

RESEARCH

Open Access

On recovery of block-sparse signals via mixed l_2/l_q ($0 < q \leq 1$) norm minimization

Yao Wang¹, Jianjun Wang^{2*} and Zongben Xu¹

Abstract

Compressed sensing (CS) states that a sparse signal can exactly be recovered from very few linear measurements. While in many applications, real-world signals also exhibit additional structures aside from standard sparsity. The typical example is the so-called block-sparse signals whose non-zero coefficients occur in a few blocks. In this article, we investigate the mixed l_2/l_q ($0 < q \leq 1$) norm minimization method for the exact and robust recovery of such block-sparse signals. We mainly show that the non-convex l_2/l_q ($0 < q < 1$) minimization method has stronger sparsity promoting ability than the commonly used l_2/l_1 minimization method both practically and theoretically. In terms of a block variant of the restricted isometry property of measurement matrix, we present weaker sufficient conditions for exact and robust block-sparse signal recovery than those known for l_2/l_1 minimization. We also propose an efficient Iteratively Reweighted Least-Squares (IRLS) algorithm for the induced non-convex optimization problem. The obtained weaker conditions and the proposed IRLS algorithm are tested and compared with the mixed l_2/l_1 minimization method and the standard l_q minimization method on a series of noiseless and noisy block-sparse signals. All the comparisons demonstrate the outperformance of the mixed l_2/l_q ($0 < q < 1$) method for block-sparse signal recovery applications, and meaningfulness in the development of new CS technology.

Keywords: Block-sparse recovery, Compressed sensing, Block-RIP, l_2/l_q minimization, IRLS

1 Introduction

According to the Shannon/Nyquist sampling theorem [1,2], if we would like to avoid lose of information when capturing a signal, we must sample the signal at the so-called Nyquist rate, which means twice the highest frequency of the signal. Since the theorem only exploits the bandlimitedness of a signal and most real-world signals are sparse or compressible, the process of massive data acquisition based on Shannon/Nyquist sampling theorem usually samples too many useless information and eventually we have to compress to store or encode a very few essential information of the signal. Obviously, this process is extremely wasteful and therefore, a more effective sampling way to directly acquire the essential information of a signal has been expected.

Compressed sensing (CS) [3-5] was motivated by this purpose and it can completely acquire the essential

information of a signal by exploiting its compressibility. In a word, the main contribution of CS is that it presents a new efficient scheme to capture and recover the compressible or sparse signals at a reduced sampling rate far below the Nyquist rate. The basic principle of CS is that it first employs non-adaptive linear projections to preserve the structure of the signal; then one can exactly recover these signals from a surprisingly small number of random linear measurements through a non-linear optimization procedure (such as l_1 -minimization) if the measurement matrix satisfies some suitable sufficient conditions in terms of *restricted isometry property* (RIP, [6]). Consequently, CS implies that it is indeed possible to acquire the data in already compressed form. Nowadays, driven by a wide range of applications, CS and other related problems have attracted much interest in various communities, such as in signal processing, machine learning, and statistics.

Different from general sparse signals in conventional sense, some real-world signals may exhibit some additional structures, i.e., the non-zero coefficients appear in

*Correspondence: wjj@swu.edu.cn

²School of Mathematics and Statistics, Southwest University, Chongqing 400715, China

Full list of author information is available at the end of the article

a few fixed blocks, we refer to these signals as block-sparse signals in this article. Such block-sparse signals arise in various application problems, say, DNA microarrays [7,8], equalization of sparse communication channels [9], source localization [10], wideband spectrum sensing [11], and color imaging [12].

Using standard convex relaxation (l_1 -minimization) as conventional CS framework for recovering the block-sparse signal does not exploit the fact that the non-zero elements of the signal appear in consecutive positions. Therefore, one natural idea is considering the block version of l_1 -minimization, i.e., the mixed l_2/l_1 -minimization, to exploit the block-sparsity. Many previous works have shown that the mixed l_2/l_1 -minimization is superior to standard l_1 -minimization when dealing with such block-sparse signals [13-15]. Huang and Zhang [13] developed a theory for the mixed l_2/l_1 -minimization by using a concept called *strong group sparsity* and they demonstrated that the mixed norm minimization is very efficient for recovering strongly group-sparse signals. Stojnic et al. [15] obtained an optimal number of Gaussian measurements for uniquely recovering a block-sparse signal through the mixed l_2/l_1 norm minimization. By generalizing the conventional RIP notion to the block-sparse case, Eldar and Mishali [14] showed that if the measurement matrix D has the same restricted isometry constant as that in the l_1 case, then the mixed norm method is guaranteed to exactly recover any block-sparse signals in the noiseless case. Furthermore, they also showed that the block-sparse signal recovery would be robust in the noisy case under the same recovery condition. Another common approach to deal with the block-sparsity problem is by suitably extending the standard greedy methods, such as orthogonal matching pursuit, iterative hard thresholding (IHT), and compressive sampling matching pursuit (CoSaMP), to the block-sparse case. In [16], the CoSaMP algorithm and the IHT algorithm were extended to the model-based setting, treating the block-sparsity as a special case. It was also shown that the new recovery algorithms demonstrate provable performance guarantees and robustness properties. Eldar et al. [17] generalized the notion of coherence to the block-sparse setting and proved that a block-version of the orthogonal matching pursuit (BOMP) algorithm can exactly recover any block-sparse signal if the block-coherence is sufficiently small. In addition, the mixed l_2/l_1 -minimization approach was certified to guarantee successful recovery with the same condition on block-coherence. Ben-Haim and Eldar [18] examined the ability of greedy algorithms to estimate a block-sparse signal from noisy measurements. They derived some near-oracle results for the block-sparse version of greedy pursuit algorithms both in the adversarial noise and the Gaussian noise cases. Majumdar and Ward [19] used BOMP to deal with the block-sparse

representation-based classification problem. The validity and robustness of these new methods were theoretically proved.

In recent years, several studies [20-27] have showed that the non-convex l_q ($0 < q < 1$) minimization allows the exact recovery of sparse signals from fewer linear measurements than that by l_1 -minimization. Chartrand and Staneva [21] provided a weaker condition to guarantee perfect recovery for the non-convex l_q ($0 < q < 1$) minimization method using a l_q variant of the RIP. They obtained the number of random Gaussian measurements necessary for the successful recovery of sparse signals via l_q ($0 < q < 1$) minimization with high probability. Sun [27] used the conventional RIP as in the l_1 case to prove that whenever q is chosen to be about $0.6796 \times (1 - \delta_{2k})$, every k -sparse signal can exactly be recovered via l_q minimization, here δ_{2k} is the restricted isometry constant for the measurement matrix. Xu et al. [24-26] considered a specially important case ($q = 1/2$) of l_q minimization. They developed a thresholding representation theory for $l_{1/2}$ minimization and conducted a phase diagram study to demonstrate the merits of $l_{1/2}$ minimization.

This article presents an ongoing effort to extend the non-convex l_q ($0 < q < 1$) minimization methodology to the setting of block-sparsity. Specifically, we will study the performance of the block-sparse signal recovery via the mixed l_2/l_q ($0 < q < 1$) norm minimization by means of the block RIP (block-RIP). We first exhibit that under similar RIP conditions with that in the standard l_q case, the mixed l_2/l_q recovery method can assuredly to recover any block-sparse signal, irrespective to the locations of non-zero blocks. In addition, the method is robust in the presence of noise. Our formulated recovery conditions will show that the non-convex l_2/l_q ($0 < q < 1$) minimization is superior to the convex l_2/l_1 minimization within block-RIP framework. Furthermore, we will compare the sparse signal recovery ability of the non-convex l_2/l_q ($0 < q < 1$) method to the convex l_2/l_1 method and the standard l_q method by conducting a series of simulation studies. To the best of the authors' knowledge, although Majumdar and Ward [12] first proposed the non-convex l_2/l_q ($0 < q < 1$) method in CS literature for color imaging and showed that the $l_2/l_{0.4}$ minimization has the best performance on some imaging experiments, their works were experimental in feature and lack of convincing theoretical assessment. As compared, our work not only highlights theoretical merits of the non-convex block optimization method, but also makes a more intensive study on the block-sparse signal recovery capabilities for some different values of q via numerical experiments.

We begin with our study in Section 2,3 and 4 by presenting the problem setting. In Section 5, we establish the sufficient conditions for the mixed l_2/l_q ($0 < q < 1$) optimization approach to guarantee exact and robust

recovery of block-sparse signals in terms of block-RIP. We also develop an efficient Iteratively Reweighted Least-Squares (IRLS) algorithm to recover block-sparse signals from given fewer measurements, which generalizes the algorithm of [28] to the unconstrained l_2/l_q ($0 < q \leq 1$) norm minimization case in Section 6. In Section 7, we show that the non-convex l_2/l_q ($0 < q < 1$) method has stronger block-sparsity promoting ability than the convex l_2/l_1 method and the standard l_q method through a series of simulations. Finally, we conclude the article in Section 8 with some useful remarks.

2 Block-sparsity

The conventional CS only consider the sparsity that the signal \mathbf{x} has at most k non-zero elements, and it does not take into account any further structure. However, in many practical scenarios, the non-zero elements are aligned to blocks, meaning that they appear in regions. These signals are referred to the block-sparse signals. Mathematically speaking, a block-sparse signal $\mathbf{x} \in \mathbb{R}^N$ over block index set $\mathcal{I} = \{d_1, \dots, d_m\}$ can be modeled as follows:

$$\mathbf{x} = [\underbrace{x_1 \cdots x_{d_1}}_{\mathbf{x}[1]} \underbrace{x_{d_1+1} \cdots x_{d_1+d_2}}_{\mathbf{x}[2]} \cdots \underbrace{x_{N-d_m+1} \cdots x_N}_{\mathbf{x}[m]}]^T. \quad (1)$$

Here, $\mathbf{x}[i]$ denotes the i th block of \mathbf{x} and d_i is the block size for the i th block. The block-sparsity we consider in this article means that there are at most $k < m$ non-zero blocks. Obviously, if $d_1 = \dots = d_m = 1$, the block-sparse signals degenerate to the conventional sparse signals well studied in CS.

Definition 1. ([14]). A block k -sparse signal over index set $\mathcal{I} = \{d_1, \dots, d_m\}$ is a signal of the form (1) in which $\mathbf{x}[i]$ is non-zero for at most k indices i , $i \in \{1, 2, \dots, m\}$.

The main focus of this study is to recover a block-sparse signal \mathbf{x} from random linear measurement $\mathbf{y} = \Phi \mathbf{x}$ (noiseless case) or $\mathbf{y} = \Phi \mathbf{x} + \mathbf{z}$ (noisy case). Here, $\mathbf{y} \in \mathbb{R}^M$ is a vector, $\Phi \in \mathbb{R}^{M \times N}$ is a measurement matrix, whose entries are usually randomly drawn from a Gaussian or a Bernoulli distribution, and \mathbf{z} is an unknown bounded noise. We represent Φ as a concatenation of column-blocks $\Phi[i]$ of size $M \times d_i$, that is,

$$\Phi = [\underbrace{\phi_1 \cdots \phi_{d_1}}_{\Phi[1]} \underbrace{\phi_{d_1+1} \cdots \phi_{d_1+d_2}}_{\Phi[2]} \cdots \underbrace{\phi_{N-d_m+1} \cdots \phi_N}_{\Phi[m]}]. \quad (2)$$

Then we are interested in formulating sufficient conditions on the measurement matrix Φ under which a block-sparse signal \mathbf{x} can assuredly be and stably recovered from

its fewer noiseless measurements $\mathbf{y} = \sum_{i=1}^m \Phi[i] \mathbf{x}[i]$ or noisy measurements $\mathbf{y} = \sum_{i=1}^m \Phi[i] \mathbf{x}[i] + \mathbf{z}$. Denote

$$\|\mathbf{x}\|_{2,0} = \sum_{i=1}^m I(\|\mathbf{x}[i]\|_2 > 0), \quad (3)$$

where $I(\|\mathbf{x}[i]\|_2 > 0)$ is an indicator function, we then notice that a block k -sparse signal \mathbf{x} can be defined as a vector that satisfies $\|\mathbf{x}\|_{2,0} \leq k$. In the remainder of the article, we will restrict our attention to how and in what conditions these block-sparse signals can be recovered exactly and stably in noiseless and noisy scenarios respectively.

3 Block-RIP

Candes and Tao [6] first introduced the notion of RIP of a matrix to characterize the condition under which the sparsest solution of an underdetermined linear system exists and can be found. And then the RIP was used as a powerful tool to study CS in several previous works [4,5,21,29]. Let Φ be a matrix of size $M \times N$, where $M < N$, we say that matrix Φ satisfies RIP of order k if there exists a constant $\delta_k \in [0, 1)$ such that for every $\mathbf{x} \in \mathbb{R}^N$ ($\|\mathbf{x}\|_0 \leq k$),

$$(1 - \delta_k) \|\mathbf{x}\|_2^2 \leq \|\Phi \mathbf{x}\|_2^2 \leq (1 + \delta_k) \|\mathbf{x}\|_2^2 \quad (4)$$

Obviously, δ_k quantifies how close to isometric the all $M \times k$ submatrices of Φ should be. Since the block-sparse signals exhibit additional structure, Eldar and Mishali [14] extended the standard RIP to the block-sparse setting and showed that the new block-RIP constant is typically smaller than the standard RIP constant. Now we state the new definition in block-sparse setting.

Definition 2. ([14]). Let $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}^M$ be a $M \times N$ measurement matrix. Then Φ is said to have the block-RIP over $\mathcal{I} = \{d_1, \dots, d_m\}$ with constant $\delta_{k|\mathcal{I}}$ if for every vector $\mathbf{x} \in \mathbb{R}^N$ that is block k -sparse over \mathcal{I} , it satisfies

$$(1 - \delta_{k|\mathcal{I}}) \|\mathbf{x}\|_2^2 \leq \|\Phi \mathbf{x}\|_2^2 \leq (1 + \delta_{k|\mathcal{I}}) \|\mathbf{x}\|_2^2. \quad (5)$$

For convenience, in the remainder of the article, we still use δ_k , instead of $\delta_{k|\mathcal{I}}$, to represent the block-RIP constant whenever the confusion is not caused.

With the new notion, Eldar and Mishali [29] generalized the sufficient recovery conditions to the block-sparse signals both in noiseless and noisy settings. They showed that if Φ is taken random as conventional CS, it satisfies the block-RIP with overwhelming probability. All these results illustrated that one can recover a block-sparse signal exactly and stably via the convex mixed l_1/l_2 minimization method whenever the measurement matrix Φ is constructed from a random ensemble (i.e., Gaussian ensemble).

4 Non-convex recovery method

It is known from [14] that whenever Φ satisfies the block-RIP with $\delta_{2k} < 1$, there is a unique block-sparse signal \mathbf{x} which can be recovered by solving the following problem:

$$\begin{aligned} \min_{\mathbf{x}} \|\mathbf{x}\|_{2,0} \\ \text{s.t. } \mathbf{y} = \Phi\mathbf{x}. \end{aligned} \quad (6)$$

Unfortunately, the problem (6) is an NP-hard problem and finding the optimal solution of (6) has exponential complexity. In principle, one only can solve the problem exactly by searching over all possible sets of k blocks whether there exists a vector consistent with the measurements. Obviously, this approach is unable to deal with high-dimensional signals.

One natural idea to find \mathbf{x} more efficiently is to employ a convex relaxation technique, namely, to replace the l_2/l_0 norm by its closest convex surrogate l_2/l_1 norm, thus resulting in the following model:

$$\begin{aligned} \min_{\mathbf{x}} \|\mathbf{x}\|_{2,1} \\ \text{s.t. } \mathbf{y} = \Phi\mathbf{x}, \end{aligned} \quad (7)$$

where $\|\mathbf{x}\|_{2,1} = \sum_{i=1}^m \|\mathbf{x}[i]\|_2$. This model can be treated as a second-order cone program (SOCP) problem and many standard software packages can be used for the solutions very efficiently. In many practical cases, the measurements \mathbf{y} are corrupted by bounded noise, then we can apply the modified SOCP or the group version of basis pursuit denoising [30] program as the following:

$$\min_{\mathbf{x}} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_{2,1}, \quad (8)$$

where λ is a tuning parameter, which controls the tolerance of the noise term. There are also many methods to solve this optimization problem efficiently, such as the block-coordinate descent technique [31] and the Landweber iterations technique [32].

As mentioned before, recent studies on non-convex CS have indicated that one can reduce the number of required linear measurements for successful recovery of a general sparse signal by replacing the l_0 norm by a non-convex surrogate l_q ($0 < q < 1$) quasi-norm, which motivates us to generalize the better recovery capability of the non-convex CS to the block-sparse setting. Therefore, we suggest the use of the following non-convex optimization model for recovery of block-sparse signals, that is,

$$\begin{aligned} \min_{\mathbf{x}} \|\mathbf{x}\|_{2,q}^q \\ \text{s.t. } \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 \leq \epsilon, \end{aligned} \quad (9)$$

where $\epsilon \geq 0$ controls the noise error term ($\epsilon = 0$ means noiseless case) and $\|\mathbf{x}\|_{2,q} = (\sum_{i=1}^m \|\mathbf{x}[i]\|_2^q)^{1/q}$ is a generalization of standard l_q quasi-norm for $0 < q < 1$. We will show that this new non-convex recovery approach can achieve better block-sparse recovery performance

both practically and theoretically when compared with the commonly used convex l_2/l_1 minimization approach. In the following section, we will provide some sufficient conditions for exact and stable recovery of block-sparse signals through the mixed l_2/l_q ($0 < q < 1$) norm minimization, and further develop a similar IRLS algorithm as in [28,33] for solutions of such non-convex optimization problem.

5 Sufficient block-sparse recovery conditions

In this section, we first consider the recovery problem of a high-dimensional signal $\mathbf{x} \in \mathbb{R}^N$ in the noiseless setting. Thus, we propose the constrained mixed l_2/l_q norm minimization with $0 < q < 1$:

$$\begin{aligned} \min_{\mathbf{x}} \|\mathbf{x}\|_{2,q}^q \\ \text{s.t. } \mathbf{y} = \Phi\mathbf{x}, \end{aligned} \quad (10)$$

where $\mathbf{y} \in \mathbb{R}^M$ are available measurements, Φ is a known $M \times N$ measurement matrix.

To state our main results, we need more notations. We first consider the case where \mathbf{x} is exactly block k -sparse. We use $\text{Null}(\Phi)$ to denote the null space of Φ and

T to denote the block index set of non-zero blocks of the signal \mathbf{x} . Let \mathbf{x}^* be a solution of the minimization problem (10). From [15], it is known that \mathbf{x}^* is the unique sparse solution of (10) being equal to \mathbf{x} if and only if

$$\|\mathbf{h}_T\|_{2,q} < \|\mathbf{h}_{T^c}\|_{2,q}$$

for all non-zero vector \mathbf{h} in the null space of Φ . This is called the null space property (NSP). In order to characterize more accurately the NSP, one can consider the following equivalent form: There exists a smallest constant ρ satisfying $0 < \rho < 1$ such that

$$\|\mathbf{h}_T\|_{2,q} \leq \rho \|\mathbf{h}_{T^c}\|_{2,q}. \quad (11)$$

When \mathbf{x} is not exactly block sparse, Aldroubi et al. [34] also showed that NSP actually guarantees stability. Precisely, if we use T to denote the block index set over the k blocks with largest l_2 norm of \mathbf{x} , then the NSP (11) gives

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leq \frac{C \|\mathbf{x}_{T^c}\|_{2,q}}{k^{1/q-1/2}} \quad (12)$$

here C is a constant. Indeed, from (11), it is easy to see that the following equality holds

$$\begin{aligned} \sup_{\mathbf{h} \in \text{Null}(\Phi), \mathbf{h} \neq 0} \frac{(\sum_{i \in T} \|\mathbf{h}[i]\|_2^q)^{1/q}}{(\sum_{i \in T^c} \|\mathbf{h}[i]\|_2^q)^{1/q}} \\ = \max_{\mathbf{h} \in \text{Null}(\Phi), \|\mathbf{h}\|_2=1} \frac{(\sum_{i \in T} \|\mathbf{h}[i]\|_2^q)^{1/q}}{(\sum_{i \in T^c} \|\mathbf{h}[i]\|_2^q)^{1/q}} \end{aligned}$$

which is denoted by ρ . In general, for $\mathbf{h} = \mathbf{x}^* - \mathbf{x}$, we let

$$\|\mathbf{h}_T\|_{2,q} = \gamma(\mathbf{h}, q) \|\mathbf{h}_{T^c}\|_{2,q}.$$

Therefore, Our main point of the study is to show how to make $\gamma(\mathbf{h}, q) < 1$ for all non-zero vector \mathbf{h} in the null space of Φ . Our first conclusion is the following theorem.

Theorem 1. (Noiseless recovery). Let $\mathbf{y} = \Phi\mathbf{x}$ be measurements of a signal \mathbf{x} . If the matrix Φ satisfies the block RIP (5) with

$$\delta_{2k} < 1/2,$$

then there exists a number $q_0(\delta_{2k}) \in (0, 1]$ such that for any $q < q_0$, the solution \mathbf{x}^* to the mixed l_2/l_q problem (10) obeys to

$$\begin{aligned} \|\mathbf{x} - \mathbf{x}^*\|_{2,q} &\leq C_0(q, \delta_{2k}) \|\mathbf{x} - \mathbf{x}_{T_0}\|_{2,q} \\ \|\mathbf{x} - \mathbf{x}^*\|_2 &\leq C_1(q, \delta_{2k}) k^{1/2-1/q} \|\mathbf{x} - \mathbf{x}_{T_0}\|_{2,q} \end{aligned} \quad (13)$$

where $C_0(q, \delta_{2k})$ and $C_1(q, \delta_{2k})$ are positive constants dependent on q and δ_{2k} , T_0 is the block index set over the k blocks with largest l_2 norm of the original signal \mathbf{x} . In particular, if \mathbf{x} is block k -sparse, the recovery is exact.

Remark 1. Theorem 1 provides a sufficient condition for the recovery of a signal \mathbf{x} via l_2/l_q minimization with $0 < q < 1$ in the noiseless setting. Focusing on the case where \mathbf{x} is block k -sparse, it is known [14] that when the l_2/l_1 minimization scheme (7) is employed to recover \mathbf{x} , the sufficient condition on the block-RIP is that $\delta_{2k} < 0.414$. As compared, Theorem 1 says that, whenever the non-convex l_2/l_q minimization scheme (10) is used, this constant can be relaxed to $\delta_{2k} < 0.5$ for some $q < 1$. This shows that similar as the standard sparse signal recovery, when compared with the convex minimization method, the non-convex minimization method can enhance performance of block-sparse signal recovery.

To prove Theorem 1, we need the following Lemmas:

Lemma 1. ([14]).

$$|\langle \Phi\mathbf{x}, \Phi\mathbf{x}' \rangle| \leq \delta_{k+k'} \|\mathbf{x}\|_2 \|\mathbf{x}'\|_2 \quad (14)$$

for all \mathbf{x}, \mathbf{x}' supported on disjoint subsets $T, T' \subseteq \{1, 2, \dots, N\}$ with $|T| < k$ and $|T'| < k'$.

Lemma 2. ([35]). For any fixed $q \in (0, 1)$ and $\mathbf{x} \in \mathbb{R}^N$,

$$\|\mathbf{x}\|_2 \leq \frac{\|\mathbf{x}\|_q}{N^{1/q-1/2}} + \sqrt{N} \left(\max_{1 \leq i \leq N} |\mathbf{x}_i| - \min_{1 \leq i \leq N} |\mathbf{x}_i| \right). \quad (15)$$

Proof of Theorem 1. Set $\mathbf{x}^* = \mathbf{x} + \mathbf{h}$ be a solution of (10), where \mathbf{x} is the original signal we need to reconstruct. Throughout the article, \mathbf{x}_T will denote the vector equal to \mathbf{x} on an index set T and zero elsewhere. Let T_0 be the block index set over the k blocks with largest l_2

norm of \mathbf{x} . And we decompose \mathbf{h} into a series of vectors $\mathbf{h}_{T_0}, \mathbf{h}_{T_1}, \mathbf{h}_{T_2}, \dots, \mathbf{h}_{T_j}$, such that

$$\mathbf{h} = \sum_i^J \mathbf{h}_{T_i}.$$

Here \mathbf{h}_{T_i} is the restriction of \mathbf{h} to the set T_i and each T_i consists of k blocks (except possibly T_j). Rearranging the block indices such that $\|\mathbf{h}_{T_j}[1]\|_2 \geq \|\mathbf{h}_{T_j}[2]\|_2 \geq \dots \geq \|\mathbf{h}_{T_j}[k]\|_2 \geq \|\mathbf{h}_{T_{j+1}}[1]\|_2 \geq \|\mathbf{h}_{T_{j+1}}[2]\|_2 \geq \dots$, for any $j \geq 1$.

Note that

$$\|\mathbf{h}\|_2 = \|\mathbf{h}_{T_0 \cup T_1} + \mathbf{h}_{(T_0 \cup T_1)^c}\|_2 \leq \|\mathbf{h}_{T_0 \cup T_1}\|_2 + \|\mathbf{h}_{(T_0 \cup T_1)^c}\|_2. \quad (16)$$

For any $j \geq 2$, if we let $\mathbf{d}_j = (\|\mathbf{h}_{T_j}[1]\|_2, \|\mathbf{h}_{T_j}[2]\|_2, \dots, \|\mathbf{h}_{T_j}[k]\|_2)$, then we have

$$\begin{aligned} \|\mathbf{d}_j\|_2 &= \left(\sum_{i=1}^k \|\mathbf{h}_{T_j}[i]\|_2^2 \right)^{1/2} = \|\mathbf{h}_{T_j}\|_2, \\ \|\mathbf{d}_j\|_q &= \left(\sum_{i=1}^k \|\mathbf{h}_{T_j}[i]\|_2^q \right)^{1/q} = \|\mathbf{h}_{T_j}\|_{2,q}, \\ \max_{1 \leq i \leq k} |\mathbf{d}_j(i)| &= \max_{1 \leq i \leq k} \|\mathbf{h}_{T_j}[i]\|_2 = \|\mathbf{h}_{T_j}\|_{\infty, T}, \\ \min_{1 \leq i \leq k} |\mathbf{d}_j(i)| &= \min_{1 \leq i \leq k} \|\mathbf{h}_{T_j}[i]\|_2. \end{aligned} \quad (17)$$

From Lemma 2, it follows that

$$\|\mathbf{d}_j\|_2 \leq \frac{\|\mathbf{d}_j\|_q}{k^{1/q-1/2}} + k^{1/2} (\max_{1 \leq i \leq k} |\mathbf{d}_j(i)| - \min_{1 \leq i \leq k} |\mathbf{d}_j(i)|),$$

that is,

$$\|\mathbf{h}_{T_j}\|_2 \leq \frac{\|\mathbf{h}_{T_j}\|_{2,q}}{k^{1/q-1/2}} + k^{1/2} (\|\mathbf{h}_{T_j}\|_{\infty, T} - \min_{1 \leq i \leq k} \|\mathbf{h}_{T_j}[i]\|_2). \quad (18)$$

From Equation (18), we obtain

$$\begin{aligned} k^{1/q-1/2} \sum_{j \geq 2} \|\mathbf{h}_{T_j}\|_2 &\leq \sum_{j \geq 2} \|\mathbf{h}_{T_j}\|_{2,q} + k^{1/q} (\|\mathbf{h}_{T_j}\|_{\infty, T} \\ &\quad - \min_{1 \leq i \leq k} \|\mathbf{h}_{T_j}[i]\|_2) \\ &\leq \sum_{j \geq 2} \|\mathbf{h}_{T_j}\|_{2,q} + k^{1/q} \|\mathbf{h}_{T_2}\|_{\infty, T} \\ &\leq \sum_{j \geq 2} \|\mathbf{h}_{T_j}\|_{2,q} + k^{1/q} \|\mathbf{h}_{T_1}\|_{2,q} / k^{1/q} \\ &= \sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q} \\ &\leq \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \right)^{1/q}, \end{aligned} \quad (19)$$

where we have used the fact that $(a + b)^q \leq a^q + b^q$ for non-negative a and b . Therefore, we have

$$\sum_{j \geq 2} \|\mathbf{h}_{T_j}\|_2 \leq \frac{1}{k^{1/q-1/2}} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \right)^{1/q}. \quad (20)$$

On the other hand, let

$$\begin{aligned} \|\mathbf{h}_{T_0}\|_{2,q} &= \gamma(\mathbf{h}, q) \|\mathbf{h}_{T_0^c}\|_{2,q}, \\ \|\mathbf{h}_{T_1}\|_{2,q}^q &= t \sum_{i \geq 1} \|\mathbf{h}_{T_i}\|_{2,q}^q, \quad t \in [0, 1]. \end{aligned} \quad (21)$$

Since $\|\mathbf{h}_{T_2}[1]\|_2 \geq \|\mathbf{h}_{T_2}[2]\|_2 \geq \dots \geq \|\mathbf{h}_{T_2}[k]\|_2 \geq \|\mathbf{h}_{T_3}[1]\|_2 \geq \|\mathbf{h}_{T_3}[2]\|_2 \geq \dots$, it is easy to see that

$$\begin{aligned} \sum_{i \geq 2} \|\mathbf{h}_{T_i}\|_2^2 &\leq \|\mathbf{h}_{T_2}[1]\|_2^{2-q} \sum_{j \geq 2} \|\mathbf{h}_{T_j}\|_{2,q}^q \\ &\leq ((\|\mathbf{h}_{T_1}[1]\|_2^q + \|\mathbf{h}_{T_1}[2]\|_2^q + \dots \\ &\quad + \|\mathbf{h}_{T_1}[k]\|_2^q)/k)^{(2-q)/q} \sum_{j \geq 2} \|\mathbf{h}_{T_j}\|_{2,q}^q \\ &= (\|\mathbf{h}_{T_1}\|_{2,q}^q/k)^{\frac{2-q}{q}} \sum_{j \geq 2} \|\mathbf{h}_{T_j}\|_{2,q}^q \\ &= (\|\mathbf{h}_{T_1}\|_{2,q}^q/k)^{\frac{2-q}{q}} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q - \|\mathbf{h}_{T_1}\|_{2,q}^q \right) \\ &= (\|\mathbf{h}_{T_1}\|_{2,q}^q/k)^{\frac{2-q}{q}} \left(\frac{1}{t} \|\mathbf{h}_{T_1}\|_{2,q}^q - \|\mathbf{h}_{T_1}\|_{2,q}^q \right) \\ &= \frac{1-t}{tk^{\frac{2-q}{q}}} \|\mathbf{h}_{T_1}\|_{2,q}^2 \\ &= \frac{1-t}{t^{1-2/q} k^{\frac{2-q}{q}}} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \right)^{2/q}. \end{aligned} \quad (22)$$

By the definition of the block-RIP, Lemma 1, (20) and (22), it then implies that

$$\begin{aligned} &\|\Phi(\sum_{j \geq 2} \mathbf{h}_{T_j})\|_2^2 \\ &= \sum_{i,j \geq 2} \langle \Phi(\mathbf{h}_{T_i}), \Phi(\mathbf{h}_{T_j}) \rangle \\ &= \sum_{j \geq 2} \langle \Phi(\mathbf{h}_{T_j}), \Phi(\mathbf{h}_{T_j}) \rangle + 2 \sum_{i,j \geq 2, i < j} \langle \Phi(\mathbf{h}_{T_i}), \Phi(\mathbf{h}_{T_j}) \rangle \\ &\leq (1 + \delta_k) \sum_{j \geq 2} \|\mathbf{h}_{T_j}\|_2^2 + 2\delta_{2k} \sum_{j > i \geq 2} \|\mathbf{h}_{T_i}\|_2 \|\mathbf{h}_{T_j}\|_2 \\ &\leq \sum_{j \geq 2} \|\mathbf{h}_{T_j}\|_2^2 + \delta_{2k} \left(\sum_{j \geq 2} \|\mathbf{h}_{T_j}\|_2 \right)^2 \\ &\leq \left(\frac{1-t}{t^{1-2/q} k^{\frac{2-q}{q}}} + \delta_{2k} k^{1-2/q} \right) \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \right)^{2/q}. \end{aligned} \quad (23)$$

Since $\Phi \mathbf{x} = \Phi \mathbf{x}^*$, we have $\Phi \mathbf{h} = 0$, thus $\Phi(\mathbf{h}_{T_0} + \mathbf{h}_{T_1}) = -\Phi(\sum_{j \geq 2} \mathbf{h}_{T_j})$. Therefore,

$$\begin{aligned} \|\Phi(\mathbf{h}_{T_0} + \mathbf{h}_{T_1})\|_2^2 &\leq \left(\frac{1-t}{t^{1-2/q} k^{\frac{2-q}{q}}} + \delta_{2k} k^{1-2/q} \right) \\ &\quad \times \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \right)^{2/q}, \end{aligned} \quad (24)$$

By the definition of δ_{2k} and using Hölder's equality, we then further have

$$\begin{aligned} &\|\Phi(\mathbf{h}_{T_0} + \mathbf{h}_{T_1})\|_2^2 \\ &\geq (1 - \delta_{2k}) \|\mathbf{h}_{T_0} + \mathbf{h}_{T_1}\|_2^2 \\ &= (1 - \delta_{2k}) (\|\mathbf{h}_{T_0}\|_2^2 + \|\mathbf{h}_{T_1}\|_2^2) \\ &\geq (1 - \delta_{2k}) (k^{1-2/q} \|\mathbf{h}_{T_0}\|_{2,q}^2 + k^{1-2/q} \|\mathbf{h}_{T_1}\|_{2,q}^2) \\ &= (1 - \delta_{2k}) k^{1-2/q} \\ &\quad \times \left(\gamma(\mathbf{h}, q)^2 \|\mathbf{h}_{T_0^c}\|_{2,q}^2 + t^{2/q} \left(\sum_{i \geq 1} \|\mathbf{h}_{T_i}\|_{2,q}^q \right)^{2/q} \right) \\ &= (1 - \delta_{2k}) k^{1-2/q} (\gamma(\mathbf{h}, q)^2 + t^{2/q}) \left(\sum_{i \geq 1} \|\mathbf{h}_{T_i}\|_{2,q}^q \right)^{2/q}. \end{aligned} \quad (25)$$

By Equations (24) and (25),

$$\gamma(\mathbf{h}, q)^2 \leq \left(\frac{1-t}{t^{1-2/q}} \right) / (1 - \delta_{2k}) / -t^{2/q} \doteq f(t).$$

Through a straightforward calculation, it is easy to get that the maximum of $f(t)$ occurs at $t_0 = \frac{1-q/2}{2-\delta_{2k}}$ and

$$f(t_0) = \frac{\delta_{2k} \frac{2-q}{2-\delta_{2k}} + q \left(\frac{1-q/2}{2-\delta_{2k}} \right)^{2/q}}{\frac{2-q}{2-\delta_{2k}} (1 - \delta_{2k})}.$$

If $f(t_0) < 1$, then we have $\gamma(\mathbf{h}, q) < 1$. However, $f(t_0) < 1$ amounts to

$$\delta_{2k} \frac{2-q}{2-\delta_{2k}} + q \left(\frac{1-q/2}{2-\delta_{2k}} \right)^{2/q} < \frac{2-q}{2-\delta_{2k}} (1 - \delta_{2k}).$$

or, equivalently,

$$\delta_{2k} + \frac{q}{2^{2/q+1}} \left(\frac{2-q}{2-\delta_{2k}} \right)^{2/q-1} < \frac{1}{2} \quad (26)$$

Since the second term on the left-hand side of (26) goes to zero as $q \rightarrow 0_+$ whenever $q \leq 1$, $\delta_{2k} < 1$, and

$$\frac{q}{2^{2/q+1}} \left(\frac{2-q}{2-\delta_{2k}} \right)^{2/q-1} \leq q \left(\frac{2-q}{2} \right)^{2/q} \approx \frac{q}{e}$$

We thus obtain that for $\delta_{2k} < 1/2$, there exists a value $q_0 = q_0(\delta_{2k}) \in (0, 1]$ such that for all $q \in (0, q_0)$ the above inequality (26) is true.

From the definition of $\gamma(\mathbf{h}, q)$, we have

$$\|\mathbf{h}_{T_0}\|_{2,q}^2 \leq \frac{\delta_{2k} \frac{2-q}{2-\delta_{2k}} + q(\frac{1-q/2}{2-\delta_{2k}})^{2/q}}{\frac{2-q}{2-\delta_{2k}}(1-\delta_{2k})} \|\mathbf{h}_{T_0^c}\|_{2,q}^2.$$

As $\|\mathbf{x}^*\|_q = \|\mathbf{x} + \mathbf{h}\|_q$ is the minimum, using the equation $\mathbf{x}^* = \mathbf{x} + \mathbf{h}_{T_0} + \mathbf{h}_{T_0^c}$, we get

$$\begin{aligned} \|\mathbf{x}\|_{2,q}^q &= \|\mathbf{x}_{T_0}\|_{2,q}^q + \|\mathbf{x}_{T_0^c}\|_{2,q}^q \\ &\geq \|\mathbf{x}^*\|_{2,q}^q = \|\mathbf{x} + \mathbf{h}_{T_0} + \mathbf{h}_{T_0^c}\|_{2,q}^q \\ &= \|(\mathbf{x} + \mathbf{h}_{T_0} + \mathbf{h}_{T_0^c})_{T_0}\|_{2,q}^q + \|(\mathbf{x} + \mathbf{h}_{T_0} + \mathbf{h}_{T_0^c})_{T_0^c}\|_{2,q}^q \\ &= \|\mathbf{x} + \mathbf{h}_{T_0}\|_{2,q}^q + \|\mathbf{h}_{T_0^c}\|_{2,q}^q \\ &\geq \|\mathbf{x}\|_{2,q}^q - \|\mathbf{h}_{T_0}\|_{2,q}^q + \|\mathbf{h}_{T_0^c}\|_{2,q}^q \\ &\geq \|\mathbf{x}_{T_0}\|_{2,q}^q - \|\mathbf{h}_{T_0}\|_{2,q}^q + \|\mathbf{h}_{T_0^c}\|_{2,q}^q - \|\mathbf{x}_{T_0^c}\|_{2,q}^q. \end{aligned} \tag{27}$$

Since $\|\mathbf{x}\|_{2,q}^q = \|\mathbf{x}_{T_0}\|_{2,q}^q + \|\mathbf{x}_{T_0^c}\|_{2,q}^q$, this then implies

$$\begin{aligned} \|\mathbf{h}_{T_0^c}\|_{2,q}^q &\leq \|\mathbf{h}_{T_0}\|_{2,q}^q + 2\|\mathbf{x}_{T_0^c}\|_{2,q}^q \\ &\leq \left(\frac{\delta_{2k} \frac{2-q}{2-\delta_{2k}} + q(\frac{1-q/2}{2-\delta_{2k}})^{2/q}}{\frac{2-q}{2-\delta_{2k}}(1-\delta_{2k})} \right)^{q/2} \|\mathbf{h}_{T_0^c}\|_{2,q}^q \\ &\quad + 2\|\mathbf{x}_{T_0^c}\|_{2,q}^q, \end{aligned}$$

$$\begin{aligned} \|\mathbf{h}\|_2 &\leq \sqrt{\|\mathbf{h}_{T_0 \cup T_1}\|_2^2 + \|\mathbf{h}_{(T_0 \cup T_1)^c}\|_2^2} \\ &\leq \sqrt{\frac{\left(\frac{1-t}{t^{1-2/q}k^{\frac{2-q}{q}}} + \delta_{2k}k^{1-2/q} \right) \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \right)^{2/q}}{1-\delta_{2k}} + \frac{1-t}{t^{1-2/q}k^{\frac{2-q}{q}}} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \right)^{2/q}} \\ &= \sqrt{\left(\frac{\frac{1-t}{t^{1-2/q}k^{\frac{2-q}{q}}} + \delta_{2k}k^{1-2/q}}{1-\delta_{2k}} + \frac{1-t}{t^{1-2/q}k^{\frac{2-q}{q}}} \right) \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \right)^{2/q}} \\ &= \sqrt{\frac{k^{1-2/q} \left(\frac{1-t}{t^{1-2/q}} + \delta_{2k} + (1-\delta_{2k}) \frac{1-t}{t^{1-2/q}} \right)}{1-\delta_{2k}} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \right)^{1/q}} \\ &= k^{1/2-1/q} \sqrt{\frac{(2-\delta_{2k})(1-t)t^{2/q-1} + \delta_{2k}}{1-\delta_{2k}}} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \right)^{1/q} \\ &\leq k^{1/2-1/q} \sqrt{\frac{(2-\delta_{2k})q(2-q)^{2/q-1} + 2^{2/q}\delta_{2k}}{(1-\delta_{2k})2^{2/q}}} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \right)^{1/q} \end{aligned} \tag{31}$$

That is,

$$\|\mathbf{h}_{T_0^c}\|_{2,q}^q \leq \frac{2}{1 - \left(\frac{\delta_{2k} \frac{2-q}{2-\delta_{2k}} + q(\frac{1-q/2}{2-\delta_{2k}})^{2/q}}{\frac{2-q}{2-\delta_{2k}}(1-\delta_{2k})} \right)^{q/2}} \|\mathbf{x}_{T_0^c}\|_{2,q}^q. \tag{28}$$

Thus, we have

$$\begin{aligned} \|\mathbf{h}\|_{2,q}^q &= \|\mathbf{h}_{T_0}\|_{2,q}^q + \|\mathbf{h}_{T_0^c}\|_{2,q}^q \\ &\leq \left(\left(\frac{\delta_{2k} \frac{2-q}{2-\delta_{2k}} + q(\frac{1-q/2}{2-\delta_{2k}})^{2/q}}{\frac{2-q}{2-\delta_{2k}}(1-\delta_{2k})} \right)^{q/2} + 1 \right) \|\mathbf{h}_{T_0^c}\|_{2,q}^q \\ &\leq \frac{2 \left(\left(\frac{\delta_{2k} \frac{2-q}{2-\delta_{2k}} + q(\frac{1-q/2}{2-\delta_{2k}})^{2/q}}{\frac{2-q}{2-\delta_{2k}}(1-\delta_{2k})} \right)^{q/2} + 1 \right)}{1 - \left(\frac{\delta_{2k} \frac{2-q}{2-\delta_{2k}} + q(\frac{1-q/2}{2-\delta_{2k}})^{2/q}}{\frac{2-q}{2-\delta_{2k}}(1-\delta_{2k})} \right)^{q/2}} \|\mathbf{x}_{T_0^c}\|_{2,q}^q \\ &= C_q^q \|\mathbf{x}_{T_0^c}\|_{2,q}^q. \end{aligned} \tag{29}$$

This justifies that the first equality of (13) is proved. In the following, we further prove the second equality of (13).

In effect, from (24) and (25), we have

$$\|\mathbf{h}_{T_0 \cup T_1}\|_2^2 \leq \frac{\left(\frac{1-t}{t^{1-2/q}k^{\frac{2-q}{q}}} + \delta_{2k}k^{1-2/q} \right) \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \right)^{2/q}}{1-\delta_{2k}} \tag{30}$$

which, together with (22) and (30), then implies

where we have used the fact that

$$\begin{aligned} \max_{t \in [0,1]} (1-t)t^{2/q-1} &= (1-t)t^{2/q-1}|_{t=1-q/2} \\ &= \frac{q}{2} \left(1 - \frac{q}{2}\right)^{2/q-1} \end{aligned}$$

By (28), we further have

$$\begin{aligned} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \right)^{1/q} &\leq \frac{2^{1/q} \|\mathbf{x}_{T_0^c}\|_{2,q}}{(1 - (f(t_0))^{q/2})^{1/q}} \\ &\leq \frac{2^{2/q-1} \|\mathbf{x}_{T_0^c}\|_{2,q}}{(1 - (f(t_0))^{q/2})^{1/q}} \end{aligned} \quad (32)$$

Thus, it follows from (31) and (32) that

$$\begin{aligned} \|\mathbf{h}\|_2 &\leq k^{1/2-1/q} \sqrt{\frac{(2 - \delta_{2k})q(2 - q)^{2/q-1} + 2^{2/q}\delta_{2k}}{(1 - \delta_{2k})2^{2/q}}} \\ &\quad \frac{2^{2/q-1} \|\mathbf{x}_{T_0^c}\|_{2,q}}{(1 - (f(t_0))^{q/2})^{1/q}} \\ &= \frac{C_1(q, \delta_{2k}) \|\mathbf{x}_{T_0^c}\|_{2,q}}{k^{1/q-1/2}} \end{aligned} \quad (33)$$

That is, the second inequality in (13) holds. With this, the proof of Theorem 1 is completed.

We further consider the recovery problem of any high-dimensional signals in the presence of noise. In this situation, the measurement can be expressed as

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{z},$$

where $\mathbf{z} \in \mathbb{R}^N$ is an unknown noise term. In order to reconstruct \mathbf{x} , we adopt the unconstrained mixed l_2/l_q minimization scheme with $0 < q < 1$:

$$\begin{aligned} \min_{\mathbf{x}} \|\mathbf{x}\|_{2,q}^q \\ \text{s.t. } \|\mathbf{y} - \Phi \mathbf{x}\|_2 \leq \epsilon, \end{aligned} \quad (34)$$

where $\epsilon > 0$ is a bound on the noisy level. We below show that one can also recover \mathbf{x} stably and robustly under the same assumption as those in Theorem 1.

Theorem 2. (Noisy recovery). Let $\mathbf{y} = \Phi \mathbf{x} + \mathbf{z}$ ($\|\mathbf{z}\|_2 \leq \epsilon$) be noisy measurements of a signal \mathbf{x} . If the matrix Φ satisfies the block RIP (5) with

$$\delta_{2k} < 1/2,$$

then there exists a number $q_0(\delta_{2k}) \in (0, 1]$ such that for any $q < q_0$, the mixed l_2/l_q method (34) can stably and robustly recover the original signal \mathbf{x} . More precisely, the solution \mathbf{x}^* of (34) obeys to

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leq \frac{C_1(q, \delta_{2k}) \|\mathbf{x}_{T_0^c}\|_{2,q}}{k^{1/q-1/2}} + C_2(q, \delta_{2k})\epsilon \quad (35)$$

where $C_1(q, \delta_{2k})$ and $C_2(q, \delta_{2k})$ are positive constants dependent on q and δ_{2k} , T_0 is the block index set over the k blocks with largest l_2 norm of the original signal \mathbf{x} .

Remark 2. The inequality (35) in Theorem 2 offers an upper bound estimation on the recovery of the mixed l_2/l_q minimization ($q \in (0, q_0)$). In particular, this estimation shows that the recovery accuracy of the mixed l_2/l_q minimization can be controlled by the degree of sparsity of the signal and the exponential number q . It reveals also the close connections among the recovery precision of the mixed l_2/l_q minimization method may achieve, the sparsity of the signal and the index q used in the recovery procedure. In addition, the estimation (35) shows that the recovery precision of the method (34) can be bounded by the noise level. In this sense, Theorem 2 shows that under certain conditions, a block k -sparse signal can be robustly recovered by the method (34).

Proof of Theorem 2. The proof of Theorem 2 is similar to the procedure of the proof of Theorem 1 with minor differences.

To be more detail, we also set $\mathbf{x}^* = \mathbf{x} + \mathbf{h}$. Due to the existence of noise, in this case, \mathbf{h} is not necessarily in the null space of Φ any more. But we can still prove Theorem 2 under the same assumption.

We still denote by T_0 the block index set over the k blocks with largest l_2 norm of \mathbf{x} , and \mathbf{h}_{T_0} the restriction of \mathbf{h} onto these blocks. We also denote \mathbf{h}_{T_j} ($j \geq 1$) similar to the proof of Theorem 1. By $\|\mathbf{z}\|_2 \leq \epsilon$ and the triangle inequality, we first have

$$\|\Phi \mathbf{h}\|_2 = \|\Phi(\mathbf{x} - \mathbf{x}^*)\|_2 \leq \|\Phi \mathbf{x} - \mathbf{y}\|_2 + \|\Phi \mathbf{x}^* - \mathbf{y}\|_2 \leq 2\epsilon. \quad (36)$$

Since $\Phi(\mathbf{h}_{T_0} + \mathbf{h}_{T_1}) = \Phi(\mathbf{h}) - \Phi\left(\sum_{j \geq 2} \mathbf{h}_{T_j}\right)$, from definition of the block-RIP, we get

$$\begin{aligned} (1 - \delta_{2k}) \|\mathbf{h}_{T_0 \cup T_1}\|_2^2 &= (1 - \delta_{2k})(\|\mathbf{h}_{T_0}\|_2^2 + \|\mathbf{h}_{T_1}\|_2^2) \\ &\leq \|\Phi(\mathbf{h}_{T_0 \cup T_1})\|_2^2 \\ &= \|\Phi \mathbf{h} - \sum_{j \geq 2} \Phi(\mathbf{h}_{T_j})\|_2^2 \\ &\leq \left(\|\Phi \mathbf{h}\|_2 + \left\| \sum_{j \geq 2} \Phi(\mathbf{h}_{T_j}) \right\|_2 \right)^2 \\ &\leq \left(2\epsilon + \left\| \sum_{j \geq 2} \Phi(\mathbf{h}_{T_j}) \right\|_2 \right)^2 \end{aligned} \quad (37)$$

Hence,

$$\|\mathbf{h}_{T_0}\|_2^2 + \|\mathbf{h}_{T_1}\|_2^2 \leq \frac{\left(2\epsilon + \left\| \sum_{j \geq 2} \Phi(\mathbf{h}_{T_j}) \right\|_2 \right)^2}{1 - \delta_{2k}} \quad (38)$$

On the other hand,

$$\|\mathbf{h}_{T_0}\|_2^2 + \|\mathbf{h}_{T_1}\|_2^2 \geq k^{1-2/q}\|\mathbf{h}_{T_0}\|_{2,q}^2 + k^{1-2/q}\|\mathbf{h}_{T_1}\|_{2,q}^2 \quad (39)$$

From (38) and (39), we thus have

$$\begin{aligned} & \|\mathbf{h}_{T_0}\|_{2,q}^2 \\ & \leq \frac{k^{2/q-1}(2\epsilon + \|\sum_{j \geq 2} \Phi(\mathbf{h}_{T_j})\|_2)^2 - (1 - \delta_{2k})\|\mathbf{h}_{T_1}\|_{2,q}^2}{1 - \delta_{2k}} \\ & \leq \frac{(2\epsilon(2k)^{1/q-1/2})^2 + 4(2k)^{1/q-1/2}\epsilon \cdot (k/2)^{1/q-1/2}\|\sum_{j \geq 2} \Phi(\mathbf{h}_{T_j})\|_2 + k^{2/q-1}\|\sum_{j \geq 2} \Phi(\mathbf{h}_{T_j})\|_2^2 - (1 - \delta_{2k})\|\mathbf{h}_{T_1}\|_{2,q}^2}{1 - \delta_{2k}} \end{aligned} \quad (40)$$

Let $\|\mathbf{h}_{T_1}\|_{2,q}^q = t \sum_{i \geq 1} \|\mathbf{h}_{T_i}\|_{2,q}^q$, $t \in [0, 1]$ and we also have (23), thus

$$\begin{aligned} & (k/2)^{1/q-1/2}\|\sum_{j \geq 2} \Phi(\mathbf{h}_{T_j})\|_2 \\ & \leq \sqrt{\frac{(1-t)t^{2/q-1} + \delta_{2k}}{2^{2/q-1}}} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \right)^{1/q} \end{aligned} \quad (41)$$

If we denote

$$g(t) \doteq \frac{(1-t)t^{2/q-1} + \delta_{2k}}{2^{2/q-1}},$$

then, by a easy calculation, we can easily obtain that the maximum of $g(t)$ occurs at $t_1 = \frac{2-q}{2}$ and

$$g(t_1) = \frac{\delta_{2k} + \frac{q}{2}(\frac{2-q}{2})^{2/q-1}}{2^{2/q-1}} \leq \delta_{2k} + \frac{q}{2^{2/q}}(\frac{2-q}{2-\delta_{2k}})^{2/q-1}$$

Therefore, we have

$$\begin{aligned} & (k/2)^{1/q-1/2}\|\sum_{j \geq 2} \Phi(\mathbf{h}_{T_j})\|_2 \\ & \leq \sqrt{\delta_{2k} + \frac{q}{2^{2/q}}(\frac{2-q}{2-\delta_{2k}})^{2/q-1}} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \right)^{1/q} \end{aligned} \quad (42)$$

By (23) and the condition of $f(t)$ in the proof of Theorem 1, we also have

$$\begin{aligned} & k^{2/q-1}\|\sum_{j \geq 2} \Phi(\mathbf{h}_{T_j})\|_2^2 - (1 - \delta_{2k})\|\mathbf{h}_{T_1}\|_{2,q}^2 \\ & \leq (1 - \delta_{2k})f(t)\left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q\right)^{2/q} \\ & \leq (1 - \delta_{2k})f(t_0)\left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q\right)^{2/q} \\ & = \left(\delta_{2k} + \frac{q}{2^{2/q}}(\frac{2-q}{2-\delta_{2k}})^{2/q-1}\right)\left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q\right)^{2/q} \end{aligned} \quad (43)$$

Plugging (42) and (43) into (40), it is easy to see that

$$\|\mathbf{h}_{T_0}\|_{2,q} \leq \frac{2(2k)^{1/q-1/2}\epsilon}{\sqrt{1-\delta_{2k}}} + \sqrt{f(t_0)} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \right)^{1/q}$$

Consequently, we obtain

$$\|\mathbf{h}_{T_0}\|_{2,q}^q \leq \frac{(2(2k)^{1/q-1/2}\epsilon)^q}{(1-\delta_{2k})^{q/2}} + (f(t_0))^{q/2} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \right) \quad (44)$$

In the following, we further prove that $f(t_0) < 1$ one can obtain the conclusion of Theorem 2. Precisely, under the same condition on δ_{2k} , we can prove Theorem 2.

Note that from (27), we also have

$$\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \leq \|\mathbf{h}_{T_0}\|_{2,q}^q + 2\|\mathbf{x}_{T_0^c}\|_{2,q}^q$$

Plugging (44) into the above inequality and by $f(t_0) < 1$, one can show that

$$\begin{aligned} \sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q & \leq \frac{(2(2k)^{1/q-1/2}\epsilon)^q}{(1-\delta_{2k})^{q/2}(1-(f(t_0))^{q/2})} \\ & \quad + \frac{2\|\mathbf{x}_{T_0^c}\|_{2,q}^q}{1-(f(t_0))^{q/2}} \end{aligned}$$

Since $\|\mathbf{v}\|_q \leq 2^{1/q-1}\|\mathbf{v}\|_1$ for $\mathbf{v} \in \mathbb{R}^2$, we further have

$$\begin{aligned} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q \right)^{1/q} & \leq \frac{2^{2/q-1/2}k^{1/q-1/2}\epsilon}{\sqrt{1-\delta_{2k}}(1-(f(t_0))^{q/2})^{1/q}} \\ & \quad + \frac{2^{2/q-1}\|\mathbf{x}_{T_0^c}\|_{2,q}}{(1-(f(t_0))^{q/2})^{1/q}} \end{aligned} \quad (45)$$

and by (37), (22), and (23), we also have

$$\begin{aligned} \|\mathbf{h}\|_2^2 &= \|\mathbf{h}_{T_0 \cup T_1}\|_2^2 + \|\mathbf{h}_{(T_0 \cup T_1)^c}\|_2^2 \leq \frac{(2\epsilon + \|\sum_{j \geq 2} \Phi(\mathbf{h}_{T_j})\|_2)^2}{1 - \delta_{2k}} + \sum_{j \geq 2} \|\mathbf{h}_{T_j}\|_2^2 \\ &\leq \frac{\left(2\epsilon + \left(\frac{1-t}{t^{1-2/q}k^{\frac{2-q}{q}}} + \delta_{2k}k^{1-2/q}\right)^{1/2} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q\right)^{1/q}\right)^2}{1 - \delta_{2k}} + \frac{1-t}{t^{1-2/q}k^{\frac{2-q}{q}}} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q\right)^{2/q} \end{aligned} \quad (46)$$

since $\sqrt{(a + \sqrt{b})^2 + c} \leq a + \sqrt{b + c}$ ($a, b, c > 0$), it gives

$$\begin{aligned} \|\mathbf{h}\|_2 &\leq \sqrt{\frac{\left(2\epsilon + \left(\frac{1-t}{t^{1-2/q}k^{\frac{2-q}{q}}} + \delta_{2k}k^{1-2/q}\right)^{1/2} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q\right)^{1/q}\right)^2}{1 - \delta_{2k}} + \frac{1-t}{t^{1-2/q}k^{\frac{2-q}{q}}} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q\right)^{2/q}} \\ &\leq \frac{2\epsilon}{\sqrt{1 - \delta_{2k}}} + \sqrt{\frac{\left(\frac{1-t}{t^{1-2/q}k^{\frac{2-q}{q}}} + \delta_{2k}k^{1-2/q}\right) \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q\right)^{2/q}}{1 - \delta_{2k}} + \frac{1-t}{t^{1-2/q}k^{\frac{2-q}{q}}} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q\right)^{2/q}} \\ &= \frac{2\epsilon}{\sqrt{1 - \delta_{2k}}} + \sqrt{\frac{k^{1-2/q} \left(\frac{1-t}{t^{1-2/q}} + \delta_{2k} + (1 - \delta_{2k})\frac{1-t}{t^{1-2/q}}\right) \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q\right)^{1/q}}{1 - \delta_{2k}}} \\ &= \frac{2\epsilon}{\sqrt{1 - \delta_{2k}}} + k^{1/2-1/q} \sqrt{\frac{(2 - \delta_{2k})(1-t)t^{2/q-1} + \delta_{2k}}{1 - \delta_{2k}}} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q\right)^{1/q} \\ &\leq \frac{2\epsilon}{\sqrt{1 - \delta_{2k}}} + k^{1/2-1/q} \sqrt{\frac{(2 - \delta_{2k})q(2 - q)^{2/q-1} + 2^{2/q}\delta_{2k}}{(1 - \delta_{2k})2^{2/q}}} \left(\sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_{2,q}^q\right)^{1/q} \end{aligned} \quad (47)$$

where we have used the fact that

$$\max_{t \in [0,1]} (1-t)t^{2/q-1} \leq (1-t)t^{2/q-1}|_{t=1-q/2} = \frac{q}{2} \left(1 - \frac{q}{2}\right)^{2/q-1}$$

Thus, it then follows from (45) and (47) that

$$\begin{aligned} \|\mathbf{h}\|_2 &\leq \frac{2\epsilon}{\sqrt{1 - \delta_{2k}}} + k^{1/2-1/q} \sqrt{\frac{(2 - \delta_{2k})q(2 - q)^{2/q-1} + 2^{2/q}\delta_{2k}}{(1 - \delta_{2k})2^{2/q}}} \\ &\quad \times \left(\frac{2^{2/q-1/2}k^{1/q-1/2}\epsilon}{\sqrt{1 - \delta_{2k}}(1 - (f(t_0))^{q/2})^{1/q}} + \frac{2^{2/q-1}\|\mathbf{x}_{T_0^c}\|_{2,q}}{(1 - (f(t_0))^{q/2})^{1/q}}\right) \\ &= k^{1/2-1/q} \sqrt{\frac{(2 - \delta_{2k})q(2 - q)^{2/q-1} + 2^{2/q}\delta_{2k}}{(1 - \delta_{2k})}} \frac{2^{1/q-1}\|\mathbf{x}_{T_0^c}\|_{2,q}}{(1 - (f(t_0))^{q/2})^{1/q}} \\ &\quad + \frac{2\epsilon}{\sqrt{1 - \delta_{2k}}} \left(1 + k^{1/2-1/q} \sqrt{\frac{(2 - \delta_{2k})q(2 - q)^{2/q-1} + 2^{2/q}\delta_{2k}}{(1 - \delta_{2k})}} \frac{2^{1/q-3/2}}{(1 - (f(t_0))^{q/2})^{1/q}}\right) \\ &= \frac{C_1(q, \delta_{2k})\|\mathbf{x}_{T_0^c}\|_{2,q}}{k^{1/q-1/2}} + C_2(q, \delta_{2k})\epsilon \end{aligned} \quad (48)$$

This arrives to the conclusion of Theorem 2.

6 An IRLS algorithm

Inspired by the ideas of [28,33], in this section, we present an efficient IRLS algorithm for the solution of the mixed l_2/l_q norm minimization problem (34). We first rewrite the problem as the following regularized version of the unconstrained smoothed l_2/l_q minimization:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_{2,q}^{\epsilon,q} + \frac{1}{2\tau} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2, \quad (49)$$

where τ is an regularization parameter and

$$\|\mathbf{x}\|_{2,q}^{\epsilon,q} = \sum_{i=1}^m (\|\mathbf{x}[i]\|_2^2 + \epsilon^2)^{q/2}.$$

Let $J_{2,q}(\mathbf{x}, \epsilon, \tau)$ be the objective function associated with (49), that is,

$$J_{2,q}(\mathbf{x}, \epsilon, \tau) = \sum_{i=1}^m (\|\mathbf{x}[i]\|_2^2 + \epsilon^2)^{q/2} + \frac{1}{2\tau} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2. \quad (50)$$

Then, the first-order necessary optimality condition for the solution of \mathbf{x} is

$$\left[\frac{q\mathbf{x}[i]}{(\epsilon^2 + \|\mathbf{x}[i]\|_2^2)^{1-q/2}} \right]_{1 \leq i \leq m} + \frac{1}{\tau} \Phi^T (\Phi\mathbf{x} - \mathbf{y}) = 0.$$

Hence, we define the diagonal weighting matrix W as $W_i = \text{diag}(q^{1/2}(\epsilon^2 + \|\mathbf{x}[i]\|_2^2)^{q/4-1/2})$ for i th block, and after simple calculations, we can obtain the following necessary optimality condition:

$$(\tau W^2 + \Phi^T \Phi)\mathbf{x} = \Phi^T \mathbf{y}. \quad (51)$$

Due to the nonlinearity, there is no straightforward method to solve the above system of equations. But if we fix $W = W^{(t)}$ to be that determined already in the t th iteration step, the solution of (51), set as the $(t+1)$ th iterate, then can be found to be

$$\mathbf{x}^{(t+1)} = (W^{(t)})^{-1} (\Phi (W^{(t)})^{-1})^T \times (\Phi (W^{(t)})^{-1} + \tau \mathbf{I})^{-1} (\Phi (W^{(t)})^{-1})^T \mathbf{y}. \quad (52)$$

This defines a nature iterative method for solution of (49). We formalize such reweighted algorithm as the following:

Algorithm 1. An IRLS algorithm for the unconstrained smoothed l_2/l_q ($0 < q \leq 1$) minimization problem

Step 1. Initialize $\mathbf{x}^{(0)} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2$ and \hat{k} be the estimated block-sparsity. Set $\epsilon_0 = 1, t = 0$.

Step 2. **while** loop **do**

$$W^{(t)} \leftarrow \text{diag}(q^{1/2}(\epsilon_t^2 + \|\mathbf{x}^{(t)}[i]\|_2^2)^{q/4-1/2}),$$

$$i = 1, \dots, m$$

$$A^{(t)} \leftarrow \Phi (W^{(t)})^{-1}$$

$$\mathbf{x}^{(t+1)} \leftarrow (W^{(t)})^{-1} ((A^{(t)})^T (A^{(t)} + \tau \mathbf{I})^{-1} (A^{(t)})^T \mathbf{y}$$

$$\epsilon_{t+1} \leftarrow \min\{\epsilon_t, \alpha r(\mathbf{x}^{(t+1)})_{\hat{k}+1}/N\}$$

$$t \leftarrow t + 1$$

until $\epsilon_{t+1} = 0$ otherwise **repeat the above computation within a reasonable time**

Step 3. Output $\mathbf{x}^{(t+1)}$ to be an approximation solution.

In Algorithm 1, $r(\mathbf{x})_{\hat{k}+1}$ is the $(\hat{k}+1)$ th largest l_2 norm value of the block of \mathbf{x} in the decreasing order, $\alpha \in (0, 1)$ is a number such that $\alpha r(\mathbf{x}^{(1)})/N < 1$, and τ is an appropriately chosen parameter which controls the tolerance of noise term. Although the best τ may change continuously with respect to noise level, we used the fixed value of τ in our numerical implementations.

Obviously, from the update equation (52), we can see that $\mathbf{x}^{(t)}$ can be understood as the minimizer of

$$\frac{1}{2\tau} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 + \frac{1}{2} \|W^{(t)}\mathbf{x}\|_2^2.$$

Due to the iteratively fixing feature of $W^{(t)}$, this gives the name IRLS. Majumdar and Ward [12] first adopted the IRLS methodology to solve a non-convex l_2/l_q block-sparse optimization problem. But their algorithm is unsuitable for noisy block-sparse recovery. Lai et al. [28] proposed an IRLS algorithm for the unstrained l_q ($0 < q \leq 1$) minimization problem and they made an detailed analysis which includes convergence, local convergence rate and error bound of the algorithm. To some extent, our proposed algorithms can be considered as a generalization of their algorithm to the setting of block-sparse recovery.

Note that $\{\epsilon_t\}$ in Algorithm 1 is a bounded non-increasing sequence, which must converge to some $\epsilon_* \geq 0$. Then using a similar argument as that in [28], we can prove that $\mathbf{x}^{(t)}$ must have a convergent subsequence and the limit of the subsequence is a critical point of (50) whenever $\epsilon_* > 0$. In addition, there exists a convergent subsequence whose limit \mathbf{x}_* is a sparse vector with block-sparsity $\|\mathbf{x}_*\|_{2,0} \leq \hat{k}$ when $\epsilon_* = 0$. Furthermore, we can also verify the super-linear local convergence rate for the proposed Algorithm 1. Due to space limitations, we leave the detailed analysis to the interested reader.

7 Numerical experiments

In this section, we conduct two numerical experiments to compare the non-convex l_2/l_q ($0 < q < 1$) minimization method with the l_2/l_1 minimization method and

the standard l_q ($0 < q \leq 1$) method in the context of block-sparse signal recovery. Note that this is possible because Algorithm 1 can apply to the standard l_q ($0 < q \leq 1$) minimization method too. For all compared methods, we use the same starting point $\mathbf{x}^{(0)} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \Phi \mathbf{x}\|_2^2$.

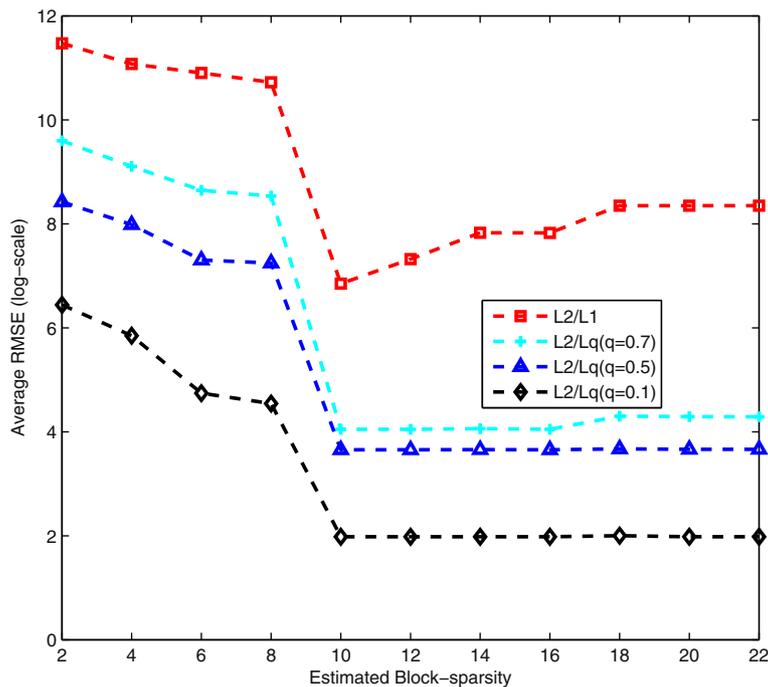
In our experiments, the measurement matrix Φ was generated by creating a $M \times N$ matrix with i.i.d draws from a standard Gaussian distribution $N(0, 1)$. We considered four different values of $q = 0.1, 0.5, 0.7, 1$ for both the l_2/l_q minimization method and the l_q minimization method. The purpose of experiments was to compare the recovery performance of the mixed l_2/l_q method and the l_q method for block-sparse signals without noise and with noise respectively.

7.1 Noiseless recovery

In this set of experiments, we considered the case that the signals were perfectly measured without noise. We first randomly generated the block-sparse signal \mathbf{x} with values chosen from a Gaussian distribution of mean 0 and standard deviation 1 and then randomly drew a measurement matrix Φ from Gaussian ensemble. Then we observed the measurements \mathbf{y} from the model $\mathbf{y} = \Phi \mathbf{x}$. In all the experiment cases, if $\epsilon_{t+1} < 10^{-7}$ or

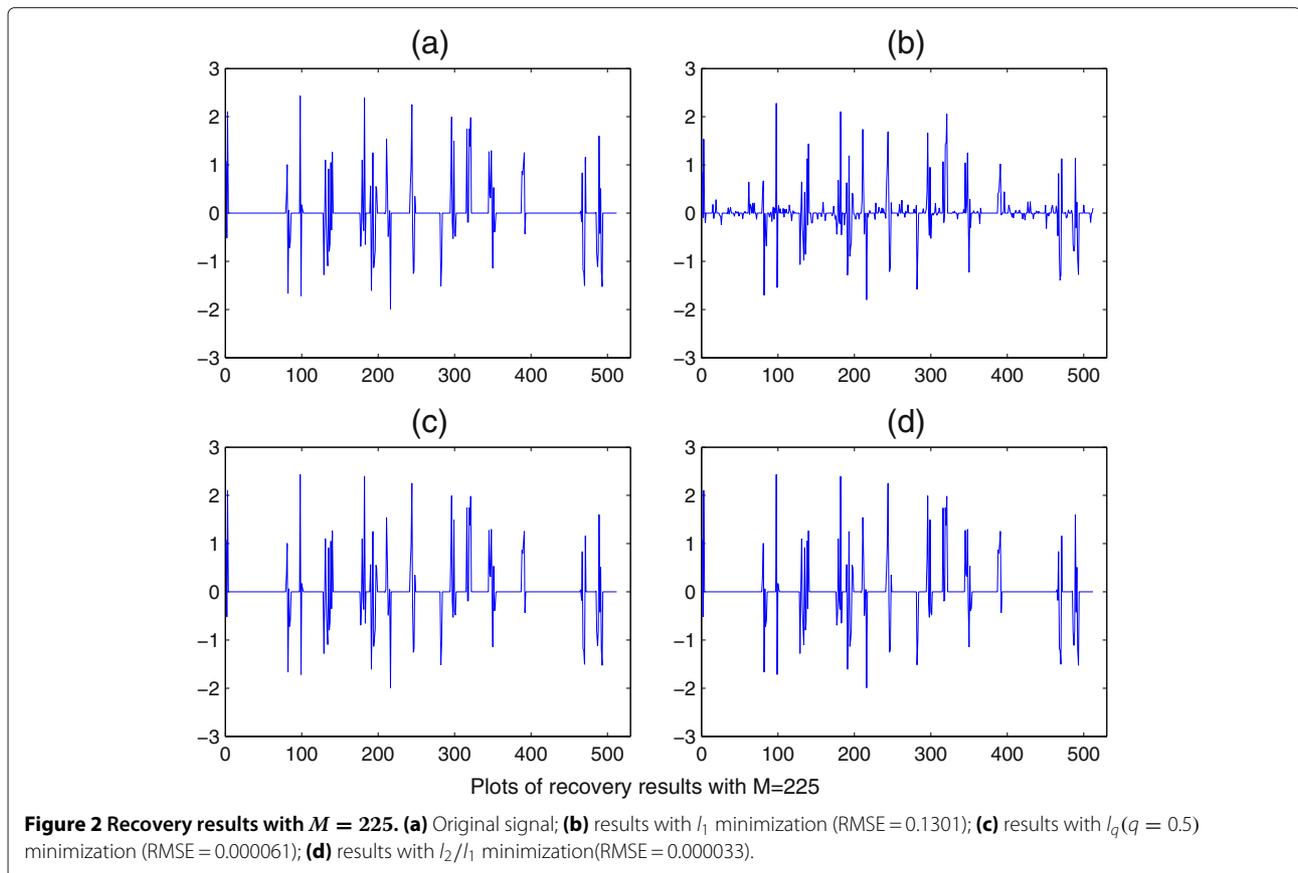
$\|\mathbf{x}^{(t+1)} - \mathbf{x}^{(t)}\|_2 < 10^{-8}$, the iteration terminates and outputs $\mathbf{x}^{(t+1)}$ as an approximation solution of original signal \mathbf{x} ; otherwise, we let the algorithms run to the maximum number of iterations $max = 2000$. We also set parameters τ and α to 10^{-5} and 0.7. We tested Algorithm 1 for different initial block-sparsity estimates and did find that any overestimated \hat{k} of k would yield similar results. A typical simulation result is shown in Figure 1. Therefore, for simplicity, we set $\hat{k} = k + 1$ in our implementation.

Figure 2a depicts an instance of the generated block-sparse signal with signal length $N = 512$. There are 128 blocks with uneven block size and 16 active blocks with the sparsity: $k_0 = \|\mathbf{x}\|_0 = 101$. Figure 2b–d shows that the recovery results by the standard l_q method with $q = 1$, the standard l_q method with $q = 0.5$, and the mixed l_2/l_1 method, respectively, when $M = 225$. Since the sample size M is only around 2.2 times the signal sparsity k_0 , the standard l_1 method does not yield good recovery results, whereas the mixed l_2/l_1 method and the non-convex l_q ($q = 0.5$) method achieve near perfect recovery of the original signal. The results illustrate that if one incorporates the block-sparsity structure into the recovery procedure, the block version of convex l_1 minimization does also reduce the number of measurement



Plots of a typical simulation result with different estimated block-sparsity

Figure 1 A typical simulation result with different estimated block-sparsity. The randomly generated signals have length $N = 256$, $m = 64$ uneven size blocks, and $k = 9$ active blocks with total sparsity $k_0 = 22$. We observed $M = 64$ measurements and the average RMSE over 100 independent random trails in the logarithmic scale is shown.

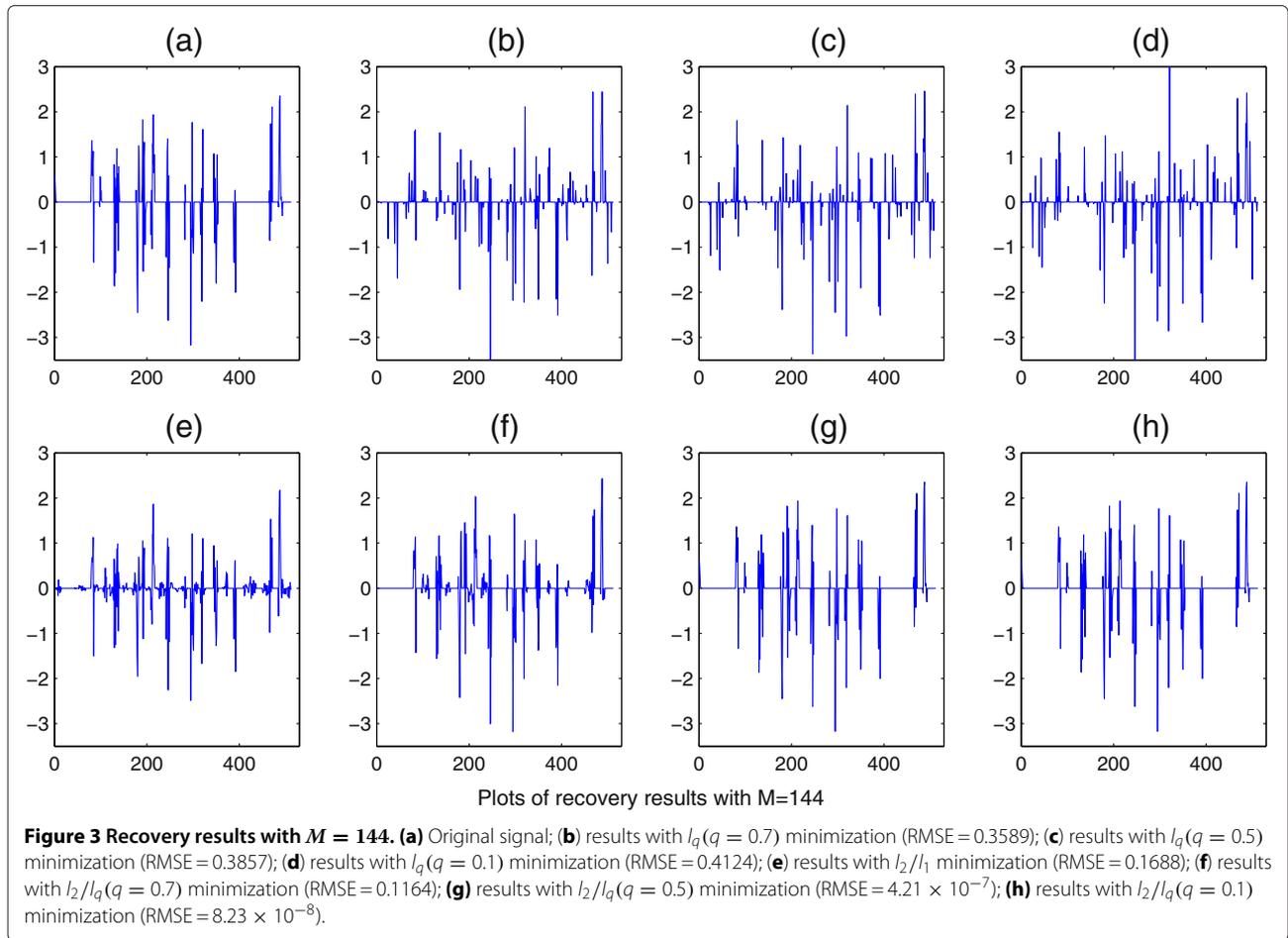


as the standard non-convex l_q minimization with some $q < 1$.

We further compared the recovery performance of the standard l_q method and the mixed l_2/l_q method for different values of q . Figure 3 shows a similar instance as Figure 2. We generated a block-sparse signal with the same non-zero block locations as in Figure 2a, and we observed $M = 144$ measurements that is only around 1.4 times the signal sparsity k_0 . From Figure 3, we can see that only the non-convex l_2/l_q method with $q = 0.5, 0.1$ achieve near optimal recovery results while other methods fail. The results illustrate that for any $q \leq 0.5$, the mixed l_2/l_q method can exactly recover the original signal. In addition, the results also demonstrate the outperformance of the non-convex mixed $l_2/l_q(0 < q < 1)$ method over the standard non-convex $l_q(0 < q < 1)$ method.

Figure 4a shows the effect of sample size, where we report the average root mean squares error (RMSE) over 100 independent random trails in the logarithmic scale for each sample size. In this case, we set signal length $N = 256$, and there are 64 blocks with uneven block size and the $k = 4$ active blocks were randomly extracted from the 64 blocks. The figure indicates the decay in

recovery error as a function of sample size for all the algorithms. We can observe that both the l_q and the mixed l_2/l_q methods improve the recovery performance as q decreases, and for a fixed q , the mixed l_2/l_q method is clearly superior to the standard l_q method in this block-sparse setting. To further study the effect of the active block number k (with k_0 fixed), we drew a matrix Φ of size 128×256 from Gaussian ensemble. We also set the signal x with the even block size and the total sparsity $k_0 = 64$. The block size was changed while keeping other parameters unchanged. Figure 4b shows the average RMSE over 100 independent random runs in the logarithmic scale. One can easily see that the recovery performance for the standard l_q method is independent to the active block number, while the recovery errors for the mixed l_2/l_q method are significantly better when the active block number k is far smaller than the total signal sparsity k_0 . As expected, the performance of the mixed l_2/l_q method becomes identical to the standard l_q method when $k = k_0$. This illustrates that the mixed method favors large sized block when the total sparsity k_0 is fixed. Moreover, similar to the standard l_q method, the mixed l_2/l_q method performs better and better as q decreases.



7.2 Noisy recovery

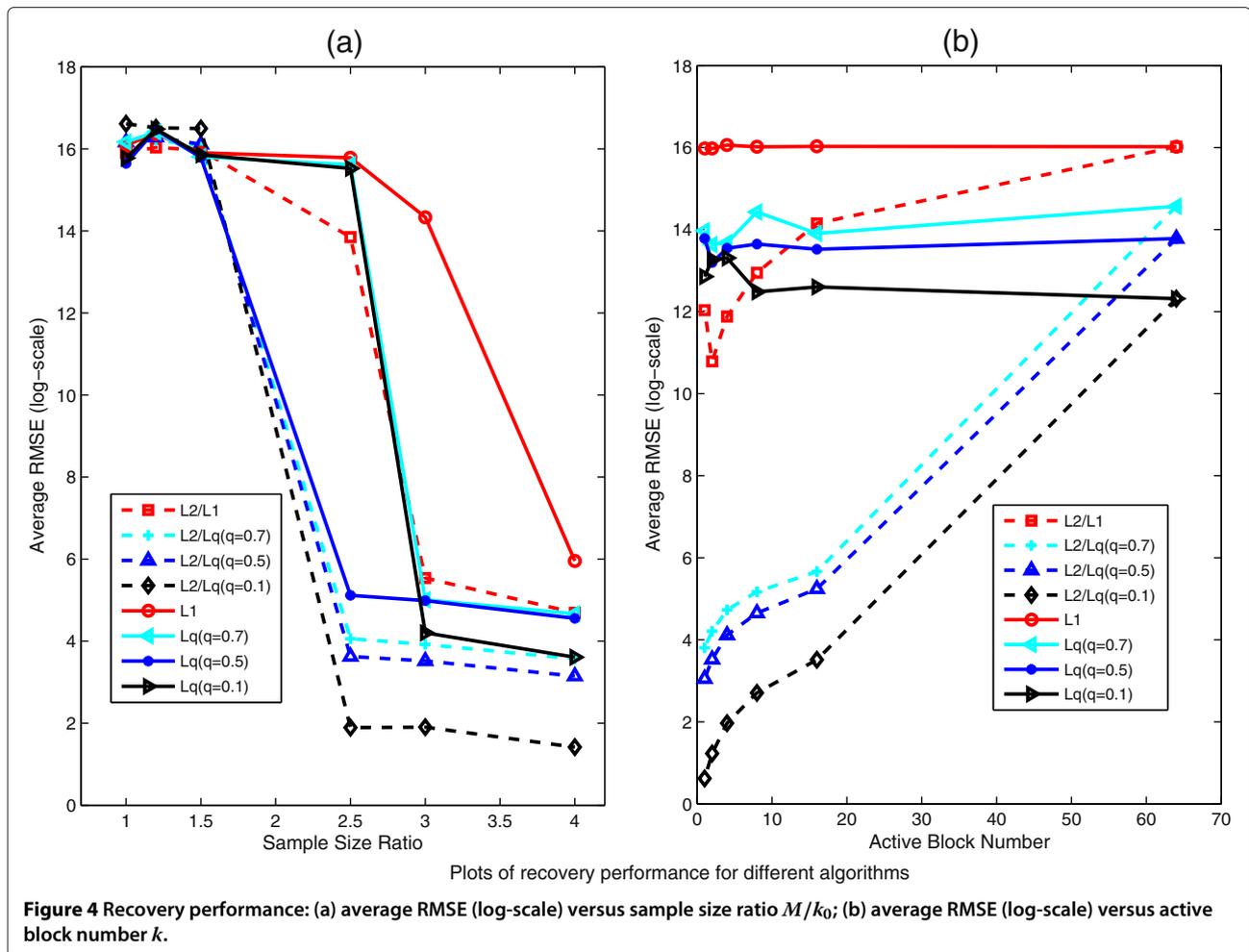
In this experiment, we considered the case of recovering the block-sparse signals in the presence of noise. We observed the measurements \mathbf{y} from the model $\mathbf{y} = \Phi \mathbf{x} + \mathbf{z}$, here Φ and \mathbf{x} were generated as the last subsection and \mathbf{z} was zero-mean Gaussian noise with standard deviation σ . In our implementation of this experiment, we set $\tau = 10^{-1} \max |\Phi^T \mathbf{y}|$ and kept other parameters unchanged.

Table 1 lists the comparison results of the relative errors of the true solutions and the approximate solutions yielded, respectively, by the mixed l_2/l_q method and the l_q method with active block number k varying in $\{4, 12\}$, sample size $r = M/k_0 \in \{3, 4\}$, and the noise level σ in $\{0.02, 0.05, 0.10\}$. Here, the relative errors are defined as $\|\mathbf{x} - \mathbf{x}^*\|_2 / \|\mathbf{x}\|_2$. It is reported in Table 1 that the average relative errors and the standard deviations over 100 random trails. From the table, it is seen that, for a fixed q , the mixed l_2/l_q method always get better results than the standard l_q method. And in the low-noisy cases (say, $\sigma = 0.02, 0.05$), as q decreases, the mixed l_2/l_q method improves the recovery performance. However, when $\sigma = 0.10$, the mixed l_2/l_q method is not always able to improve the recovery results when $q \leq 0.7$. Thus,

we may reasonably infer that there exists a $q_0 \leq 0.1$ such that for any $q < q_0$, all the l_2/l_q minimization can obtain similar recovery results when the noise level is low ($\sigma = 0.02, 0.05$); while there exists a $q_0 \leq 0.7$ such that as $q < q_0$ decreases, the mixed l_2/l_q method is unable to improve the recovery results when $\sigma = 0.10$.

8 Conclusion

In this article, we have investigated the block-sparse recovery performance of the mixed l_2/l_q minimization approach, especially for the non-convex case of $0 < q < 1$. Under the assumption that the measurement matrix Φ has the RIP with $\delta_{2k} < 1/2$, we have proved that the non-convex $l_2/l_q (0 < q < 1)$ method can exactly and stably recover original block-sparse signals in noiseless case and noisy case, respectively. The sufficient recovery condition we obtained is weaker than those of l_2/l_1 method ($\delta_{2k} < 0.414$), which implies the better block-sparse recovery ability of the mixed $l_2/l_q (0 < q < 1)$ method. We have conducted a series of numerical experiments to support the correctness of the theory and the outperformance of the mixed method.



Our study so far is only concerned with the block-sparse signal recovery without overlapping blocks. While in many real applications, such as the gene expression data in bioinformatics, the blocks of elements could potentially be overlapped. Rao et al. [36] derived some tight bounds for the number of measurements required for exact and stable recovery of block-sparse signals with overlapping blocks by the mixed l_2/l_1 minimization method. Their analysis can naturally be extended to the non-convex l_2/l_q ($0 < q < 1$) method. All these extensions will be part of our future research.

Although our simulation studies in this article demonstrate clearly that the recovery ability of the mixed l_2/l_q method would be better and better as q decreases, there is still lack of further theoretical analysis to support for such observation. The works of [37,38] addressed the block-sparse recovery ability by an accurate analysis of the breakdown behavior of the mixed l_2/l_1 method. One could then have an interest in extending these results to the case of the mixed l_2/l_q ($0 < q < 1$) method as well. It is noted that since the resultant mixed minimization problem is

Table 1 Comparison results ($\|x - x^*\|_2 / \|x\|_2$) of randomly generated signals with length $N = 256$ and $m = 64$ uneven size blocks by mixed l_2/l_q method and l_q method with four different values of q and two different sampling sizes $r = M/k_0$

	$\sigma = 0.02$	$\sigma = 0.05$	$\sigma = 0.10$
$k = 4, r = 3$			
Lasso	0.535 ± 0.153	0.543 ± 0.153	0.572 ± 0.153
$l_q(q = 0.7)$	0.467 ± 0.184	0.458 ± 0.177	0.504 ± 0.209
$l_q(q = 0.5)$	0.472 ± 0.185	0.469 ± 0.206	0.507 ± 0.225
$l_q(q = 0.1)$	0.355 ± 0.243	0.364 ± 0.251	0.461 ± 0.253
Group Lasso	0.283 ± 0.149	0.332 ± 0.175	0.426 ± 0.208
$l_2/l_q(q = 0.7)$	0.159 ± 0.125	0.189 ± 0.173	0.312 ± 0.268
$l_2/l_q(q = 0.5)$	0.145 ± 0.095	0.168 ± 0.117	0.313 ± 0.201
$l_2/l_q(q = 0.1)$	0.064 ± 0.042	0.118 ± 0.103	0.324 ± 0.263

Table 1 Comparison results ($\|x - x^*\|_2 / \|x\|_2$) of randomly generated signals with length $N = 256$ and $m = 64$ uneven size blocks by mixed l_2/l_q method and l_q method with four different values of q and two different sampling sizes $r = M/k_0$ (Continued)

$k = 4, r = 4$			
Lasso	0.363 ± 0.093	0.397 ± 0.106	0.421 ± 0.105
$l_q(q = 0.7)$	0.308 ± 0.112	0.327 ± 0.112	0.332 ± 0.102
$l_q(q = 0.5)$	0.307 ± 0.115	0.309 ± 0.105	0.320 ± 0.095
$l_q(q = 0.1)$	0.179 ± 0.083	0.177 ± 0.093	0.240 ± 0.112
Group Lasso	0.193 ± 0.078	0.222 ± 0.064	0.306 ± 0.130
$l_2/l_q(q = 0.7)$	0.126 ± 0.060	0.130 ± 0.048	0.182 ± 0.061
$l_2/l_q(q = 0.5)$	0.112 ± 0.059	0.120 ± 0.063	0.156 ± 0.052
$l_2/l_q(q = 0.1)$	0.041 ± 0.032	0.062 ± 0.024	0.178 ± 0.099
$k = 12, r = 3$			
Lasso	0.419 ± 0.063	0.434 ± 0.059	0.444 ± 0.058
$l_q(q = 0.7)$	0.382 ± 0.064	0.397 ± 0.065	0.399 ± 0.062
$l_q(q = 0.5)$	0.359 ± 0.064	0.372 ± 0.063	0.374 ± 0.057
$l_q(q = 0.1)$	0.162 ± 0.040	0.183 ± 0.047	0.221 ± 0.051
Group Lasso	0.244 ± 0.046	0.264 ± 0.052	0.291 ± 0.046
$l_2/l_q(q = 0.7)$	0.164 ± 0.036	0.182 ± 0.048	0.198 ± 0.036
$l_2/l_q(q = 0.5)$	0.129 ± 0.039	0.148 ± 0.048	0.168 ± 0.034
$l_2/l_q(q = 0.1)$	0.042 ± 0.018	0.073 ± 0.018	0.154 ± 0.048
$k = 12, r = 4$			
Lasso	0.350 ± 0.040	0.362 ± 0.054	0.373 ± 0.048
$l_q(q = 0.7)$	0.330 ± 0.048	0.336 ± 0.053	0.333 ± 0.051
$l_q(q = 0.5)$	0.309 ± 0.049	0.313 ± 0.051	0.308 ± 0.050
$l_q(q = 0.1)$	0.138 ± 0.033	0.149 ± 0.023	0.189 ± 0.037
Group Lasso	0.202 ± 0.030	0.214 ± 0.069	0.242 ± 0.038
$l_2/l_q(q = 0.7)$	0.143 ± 0.029	0.153 ± 0.028	0.175 ± 0.030
$l_2/l_q(q = 0.5)$	0.109 ± 0.029	0.124 ± 0.026	0.155 ± 0.027
$l_2/l_q(q = 0.1)$	0.038 ± 0.014	0.069 ± 0.013	0.157 ± 0.041

non-convex, it seems very difficult to make such similar theoretical analysis and perhaps we need more powerful tools of geometric functional analysis for the extension.

Abbreviations

CS: Compressed sensing; IHT: Iterative hard thresholding; IRLS: Iteratively Reweighted Least-Squares; NSP: Null space property; OMP: Orthogonal matching pursuit; RIP: Restricted isometry property.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The authors would like to thank the anonymous reviewers and the Associate Editor for their valuable comments and suggestions which have helped to improve the presentation of the article. This study was supported by the National 973 Project of China under Grant no. 2013CB329404, the Natural Science Foundation of China under Grant nos. 61273020, 11001227, 11171272, and by the Fundamental Research Funds for the Central Universities (No. XDJK2010B005).

Author details

¹Institute for Information and System Sciences, School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China. ²School of Mathematics and Statistics, Southwest University, Chongqing 400715, China.

Received: 19 August 2012 Accepted: 15 March 2013

Published: 11 April 2013

References

1. C Shannon, Communication in the presence of noise. *Proc. IRE.* **37**, 10–21 (1949)
2. A Jerri, The Shannon sampling theorem—its various extensions and applications: a tutorial review. *Proc. IEEE.* **65**(11), 1565–1596 (1977)
3. D Donoho, Compressed sensing. *IEEE Trans. Inf. Theory.* **52**(4), 1289–1306 (2006)
4. E Candes, J Romberg, T Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory.* **52**(2), 489–509 (2006)
5. E Candes, T Tao, Near-optimal signal recovery from random projections: universal encoding strategies. *IEEE Trans. Inf. Theory.* **52**(12), 5406–5425 (2006)
6. E Candes, T Tao, Decoding by linear programming. *IEEE Trans. Inf. Theory.* **51**(12), 4203–4215 (2005)
7. F Parvaresh, H Vikalo, S Misra, B Hassibi, Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays. *IEEE J. Sel. Topics Signal Process.* **2**(3), 275–285 (2008)
8. S Erickson, C Sabatti, Empirical Bayes estimation of a sparse vector of gene expression changes. *Stat. Appl. Genetics Mol. Biol.* **4**, 22 (2005)
9. S Cotter, B Rao, Sparse channel estimation via matching pursuit with application to equalization. *IEEE Trans. Commun.* **50**(3), 374–377 (2002)
10. D Malioutov, M Cetin, A Willsky, Sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. Signal Process.* **53**(8), 3010–3022 (2005)
11. Y Liu, Q Wan, Enhanced compressive wideband frequency spectrum sensing for dynamic spectrum access. *EURASIP J. Adv. Signal Process.* **2012**(177), 1–11 (2012)
12. A Majumdar, R Ward, Compressed sensing of color images. *Signal Process.* **90**(12), 3122–3127 (2010)
13. J Huang, T Zhang, The benefit of group sparsity. *Ann. Stat.* **38**(4), 1978–2004 (2010)
14. Y Eldar, M Mishali, Robust recovery of signals from a structured union of subspaces. *IEEE Trans. Inf. Theory.* **55**(11), 5302–5316 (2009)
15. M Stojnic, F Parvaresh, B Hassibi, On the reconstruction of block-sparse signals with an optimal number of measurements. *IEEE Trans. Signal Process.* **57**(8), 3075–3085 (2009)
16. R Baraniuk, V Cevher, M Duarte, C Hegde, Model-based compressive sensing. *IEEE Trans. Inf. Theory.* **56**(4), 1982–2001 (2010)
17. Y Eldar, P Kuppinger, H Bolcskei, Block-sparse signals: uncertainty relations and efficient recovery. *IEEE Trans. Signal Process.* **58**(6), 3042–3054 (2010)
18. Z Ben-Haim, Y Eldar, Near-oracle performance of greedy block-sparse estimation techniques from noisy measurements. *IEEE J. Sel. Topics Signal Process.* **5**(5), 1032–1047 (2011)
19. A Majumdar, R Ward, Fast group sparse classification. *IEEE Can. J. Electr. Comput. Eng.* **34**(4), 136–144 (2009)
20. S Foucart, M Lai, Sparsest solutions of underdetermined linear systems via l_q -minimization for $0 < q \leq 1$. *Appl. Comput. Harmonic Anal.* **26**(3), 395–407 (2009)
21. R Chartrand, V Staneva, Restricted isometry properties and nonconvex compressive sensing. *Inverse Problems.* **24**(03520), 1–14 (2008)
22. R Saab, R Chartrand, O Yilmaz, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Stable sparse approximations via nonconvex optimization (IEEE, 2008), pp. 3885–3888
23. R Chartrand, Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Process. Lett.* **14**(10), 707–710 (2007)
24. Z Xu, H Zhang, Y Wang, X Chang, Y Liang, $l_{1/2}$ regularization. *Sci. China Inf. Sci.* **53**(6), 1159–1169 (2010)
25. Z Xu, H Guo, Y Wang, H Zhang, The representative of $l_{1/2}$ regularization among l_q ($0 < q \leq 1$) regularizations: an experimental study based on phase diagram. *Acta Autom. Sinica.* **38**(7), 1225–1228 (2012)

26. Z Xu, X Chang, F Xu, H Zhang, $L_{1/2}$ regularization: a thresholding representation theory and a fast solver. *IEEE Trans. Neural Netw. Learn. Syst.* **23**(7), 1013–1027 (2012)
27. Q Sun, Recovery of sparsest signals via l_q -minimization. *Appl. Comput. Harmonic Anal.* **32**, 329–341 (2012)
28. M Lai, Y Xu, W Yin, Improved iteratively reweighted least squares for unconstrained smoothed l_q minimization. *SIAM J. Numer. Anal.* (2013). [http://www.math.uga.edu/~mjlai/papers/LXY2013.pdf]
29. E Candes, The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique.* **346**(9–10), 589–592 (2008)
30. S Chen, D Donoho, M Saunders, Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20**, 33–61 (1998)
31. M Yuan, Y Lin, Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc.: Series B (Statistical Methodology).* **68**, 49–67 (2006)
32. M Fornasier, H Rauhut, Recovery algorithms for vector-valued data with joint sparsity constraint. *SIAM J. Numer. Anal.* **46**(2), 577–613 (2009)
33. I Daubechies, R DeVore, M Fornasier, C Güntürk, Iteratively reweighted least squares minimization for sparse recovery. *Commun. Pure Appl. Math.* **63**, 1–38 (2010)
34. A Aldroubi, X Chen, A Powell, in *Proceedings of SampTA*. Stability and robustness of l_q minimization using null space property (Singapore, 2011)
35. M Lai, L Liu, A new estimate of restricted isometry constants for sparse solutions (2011). http://www.math.uga.edu/~mjlai/papers/LaiLiu11.pdf
36. N Rao, B Recht, R Nowak, Tight measurement bounds for exact recovery of structured sparse signals. *Arxiv preprint arXiv.* **1106**, 4355 (2011)
37. M Stojnic, l_2/l_1 -optimization in block-sparse compressed sensing and its strong thresholds. *IEEE J. Sel. Topics Signal Process.* **4**(2), 350–357 (2010)
38. M Stojnic, Block-length dependent thresholds in block-sparse compressed sensing. *Arxiv preprint arXiv.* **0907**, 3679 (2009)

doi:10.1186/1687-6180-2013-76

Cite this article as: Wang et al.: On recovery of block-sparse signals via mixed l_2/l_q ($0 < q \leq 1$) norm minimization. *EURASIP Journal on Advances in Signal Processing* 2013 **2013**:76.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
