we have got the result $\varphi'(p) = (\frac{1}{p^2} \ln(1 + \frac{p}{2 - p}) - \frac{1}{2p}) \cdot \varphi(p)$, and it is easy to get

$$\begin{aligned}
\varphi''(p) &= \left(\left(\frac{-2}{p} + \frac{1}{p^2}\right)(\ln 2 - \ln(2 - p)) + \frac{1}{p^2(2 - p)} + \frac{1}{2p^2} - \frac{1}{2p}\right) \cdot \varphi(p) \\
&= \frac{1}{p} \cdot g(p) \cdot \varphi(p),
\end{aligned} \tag{25}$$

where $g(p) = \frac{1}{p(2 - p)} + \frac{1}{2p} - (\frac{2 - p}{p^2} \ln(1 + \frac{p}{2 - p}) + \frac{1}{2})$.

Due to the fact $\ln(1 + x) \leq x - \frac{x^2}{2} + \frac{x^3}{3}$, we have that

$$\begin{aligned}
\ln \left(1 + \frac{p}{2 - p}\right) + \frac{1}{2} &\leq \frac{p}{2 - p} - \frac{p^2}{2(2 - p)} + \frac{p^3}{3(2 - p)} + \frac{1}{2}, \\
&= \frac{3p^3 - p^2 - p + 24}{6p(2 - p)^2}.
\end{aligned} \tag{26}$$

Furthermore, it is easy to get that $\frac{1}{p(2 - p)} + \frac{1}{2p} = \frac{24 - 18p + 3p^2}{6p(2 - p)^2}$ and $g(p) \geq 0$ since $4 > 3p$.

Therefore, $\varphi(p)$ is an increasing convex function in $(0, 1]$. Furthermore, it is easy to get that $\varphi(0) = e^{-\frac{1}{2}}$ and $\varphi(1) = \frac{\sqrt{2}}{2}$. By the property of the convex function, we can conclude that

$$\left(1 - \frac{p}{2}\right)^{\frac{1}{p} - \frac{1}{2}} \leq \left(\frac{\sqrt{2}}{2} - e^{-\frac{1}{2}}\right)p + e^{-\frac{1}{2}}. \tag{27}$$

$\square$

**Lemma 5** *Given an underdetermined matrix $A \in \mathbb{R}^{m \times n}$ which satisfies RIP of order $2k$, we have that*

$$\left| \langle AX_1, AX_2 \rangle \right| \leq \delta_{2k} \|X_1\|_F \|X_2\|_F \tag{28}$$

*for any $k$-sparse matrices $X_1, X_2 \in \mathbb{R}^{n \times r}$ with $\mathrm{support}(X_1) \cap \mathrm{support}(X_2) = \varnothing$.*

*Proof* According to the definition of inner product of matrixes, it is easy to get that

$$\left\| A(X_1 + X_2) \right\|_F^2 = \left\langle A(X_1 + X_2), A(X_1 + X_2) \right\rangle \tag{29}$$

and

$$\left\| \frac{1}{\|X_1\|_F} X_1 + \frac{1}{\|X_2\|_F} X_2 \right\|_F^2 = \left\| \frac{1}{\|X_1\|_F} X_1 - \frac{1}{\|X_2\|_F} X_2 \right\|_F^2 = 2 \tag{30}$$

since $\mathrm{support}(X_1) \cap \mathrm{support}(X_2) = \varnothing$.

Therefore, we can get that

$$
\begin{aligned}
\frac{|\langle AX_1, AX_2 \rangle|}{\|X_1\|_F \|X_2\|_F} &= \left| \left\langle A \frac{1}{\|X_1\|_F} X_1, A \frac{1}{\|X_2\|_F} X_2 \right\rangle \right| \\
&= \frac{1}{4} \left| \left\| A \frac{1}{\|X_1\|_F} X_1 + A \frac{1}{\|X_2\|_F} X_2 \right\|_F^2 - \left\| A \frac{1}{\|X_1\|_F} X_1 - A \frac{1}{\|X_2\|_F} X_2 \right\|_F^2 \right| \\
&\leq \frac{1}{4} \left( (1 + \delta_{2k}) \cdot 2 - (1 - \delta_{2k}) \cdot 2 \right) \\
&= \delta_{2k}.
\end{aligned}
\tag{31}
$$

The proof is completed. $\qquad\square$

Now, we present another main contribution in this paper.

**Theorem 2** *Given an underdetermined matrix $A \in \mathbb{R}^{m \times n}$ which satisfies RIP of order $2k$, for any $p \in (0, 1]$, we can get an upper bound of $h(p, A, r, k)$.*

$$h(p, A, r, k) \leq M(\delta_{2k}, p) = \left( \frac{(\sqrt{2} + 1)\delta_{2k}}{1 - \delta_{2k}} \left( \frac{p}{2} \right)^{\frac{1}{2}} \left( 1 - \frac{p}{2} \right)^{\frac{1}{p} - \frac{1}{2}} \right)^p. \tag{32}$$

*Proof* For any $X \in (N(A))^r \setminus \{(\mathbf{0}, \mathbf{0}, \dots, \mathbf{0})\}$, we define a vector $x \in \mathbb{R}^n$ as

$$x = \left[ \|X_{\mathrm{row}\,1}\|_2, \|X_{\mathrm{row}\,2}\|_2, \dots, \|X_{\mathrm{row}\,n}\|_2 \right]^T, \tag{33}$$

and we consider the index set

$S_0 = \{$indices of the largest $k$ values component of $x\}$,

$S_1 = \{$indices of the largest $k1$ values component of $x$ except $S_0\}$,

$S_2 = \{$indices of the largest $k$ values component of $x$ except $S_0$ and $S_1\}$,

...

$S_t$ = {indices of the rest components of $x$}.

Since $X \in (N(A))^r$, it is obvious that

$$AX = A(X_{S_0} + X_{S_1} + \cdots + X_{S_t}) = (\mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}), \tag{34}$$

and it is easy to get that

$$\begin{aligned}
\left\| A(X_{S_0} + X_{S_1}) \right\|_F^2 &= \left\langle A(X_{S_0} + X_{S_1}), -A(X_{S_2} + X_{S_3} + \cdots + X_{S_t}) \right\rangle \\
&\le \delta_{2k} \left( \|X_{S_0}\|_F + \|X_{S_1}\|_F \right) \cdot \sum_{i=2}^{t} \|X_{S_i}\|_F.
\end{aligned} \tag{35}$$

The second inequality uses the result from Lemma 5. Next, we will give an estimate of $\sum_{i=2}^{t} \|X_{S_i}\|_F$. On the one hand, by (35), it is obvious that

$$\begin{aligned}
\|X_{S_0}\|_F^2 + \|X_{S_1}\|_F^2 &= \|X_{S_0} + X_{S_1}\|_F^2 \\
&\le \frac{1}{1 - \delta_{2k}} \left\| A(X_{S_0} + X_{S_1}) \right\|_F^2 \\
&\le \frac{\delta_{2k}}{1 - \delta_{2k}} \left( \|X_{S_0}\|_F + \|X_{S_1}\|_F \right) \cdot \sum_{i=2}^{t} \|X_{S_i}\|_F.
\end{aligned} \tag{36}$$

On the other hand, we have that

$$\|X_{S_i}\|_F + \|X_{S_{i+1}}\|_F + \|X_{S_{i+2}}\|_F + \|X_{S_{i+3}}\|_F \le 2 \cdot \|X_{S_i \cup S_{i+1} \cup S_{i+2} \cup S_{i+3}}\|_F \tag{37}$$

for any $2 \le i \le t$. Therefore,

$$\sum_{i=2}^{t} \|X_{S_i}\|_F \le 2 \cdot \sum_{i \ge 0} \|X_{S_{4i+2} \cup S_{4i+3} \cup S_{4i+4} \cup S_{4i+5}}\|_F. \tag{38}$$

By Lemma 2, we can get that

$$\|X_{S_i \cup S_{i+1} \cup S_{i+2} \cup S_{i+3}}\|_F \le C(p) \|X_{S_{i-1} \cup S_i \cup S_{i+1} \cup S_{i+2}}\|_{2,p}, \tag{39}$$

where

$$C(p) = \max \left\{ (4k)^{\frac{1}{2} - \frac{1}{p}}, \left( \frac{p}{2} \right)^{\frac{1}{2}} (2-p)^{\frac{1}{p} - \frac{1}{2}} (2k)^{\frac{1}{2} - \frac{1}{p}} \right\}.$$

By Lemma 3, we can conclude that

$$C(p) = \left( \frac{p}{2} \right)^{\frac{1}{2}} (2-p)^{\frac{1}{p} - \frac{1}{2}} (2k)^{\frac{1}{2} - \frac{1}{p}} \tag{40}$$

since $\left( \frac{p}{2} \right)^{\frac{1}{2}} \left( \frac{1}{2-p} \right)^{\frac{1}{2} - \frac{1}{p}} \ge \frac{\sqrt{2}}{2}$.

Therefore, we can get that

$$\sum_{i=2}^{t} \|X_{S_i}\|_F \le 2 \sum_{i \ge 0} \|X_{S_{4i+2} \cup S_{4i+3} \cup S_{4i+4} \cup S_{4i+5}}\|_F$$

$$\le 2 \cdot \left(\frac{p}{2}\right)^{\frac{1}{2}} (2-p)^{\frac{1}{p}-\frac{1}{2}} (2k)^{\frac{1}{2}-\frac{1}{p}} \sum_{i \ge 0} \|X_{S_{4i+1} \cup S_{4i+2} \cup S_{4i+3} \cup S_{4i+4}}\|_{2,p}$$

$$\le 2 \cdot \left(\frac{p}{2}\right)^{\frac{1}{2}} (2-p)^{\frac{1}{p}-\frac{1}{2}} (2k)^{\frac{1}{2}-\frac{1}{p}} \|X_{S_0^C}\|_{2,p}. \tag{41}$$

Therefore, we have that

$$\|X_{S_0}\|_F^2 + \|X_{S_1}\|_F^2 \le \frac{\delta_{2k}}{1-\delta_{2k}} \left(\|X_{S_0}\|_F + \|X_{S_1}\|_F\right) \left(\sum_{i=2}^{t} \|X_{S_i}\|_F\right)$$

$$\le \Phi \cdot \left(\|X_{S_0}\|_F + \|X_{S_1}\|_F\right), \tag{42}$$

where $\Phi = \frac{2\delta_{2k}}{1-\delta_{2k}} \left(\frac{p}{2}\right)^{\frac{1}{2}} (2-p)^{\frac{1}{p}-\frac{1}{2}} (2k)^{\frac{1}{2}-\frac{1}{p}} \|X_{S_0^C}\|_{2,p}$.

By (42), it is easy to get that

$$\left(\|X_{S_0}\|_F - \frac{\Phi}{2}\right)^2 + \left(\|X_{S_1}\|_F - \frac{\Phi}{2}\right)^2 \le \frac{\Phi^2}{2}. \tag{43}$$

Therefore, we can get that

$$\|X_{S_0}\|_F \le \frac{\sqrt{2}+1}{2} \Phi. \tag{44}$$

By Lemma 1,

$$\|X_{S_0}\|_{2,p} \le k^{\frac{1}{p}-\frac{1}{2}} \|X_{S_0}\|_F \le k^{\frac{1}{p}-\frac{1}{2}} \frac{\sqrt{2}+1}{2} \Phi. \tag{45}$$

By the definition of M-NSC, we can conclude that

$$h(p, A, r, k) \le M(\delta_{2k}, p) = \left(\frac{(\sqrt{2}+1)\delta_{2k}}{1-\delta_{2k}} \left(\frac{p}{2}\right)^{\frac{1}{2}} \left(1-\frac{p}{2}\right)^{\frac{1}{p}-\frac{1}{2}}\right)^p. \tag{46}$$

**Theorem 3** *Given a matrix $A \in R^{m \times n}$ with $m \le n$, every $k$-sparse $X^*$ can also be recovered by $l_{2,p}$-minimization, for any $0 < p \le p^*(A)$, where*

$$p^*(A) = \left(\sqrt[3]{-\frac{w_2}{2} + \sqrt{\frac{w_2^2}{4} + \frac{w_1^3}{27}}} + \sqrt[3]{-\frac{w_2}{2} - \sqrt{\frac{w_2^2}{4} + \frac{w_1^3}{27}}}\right)^2, \tag{47}$$

*and $w_1 = \frac{e^{-\frac{1}{2}}}{\frac{\sqrt{2}}{2} - e^{-\frac{1}{2}}}$ and $w_2 = -\frac{(1-\delta_{2k})^2}{(2+\sqrt{2})(\frac{\sqrt{2}}{2} - e^{-\frac{1}{2}})\delta_{2k}}$.*

*Proof* According to Theorem 2, we can get the equivalence between $l_{2,0}$-minimization and $l_{2,p}$-minimization as soon as $M(\delta_{2k}, p) < 1$, where $M(\delta_{2k}, p)$ is defined in Theorem 2.

$$\frac{(\sqrt{2}+1)\delta_{2k}}{1-\delta_{2k}}\left(\frac{p}{2}\right)^{\frac{1}{2}}\left(1-\frac{p}{2}\right)^{\frac{1}{p}-\frac{1}{2}} < 1. \tag{48}$$

However, we realize the inequality $M(\delta_{2k}, p) < 1$ is difficult to solve, so we need to change this equality into another form. By Lemma 4, it is easy to get that

$$M(\delta_{2k}, p) \leq \frac{(\sqrt{2}+2)\delta_{2k}}{2(1-\delta_{2k})}\sqrt{p}\left(\left(\frac{\sqrt{2}}{2}-e^{-\frac{1}{2}}\right)p + e^{-\frac{1}{2}}\right). \tag{49}$$

Let $\sqrt{p} = \eta$, and we consider the following inequality:

$$\eta^3 + w_1\eta + w_2 < 0, \tag{50}$$

where $w_1 = \frac{e^{-\frac{1}{2}}}{\frac{\sqrt{2}}{2}-e^{-\frac{1}{2}}}$ and $w_2 = -\frac{(1-\delta_{2k})^2}{(2+\sqrt{2})\delta_{2k}(\frac{\sqrt{2}}{2}-e^{-\frac{1}{2}})}$. This cubic inequality is easy to be solved. We can get the solution of this inequality by Cardano's formula, and we can conclude $M(\delta_{2k}, p) < 1$ when $0 < p < p^*(A)$, where

$$p < p^*(A) = \left(\sqrt[3]{-\frac{w_2}{2}+\sqrt{\frac{w_2^2}{4}+\frac{w_1^3}{27}}}+\sqrt[3]{-\frac{w_2}{2}-\sqrt{\frac{w_2^2}{4}+\frac{w_1^3}{27}}}\right)^2. \tag{51}$$

Therefore, for a given $k$-sparse matrix $X^* \in \mathbb{R}^{n\times r}$ and any $0 < p < p^*$, we have that

$$\begin{aligned}\left\|X^*+H\right\|_{2,p}^p &= \left\|\left(X^*+H\right)_{S^*}\right\|_{2,p}^p + \|H_{S^{*C}}\|_{2,p}^p \\ &\geq \left\|X^*\right\|_{2,p}^p - \|H_{S^*}\|_{2,p}^p + \|H_{S^{*C}}\|_{2,p}^p > \left\|X^*\right\|_{2,p}^p.\end{aligned} \tag{52}$$

The last inequality is the result from the fact $M(\delta_{2k}, p) < 1$.

The proof is completed. $\qquad\square$

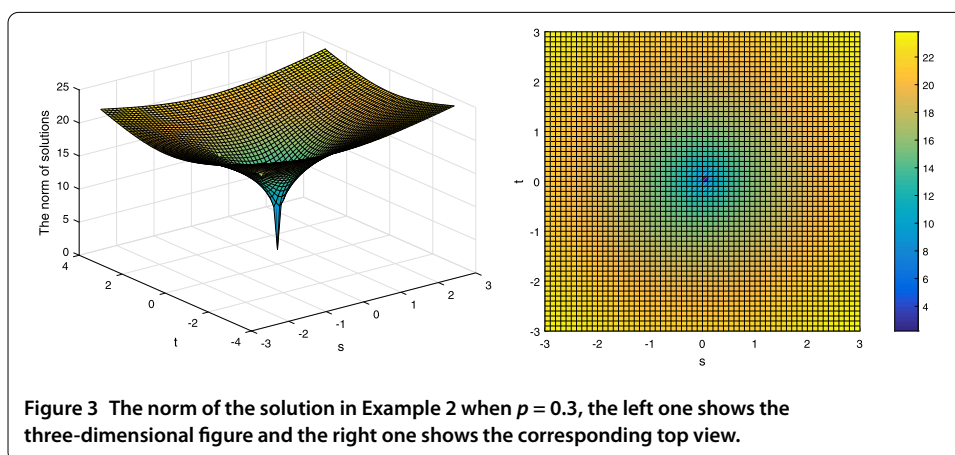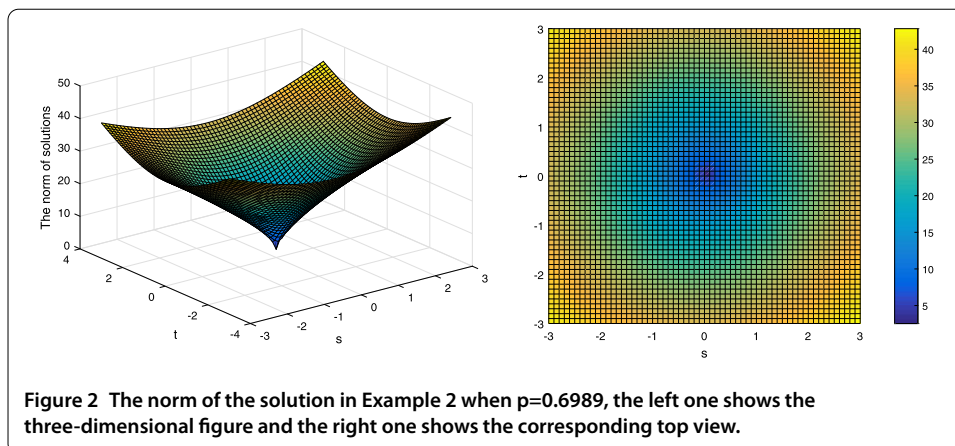Now, we present one example to demonstrate the validation of our main contribution in this paper.

**Example 2** We consider an underdetermined system $AX = B$, where $A \in \mathbb{R}^{12\times 13}$

$$A = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0.3 \\ 0 & 1 & \ldots & 0 & 0.3 \\ 0 & \ddots & \ldots & \vdots & 0.3 \\ 0 & 0 & \ldots & 1 & 0.3 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}.$$

It is easy to verify that the unique sparse solution to $l_{2,0}$-minimization is

$$X^* = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} \tag{53}$$

and $N(A) = \{ax : a \in \mathbb{R}, x = [1, 1, \ldots, 1, 10/3]^T\}$.

**Figure 2** **The norm of the solution in Example 2 when p=0.6989, the left one shows the three-dimensional figure and the right one shows the corresponding top view.**



**Figure 3** **The norm of the solution in Example 2 when $p = 0.3$, the left one shows the three-dimensional figure and the right one shows the corresponding top view.**

So the solution to $AX = B$ can be expressed as the following form:

$$X = \begin{pmatrix} 1+s & 1+t \\ 1+s & -1+t \\ s & t \\ \vdots & \vdots \\ \frac{10}{3}s & \frac{10}{3}t \end{pmatrix}, \tag{54}$$

where $s, t \in \mathbb{R}$ such that

$$\|X\|_{2,p}^p = \left(\sqrt{(1+s)^2 + (1+t)^2}\right)^p + \left(\sqrt{(1+s)^2 + (-1+t)^2}\right)^p + \frac{40}{3}\left(\sqrt{s^2 + t^2}\right)^p.$$

Then we will verify the result in Theorem 3. It is easy to get that $p^* = 0.6989$ since $\delta_4 = 0.5612$, and we show the cases when $p = 0.6989$ and $p = 0.3$ in Figures 2 and 3.

It is obvious that $\|X\|_{2,p}$ has the minimum point at $s = t = 0$, which is the original solution to $l_{2,0}$-minimization.

## 4 Numerical experiment

Although $l_{2,p}$-minimization is difficult to be solved, there are a lot algorithms designed for this problem. In this section, we adopt two excellent algorithms presented in [18] and [21].

---

**Algorithm 1** [21]

---

**Input:** Given $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{m \times r}$.

1: For $k = 1, 2, \ldots r$

2: Solve $l_1$-minimization problem,

$\quad x_k = \arg \min_{x \in \mathbb{R}^n} \|x\|_0$ s.t. $Ax = b_{\text{col},k}$.

$\quad S_k = \text{support}(x_k)$.

3: $|S_k| < \frac{m}{2}$

4: $X = \arg \min_{\text{support}(X) \subset S} \|X\|_F$ s.t. $AX = B$.

5: $A_S X = B$

6: $X^* = X$.

7: Else

8: Return fail.

9: Else

10: Return fail.

11: Return solution $X^*$.

---

**Algorithm 2** [18]

---

**Input:** Given $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{m \times r}$ and $p \in (0, 1]$.

1: Set $D_1 = I_n$.

2: For $k = 1, 2, \ldots$ until convergence

$\quad X_{k+1} = D_k^{-1} A^T (A D_k^{-1} A^T)^{-1} B$.

$\quad$ Update $D_{k+1}^{-1}$ with diagonal entries: $\|(X_{k+1})\text{row } i\|_2^{2-p}$, $i = 1, 2, \ldots, n$.

---

**Algorithm 3** [18]

---

**Input:** Given $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{m \times r}$ and $p \in (0, 1]$.

1: QR decompose $A^T = Q_1 R$, where $Q_1 \in \mathbb{R}^{n \times m}$ and $R \in \mathbb{R}^{m \times m}$.

2: $P = I_n - Q_1 Q_1^T$ and $X_1 = Q_1 (R^{-1})^T B$.

3: For $k = 1, 2, \ldots$ until convergence

$\quad D_k = \text{diag}\{\|(X_k)_{\text{row } 1}\|_2^{p-2}, \|(X_k)_{\text{row } 2}\|_2^{p-2}, \ldots, \|(X_k)_{\text{row } n}\|_2^{p-2}\}$.
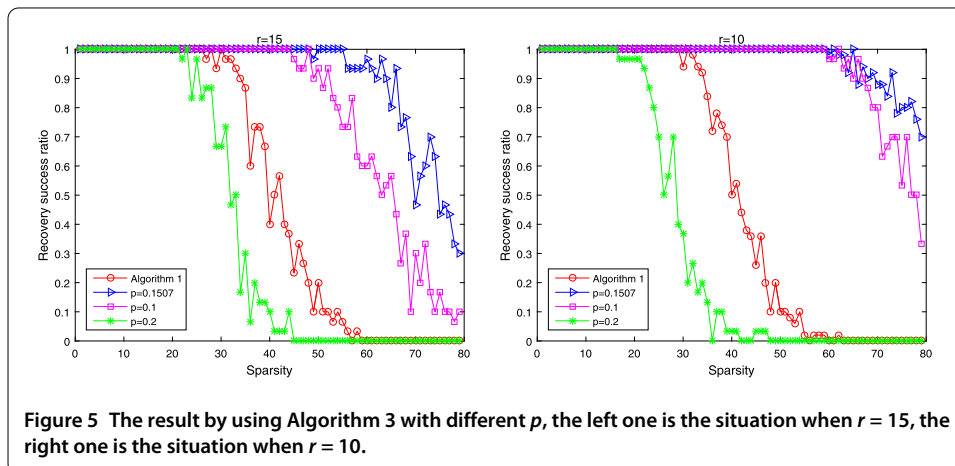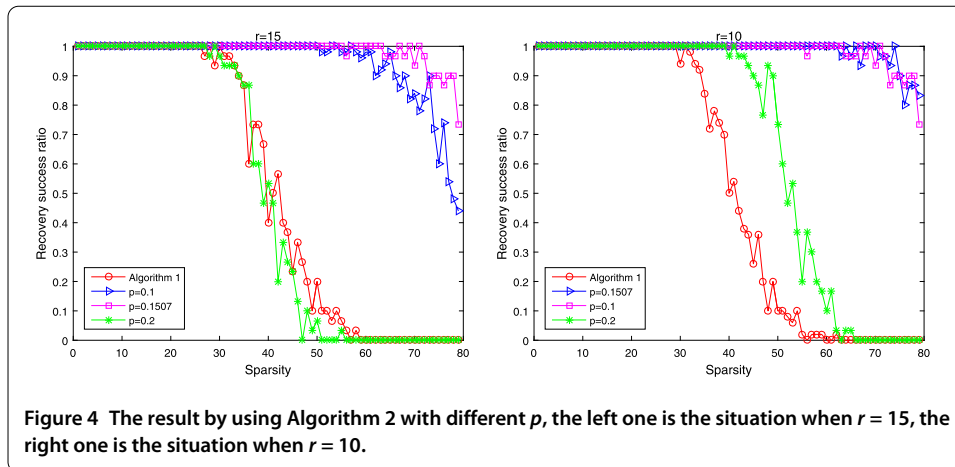
$\quad S_k = -P D_k X_k$.

$\quad \alpha_k = -\frac{\text{Tr}(S_k^T D_k X_k)}{\text{Tr}(S_k^T D_k S_k)}$.

$\quad X_{k+1} = X_k + \alpha_k S_k$.

---

The main reason why we choose these algorithms is not only their efficient performance and theoretical results but also the feature that any $p \in (0, 1)$ can be applied in these two algorithms. The details of these algorithms are presented in Algorithms 1, 2 and 3. However, we need to underline the choice of the parameters in Algorithms 2 and 3, not the smaller the better, a reasonable $p$ has a key role in these algorithms. Therefore, it is urgent for us to see whether these algorithms can do better by using our result.

In order to set reference standards for these $l_{2,p}$-minimization algorithms, we will consider $l_{2,1}$-minimization

$$\min_{X \in \mathbb{R}^{n \times r}} \|X\|_{2,1} = \sum_{i=1}^{n} \|X_{\text{row } i}\|_2 \quad \text{s.t. } AX = B. \tag{55}$$

**Figure 4** The result by using Algorithm 2 with different *p*, the left one is the situation when *r* = 15, the right one is the situation when *r* = 10.



**Figure 5** The result by using Algorithm 3 with different *p*, the left one is the situation when *r* = 15, the right one is the situation when *r* = 10.

Different to other $p \in (0, 1)$, $l_{2,1}$-minimization is a convex optimization problem which can be solved efficiently; especially, this problem can be transformed into a linear programming problem when $r = 1$. In this section, we will adopt Algorithm 1 designed for $l_{2,1}$-minimization, and we will adopt a $256 \times 1024$ measurement $A$ with $\delta_{100} = 0.8646$, and we can get $p^*(A) = 0.1507$ by our result.

Firstly, we look at the relations between sparsity and recovery success ratio. As shown in Figures 4 and 5, the performance of $p = 0.1507$ is better than the choice $p = 0.2$ and $l_{2,1}$-minimization in both of these two algorithms. The results show us that our result helps these algorithms to increase the efficiency.

Secondly, we look at the relations between the relative error (RE) and sparsity. We define the relative error by

$$RE = \frac{\|X^* - X_{sol}\|_F}{\|X^*\|_F}.$$

As shown in Figure 6, our result performs better than a larger choice. However, we need to emphasize the fact that our result may not be the optimal choice, only a little larger is permitted. In our experiments, the choice of $p = 0.16$ is still good, but its performance begins to get worse when $p = 0.19$.
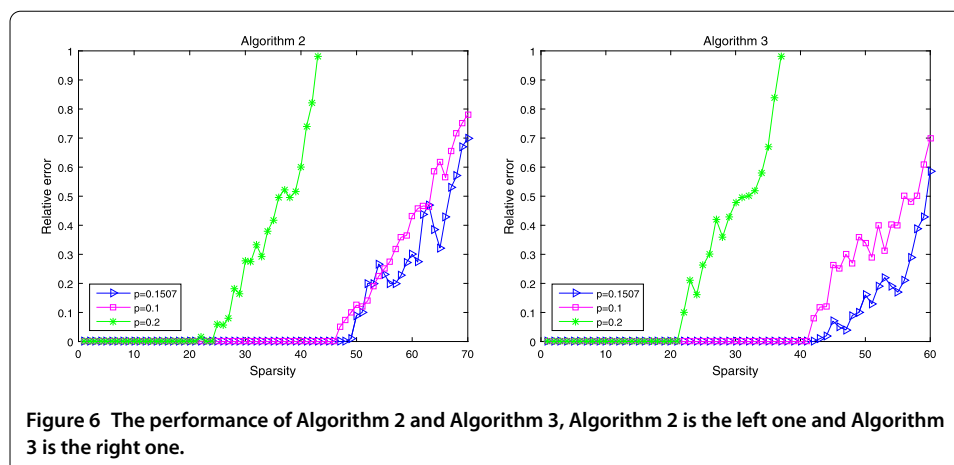
**Figure 6 The performance of Algorithm 2 and Algorithm 3, Algorithm 2 is the left one and Algorithm 3 is the right one.**

## 5 Conclusion

In this paper we have studied the equivalence relationship between $l_{2,0}$-minimization and $l_{2,p}$-minimization, and we have given an analysis expression of such $p^*$.

Furthermore, it needs to be pointed out that the conclusion in Theorems 2 and 3 is valid in a single measurement vector problem, i.e., $l_p$-minimization also can recover the original unique solution to $l_0$-minimization when $0 < p < p^*$.

However, the analysis expression of such $p^*$ in Theorem 3 may not be the optimal result. In this paper, we have considered all the underdetermined matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{m \times r}$ from a theoretical point of view. So the result can be improved with a particular structure of the matrices $A$ and $B$. The authors think the answer to this problem will be an important improvement of the application of $l_{2,p}$-minimization. In conclusion, the authors hope that in publishing this paper a brick will be thrown out and be replaced with a gem.

**Competing interests**
The authors declare that they have no competing interests.

**Authors' contributions**
All authors contributed equally to this work. All authors read and approved the final manuscript.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**References**
1. Olshausen, BA, Field, DJ: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature **381**(6583), 607 (1996)
2. Candès, EJ, Recht, B: Exact matrix completion via convex optimization. Found. Comput. Math. **9**(6), 717 (2009)
3. Malioutov, D, Cetin, M, Willsky, AS: A sparse signal reconstruction perspective for source localization with sensor arrays. IEEE Trans. Signal Process. **53**(8), 3010-3022 (2005)
4. Wright, J, Ganesh, A, Zhou, Z, Wanger, A, Ma, Y: Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. **31**(2), 210 (2009)
5. Cotter, SF, Rao, BD: Sparse channel estimation via matching pursuit with application to equalization. IEEE Trans. Commun. **50**(3), 374-377 (2002)
6. Fevrier, IJ, Gelfand, SB, Fitz, MP: Reduced complexity decision feedback equalization for multipath channels with large delay spreads. IEEE Trans. Commun. **47**(6), 927-937 (1999)
7. Natarajan, BK: Sparse approximate solutions to linear systems. SIAM J. Comput. **24**(2), 227-234 (1995)

8. Liao, A, Yang, X, Xie, J, Lei, Y: Analysis of convergence for the alternating direction method applied to joint sparse recovery. Appl. Math. Comput. **269**, 548-557 (2015)

9. Foucart, S, Gribonval, R: Real versus complex null space properties for sparse vector recovery. C. R. Math. **348**(15), 863-865 (2010)

10. Lai, MJ, Liu, Y: The null space property for sparse recovery from multiple measurement vectors. Appl. Comput. Harmon. Anal. **30**(3), 402-406 (2011)

11. Van Den Berg, E, Friedlander, MP: Theoretical and empirical results for recovery from multiple measurements. IEEE Trans. Inf. Theory **56**(5), 2516-2527 (2010)

12. Eldar, YC, Michaeli, T: Beyond bandlimited sampling. IEEE Signal Process. Mag. **26**(3), 48-68 (2009)

13. Unser, M: Sampling 50 years after Shannon. Proc. IEEE **88**(4), 569-587 (2000)

14. Hyder, MM, Mahata, K: Direction-of-arrival estimation using a mixed $l_{2,0}$ norm approximation. IEEE Trans. Signal Process. **58**(9), 4646-4655 (2010)

15. Hyder, MM, Mahata, K: A robust algorithm for joint-sparse recovery. IEEE Signal Process. Lett. **16**(12), 1091-1094 (2009)

16. Tropp, JA, Gilbert, AC: Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans. Inf. Theory **53**(12), 4655-4666 (2007)

17. Milzarek, A, Ulbrich, M: A semismooth Newton method with multidimensional filter globalization for $l_1$-optimization. SIAM J. Optim. **24**(1), 298-333 (2014)

18. Wang, L, Chen, S, Wang, Y: A unified algorithm for mixed, $l_{2,p}$ -minimizations and its application in feature selection. Comput. Optim. Appl. **58**(2), 409-421 (2014)

19. Kowalski, M: Sparse regression using mixed norms. Appl. Comput. Harmon. Anal. **27**(3), 303-324 (2009)

20. Zhao, M, Zhang, H, Cheng, W, Zhang, Z: Joint $l_p$ and $l_{2,p}$-norm minimization for subspace clustering with outlier pursuit. In: International Joint Conference on Neural Networks IEEE, pp. 3658-3665 (2016)

21. Van Den Berg, E, Friedlander, MP: Joint-sparse recovery from multiple measurements. Mathematics (2009)

22. Cotter, SF, Rao, BD, Engan, K, Kreutz-Delgado, K: Sparse solutions to linear inverse problems with multiple measurement vectors. IEEE Trans. Signal Process. **53**(7), 2477-2488 (2005)

23. Peng, J, Yue, S, Li, H: NP/CMP equivalence: a phenomenon hidden among sparsity models, $l_0$-minimization and $l_p$-minimization for information processing. IEEE Trans. Inf. Theory **61**(7), 4028-4033 (2015)

24. Candès, EJ, Tao, T: Decoding by linear programming. IEEE Trans. Inf. Theory **51**(12), 4203-4215 (2005)

25. Foucart, S: Sparse recovery algorithms: sufficient conditions in terms of restricted isometry constants. Springer Proc. Math. Stat. **13**, 65-77 (2012)