Signal Processing 104 (2014) 188–196

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Fast communication

Restricted p-isometry properties of nonconvex block-sparse compressed sensing ☆

Yao Wang [a], Jianjun Wang [b,*], Zongben Xu [a]

[a] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China
[b] School of Mathematics and Statistics, Southwest University, Chongqing 400715, China

ARTICLE INFO

Article history:
Received 22 July 2013
Received in revised form
21 February 2014
Accepted 27 March 2014
Available online 12 April 2014

Keywords:
Compressed sensing
Block-sparse signal recovery
Mixed l_2/l_p-minimization
Restricted p-isometry properties
Gaussian measurements

ABSTRACT

In this paper, by generalizing the notion of restricted p-isometry constant $(0<p\leq 1)$ defined by Chartrand and Staneva [1] to the setting of block-sparse signal recovery, we establish a general restricted p-isometry property (p-RIP) condition for recovery of (nearly) block-sparse signals via mixed l_2/l_p-minimization. Moreover, we derive a lower bound on the necessary number of Gaussian measurements for the p-RIP condition to hold with high probability, which shows clearly that fewer measurements with smaller p are needed for exact recovery of block-sparse signals via mixed l_2/l_p-minimization than when $p=1$.

1. Introduction

The problem of block-sparse signal recovery naturally arises in a number of genetics, communications, image processing and computer vision applications. Prominent examples include DNA microarrays [2], equalization of sparse communication channels [3], color imaging [4], motion segmentation [5], and more. In such contexts, we often require to recover an unknown signal $\mathbf{x}\in\mathbb{R}^N$ from an underdetermined system of linear equations $\mathbf{y}=\Phi\mathbf{x}$, where $\mathbf{y}\in\mathbb{R}^M$ are available measurements, and Φ is an $M\times N$ $(M<N)$ measurement matrix. Unlike previous works in compressed sensing [6,7], the unknown signal $\mathbf{x}$ is not only sparse but also exhibits additional structure in the form that the nonzero coefficients appear in some fixed blocks. We refer to such signal as a block-sparse signal in this paper. Following [8,9], we model a block-sparse signal $\mathbf{x}\in\mathbb{R}^N$ over $\mathcal{I}=\{d_1,d_2,\ldots,d_m\}$ as concatenation of $\mathbf{x}$ in m blocks of length d_i, i.e.,

$$\mathbf{x}=\Big[\underbrace{x_1\ldots x_{d_1}}_{\mathbf{x}[1]}\underbrace{x_{d_1+1}\ldots x_{d_1+d_2}}_{\mathbf{x}[2]}\ldots\underbrace{x_{N-d_m+1}\ldots x_N}_{\mathbf{x}[m]}\Big]^T, \tag{1}$$

where $\mathbf{x}[i]$ denotes the ith block of $\mathbf{x}$ and $N=\sum_{i=1}^m d_i$. In this term, we say that $\mathbf{x}$ is a block k-sparse if $\mathbf{x}[i]$ has nonzero Euclidean norm for at most k blocks. If $d_i=1$ for all $i\in\{1,2,\ldots,m\}$, the block-sparse signal degenerates to the conventional sparse signal well studied in compressed sensing. Denoting $\|\mathbf{x}\|_{2,0}=\sum_{i=1}^m I(\|\mathbf{x}[i]\|_2>0)$ with an indicator function $I(\cdot)$, a block k-sparse signal $\mathbf{x}$ thus can be defined as a signal that satisfies $\|\mathbf{x}\|_{2,0}\leq k$. It is known that under certain conditions on measurement matrix Φ (i.e., [8]), there is a unique block-sparse signal that obeys to the observation $\mathbf{y}=\Phi\mathbf{x}$ and can be exactly recovered by solving the mixed l_2/l_0 norm minimization:

$$\min_{\mathbf{x}}\|\mathbf{x}\|_{2,0}\quad \text{subject to } \mathbf{y}=\Phi\mathbf{x}. \tag{2}$$

☆ This work was supported in part by Natural Science Foundation of China under grant numbers 61273020 and 61075054 and National 973 Project of China under grant number 2013CB329404.
* Corresponding author.
E-mail addresses: yao.s.wang@gmail.com (Y. Wang), wjj@swu.edu.cn (J. Wang), zbxu@mail.xjtu.edu.cn (Z. Xu).

http://dx.doi.org/10.1016/j.sigpro.2014.03.040
0165-1684/

If the measurement $\mathbf{y}$ is moderately flawed, the above problem (2) turns to be the following noisy version:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_{2,0} \quad \text{subject to } \|\Phi\mathbf{x}-\mathbf{y}\|_2 \le \epsilon, \tag{3}$$

where ϵ represents the noise level. Similar as the standard l_0-minimization problem, problem (2) is also NP-hard and computationally intractable except for very small size. Motivated by the study of compressed sensing, one then commonly uses the strategy to replace the l_2/l_0 norm with its closest convex surrogate l_2/l_1 norm, then solve the mixed l_2/l_1 norm minimization problem:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_{2,1} \quad \text{subject to } \mathbf{y}=\Phi\mathbf{x}, \tag{4}$$

where $\|\mathbf{x}\|_{2,1}=\sum_{i=1}^{m}\|\mathbf{x}[i]\|_2$. (4) is a convex optimization problem and can be recast as a second-order cone program, thus can be solved very efficiently.

To investigate the theoretical performance of mixed l_2/l_1-minimization, Eldar and Mishali [8] introduced the notion of the block-restricted isometry constant (block-RIC), $\delta_{k|\mathcal{I}}$, of a measurement matrix $\Phi\in\mathbb{R}^{M\times N}$, i.e., $\delta_{k|\mathcal{I}}$ is the smallest non-negative number such that

$$(1-\delta_{k|\mathcal{I}})\|\mathbf{c}\|_2^2 \le \|\Phi\mathbf{c}\|_2^2 \le (1+\delta_{k|\mathcal{I}})\|\mathbf{c}\|_2^2, \tag{5}$$

for all vectors $\mathbf{c}\in\mathbb{R}^N$ that are block k-sparse over $\mathcal{I}$. In terms of block-RIC, they established a sufficient condition, i.e., the block-restricted isometry property (block-RIP) condition for exact recovery of block-sparse signals. More precisely, if the measurement matrix Φ satisfies $\delta_{2k|\mathcal{I}}<\sqrt{2}-1$, then the mixed l_2/l_1 minimization method is guaranteed to recover any block k-sparse signal exactly. Furthermore, it is also proved that random matrices with Gaussian entries satisfy the block-RIP with overwhelming probability. Recently, Lin and Li [10] improved the sufficient condition on $\delta_{2k|\mathcal{I}}$ from 0.414 to 0.4931, and established condition $\delta_{k|\mathcal{I}}<0.307$ for exact recovery. There are also a number of works based on non-RIP analysis to characterize the theoretical performance of mixed l_2/l_1-minimization. For example, Eldar et al. [9] provided the exact and robust recovery conditions based on block coherence; Huang and Zhang [11] developed a theory for the mixed l_2/l_1 minimization method by using a concept called strong group sparsity.

Among the latest researches in standard compressed sensing, many authors [12–14] have showed that l_p-minimization with $0<p<1$ allows exact recovery of conventional sparse signals from much fewer linear measurements than that by l_1 minimization. Naturally, it would be interesting to make an ongoing effort to extend the l_p $(0<p<1)$ norm minimization to the setting of block-sparse signal recovery. Specifically, one can replace (4) with the following mixed $l_2/l_p(0<p\le 1)$ norm minimization problem:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_{2,p}^p \quad \text{subject to } \mathbf{y}=\Phi\mathbf{x}, \tag{6}$$

where $\|\mathbf{x}\|_{2,p}=(\sum_{i=1}^{m}\|\mathbf{x}[i]\|_2^p)^{1/p}$. Some numerical experiments [4,15] demonstrated that fewer measurements are needed for exact recovery when $0<p<1$ than when $p=1$. Moreover, exact recovery conditions based on block-RIP have also been studied [15].

In this paper, we further investigate the exact recovery for block-sparse signals via mixed l_2/l_p-minimization. We will introduce a block variant of the restricted p-isometry constant defined by Chartrand and Staneva [1]. With this notion, more generally, we will establish a p-RIP condition for stable recovery of nearly block-sparse signals via the following noisy model:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_{2,p}^p \quad \text{subject to } \|\Phi\mathbf{x}-\mathbf{y}\|_2 \le \epsilon, \tag{7}$$

and derive an error bound between the solution of (7) and the unknown original signal $\mathbf{x}$. It is obvious that when $\mathbf{x}$ is exactly block-sparse and $\epsilon=0$ (where $\mathbf{y}=\Phi\mathbf{x}$), we will obtain the exact recovery condition. In particular, we will determine how many rows of Gaussian random matrix are sufficient for the p-RIP condition to be satisfied with high probability.

2. Restricted p-isometry properties

We begin with introducing the notion of block restricted p-isometry constant, which is a natural extension of restricted p-isometry constant.

Definition 2.1. Given a measurement matrix $\Phi\in\mathbb{R}^{M\times N}$ and $0<p\le 1$. Let k be a positive integer. Then the block restricted p-isometry constant (block p-RIC) $\delta_{k|\mathcal{I}}$ of order k is defined to be the smallest positive number such that

$$(1-\delta_{k|\mathcal{I}})\|\mathbf{c}\|_2^p \le \|\Phi\mathbf{c}\|_p^p \le (1+\delta_{k|\mathcal{I}})\|\mathbf{c}\|_2^p \tag{8}$$

for all $\mathbf{c}\in\mathbb{R}^N$ that are block k-sparse over block index set $\mathcal{I}$.

For convenience, in the remainder of this paper, we write δ_k for the block p-RIC $\delta_{k|\mathcal{I}}$ whenever the confusion is not caused. With this new notion, we will establish a restricted p-isometry (p-RIP) condition for stable recovery of nearly block-sparse signals via mixed l_2/l_p minimization (7). Hereafter, we denote the best block k-sparse approximation by $\mathbf{x}_k$, i.e, the vector consisting of the k-largest blocks over $\mathcal{I}$ of $\mathbf{x}\in\mathbb{R}^N$ in l_2 norm. More precisely, $\mathbf{x}_k$ can be defined as

$$\mathbf{x}_k = \arg\min_{\|\mathbf{v}\|_{2,0}\le k} \|\mathbf{x}-\mathbf{v}\|_{2,1}$$

Theorem 2.1. *Let $\Phi\in\mathbb{R}^{M\times N}$ be a measurement matrix, $\mathbf{x}\in\mathbb{R}^N$ be a nearly block k-sparse signal, and $0<p\le 1$. Let $b>1$, $a=b^{2/(2-p)}$, rounded up so that ak is an integer (that is, $a=\lceil b^{2/(2-p)}k\rceil/k$). If Φ satisfies*

$$\delta_{ak}+b\delta_{(a+1)k}<b-1, \tag{9}$$

then a minimizer $\mathbf{x}^*$ of problem (7) obeys

$$\|\mathbf{x}-\mathbf{x}^*\|_2 \le C_1\frac{\|\mathbf{x}-\mathbf{x}_k\|_{2,p}}{k^{1/p-1/2}}+C_2\epsilon, \tag{10}$$

for some constants C_1 and C_2, which are given explicitly below.

Proof. The following proof makes use of the ideas from [1,13]. Set $\mathbf{x}^*=\mathbf{x}+\mathbf{h}$ be a solution of problem (7), where $\mathbf{x}$ is the original signal we need to recover. Throughout the paper, $\mathbf{x}_T$ will denote the vector equal to $\mathbf{x}$ on an block 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 $\mathbf{h}_{T_0}$, $\mathbf{h}_{T_1}$, $\mathbf{h}_{T_2},\ldots,\mathbf{h}_{T_J}$, such that $\mathbf{h}=\sum_{i=0}^{J}\mathbf{h}_{T_i}$, and $\mathbf{h}_{T_i}$ is the restriction of $\mathbf{h}$ to the set T_i, T_0 consists of k blocks and each $T_i (i\geq 1)$ consists of ak blocks (except possibly T_J). Rearranging the block indices such that $\|\mathbf{h}_{T_j}[1]\|_2 \geq \|\mathbf{h}_{T_j}[2]\|_2 \geq \cdots \geq \|\mathbf{h}_{T_j}[ak]\|_2 \geq \|\mathbf{h}_{T_{j+1}}[1]\|_2 \geq \|\mathbf{h}_{T_{j+1}}[2]\|_2 \geq \cdots$, for any $j\geq 1$. Using Holder's inequality, we obtain

$$\|\Phi\mathbf{h}\|_p^p \leq \left(\sum_{i=1}^{M}\left(|(\Phi\mathbf{h})_i|^p\right)^{2/p}\right)^{p/2}\cdot\left(\sum_{i=1}^{M}1\right)^{1-p/2} = M^{1-p/2}\|\Phi\mathbf{h}\|_2^p.$$

By $\|\Phi\mathbf{x}-\mathbf{y}\|_2\leq\epsilon$ and the triangle inequality, we have

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

Thus,

$$\|\Phi\mathbf{h}\|_p^p \leq M^{1-p/2}\|\Phi\mathbf{h}\|_2^p \leq M^{1-p/2}(2\epsilon)^p. \tag{11}$$

Since $\mathbf{x}^*$ is a minimizer of problem (7), we have

$$\begin{aligned}\|\mathbf{x}_{T_0}\|_{2,p}^p+\|\mathbf{x}_{T_0^C}\|_{2,p}^p &= \|\mathbf{x}\|_{2,p}^p \geq \|\mathbf{x}^*\|_{2,p}^p = \|\mathbf{x}+\mathbf{h}\|_{2,p}^p\\ &= \|\mathbf{x}_{T_0}+\mathbf{h}_{T_0}\|_{2,p}^p+\|\mathbf{x}_{T_0^C}+\mathbf{h}_{T_0^C}\|_{2,p}^p\\ &\geq \|\mathbf{x}_{T_0}\|_{2,p}^p-\|\mathbf{h}_{T_0}\|_{2,p}^p+\|\mathbf{h}_{T_0^C}\|_{2,p}^p-\|\mathbf{x}_{T_0^C}\|_{2,p}^p,\end{aligned} \tag{12}$$

where T_0^C denotes the complement of T_0 in $\{1,\ldots,N\}$. Therefore,

$$\|\mathbf{h}_{T_0^C}\|_{2,p}^p \leq \|\mathbf{h}_{T_0}\|_{2,p}^p+2\|\mathbf{x}_{T_0^C}\|_{2,p}^p. \tag{13}$$

From the decomposition of $\mathbf{h}$, for each $i\in T_j (j\geq 2)$, it is easy to see that

$$\|\mathbf{h}_{T_j}[i]\|_2^p \leq \frac{\|\mathbf{h}_{T_{j-1}}[1]\|_2^p+\cdots+\|\mathbf{h}_{T_{j-1}}[ak]\|_2^p}{ak} = \frac{\|\mathbf{h}_{T_{j-1}}\|_{2,p}^p}{ak}. \tag{14}$$

Then

$$\|\mathbf{h}_{T_j}[i]\|_2^2 \leq \frac{\|\mathbf{h}_{T_{j-1}}\|_{2,p}^2}{(ak)^{2/p}},\ \|\mathbf{h}_{T_j}\|_2^2 \leq \frac{ak\|\mathbf{h}_{T_{j-1}}\|_{2,p}^2}{(ak)^{2/p}},$$

$$\|\mathbf{h}_{T_j}\|_2^p \leq \frac{\|\mathbf{h}_{T_{j-1}}\|_{2,p}^p}{(ak)^{1-p/2}}.$$

So that

$$\sum_{j\geq 2}\|\mathbf{h}_{T_j}\|_2^p \leq (ak)^{p/2-1}\sum_{j\geq 2}\|\mathbf{h}_{T_{j-1}}\|_{2,p}^p \leq (ak)^{p/2-1}\|\mathbf{h}_{T_0^C}\|_{2,p}^p. \tag{15}$$

From the Definition 2.1, we have

$$\begin{aligned}\|\Phi\mathbf{h}\|_p^p &= \|\Phi(\mathbf{h}_{T_0}+\mathbf{h}_{T_1})+\Phi(\mathbf{h}_{T_0^C}-\mathbf{h}_{T_1})\|_p^p \geq \|\Phi(\mathbf{h}_{T_0}\\ &\quad+\mathbf{h}_{T_1})\|_p^p-\sum_{j\geq 2}\|\Phi(\mathbf{h}_{T_j})\|_p^p\\ &\geq (1-\delta_{(a+1)k})\|\mathbf{h}_{T_0}+\mathbf{h}_{T_1}\|_2^p-(1+\delta_{ak})\sum_{j\geq 2}\|\mathbf{h}_{T_j}\|_2^p.\end{aligned}$$

Then, it is not hard to get

$$\begin{aligned}\|\Phi\mathbf{h}\|_p^p &\geq (1-\delta_{(a+1)k})\|\mathbf{h}_{T_0}+\mathbf{h}_{T_1}\|_2^p-(1+\delta_{ak})(ak)^{p/2-1}\|\mathbf{h}_{T_0^C}\|_{2,p}^p\\ &\geq (1-\delta_{(a+1)k})\|\mathbf{h}_{T_0}+\mathbf{h}_{T_1}\|_2^p\\ &\quad-(1+\delta_{ak})(ak)^{p/2-1}(\|\mathbf{h}_{T_0}\|_{2,p}^p+2\|\mathbf{x}_{T_0^C}\|_{2,p}^p)\\ &\geq (1-\delta_{(a+1)k}-(1+\delta_{ak})/b)\|\mathbf{h}_{T_0}\\ &\quad+\mathbf{h}_{T_1}\|_2^p-2(1+\delta_{ak})(ak)^{p/2-1}\|\mathbf{x}_{T_0^C}\|_{2,p}^p,\end{aligned} \tag{16}$$

where first inequality holds from (15), the second inequality holds from (13) and the last inequality follows from the fact that $\|\mathbf{h}_{T_0}\|_{2,p}^p \leq k^{1-p/2}\|\mathbf{h}_{T_0}\|_2^p \leq k^{1-p/2}\|\mathbf{h}_{T_0}+\mathbf{h}_{T_1}\|_2^p$.

Therefore, if $\delta_{ak}+b\delta_{(a+1)k}<b-1$, by combining (11) and (16), we obtain

$$\|\mathbf{h}_{T_0}+\mathbf{h}_{T_1}\|_2^p \leq \frac{2(1+\delta_{ak})k^{p/2-1}\|\mathbf{x}-\mathbf{x}_k\|_{2,p}^p}{b-b\delta_{(a+1)k}-1-\delta_{ak}} + \frac{bM^{1-p/2}(2\epsilon)^p}{b-b\delta_{(a+1)k}-1-\delta_{ak}}. \tag{17}$$

On the other hand,

$$\begin{aligned}\|\mathbf{h}_{(T_0\cup T_1)^c}\|_2 &= \|\sum_{j\geq 2}\mathbf{h}_{T_j}\|_2 \leq \sum_{j\geq 2}\|\mathbf{h}_{T_j}\|_2\\ &\leq (ak)^{1/2-1/p}\sum_{j\geq 2}\|\mathbf{h}_{T_{j-1}}\|_{2,p}\\ &= (ak)^{1/2-1/p}\sum_{j\geq 1}\|\mathbf{h}_{T_j}\|_{2,p}.\end{aligned} \tag{18}$$

Hence

$$\begin{aligned}\|\mathbf{h}_{(T_0\cup T_1)^c}\|_2^p &\leq (ak)^{p/2-1}\left(\sum_{j\geq 1}\|\mathbf{h}_{T_j}\|_{2,p}\right)^p\\ &\leq (ak)^{p/2-1}\sum_{j\geq 1}\|\mathbf{h}_{T_j}\|_{2,p}^p = (ak)^{p/2-1}\|\mathbf{h}_{T_0^C}\|_{2,p}^p,\end{aligned}$$

where the second inequality follows from the fact that $(a_1+\cdots+a_n)^p \leq a_1^p+\cdots+a_n^p$ holds for nonnegative constants $a_1,\ldots,a_n$. Therefore, we get

$$\begin{aligned}\|\mathbf{h}_{(T_0\cup T_1)^c}\|_2^p &\leq (ak)^{p/2-1}\|\mathbf{h}_{T_0^C}\|_{2,p}^p\\ &\leq (ak)^{p/2-1}\left(\|\mathbf{h}_{T_0}\|_{2,p}^p+2\|\mathbf{x}_{T_0^C}\|_{2,p}^p\right)\\ &\leq a^{p/2-1}\|\mathbf{h}_{T_0}\|_2^p+2(ak)^{p/2-1}\|\mathbf{x}_{T_0^C}\|_{2,p}^p\\ &\leq \frac{1}{b}\|\mathbf{h}_{T_0}+\mathbf{h}_{T_1}\|_2^p+2\frac{k^{p/2-1}}{b}\|\mathbf{x}_{T_0^C}\|_{2,p}^p\\ &\leq \frac{2(1-\delta_{(a+1)k})k^{p/2-1}}{b-b\delta_{(a+1)k}-1-\delta_{ak}}\|\mathbf{x}_{T_0^C}\|_{2,p}^p\\ &\quad+\frac{2^pM^{1-p/2}\epsilon^p}{b-b\delta_{(a+1)k}-1-\delta_{ak}},\end{aligned} \tag{19}$$

where the third inequality is a result of the fact that $\|\mathbf{h}_{T_0}\|_{2,p}^p \leq k^{1-p/2}\|\mathbf{h}_{T_0}\|_2^p$, and the last inequality follows from (17).

Since $\|\mathbf{v}\|_p \leq 2^{1/p-1}\|\mathbf{v}\|_1$ for $\mathbf{v}\in\mathbb{R}^2$, it is not hard to see that

$$\begin{aligned}\|\mathbf{h}\|_2 &\leq \|\mathbf{h}_{T_0}+\mathbf{h}_{T_1}\|_2+\|\mathbf{h}_{(T_0\cup T_1)^c}\|_2\\ &\leq 2^{1/p-1}\left(\frac{2^{1/p}(1+\delta_{ak})^{1/p}k^{1/2-1/p}\|\mathbf{x}-\mathbf{x}_k\|_{2,p}}{(b-b\delta_{(a+1)k}-1-\delta_{ak})^{1/p}}\right.\\ &\quad\left.+\frac{2b^{1/p}M^{1/p-1/2}\epsilon}{(b-b\delta_{(a+1)k}-1-\delta_{ak})^{1/p}}\right)\\ &\quad+2^{1/p-1}\left(\frac{2^{1/p}(1-\delta_{(a+1)k})^{1/p}k^{1/2-1/p}}{(b-b\delta_{(a+1)k}-1-\delta_{ak})^{1/p}}\|\mathbf{x}-\mathbf{x}_k\|_{2,p}\right.\\ &\quad\left.+\frac{2\,M^{1/p-1/2}\epsilon}{(b-b\delta_{(a+1)k}-1-\delta_{ak})^{1/p}}\right)\\ &\leq \frac{2^{2/p-1}[(1-\delta_{(a+1)k})^{1/p}+(1+\delta_{ak})^{1/p}]}{(b-b\delta_{(a+1)k}-1-\delta_{ak})^{1/p}}\frac{\|\mathbf{x}-\mathbf{x}_k\|_{2,p}}{k^{1/p-1/2}}\\ &\quad+\frac{2^{1/p}M^{1/p-1/2}(1+b^{1/p})\epsilon}{(b-b\delta_{(a+1)k}-1-\delta_{ak})^{1/p}} = C_1\frac{\|\mathbf{x}-\mathbf{x}_k\|_{2,p}}{k^{1/p-1/2}}+C_2\epsilon.\end{aligned} \tag{20}$$

This arrives to the conclusion of Theorem 2.1. □

Remark 2.1. Though we have only considered l_2 bounded noise in Theorem 2.1, the conclusion, however, can be applied directly to Gaussian noise.

Corollary 2.1. *Let $\Phi \in \mathbb{R}^{M\times N}$ be a measurement matrix, $\mathbf{x}\in\mathbb{R}^N$ be a block k-sparse signal with $\mathbf{y}=\Phi\mathbf{x}$, and $0<p\le 1$. Let $b>1$, $a=b^{2/(2-p)}$, rounded up so that ak is an integer (that is, $a=\lceil b^{2/(2-p)}k\rceil/k$). If Φ satisfies*

$$\delta_{ak}+b\delta_{(a+1)k}<b-1,$$

then the unique minimizer of problem (6) *is exactly* $\mathbf{x}$.

Remark 2.2. Note that if we define δ_k^c in the slightly stronger version as the smallest number such that

$$(1-\delta_k^c)\|\mathbf{c}\|_2^p \le (1/c)\|\Phi\mathbf{c}\|_p^p \le (1+\delta_k^c)\|\mathbf{c}\|_2^p \tag{21}$$

holds for all block k-sparse signals $\mathbf{x}\in\mathbb{R}^N$ and $c>0$, then with the similar argument used in [1], we have the exact recovery condition as

$$\delta_{ak}^c+b\delta_{(a+1)k}^c<b-1. \tag{22}$$

3. Gaussian random matrices

In this section, we will determine how many random Gaussian measurements are needed for (22) to be satisfied with high probability. In the sequel, we denote by Φ an $M\times N$ matrix with i.i.d Gaussian random entries, specifically, $\Phi\sim\mathcal{N}(0,\sigma^2)$. As in [1], for a given p, let $\mu_p := \sigma^p 2^{p/2}\Gamma(p+1/2)/\sqrt{\pi}$.

Lemma 3.1 (*Chartrand and Staneva [1, Lemma 3.3]*). *Let $0<p\le 1$ and ϕ be an $M\times L$ submatrix of Φ. Suppose $\delta>0$. Choose $\eta,\tau>0$ such that $\eta+\tau^p/1-\tau^p\le\delta$. Then*

$$(1-\delta)M\mu_p\|\mathbf{c}\|_2^p \le \|\phi\mathbf{c}\|_p^p \le (1+\delta)M\mu_p\|\mathbf{c}\|_2^p \tag{23}$$

holds uniformly for $\mathbf{c}\in\mathbb{R}^L$ with probability exceeding $1-2(1+2/\tau)^L e^{-\eta^2 M/2pc_p^2}$, where

$$c_p=(31/40)^{1/4}\left[1.13+\sqrt{p}\left(\frac{\Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}}\right)^{-1/p}\right]. \tag{24}$$

The above Lemma 3.1 will be very useful for our proof of the following Theorem 3.1.

Theorem 3.1. *Let Φ be an $M\times N$ ($M<N$) matrix whose entries are i.i.d random variable distributed normally with mean zero and variance σ^2. Then there exist constants $C_3(p)$ and $C_4(p)$ such that whenever $0<p\le 1$ and $M\ge C_3(p)kd+pC_4(p)k\ln(m/k)$, the following is true with probability exceeding $1-2e^{-\beta(p)M}$: for any block k-sparse signal $\mathbf{x}\in\mathbb{R}^N$ over $\mathcal{I}=\{d_1=d,d_2=d,\dots,d_m=d\}$ with $N=md$ for some integer m, $\mathbf{x}$ is the unique solution of problem* (6), *where $\beta(p)$ is a positive constant.*

Proof. The proof is similar to the procedure of proof of Theorem 1.1 in [16]. We will make use of the stronger version (21) with $c=M\mu_p$. Theorem 2.1 states that under a stronger condition $\delta_{(a+1)k}^c<(b-1)/(b+1)$, there exists unique solution of (6). To this end, let $L=(a+1)kd=(\lceil b^{2/(2-p)}\rceil+1)kd$, $b>1$. Choose $\eta=r(b-1)/(b+1)$ for $r\in(0,1)$ and $\tau^p=(1-r)(b-1)/2b<1$ to satisfy

$$\frac{\eta+\tau^p}{1-\tau^p}\le\delta_{(a+1)k}^c\le\frac{b-1}{b+1}.$$

From Lemma 3.1, an upper bound for the probability that an $M\times L$ submatrix of Φ fails to satisfy (21) is $2(1+2/\tau)^L e^{-\eta^2M/2pc_p^2}$. As discussed in [8], a block-sparse signal can be treated as the vector that lies in a structured union of subspaces. Specifically, for a block k-sparse signal $\mathbf{x}\in\mathbb{R}^N$ over $\mathcal{I}=\{d_1=d,\dots,d_m=d\}$ with $N=md$, there are $\binom{N/d}{k}$ subspaces in the union. Thus, the union bound tells us that Φ fails to satisfy (21) with probability

$$\le\binom{m}{(\lceil b^{2/(2-p)}\rceil+1)k}2\left(1+\frac{2}{\tau}\right)^L e^{-\eta^2M/2pc_p^2},$$

where c_p is given in (24). Since $\binom{u}{v}\le(eu/v)^v$ always holds for integers $u>v>0$ and $\tau<1$, it is not hard to see that

$$\binom{m}{(\lceil b^{2/(2-p)}\rceil+1)k}2\left(1+\frac{2}{\tau}\right)^L e^{-\eta^2M/2pc_p^2}$$
$$\le 2\left(\frac{em}{\lceil b^{2/(2-p)}\rceil k+k}\right)^{(\lceil b^{2/(2-p)}\rceil k+k)}\left(\frac{3}{\tau}\right)^L e^{-\eta^2M/2pc_p^2}.$$

It is sufficient to show that the right hand side of above quantity can be bounded by $2e^{-\eta^2M/4pc_p^2}$. For this it suffices that

$$\begin{aligned}
M\ge{}&\frac{4pc_p^2}{\eta^2}\left[k\left(\lceil b^{2/(2-p)}\rceil+1\right)\ln\left(\frac{em}{\lceil b^{2/(2-p)}\rceil k+k}\right)\right.\\
&\left.+kd\left(\lceil b^{2/(2-p)}\rceil+1\right)\ln\left(\frac{3}{\tau}\right)\right]\\
={}&\frac{4pc_p^2}{\eta^2}\left[k\left(\lceil b^{2/(2-p)}\rceil+1\right)\left(1+\ln\left(\frac{m}{k}\right)\right.\right.\\
&\left.\left.-\ln(\lceil b^{2/(2-p)}\rceil+1)\right)+kd\left(\lceil b^{2/(2-p)}\rceil+1\right)(\ln 3-\ln\tau)\right]\\
={}&\frac{4c_p^2(b+1)^2}{r^2(b-1)^2}kd\left(\lceil b^{2/(2-p)}\rceil+1\right)\left(p\ln 3+\ln\frac{2b}{(1-r)(b-1)}\right)\\
&+\frac{4pc_p^2(b+1)^2}{r^2(b-1)^2}k\left(\lceil b^{2/(2-p)}\rceil+1\right)\\
&\times\left(\ln\frac{m}{k}+1-\ln\left(\lceil b^{2/(2-p)}\rceil+1\right)\right).
\end{aligned}$$

With the same arguments as in [1,16], we choose $r=0.849$ and $b=5$. Then we obtain

$$\begin{aligned}
M\ge{}&(35.1+13.7p)c_p^2\left(\lceil 5^{2/(2-p)}\rceil+1\right)kd\\
&+p12.5c_p^2\left(\lceil 5^{2/(2-p)}\rceil+1\right)\\
&\times\left[k\ln\frac{m}{k}+k\left(1-\ln\left(\lceil 5^{2/(2-p)}\rceil+1\right)\right)\right]\\
\ge{}&C_3(p)kd+pC_4(p)k\ln\frac{m}{k}
\end{aligned}\tag{25}$$

measurements are sufficient to yield the p-RIP condition (21) with probability exceeding

$$1-2e^{-\eta^2M/4pc_p^2}=1-2e^{-0.7049M/pc_p^2}\ge 1-2e^{-\beta(p)M},$$

where $\beta(p)=1/2pc_p^2$. □

Remark 3.1. It is easy to check that for a given $p \in (0,1]$, $C_3(p)$ and $C_4(p)$ are finite constants, and the second term of (25) has the dominant impact on the number of measurements in an asymptotic sense. When $p \to 0$, the second term of (25) vanishes, (25) thus turns to be $M \ge C_3(0)kd$. And when $p=1$, (25) turns to be $M \ge C_3(1)kd + C_4(1)k\ln(m/k)$, which implies fewer measurements are required with smaller p for exact recovery via mixed l_2/l_p minimization than when $p=1$. Meanwhile, when $p=1$, (25) has the same order with the result of proposition 4 in [8].

Remark 3.2. Note that when $p \to 0$, $c_p \le 1.13(31/40)^{1/4} \approx 1.062$. Thus, Theorem 3.1 gives an estimate of $C_3(0) \le 237.5$, while numerical experiments (see Section 4) suggest that $C_3(0)$ should be less than 3.5. As discussed in [1], there are several ways to obtain sharp constants. We leave it to the interested readers.

Remark 3.3. Though we have only considered the case in which $d_1 = \cdots = d_m = d$, the proof of Theorem 3.1 can be adapted to the case in which d_i are not equal. In this case, we need to consider the worst case scenario corresponding to the maximal block length in $\mathcal{I}$. Thus, Theorem 3.1 holds for $d = \max(d_i)$.

4. Numerical experiments

In this section, we conduct several numerical experiments to demonstrate the validation of our presented theoretical results. More precisely, the main purpose of this section is two-fold: first, to check how many random Gaussian measurements are needed for mixed l_2/l_p-minimization to recover a (nearly) block-sparse signal in both noiseless and noisy cases; second, to empirically investigate the solution of (7) in the presence of noise, and how it depends on p. We adopt the iteratively reweighted least squares (IRLS) approach to solve the nonconvex optimization problem (7). The IRLS methodology has been widely used for recovering sparse signals and low-rank matrices [15,17,18]. We begin with $\mathbf{x}^{(0)} = \arg\min \|\mathbf{y} - \Phi\mathbf{x}\|_2^2$, and set $\gamma_0 = 1$. Then let $\mathbf{x}^{(t+1)}$ be the solution of

$$\min_{\mathbf{x}} \frac{1}{2\lambda}\|\mathbf{y} - \Phi\mathbf{x}\|_2^2 + \frac{1}{2}\|W^{(t)}\mathbf{x}\|_2^2, \tag{26}$$

where $\lambda > 0$ is a regularization parameter, and the weighting matrix $W^{(t)}$ is defined as $W_i^{(t)} = diag(p^{1/2}(\gamma_t^2 + \|\mathbf{x}^{(t)}[i]\|_2^2)^{p/4-1/2})$ for i-th block. It is easy to obtain that the solution of (26) can be given explicitly as

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

The value of γ is decreased according to the rule $\gamma_{t+1} = 0.99\gamma_t$ and this iteration is continued until γ becomes very small, i.e, $\gamma \le 10^{-8}$.

In our experiments, the measurement matrix Φ was generated by creating an $M \times N$ matrix with i.i.d draws from a standard Gaussian distribution. For a generated (nearly) block-sparse signal $\mathbf{x}$, the measurements $\mathbf{y}$ were observed from the noisy model $\mathbf{y} = \Phi\mathbf{x} + \sigma\mathbf{z}$, where $\mathbf{z}$ was Gaussian white noise which can be generated by MATLAB command *randn*$(M,1)$. In order to verify the validity of our presented theoretical results, we consider several different values of p for the mixed l_2/l_p method. In each experiment, we report the average results over 100 independent random trails.

4.1. The exactly block-sparse case

We first consider the case that the signal $\mathbf{x}$ is exactly block-sparse. In this set of experiments, the signals with length $N=192$ were generated by choosing k blocks uniformly at random, and then choosing the nonzero values from a standard Gaussian distribution for these k blocks. For IRLS, we set $\lambda = 10^{-6}$ in noiseless case ($\sigma = 0$), and manually adjusted λ in noisy case ($\sigma > 0$).

In Fig. 1(a)–(c), exact recovery frequency is plotted versus measurement level M for three different block sizes: $d=2$, $d=4$, $d=8$. In this test, the number of nonzero blocks k was fixed to 8, and the measurements y were observed without noise ($\sigma = 0$). The recovery was regarded exact if $\|\mathbf{x}^* - \mathbf{x}\|_2 / \|\mathbf{x}\|_2 \le 10^{-4}$. One can easily see that reducing p below 1 clearly reduces the number of measurements required for exact recovery, which is expected by Theorem 3.1. However, it is also shown that there is almost no improvement when $p \le 0.5$. We emphasize that the phenomenon do not counter our theoretical results that smaller p should need fewer measurements for exact recovery. As noted in [1,18], a smaller p makes the minimizing functional more nonconvex and thus more measurements may be needed for the IRLS algorithm to converge to the global solution. It is also easy to see that when $p \le 0.5$, the sample size M/kd needed for exact recovery is always less than 3.5 and decreases as d increases. This suggests that the constant $C_3(0)$ should be less than 3.5.

Fig. 2(a)–(c) depict exact recovery frequency versus the block sparsity k for three different block sizes: $d=2$, $d=4$, $d=8$. Similar to the above test, the signals were perfectly measured. The number of measurements M was fixed to 130. It is evident that reducing p below 1 recovers more sparse signals, which is expected. It is also observed that, $p=0.5$ performed similar with $p = 0.2, 0.01$.

We now present some simulation results related to Theorem 2.1. Several curves of theoretical recovery error $\|\mathbf{x} - \mathbf{x}^*\|_2$ versus different values of $p = 0.01, 0.1, 0.2, \ldots, 0.9, 1$ are shown in Fig. 3. In this test, σ varied among $\{0.01, 0.03, 0.1\}$. It is shown that, for relatively small noise ($\sigma = 0.01, 0.03$), except $p=0.01$, the error changes little, and $p<1$ performed a little bit better than $p=1$; for relatively large noise ($\sigma = 0.1$), various different values of p provided worse recovery performance than $p=1$, and the error is least for $p=0.5$ in Fig. 3(a) and $p=0.9$ in Fig. 3(b). Recall that [19], for a given noise level σ, the error $\|\mathbf{y} - \Phi\mathbf{x}\|_2^2$ can be bounded by term $\sigma^2(M + \tau\sqrt{2M})$ with high probability, where $\tau > 1$ is a constant. Thus, for demonstrating the behavior of the constants in (10), we can set $\epsilon = \sigma\sqrt{M + \tau\sqrt{2M}}$. Let $\tau = 2$ used in [19], one can easily calculate ϵ=0.140, 0.420, 1.399 for block 16-sparse signals (Fig. 3(a)) and ϵ=0.123, 0.369, 1.229 for block 6-sparse signals (Fig. 3(b)), varying the noise level σ=0.01, 0.03, 0.1. Therefore, one can further obtain that the constant C_2 in (10) appears to be less than 2 in most cases.

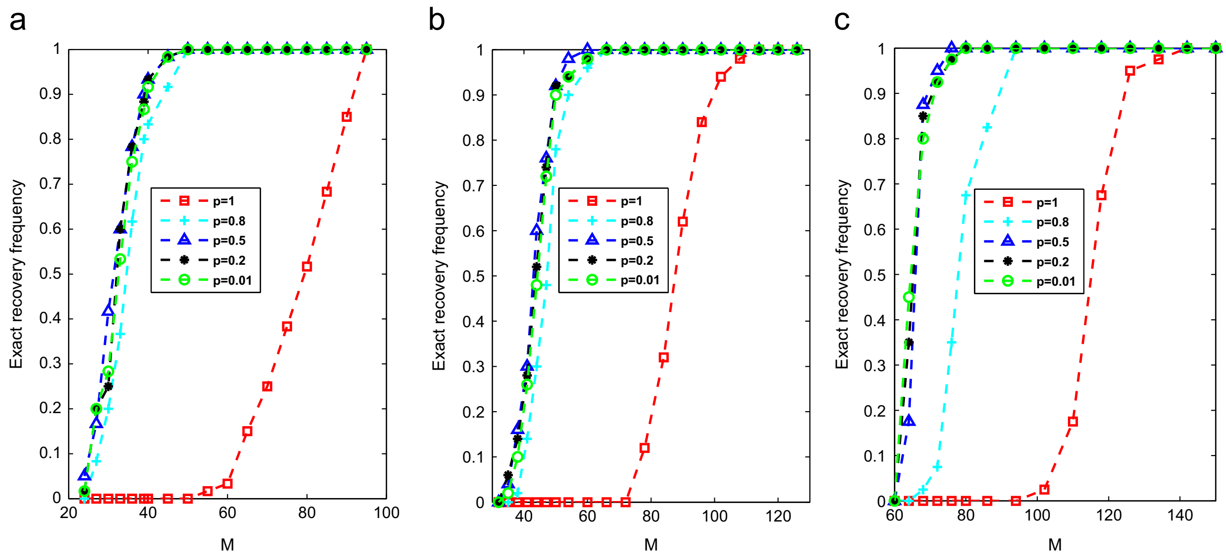

Fig. 1. Exact recovery performance of mixed l_2/l_p-minimization, varying the number of measurements for (a) $d=2$, (b) $d=4$ and (c) $d=8$. The signals have $k=8$ nonzero blocks.

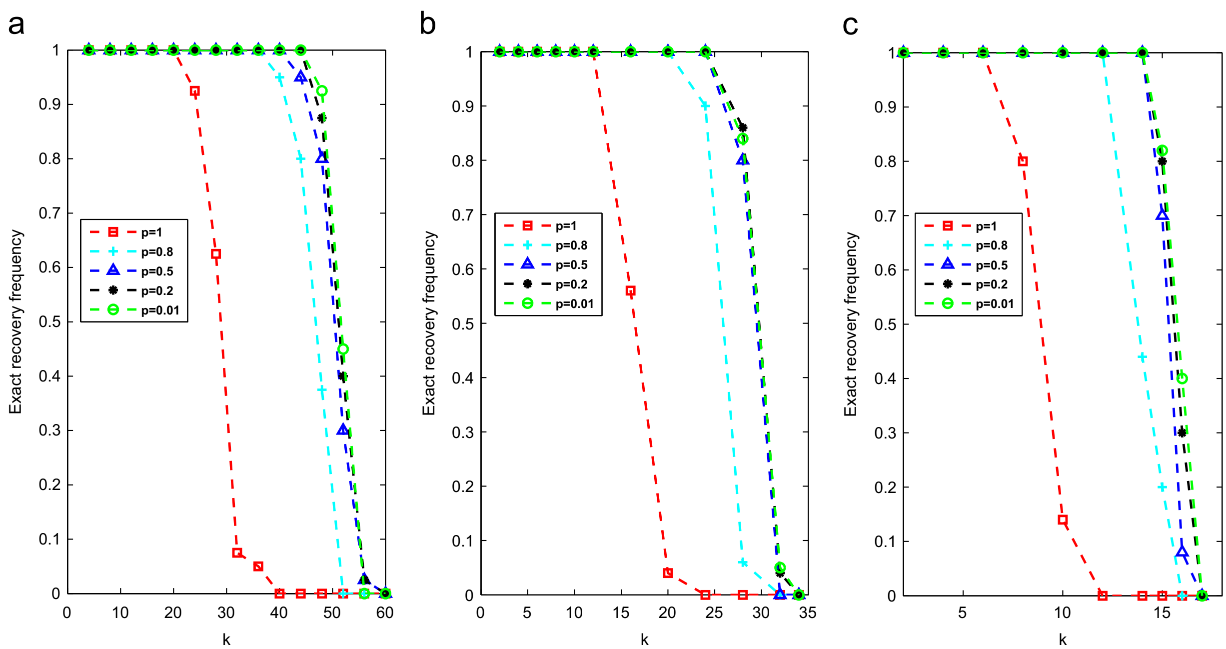

Fig. 2. Exact recovery performance of mixed l_2/l_p-minimization, varying the number of block sparsity for (a) $d=2$, (b) $d=4$ and (c) $d=8$. The number of measurements M was fixed to 130.

These illustrate that the constant C_2 is well behaved for a wide range of p values, and the mixed l_2/l_p minimization method guarantees a stable recovery of block-sparse signals in the presence of noise.

4.2. The nearly block-sparse case

As noted in [20], the signals whose l_2 norm of blocks have a power-law decay rate are nearly block-sparse signals. Thus, in this case, we generated $\mathbf{x}$ whose l_2 norm of blocks decay like $i^{-\alpha}$ where $i \in \{1,\ldots,m\}$ and $\alpha > 1$, and then observed the noisy measurements $\mathbf{y}$ from the model described before. In this set of experiments, the signal length N was fixed to 192.

Fig. 4 shows the recovery performance of mixed l_2/l_p minimization in terms of signal to noise ratio (SNR). For a recovered signal $\mathbf{x}^*$, the SNR is calculated as $\mathrm{SNR} = 20\log_{10}(\|\mathbf{x}\|_2/\|\mathbf{x}-\mathbf{x}^*\|_2)$. For simplicity, we only consider one

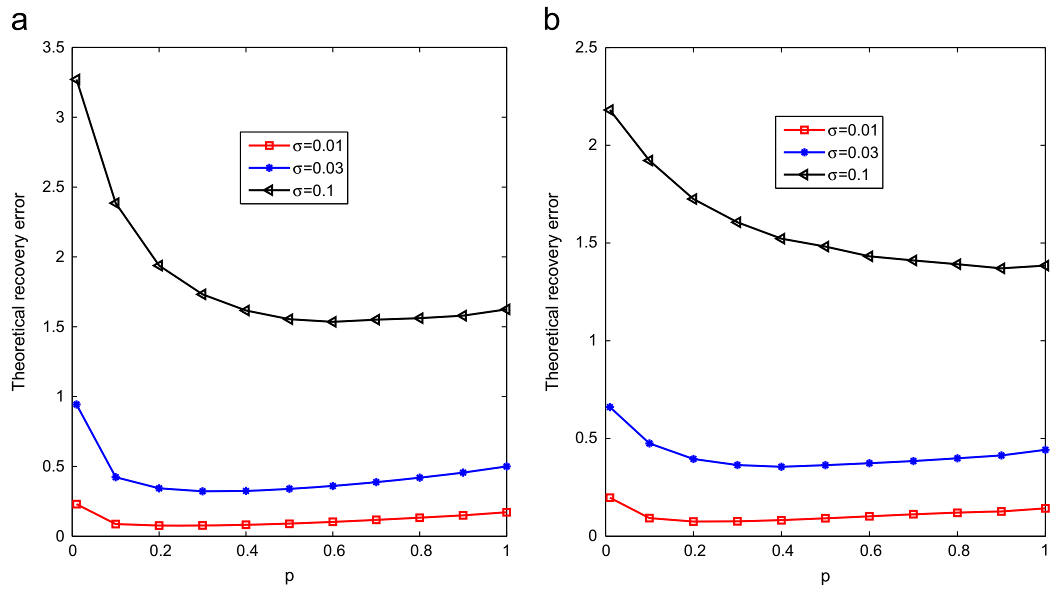

Fig. 3. Theoretical recovery error versus p for the solution of mixed l_2/l_p-minimization (7) for different exactly block-sparse signals: (a) $d=2$, $k=16$ and (b) $d=4$, $k=6$. The number of measurements M was fixed to 160 and 120 respectively.

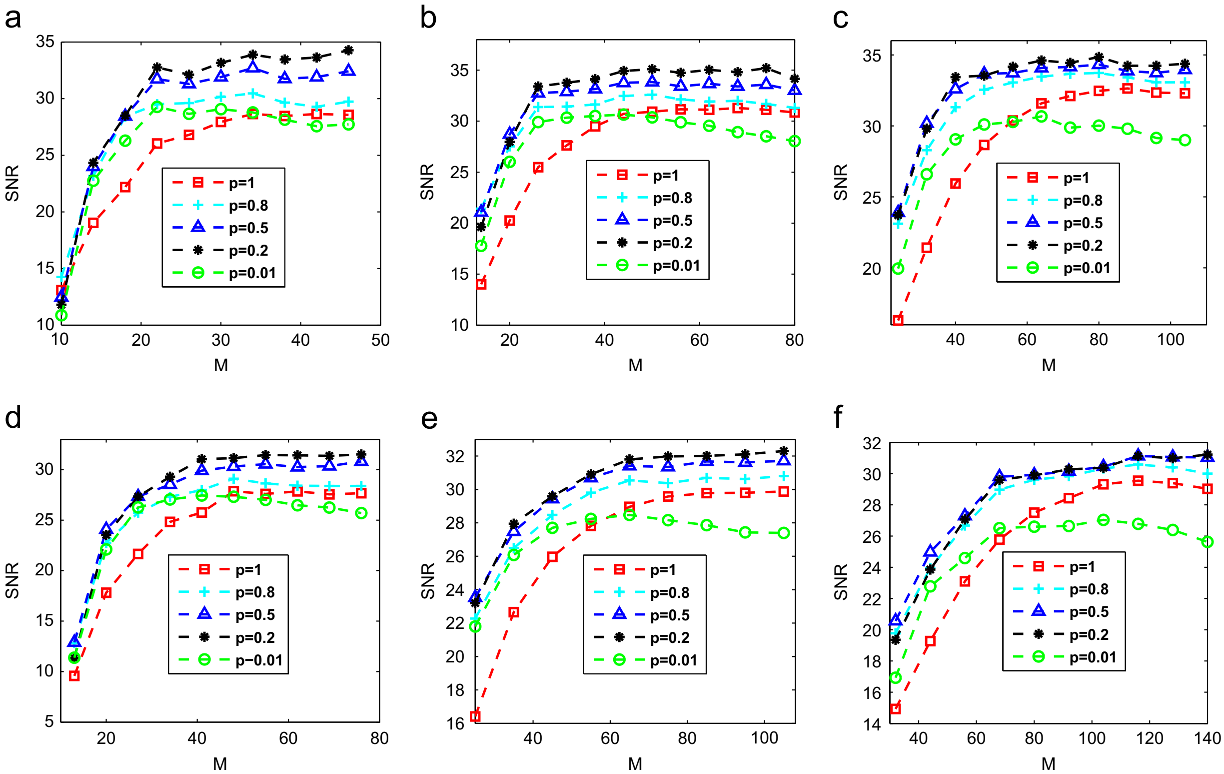

Fig. 4. Robust recovery performance of mixed l_2/l_p-minimization in terms of SNR, varying the number of measurements for different nearly block-sparse signals: (a) $\alpha=3.5$, $d=2$; (b) $\alpha=3.5$, $d=4$; (c) $\alpha=3.5$, $d=8$; (d) $\alpha=2.5$, $d=2$; (e) $\alpha=2.5$, $d=4$; and (f) $\alpha=2.5$, $d=8$.

noise level: $\sigma=0.01$. For IRLS, the parameter $\lambda=0.002$ was used. We plotted SNR for nearly block 3-sparse signals ($\alpha=3.5$) in Fig. 3(a)–(c) and for nearly block 6-sparse signals ($\alpha=2.5$) in Fig. 3(d)–(f), varying the number of measurements. It is easy to see from Fig. 3 that, except $p=0.01$, the mixed l_2/l_p method improves the recovery performance as p decreases. This suggests that, on the one hand, for small values of σ, decreasing p improves

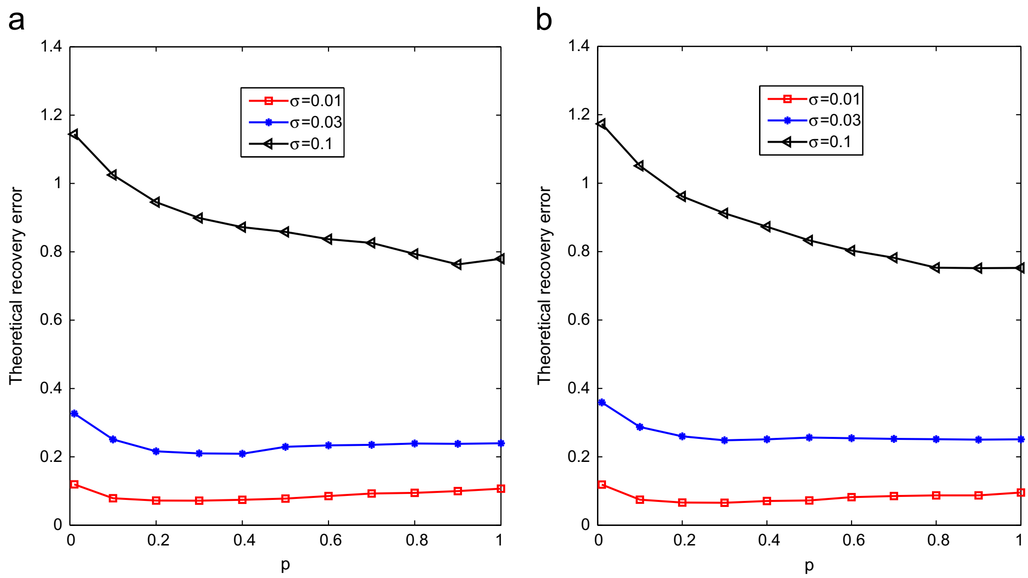

Fig. 5. Theoretical recovery error versus p for the solution of mixed l_2/l_p-minimization (7) for different nearly block-sparse signals: (a) $\alpha=2.5$, $d=2$ and (b) $\alpha=3.5$, $d=4$. The number of measurements M was fixed to 60 in both situations.

robustness to noise, which is also shown in Fig. 3; on the other hand, in noisy case, the IRLS algorithm may usually find bad local minimizer when p is very small. It is also clear that fewer measurements with smaller p are needed for robust recovery than when $p=1$, which is consistent with our theoretical results. In addition, as in the exact recovery case, the sample size M/kd needed for robust recovery of nearly block-sparse signals is less than 3.5 when $p \le 0.5$.

Fig. 5 shows some sample curves of theoretical recovery error $||\mathbf{x}-\mathbf{x}^*||_2$ versus p for nearly block-sparse signals. It is shown that there is little difference between the each curve in Fig. 5 and the corresponding curve in Fig. 3. Similarly, we obtained $\epsilon=0.091, 0.272, 0.905$, varying the noise level $\sigma=0.01, 0.03, 0.1$ in both situations. One can also check the approximation error – the first term on the right side of (10). As a reference, the average best k-block approximation error equals to 0.106 for nearly block 6-sparse signals ($\alpha=2.5$), and 0.071 for nearly block 3-sparse signals ($\alpha=3.5$) respectively. Therefore, one can obtain from Fig. 5 that the constants in (10) seem to be quite low, and the theoretical recovery error is dominated by the observation error in relatively strong noisy case ($\sigma=0.1$). To sum up, as expected, there is a wide range of p values for which the constants in (10) are well behaved, and mixed l_2/l_p-minimization achieves a robust recovery.

5. Conclusion

In this paper, we studied the problem of recovering an unknown nearly block k-sparse signal $\mathbf{x}$ from a given set of noisy linear measurements. By extending the notion of restricted p-isometry constant defined in [1] to the setting of block-sparse signal recovery, we established a p-RIP condition for robust recovery of nearly block-sparse signals via mixed l_2/l_p-minimization in the presence of noise. In particular, we obtained a p-RIP condition for exact recovery and determined how many random Gaussian measurements are needed for the p-RIP condition to be satisfied with high probability. Finally, a series of numerical experiments have been carried out to prove the validation of the theoretical derivations.

References

[1] R. Chartrand, V. Staneva, Restricted isometry properties and non-convex compressive sensing, Inverse Probl. 24 (3) (2008) 1–14.

[2] F. Parvaresh, H. Vikalo, S. Misra, B. Hassibi, Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays, IEEE J. Sel. Top. Signal Process. 2 (3) (2008) 275–285.

[3] S. Cotter, B. Rao, Sparse channel estimation via matching pursuit with application to equalization, IEEE Trans. Commun. 50 (3) (2002) 374–377.

[4] A. Majumdar, R.K. Ward, Compressed sensing of color images, Signal Process. 90 (12) (2010) 3122–3127.

[5] R. Vidal, Y. Ma, A unified algebraic approach to 2-D and 3-D motion segmentation and estimation, J. Math. Imaging Vis. 25 (3) (2006) 403–421.

[6] E.J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inf. Theory 52 (2) (2006) 489–509.

[7] D.L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (4) (2006) 1289–1306.

[8] Y.C. Eldar, M. Mishali, Robust recovery of signals from a structured union of subspaces, IEEE Trans. Inf. Theory 55 (11) (2009) 5302–5316.

[9] Y.C. Eldar, P. Kuppinger, H. Bolcskei, Block-sparse signals: uncertainty relations and efficient recovery, IEEE Trans. Signal Process. 58 (6) (2010) 3042–3054.

[10] J.H. Lin, S. Li, Block sparse recovery via mixed l_2/l_1 minimization, Acta Math. Sin. 29 (7) (2013) 1401–1412.

[11] J. Huang, T. Zhang, The benefit of group sparsity, Ann. Stat. 38 (4) (2010) 1978–2004.

[12] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, IEEE Signal Process. Lett. 14 (10) (2007) 707–710.

[13] S. Foucart, M.J. Lai, Sparsest solutions of underdetermined linear systems via l_q-minimization for $0<q\le 1$, Appl. Comput. Harmon. Anal. 26 (3) (2009) 395–407.

[14] Z.B. Xu, X.Y. Chang, F.M. Xu, H. Zhang, $l_{1/2}$ regularization: a thresholding representation theory and a fast solver, IEEE Trans. Neural Netw. Learn. Syst. 23 (7) (2012) 1013–1027.
[15] Y. Wang, J.J. Wang, Z.B. Xu, On recovery of block-sparse signals via mixed $l_2/l_q(0<q\leq 1)$ norm minimization, EURASIP J. Adv. Signal Process. 76 (2013) 1–17.
[16] Y. Shen, S. Li, Restriced p-isometry property and its application for nonconvex compressive sensing, Adv. Comput. Math. 37 (2012) 441–452.
[17] R. Chartrand, W.T. Yin, Iteratively reweighted algorithms for compressive sensing, in: IEEE International Conference on Acoustics Speech and Signal Processing, 2008, pp. 3869–3872.
[18] M.J. Lai, Y.Y. Xu, W.T. Yin, Improved iteratively reweighted least squares for unconstrained smoothed l_p minimization, SIAM J. Numer. Anal. 5 (2) (2013) 927–957.
[19] E.J. Candès, J. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Commun. Pure Appl. Math. 59 (8) (2006) 1207–1223.
[20] R.G. Baraniuk, V. Cevher, M.F. Duarte, C. Hegde, Model-based compressive sensing, IEEE Trans. Inf. Theory 56 (4) (2010) 1982–2001.